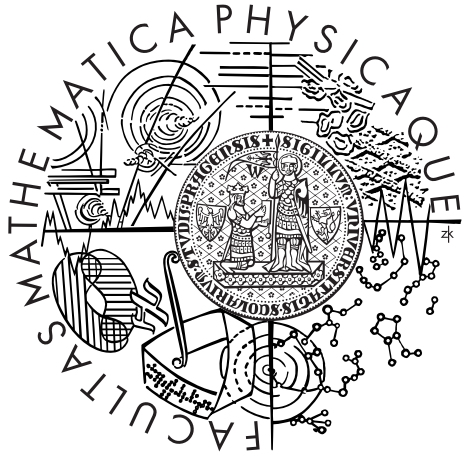
Charles University in Prague
Faculty of Mathematics and Physics

Ruprecht-Karls-University Heidelberg
Faculty of Mathematics and Computer Sciences

DOCTORAL THESIS



Jaroslava Hasnedlová

Fluid-structure interaction of compressible flow

Department of Numerical Mathematics
Institute of Applied Mathematics

Supervisors: Prof. RNDr. Miloslav Feistauer, DrSc., Dr. h. c.
Prof. Dr. Dr. h. c. Rolf Rannacher

Study programme: Mathematics
Specialization: Scientific and Technical Calculations

Prague 2012

I would like to thank all those who supported me in my doctoral study and the work on my thesis. I very appreciate the help and guidance received from my supervisors Prof. RNDr. Miloslav Feistauer, DrSc., Dr. h. c. and Prof. Dr. Dr. h. c. Rolf Rannacher and I am grateful for numerous remarks, corrections and advices they gave me throughout my work. I would like to thank Ing. Jaromír Horáček, DrSc. for my admission to his team and a lot of fruitful discussions and practical advices touching the problem of fluid-structure interaction in human vocal tract. I am also obliged to RNDr. Václav Kučera, Ph.D. for the help he provided me during the work on this thesis and the DGFEM program as well as to Mgr. Adam Kosík for his provided structure program and a great cooperation in our working group. My thanks go also to all other members of working groups of Prof. RNDr. Miloslav Feistauer, DrSc., Dr. h. c., Prof. Dr. Dr. h. c. Rolf Rannacher and Ing. Jaromír Horáček, DrSc. who provided me everytime fruitful creative surroundings.

Last but not least, I am in debt to my parents and husband, whose support and patience made this work possible.

My thanks also go to institutions that provided financial support for my research work. Through my doctoral study, my work was partially supported by the Grants GAChU 549912 and GAChU 12810, the Project OC 09019 "Modelling of voice production based on biomechanics" within the program COST of the Ministry of Education of the Czech Republic and EU project COST 2103 "Advanced Voice Assessment", by the project P101/11/0207 of the Czech Science Foundation,by the grants SVV-2010-261316, SVV-2011-263316 and SVV-2012-265316.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, May 18, 2012 .....

Jaroslava Hasnedlová

Title: Fluid-structure interaction of compressible flow

Author: RNDr. Jaroslava Hasnedlová
Department: Department of Numerical Mathematics,
Institute of Applied Mathematics
Supervisors: Prof. RNDr. Miloslav Feistauer, DrSc., Dr. h. c.,
Prof. Dr. Dr. h. c. Rolf Rannacher
Supervisors' e-mail addresses: feist@karlin.mff.cuni.cz,
rannacher@iwr.uni-heidelberg.de

Abstract: The presented work is split into two parts. The first part is devoted to the theory of the discontinuous Galerkin finite element (DGFE) method for the space-time discretization of a nonstationary convection-diffusion initial-boundary value problem with nonlinear convection and linear diffusion. The DGFE method is applied separately in space and time using, in general, different space grids on different time levels and different polynomial degrees p and q in space and time discretization. The main result is the proof of error estimates in $L^2(L^2)$-norm and in DG-norm formed by the $L^2(H^1)$-seminorm and penalty terms. The second part of the thesis deals with the realization of fluid-structure interaction problem of the compressible viscous flow with the elastic structure. The time-dependence of the domain occupied by the fluid is treated by the ALE (Arbitrary Lagrangian-Eulerian) method, when the compressible Navier-Stokes equations are formulated in the ALE formulation. The deformation of the elastic body, caused by the aeroelastic forces, is described by the dynamical elasticity equations. Both these systems are coupled by the transmission conditions. For the space discretization of the flow problem the DGFE method is used. The time-discretization is realized by the backward difference formula. The structural problem is discretized by conforming finite element method and the Newmark method. The fluid-structure interaction is realized via weak or strong coupling algorithms. The developed technique is tested by numerical experiments and applied to the simulation of vibrations of vocal folds during phonation onset.

Keywords: Discontinuous Galerkin method, nonstationary convection-diffusion problem, space-time discretization, error estimates, fluid-structure interaction, compressible Navier-Stokes equations, dynamical elasticity equations, ALE method, coupling algorithms.

Název práce: Interakce stlačitelného proudění a struktur
Autor: RNDr. Jaroslava Hasnedlová
Katedra: Katedra numerické matematiky,
Institute of Applied Mathematics
Vedoucí práce: Prof. RNDr. Miloslav Feistauer, DrSc., Dr. h. c.,
Prof. Dr. Dr. h. c. Rolf Rannacher
e-mail vedoucího: feist@karlin.mff.cuni.cz,
rannacher@iwr.uni-heidelberg.de

Abstrakt: Předkládaná práce je rozdělena do dvou částí. První část se zabývá teorií nespojité Galerkinovy metody konečných prvků (DGFEM) pro časoprostorovou diskretizaci nestacionárního problému konvekce-difuze s nelinearní konvekcí a linearní difuzí. DGFEM je aplikována odděleně v čase a prostoru s užitím obecně rozdílných sítí na různých časových úrovních a polynomů obecně rozdílných řádů p a q pro prostorovou a časovou diskretizaci. Hlavním zájmem této části je důkaz odhadu chyby metody v $L^2(L^2)$-normě a v DG-normě. Druhá část práce pojednává o problému interakce stlačitelného vazkého proudění s elastickým tělesem. Časová závislost oblasti vyplněné tekutinou je brána v potaz pomocí ALE metody a stlačitelné Navierovy-Stokesovy rovnice jsou formulovány v ALE tvaru. Deformace elastického tělesa způsobená aerodymickými silami je popsána pomocí dynamických rovnic elastického tělesa. Oba systémy jsou propojeny přechodovými podmínkami. Diskretizece proudění je v prostoru provedena pomocí DGFEM a v čase s využitím metody zpětných diferencí. Problém elastické struktury je diskretizován pomocí metody konečných prvků a Newmarkovy metody. Interakce je realizována pomocí silné a slabé vazby. Vyvinutá technika je testována na numerických experimentech a aplikována na simulaci vibrací lidských hlasivek na začátku fonace.

Klíčová slova: Nespojitá Galerkinova metoda, nestacionární problém konvekce-difuze, časoprostorová diskretizace, odhady chyb, interakce proudění a struktury, stlačitelné Navierovy-Stokesovy rovnice, dynamické rovnice elastického tělesa, ALE metoda, algoritmy vazby.

Contents

List of Figures

List of Tables

Introduction

At the current speed of technology progress, the coupled problems describing the interactions of fluid flow with elastic structure motion are of great importance in many fields of physical and technical sciences such as biomechanics, aerospace, civil and mechanical engineering, etc.. The need of the modeling of flow around flexible structures leads to the developement of a new scientifical and technical discipline: the aeroelasticity. The aeroelasticity has many important engineering and scientific applications (e.g. in aerospace industry - aircraft design and safety; in civil engineering - stability of bridges, towers, smokestacks or skyscrapers; in mechanical engineering - bladed machines, ect.). The consequence of the aeroelastic effects can positively (the flow-induced vibration producing voice in human vocal folds) or negatively (the flow-induced vibration leading to material fatigue or inducing excessive noise generation) affect the operation of the system. The problems of the interaction of fluid flow with elastic structures were studied by a number of different methods in several books (e.g. [36], [27], [51], [28], [14], [54], [55]). Mostly, simplified linearized problems applied in technology are used. Recently, the research focuses also on mathematical and numerical modeling of nonlinear coupled problems. This represents complicated mathematical problems caused by the time-dependence of the computational domain and by the necessity of coupling of the flow problem with the elasticity problem. Here, we can mention for example the papers [42], [38], [40], [52]. In the case of overcoming the problems of coupling elasticity of the body with the flow problem we need to solve difficulties linked with the simulation of compressible flow. Due to the simulation of compressible flow in the time dependent domain, which is affected by the behaviour of the elastic structure, it is necessary to treat problems caused by nonlinear convection dominating over diffusion, i.e. boundary layers and wakes for large Reynolds numbers and instabilities caused by acoustic effects for low Mach numbers. A suitable numerical method for the solution of compressible flow suffering from mentioned difficulties is the discontinuous Galerkin finite element (DGFE) method.

This thesis is split into two main parts. The first part is devoted to the theoretical analysis of the space-time discontinuous Galerkin method for nonlinear convection-diffusion problems. In the second part we are concerned with the simulation of flow-induces vibrations of human vocal folds.

The discontinuous Galerkin finite element method represents a good technique allowing to realize numerical approximations of nonlinear differential equations in domains with a complex geometry, whose solutions have a complicated structure. The use of piecewise polynomial approximations of the sought solution on finite element meshes without any requirement on the continuity between neighbouring elements can be considered as an generalization of the finite volume and finite element methods.

This gives us the advantages of the both mentioned methods and the possibility of applications on unstructured grids used in the case of complex geometries. The DGFE methods allow also to construct higher order schemes in a natural way, which is suitable for the approximation of discontinuous solutions of conservation laws or solutions of singularly perturbed convection-diffusion problems having steep gradients.

The original DGFE method was first used in [59] for the solution of a neutron transport linear equation and analyzed theoretically in [50] and later in [46]. The developement of the DGFE techniques for the numerical solution of second-order elliptic problems or parabolic problems ([5], [66]) and a biharmonic problem ([10]) comes nearly simultaneously. Further, the DGFE method was applied to a wide range of problems, e.g. convection-diffusion linear or nonlinear problems ([15], [17], [18], [35], [32]), compressible flow ([11], [12], [13], [21], [22], [41], [65]), etc.. Theoretical analysis of various types of the DGFE method can be found, e.g. in [6], [7], [8], [60], [45].

In the first part of this thesis we are concerned with the space-time discontinuous Galerkin discretization applied separately in space and in time for the numerical solution of a nonstationary nonlinear convection-diffusion equation with a linear diffusion and nonlinear convection. The diffusion coefficient is a fixed positive constant. A singularly perturbed case with dominating convection is not considered. The main subject of this part of the thesis is the derivation of error estimates of the space-time DGFE method. Error estimates are derived under the assumption that the triangulations on all time levels are uniformly shape regular and the exact solution has some regularity properties. The time interval is split into subintervals and on each time level a different space mesh may be used in general. This is a suitable approach particularly in the case of the use of the space mesh adaptivity in the course of increasing time. Moreover, the triangulations used for the space discretization may be nonconforming with hanging nodes. The nonsymmetric, symmetric and incomplete versions of the discretization of the diffusion terms and interior and boundary penalty (i.e. NIPG, SIPG or IIPG versions) are used in the discontinuous Galerkin formulation. Piecewise polynomial approximations of different degrees p and q are applied for the space and time discretization. The error estimates are optimal in time, if the Dirichlet boundary conditions have behaviour in time as a polynomial of degree $\leq q$.

The structure of the first part of the thesis is the following. The continuous problem is formulated together with the main assumptions. Then the discontinuous Galerkin discretization in space and time is described. Further, some auxiliary results concerning properties of the forms appearing in the definition of the approximate solution are obtained. On the basis of these results the abstract error estimate is derived and the error estimates of the discontinuous Galerkin space-time discretization in terms of the sizes of the space and time meshes are proven.

In this part of the thesis we often cite and use techniques from articles [23], [24], [25] and [26]. In these articles only the space discretization of the problem is carried out by the discontinuous Galerkin finite element method. By contrast we deal with both space and time discontinuous Galerkin discretization.

The second part of this thesis is devoted to the numerical simulation of fluid-structure interaction. Especially we are focused on the modeling of flow-induced vibrations of the human vocal folds during the phonation onset. It means that we need to take into account the simulation of compressible viscous flow in a time-dependent

domain together with the elasticity behaviour of the channel walls formed by an elastic structure.

Our goal is the numerical finite element (FE) simulation of interaction of 2D compressible viscous flow in the glottal region with a compliant tissue of the human vocal folds modeled by a 2D elastic layered structure. A question is the mathematical and physical description of the mechanism for transforming the airflow energy in the glottis into the acoustic energy representing the voice source in humans. The primary voice source is given by the airflow coming from the lungs that causes self-oscillations of the vocal folds. The voice source signal travels from the glottis to the mouth, exciting the acoustic supraglottal spaces, and becomes modified by acoustic resonance properties of the vocal tract [62].

In [3] we can find an overview of the current state of mathematical models for the human phonation process. Such models are valuable tools for providing insight into the basic mechanisms of phonation and in future could help with surgical planning, diagnostics and voice rehabilitation. In current publications various simplified glottal flow models are used. They are based on the Bernoulli equation ([62]), 1D models for an incompressible inviscid fluid ([43]), 2D incompressible Navier-Stokes equations solved by the finite volume method ([4]) or finite element method ([20]). Acoustic wave propagation in the vocal tract is usually modeled separately using linear acoustic perturbation theory ([63]). Also the work [57], which is concerned with the finite volume solution of the Navier-Stokes equations for a compressible fluid with prescribed periodic changes of the channel cross-section of the glottal channel, can be found. The phonation onset was studied by using the potential flow model and three-mass lumped model for the vibrating vocal folds in [44] and for a 2D isotropic elastic model of the vocal folds in [67].

In the second part of this thesis we shall describe step by step a technique of the numerical simulation of vocal folds vibrations induced by compressible viscous flow. The first chapter of this part will be devoted to the description of the airflow by the Navier-Stokes equations or by the Euler equations. The dimensionless governing equations will be derived and the transformation of the governing equations to the arbitrary Lagrangian-Eulerian (ALE) form will be presented, which allows us to treat the time-dependency of the domain occupied by air. In the second chapter of this part we shall pay attention to the model of human vocal folds that are considered as isotropic elastic bodies. The linear elasticity equations used for the description of vibrations of vocal folds are presented. In the next chapter we shall take into account the coupled problem, which presents a strongly nonlinear dynamical system. Coupling conditions together with the continuous coupled problem will be described.

Further two chapters will be focused on the discretization of the flow and structural problems separately. The flow problem is discretized in space by the discontinuous Galerkin finite element method, using piecewise polynomial approximations, in general discontinuous on interfaces between neighbouring elements. The time discretization is carried out by the backward difference formula (BDF) in time. The structural problem is approximated by conforming finite elements and the Newmark method.

The next chapter will present the construction of the ALE mapping with the aid of a stationary elasticity problem together with its discretization and the formulation of

the discrete fluid-structure interaction problem. The weak (loose) and strong coupling algorithms will be described.

Further, the algorithmization of the developed method and the description of the programme which was worked out are presented. The described method was applied to the solution of several problems. First we present the results obtained in a simplified computational domain with the use of three different meshes with different numbers of elements. The comparison of these results allows us to demonstrate the convergence tendency of the method. The applicability and robustness of the developed method are shown on the second example representing approximate human vocal folds region with a more realistic model of an elastic part. At the end we mention some open problems and specify subjects for a further work.

Part I

Analysis of space-time discontinuous Galerkin method for nonlinear convection-diffusion problems

Chapter 1

Continuous problem

In Part I we shall be concerned with the theoretical analysis of the space-time discontinuous Galerkin method for the numerical solution of a nonstationary convection-diffusion equation. This problem represents a very simplified model of the flow problem, which will be an ingredient of the problem simulated in Part II.

In this chapter the continuous problem for the mentioned model equation is formulated and the main assumptions are introduced. The presented problem is simplified due to the targeting the error analysis.

In what follows we denote by $\mathbb{R}$ the set of all real numbers and by $\mathbb{N}$ the set of all natural numbers.

Let us consider a bounded polyhedral domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) and a time interval $(0,T) \subset \mathbb{R}$ with $T > 0$. Then we formulate the following initial-boundary value problem: Find $u:\ Q_T = \Omega \times (0,T) \longrightarrow \mathbb{R}$ such that

$$\frac{\partial u}{\partial t} + \sum_{s=1}^{d} \frac{\partial f_s(u)}{\partial x_s} - \varepsilon \Delta u = g \text{ in } Q_T = \Omega \times (0,T), \tag{1.1}$$

$$u|_{\partial\Omega\times(0,T)} = u_D, \tag{1.2}$$

$$u(x,0) = u^0(x),\ x \in \Omega. \tag{1.3}$$

We assume that $\varepsilon > 0$ is a constant, $g : Q_T \longrightarrow \mathbb{R}$ and $\boldsymbol{f} = (f_1, \ldots, f_d),\ f_s \in C^1(\mathbb{R}),\ |f_s'| \le C,\ s = 1, \ldots, d$. This means that the fluxes f_s are Lipschitz-continuous in $\mathbb{R}$, which simplifies the problem and makes some further estimates possible. As an example of an application of this assumption we can mention estimate (3.10).

It is possible to prove the existence and uniqueness of a weak solution to problem (1.1)-(1.3) using techniques presented in [61].

In the following part we use the standard notation of function spaces (see, e.g. [49]). If ω is a bounded domain, we define the Lebesgue spaces

$$L^\infty(\omega) = \left\{\text{measurable functions } \varphi;\ \|\varphi\|_{L^\infty(\omega)} = \operatorname{ess\,sup}_{x\in\omega} |\varphi(x)| < \infty\right\},$$

$$L^2(\omega) = \left\{\text{measurable functions } \varphi;\ \|\varphi\|_{L^2(\omega)} = \left(\int_\omega |\varphi|^2\right)^{1/2} < \infty\right\}$$

and the Sobolev space

$$H^k(\omega) = \left\{ \varphi \in L^2(\omega);\ \|\varphi\|_{H^k(\omega)} = \left(\sum_{|\alpha| \le k} \|D^\alpha \varphi\|^2_{L^2(\omega)} \right)^{1/2} < \infty \right\},$$

with the seminorm

$$|\varphi|_{H^k(\omega)} = \left(\sum_{|\alpha| = k} \|D^\alpha \varphi\|^2_{L^2(\omega)} \right)^{1/2}.$$

We also use the Bochner space. Let X be a Banach space with a norm $\|\cdot\|_X$ and a seminorm $|\cdot|_X$ and let s be an integer. Then we define:

$$\begin{aligned}
C([0,T];X) &= \left\{ \varphi : [0,T] \to X,\ \text{continuous}, \|\varphi\|_{C([0,T];X)} = \sup_{t\in[0,T]} \|\varphi(t)\|_X < \infty \right\}, \\
L^2(0,T;X) &= \left\{ \varphi : (0,T) \to X,\ \text{strongly measurable}, \|\varphi\|^2_{L^2(0,T;X)} = \int_0^T \|\varphi\|^2_X\ dt < \infty \right\}, \\
H^s(0,T;X) &= \left\{ \varphi \in L^2(0,T;X); \|\varphi\|^2_{H^s(0,T;X)} = \int_0^T \sum_{\alpha=0}^{s} \left\| \frac{\partial^\alpha \varphi}{\partial t^\alpha} \right\|^2_X dt < \infty \right\}.
\end{aligned}$$

Moreover, we set

$$\begin{aligned}
|\varphi|_{C([0,T];X)} &= \sup_{t\in[0,T]} |\varphi|_X, \\
|\varphi|_{L^2(0,T;X)} &= \left(\int_0^T |\varphi|^2_X\ dt \right)^{1/2}, \\
|\varphi|_{H^s(0,T;X)} &= \left(\int_0^T \left| \frac{\partial^s \varphi}{\partial t^s} \right|^2_X dt \right)^{1/2}.
\end{aligned}$$

We say that u satisfying (1.1)-(1.3) is a *strong solution*, if

$$u \in L^2(0,T;H^2(\Omega)) \tag{1.4}$$
$$\tfrac{\partial u}{\partial t} \in L^2(0,T;H^1(\Omega)). \tag{1.5}$$

The strong solution satisfies equation (1.1) pointwise (almost anywhere).

Chapter 2

Discretization

The aim of this chapter is the derivation of the discrete problem. First we need to explain our requirements on the construction of a mesh in Q_T and used notation. Also spaces of discontinuous functions and forms defined on this spaces will be presented.

2.1 Construction of a mesh in Q_T

In the time interval $[0, T]$ we shall construct a partition formed by time instants $0 = t_0 < \ldots < t_M = T$ and denote the interval $I_m = (t_{m-1}, t_m)$ with the time step $\tau_m = t_m - t_{m-1}$. Then we have $[0, T] = \bigcup_{i=1}^{M} \bar{I}_m, I_m \cap I_n = \emptyset$ for m $\neq$ n.

For each I_m we consider a partition $\mathcal{T}_{h,m}$ of the closure $\bar{\Omega}$ of the domain Ω into a finite number of closed d-dimensional simplices (triangles for $d = 2$ and tetrahedra for $d = 3$) with mutually disjoint interiors. We shall call $\mathcal{T}_{h,m}$ a triangulation of Ω and the standard properties of $\mathcal{T}_{h,m}$ used in the finite element method are not required. This means that the so-called hanging nodes (and in 3D also hanging edges) are admitted. The partitions $\mathcal{T}_{h,m}$ can be in general different for different m.

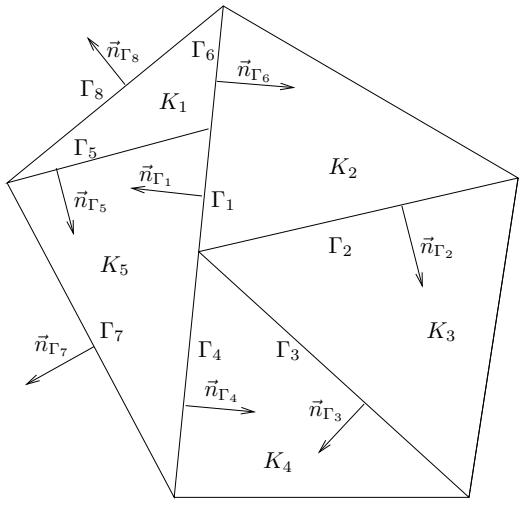


Figure 2.1: Example of elements $K_l,\ l = 1, \ldots, 5$, and faces $\Gamma_l,\ l = 1, \ldots, 8$, with the corresponding normals $\boldsymbol{n}_{\Gamma_l}$

Let K, $K' \in \mathcal{T}_{h,m}$. We say that K and K' are *neighbouring elements (neighbours)*, if the set $\partial K \cap \partial K'$ has positive $(d-1)$-dimensional measure. We call $\Gamma \subset K$ a *face* of K, if it is a maximal connected open subset either of $\partial K \cap \partial K'$, where K' is a neighbour of K, or of $\partial K \cap \partial\Omega$. $\mathcal{F}_{h,m}$ denotes the system of all faces of all elements $K \in \mathcal{T}_{h,m}$. Further, by $\mathcal{F}_{h,m}^B = \{\Gamma \in \mathcal{F}_{h,m}; \Gamma \subset \partial\Omega\}$ we denote the set of all boundary faces and set $\mathcal{F}_{h,m}^I = \mathcal{F}_{h,m} \backslash \mathcal{F}_{h,m}^B$, which is the set of all inner faces. Obviously, $\mathcal{F}_{h,m} = \mathcal{F}_{h,m}^I \cup \mathcal{F}_{h,m}^B$.

For each $\Gamma \in \mathcal{F}_{h,m}$ we define a unit normal vector $\boldsymbol{n}_\Gamma$. We assume that for $\Gamma \in \mathcal{F}_{h,m}^B$ the normal $\boldsymbol{n}_\Gamma$ has the same orientation as the outer normal to $\partial\Omega$. For each face $\Gamma \in \mathcal{F}_{h,m}^I$ the orientation of $\boldsymbol{n}_\Gamma$ is arbitrary but fixed. See Figure 2.1.

In our further considerations we shall use the following notation. For an element $K \in \mathcal{T}_{h,m}$ we set $h_K = \mathrm{diam(K)}$, $h_m = \max_{\mathrm{K}\in\mathcal{T}_{\mathrm{h,m}}} \mathrm{h_K}$, $h = \max_{\mathrm{m}=1,\ldots,\mathrm{M}} \mathrm{h_m}$. By ρ_K we denote the radius of the largest d-dimensional ball inscribed into K and by $|K|$ we denote the d-dimensional Lebesgue measure of K. $d(\Gamma)$ denotes the diameter of $\Gamma \in \mathcal{F}_{h,m}$. Finally, we set $\tau = \max_{\mathrm{m}=1,\ldots,\mathrm{M}} \tau_{\mathrm{m}}$.

2.2 Spaces of discontinuous functions and forms defined on these spaces

Let a function φ be defined in $\bigcup_{m=1}^M I_m$. Than we denote

$$\varphi_m^{\pm} = \varphi(t_m\pm) = \lim_{t\to t_m\pm} \varphi(t), \quad \{\varphi\}_m = \varphi(t_m+) - \varphi(t_m-). \tag{2.1}$$

Let $k > 0$ be an integer. Over a triangulation $\mathcal{T}_{h,m}$ we define the *Broken Sobolev spaces*

$$H^k(\Omega, \mathcal{T}_{h,m}) = \left\{v;\ v|_K \in H^k(K)\ \forall K \in \mathcal{T}_{h,m}\right\} \tag{2.2}$$

equipped with the seminorm

$$|v|_{H^k\left((\Omega,\mathcal{T}_{h,m}\right)} = \left(\sum_{K\in\mathcal{T}_{h,m}} |v|^2_{H^k(K)}\right)^{1/2}. \tag{2.3}$$

For each face $\Gamma \in \mathcal{F}_{h,m}^I$ there exist two neighbouring elements $K_\Gamma^L,\ K_\Gamma^R \in \mathcal{T}_{h,m}$ such that $\Gamma \subset \partial K_\Gamma^L \cap \partial K_\Gamma^R$. We use convention that $\boldsymbol{n}_\Gamma$ is the outer normal to the element K_Γ^L and the inner normal to the element K_Γ^R. For $v \in H^1(\Omega, \mathcal{T}_{h,m})$ and $\Gamma \in \mathcal{F}_{h,m}^I$ we introduce the following notation:

$$\begin{aligned} v|_\Gamma^L &= \text{the trace of } v|_{K_\Gamma^L} \text{ on } \Gamma, \\ v|_\Gamma^R &= \text{the trace of } v|_{K_\Gamma^R} \text{ on } \Gamma, \\ \langle v\rangle_\Gamma &= \tfrac{1}{2}\left(v|_\Gamma^L + v|_\Gamma^R\right), \\ [v]_\Gamma &= v|_\Gamma^L - v|_\Gamma^R. \end{aligned} \tag{2.4}$$

We can see that the value $[v]_\Gamma$ depends on the orientation of $\boldsymbol{n}_\Gamma$, but the value $[v]_\Gamma \boldsymbol{n}_\Gamma$ is independent of this orientation.

Now, we assume that $\Gamma \in \mathcal{F}_{h,m}^B$ and $K_\Gamma^L \in \mathcal{T}_{h,m}$ is such an element that $\Gamma \subset K_\Gamma^L \cap \partial\Omega$. For $v \in H^1(\Omega, \mathcal{T}_{h,m})$ we define $v|_\Gamma^R$ by extrapolation, i.e.

$$v|_\Gamma^R := v|_\Gamma^L = \text{the trace of } v|_{K_\Gamma^L} \text{ on } \Gamma. \tag{2.5}$$

In case of an appearance of $[\cdot]_\Gamma$ and $\langle\cdot\rangle_\Gamma$ in an integral $\int_\Gamma \ldots \mathrm{dS}$, where $\Gamma \in \mathcal{F}_{h,m}$, we omit the subscript Γ and write simply $[\cdot]$ and $\langle\cdot\rangle$. If $\Gamma \in \mathcal{F}_{h,m}^B$ and $v \in H^1(\Omega, \mathcal{T}_{h,m})$, then $\int_\Gamma v\mathrm{dS}$ means $\int_\Gamma v|_\Gamma^L \mathrm{dS}$.

For a fixed constant $C_W > 0$ we introduce the notation

$$\begin{aligned} h(\Gamma) &= \frac{h_{K_\Gamma^L} + h_{K_\Gamma^R}}{2C_W} \quad \text{for } \Gamma \in \mathcal{F}_{h,m}^I, \\ h(\Gamma) &= \frac{h_{K_\Gamma^L}}{C_W} \quad \text{for } \Gamma \in \mathcal{F}_{h,m}^B. \end{aligned} \tag{2.6}$$

In order to derive the DG scheme for the numerical solution of problem (1.1)-(1.3), we shall assume that u is a strong solution of this problem. We proceed in the following way: We multiply (1.1) by an arbitrary function $\varphi \in H^2(\Omega, \mathcal{T}_{h,m})$, integrate over $K \in \mathcal{T}_{h,m}$, use Green's theorem, summing over all $K \in \mathcal{T}_{h,m}$ and obtain the identity

$$\begin{aligned} &\sum_{K\in\mathcal{T}_{h,m}} \int_K \frac{\partial u(t)}{\partial t}\varphi \mathrm{d}x + \sum_{K\in\mathcal{T}_{h,m}} \int_{\partial K} \boldsymbol{f}(u(t))\cdot\boldsymbol{n}\varphi \mathrm{d}S - \sum_{K\in\mathcal{T}_{h,m}} \int_K \boldsymbol{f}(u(t))\cdot\nabla\varphi \mathrm{d}x \\ &+ \sum_{K\in\mathcal{T}_{h,m}} \varepsilon \int_K \nabla u(t)\cdot\nabla\varphi \mathrm{d}x - \sum_{K\in\mathcal{T}_{h,m}} \varepsilon \int_{\partial K} (\nabla u(t)\cdot\boldsymbol{n})\varphi \mathrm{d}S = \int_\Omega g\varphi \mathrm{d}x. \end{aligned} \tag{2.7}$$

Here $\boldsymbol{n}$ denotes the unit outer normal to ∂K. The surface integrals over ∂K make sense due to the regularity of u. We split them according to the type of faces Γ that form the boundaries of the elements $K \in \mathcal{T}_{h,m}$:

$$\begin{aligned} \sum_{K\in\mathcal{T}_{h,m}} \int_{\partial K} (\boldsymbol{n}\cdot\nabla u)\varphi \mathrm{d}S &= \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma (\boldsymbol{n}_\Gamma\cdot\nabla u)\varphi \mathrm{d}S \\ &+ \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \boldsymbol{n}_\Gamma\cdot\left((\nabla u|_\Gamma^L)\varphi|_\Gamma^L - (\nabla u|_\Gamma^R)\varphi|_\Gamma^R\right)\mathrm{d}S. \end{aligned} \tag{2.8}$$

Due to the assumption that $u \in L^2(0,T;H^2(\Omega))$,

$$[u] = 0 = [\nabla u], \quad \nabla u|_\Gamma^L = \nabla u|_\Gamma^R = \langle\nabla u\rangle, \quad \Gamma \in \mathcal{F}_{h,m}^I. \tag{2.9}$$

Thus, the integrand of the second integral in (2.8) can be written in the form

$$\boldsymbol{n}\cdot(\nabla u|_\Gamma^L)\varphi|_\Gamma^L - \boldsymbol{n}\cdot(\nabla u|_\Gamma^R)\varphi|_\Gamma^R = \boldsymbol{n}\cdot\langle\nabla u\rangle[\varphi]. \tag{2.10}$$

Due to above relations we have

$$\sum_{K\in\mathcal{T}_{h,m}}\int_K \frac{\partial u(t)}{\partial t}\varphi \mathrm{x} + \sum_{K\in\mathcal{T}_{h,m}}\int_{\partial K} \boldsymbol{f}(u(t))\cdot\boldsymbol{n}\varphi \mathrm{S} - \sum_{K\in\mathcal{T}_{h,m}}\int_K \boldsymbol{f}(u(t))\cdot\nabla\varphi \mathrm{d}x$$
$$+\varepsilon\sum_{K\in\mathcal{T}_{h,m}}\int_K \nabla u(t)\cdot\nabla\varphi \mathrm{d}x - \varepsilon\sum_{\Gamma\in\mathcal{F}^I_{h,m}}\int_{\partial K}\boldsymbol{n}\cdot\langle\nabla u(t)\rangle\,[\varphi]\,\mathrm{d}S$$
$$-\varepsilon\sum_{\Gamma\in\mathcal{F}^B_{h,m}}\int_{\partial K}\boldsymbol{n}\cdot\nabla u(t)\varphi \mathrm{d}S = \int_\Omega g\varphi \mathrm{d}x. \tag{2.11}$$

If $u\in L^2(0,T;H^2(\Omega))\cap L^2(0,T;H^2(\Omega,\mathcal{T}_{h,m}))$ and u satisfies the Dirichlet boundary condition, then

$$\sum_{\Gamma\in\mathcal{F}_{h,m}}\int_\Gamma \boldsymbol{n}\cdot\langle\nabla\varphi\rangle\,[u]\,\mathrm{d}S = \sum_{\Gamma\in\mathcal{F}^B_{h,m}}\int_\Gamma \boldsymbol{n}\cdot\nabla\varphi u_D \mathrm{d}S \quad \forall\varphi\in H^2(\Omega,\mathcal{T}_{h,m}) \tag{2.12}$$

and

$$\sum_{\Gamma\in\mathcal{F}_{h,m}}\int_\Gamma h(\Gamma)^{-1}\int_\Gamma [u]\,[\varphi]\,\mathrm{d}S = \sum_{\Gamma\in\mathcal{F}^B_{h,m}} h(\Gamma)^{-1}\int_\Gamma u\varphi \mathrm{d}S \quad \forall\varphi\in H^2(\Omega,\mathcal{T}_{h,m}), \tag{2.13}$$

since $[u]=0$ for $\Gamma\in\mathcal{F}^I_{h,m}$ and $[u]=u|_\Gamma=u_D$ for $\Gamma\in\mathcal{F}^B_{h,m}$. We called $\sum_{\Gamma\in\mathcal{F}_{h,m}}\int_\Gamma h(\Gamma)^{-1}\int_\Gamma [u]\,[\varphi]\,\mathrm{d}S$ and $\sum_{\Gamma\in\mathcal{F}^B_{h,m}} h(\Gamma)^{-1}\int_\Gamma u\varphi \mathrm{d}S \quad \forall\varphi\in H^2(\Omega,\mathcal{T}_{h,m})$ *interior* and *boundary penalty.*

Then we approximate fluxes through the faces Γ with the aid of numerical flux $\boldsymbol{H}=\boldsymbol{H}(u,\varphi,\boldsymbol{n})$ in the form

$$\int_\Gamma \boldsymbol{f}(u(t))\cdot\boldsymbol{n}\varphi \mathrm{S} \approx \int_\Gamma \boldsymbol{H}(u|^L_\Gamma, u|^R_\Gamma,\boldsymbol{n})\varphi \mathrm{d}S. \tag{2.14}$$

If we apply the described technique and sum (2.11) with θ-multiple ($\theta=-1,\ 0$ or 1) of (2.12), we define the forms for $u,\varphi\in H^1(\Omega,\mathcal{T}_{h,m}),\ u\in L^\infty(Q_T)$

$$\begin{aligned} a_{h,m}(u,\varphi) &= \varepsilon\sum_{K\in\mathcal{T}_{h,m}}\int_K \nabla u\cdot\nabla\varphi \mathrm{d}x \qquad (2.15)\\ &\quad -\varepsilon\sum_{\Gamma\in\mathcal{F}^I_{h,m}}\int_\Gamma \left(\langle\nabla u\rangle\cdot\boldsymbol{n}_\Gamma\,[\varphi]+\theta\,\langle\nabla\varphi\rangle\cdot\boldsymbol{n}_\Gamma\,[u]\right)\mathrm{d}S\\ &\quad -\varepsilon\sum_{\Gamma\in\mathcal{F}^B_{h,m}}\int_\Gamma \left(\nabla u\cdot\boldsymbol{n}_\Gamma\varphi+\theta\nabla\varphi\cdot\boldsymbol{n}_\Gamma u\right)\mathrm{d}S,\\ J_{h,m}(u,\varphi) &= \sum_{\Gamma\in\mathcal{F}^I_{h,m}} h(\Gamma)^{-1}\int_\Gamma [u]\,[\varphi]\,\mathrm{d}S + \sum_{\Gamma\in\mathcal{F}^B_{h,m}} h(\Gamma)^{-1}\int_\Gamma u\varphi \mathrm{d}S, \qquad (2.16)\\ A_{h,m} &= a_{h,m}+\varepsilon J_{h,m}, \qquad (2.17) \end{aligned}$$

$$
\begin{aligned}
b_{h,m}(u,\varphi) \;=\; & -\sum_{K\in\mathcal{T}_{h,m}}\int_K \sum_{s=1}^{d} f_s(u)\frac{\partial\varphi}{\partial x_s}\mathrm{d}x \\
& +\sum_{\Gamma\in\mathcal{F}^I_{h,m}}\int_\Gamma H\left(u|^L_\Gamma, u|^R_\Gamma, \boldsymbol{n}_\Gamma\right)[\varphi]_\Gamma\,\mathrm{d}S \\
& +\sum_{\Gamma\in\mathcal{F}^B_{h,m}}\int_\Gamma H\left(u|^L_\Gamma, u|^R_\Gamma, \boldsymbol{n}_\Gamma\right)\varphi|^L_\Gamma\mathrm{d}S.
\end{aligned}
\tag{2.18}
$$

By $(\cdot,\cdot)$ we denote the scalar product in $L^2(\Omega)$ and by $\|\cdot\|$ we denote the norm in $L^2(\Omega)$.

H is a numerical flux. We assume that it has the following properties.

(H1) $H(u,v,\boldsymbol{n})$ is defined in $I\!R^2\times B_1$, where $B_1=\boldsymbol{n}\in I\!R^d; |\boldsymbol{n}|=1$, and is *Lipschitz-continuous* with respect to u,v :

$$
|H(u,v,\boldsymbol{n})-H(u^*,v^*,\boldsymbol{n})|\le L_H\left(|u-u^*|+|v-v^*|\right),\quad u,v,u^*,v^*\in I\!R,\ \boldsymbol{n}\in B_1.
$$

(H2) $H(u,v,\boldsymbol{n})$ is *consistent*:

$$
H(u,v,\boldsymbol{n})=\sum_{s=1}^{d} f_s(u)n_s,\quad u\in I\!R,\ \boldsymbol{n}=(n_1,\ldots,n_d)\in B_1.
$$

(H3) $H(u,v,\boldsymbol{n})$ is *conservative*:

$$
H(u,v,\boldsymbol{n})=-H(u,v,-\boldsymbol{n}),\quad u,v\in I\!R,\ \boldsymbol{n}\in B_1.
$$

Finally, the right-hand side form is defined on the basis of data:

$$
l_{h,m}(\varphi)=(g,\varphi)+\varepsilon\sum_{\Gamma\in\mathcal{F}^B_{h,m}}\left(h(\Gamma)^{-1}\int_\Gamma u_D\varphi\mathrm{d}S-\theta\int_\Gamma\nabla\varphi\cdot\boldsymbol{n}_\Gamma u_D\mathrm{d}S\right).
\tag{2.19}
$$

In the above forms we take $\theta=-1,\ \theta=0,\ \theta=1$ and we obtain the nonsymmetric (NIPG), incomplete (IIPG) and symmetric (SIPG) variants of the approximation of the diffusion terms.

Base on it we will use the following norm in space $H^1(\Omega,\mathcal{T}_{h,m})$:

$$
\|\varphi\|_{DG,m}=\left(\sum_{K\in\mathcal{T}_{h,m}}|\varphi|^2_{H^1(K)}+J_{h,m}(\varphi,\varphi)\right)^{1/2}.
\tag{2.20}
$$

2.3 Discrete problem

Let $p,q\ge 1$ be integers. For each $m=1,\ldots,M$ we define the finite-dimensional space

$$
S^p_{h,m}=\left\{\varphi\in L^2(\Omega);\varphi|_K\in P^p(K)\ \forall K\in\mathcal{T}_{h,m}\right\}.
\tag{2.21}
$$

By Π_m we denote the $L^2(\Omega)-$projection on $S_{h,m}^p$, i.e., if $\varphi \in L^2(\Omega)$, then $\Pi_m\varphi \in S_{h,m}^p$ and

$$(\Pi_m\varphi - \varphi, \psi) = 0, \quad \forall\psi \ S_{h,m}^p. \tag{2.22}$$

The approximate solution will be sought in space

$$S_{h,\tau}^{p,q} = \left\{ \varphi \in L^2(Q_T); \varphi|_{I_m} = \sum_{i=0}^{q} t^i \varphi_i \text{ with } \varphi_i \in S_{h,m}^p, m = 1, \ldots, M \right\}. \tag{2.23}$$

In what follows we shall use the notation $U' = \partial U/\partial t$, $u' = \partial u/\partial t$, $D^{q+1} = \partial^{q+1}/\partial t^{q+1}$.

Definition 2.1: *We say that the function U is an approximate solution of problem (1.1)-(1.3), if $U \in S_{h,\tau}^{p,q}$ and*

$$\begin{aligned} &\int_{I_m} \left((U', \varphi) + A_{h,m}(U, \varphi) + b_{h,m}(U, \varphi) \right) \mathrm{d}t + \left(\{U\}_{m-1}, \varphi_{m-1}^+ \right) \\ &\quad = \int_{I_m} l_{h,m}(\varphi) \mathrm{d}t, \quad \forall \varphi \in S_{h,\tau}^{p,q}, \ \forall m = 1, \ldots, M, \\ &\quad (U_0^-, \varphi) = (u^0, \varphi), \quad \forall \varphi \in S_{h,1}^p. \end{aligned} \tag{2.24}$$

It is possible to see that the exact strong solution u satisfies the identity

$$\begin{aligned} &\int_{I_m} \left((u', \varphi) + A_{h,m}(u, \varphi) + b_{h,m}(u, \varphi) \right) \mathrm{d}t + \left(\{u\}_{m-1}, \varphi_{m-1}^+ \right) \\ &\quad = \int_{I_m} l_{h,m}(\varphi) \mathrm{d}t, \quad \forall \varphi \in S_{h,\tau}^{p,q}, \ \forall m = 1, \ldots, M, \end{aligned} \tag{2.25}$$

if we set $u(0-) = u(0)$.

Remark 2.1: *It is also possible to consider $q = 0$. In this case, scheme (2.24) represents a version of the backward Euler method. Since it can be analyzed in a similar way as, for example, in [24], we shall be concerned only with $q \geq 1$.*

In the error analysis we shall use the $S_{h,\tau}^{p,q}-$interpolation π of function $v \in H^1(0, T; L^2(\Omega))$ defined by

$$\begin{aligned} &\text{(a)} \quad \pi v \in S_{m,\tau}^{p,q}, \\ &\text{(b)} \quad (\pi v)(t_m-) = \Pi_m v(t_m-), \\ &\text{(c)} \quad \int_{I_m} (\pi v - v, \varphi^*) \mathrm{d}t = 0, \quad \forall \varphi^* \in S_{h,\tau}^{p,q-1}, \ \forall m = 1, \ldots, M. \end{aligned} \tag{2.26}$$

In [32], Lemma 4, it was proven that πu is uniquely determined. Moreover, by [32], Lemma 9,

$$\pi u|_{I_m} = \pi(\Pi_m u)|_{I_m}. \tag{2.27}$$

The main goal is the derivation of the estimation of the error $e = U - u$, which can be expressed in the form

$$e = \xi + \eta, \tag{2.28}$$

where

$$\begin{aligned} \xi &= U - \pi u, \quad \xi \in S_{h,\tau}^{p,q}, \\ \eta &= \pi u - u. \end{aligned} \tag{2.29}$$

As we see, the function η is the error of the interpolation of the exact solution u. Then, in virtue of (2.24) and (2.25),

$$\int_{I_m} \left((\xi', \varphi) + A_{h,m}(\xi, \varphi) \right) \mathrm{d}t + \left(\{\xi\}_{m-1}, \varphi_{m-1}^{+} \right) = \int_{I_m} \left(b_{h,m}(u, \varphi) - b_{h,m}(U, \varphi) \right) \mathrm{d}t$$
$$- \int_{I_m} \left((\eta', \varphi) + A_{h,m}(\eta, \varphi) \right) \mathrm{d}t - \left(\{\eta\}_{m-1}, \varphi_{m-1}^{+} \right), \quad \forall \varphi \in S_{h,\tau}^{p,q}. \tag{2.30}$$

Chapter 3

Derivation of an abstract error estimate

This chapter will be devoted to the derivation of estimates of the function ξ in terms of the interpolation error η.

3.1 Assumption on the triangulation

In our further considerations, by C and c we shall denote positive generic constants, independent of $h, \tau,\ \varepsilon,\ u,\ U$, which can attain different values in different places. In the sequel, we shall consider a system of triangulations $\mathcal{T}_{h,m},\ m = 1, \ldots, M,\ h \in (0, h_0),\ h_0 > 0$, which is *shape regular* and *locally quasiuniform*: there exist constants $C_R,\ C_Q > 0$ such that

$$\frac{h_K}{\rho_K} \leq C_R, \quad K \in \mathcal{T}_{h,m},\ m = 1, \ldots, M,\ h \in (0, h_0), \tag{3.1}$$

$$h_K \leq C_Q h_{K'}, \quad \text{for neighbouring elements } K, K' \in \mathcal{T}_{h,m}. \tag{3.2}$$

Then there exist positive constants C_-, C_+ such that

$$C_- h_K \leq h(\Gamma) \leq C_+ h_{K'}, \quad \Gamma \in \mathcal{F}_{h,m},\ \Gamma \subset K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \tag{3.3}$$

3.2 Auxiliary results

In the analysis of the discontinuous Galerkin finite element method (DGFEM) we use the following important tools.

Multiplicative trace inequality: There exists a constant $C_M > 0$ independent of $v,\ h,\ K$ and M such that

$$\begin{aligned} &\|v\|^2_{L^2(\partial K)} \leq C_M \left(\|v\|_{L^2(K)} |v|_{H^1(K)} + h_K^{-1} \|v\|^2_{L^2(K)} \right), \\ &v \in H^1(K),\ K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \end{aligned} \tag{3.4}$$

Inverse inequality: There exists a constant $C_I > 0$ independent of $v,\ h,\ K$ and M such that

$$|v|_{H^1(K)} \leq C_I h_K^{-1} \|v\|_{L^2(K)}, \quad v \in P^p(K),\ K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \tag{3.5}$$

(For proofs, see, e.g. [16], [25].)

Let us remind two widely used inequalities:
Young's inequality: For arbitrary $a,\ b,\ c > 0$ we have

$$ab \leq 2ab \leq ca^2 + \frac{b^2}{c}. \tag{3.6}$$

Discrete Cauchy inequality: For arbitrary $a_i,\ b_i \in I\!R,\ i = 1, \ldots, n$ it holds

$$\sum_{i=1}^{n} a_i b_i \leq \sqrt{\sum_{i=1}^{n} a_i^2} \sqrt{\sum_{i=1}^{n} b_i^2}. \tag{3.7}$$

Coercivity of the form $A_{h,m}$: It holds

$$A_{h,m}(\xi, \xi) \geq \frac{\varepsilon}{2} \|\xi\|_{DG,m}^2 \tag{3.8}$$

provided

$$\begin{aligned} &C_W > 0 \quad \text{for NIPG}, \\ &C_W \geq C_M(1 + C_I)(1 + C_Q) \quad \text{for IIPG}, \\ &C_W \geq 2C_M(1 + C_I)(1 + C_Q) \quad \text{for SIPG}. \end{aligned} \tag{3.9}$$

(See, [30].)
Consistency of $b_{h,m}$: For any $\varphi \in S_{h,\tau}^{p,q}$ and $k > 0$,

$$\begin{aligned} |b_{h,m}(u, \varphi) - b_{h,m}(U, \varphi)| &\leq C \|\varphi\|_{DG,m} \left(\|\xi\|^2 + \tilde{\sigma}_m^2(\eta) \right)^{1/2} \\ &\leq \frac{\varepsilon}{k} \|\varphi\|_{DG,m}^2 + \frac{C_k}{\varepsilon} \left(\|\xi\|^2 + \tilde{\sigma}_m^2(\eta) \right), \end{aligned} \tag{3.10}$$

where

$$\tilde{\sigma}_m^2(\eta) = \sum_{K \in \mathcal{T}_{h,m}} \left(\|\eta\|_{L^2(K)}^2 + h_K^2 |\eta|_{H^1(K)}^2 \right). \tag{3.11}$$

It should be mention that the constant C_k in the last expression depends on k. The proof can be carried out in a similar way as in [23] or [26].

At the end of this section we add some useful lemmas that will be used in following sections.

Lemma 3.1: *There exists constant $C_{MI} > 0$ independent of $v,\ h,\ K$ such that it holds*

$$h_K \|v\|_{L^2(\partial K)}^2 \leq C_{MI} \|v\|_{L^2(K)}^2 \quad v \in P^p(K),\ K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \tag{3.12}$$

Proof: Starting by (3.4) and using the inverse inequality (3.5) for estimation of $|v|_{H^1(K)}$ we get

$$\begin{aligned} \|v\|^2_{L^2(\partial K)} &\leq C_M \left(\|v\|_{L^2(K)} C_I h_K^{-1} \|v\|_{L^2(K)} + h_K^{-1} \|v\|^2_{L^2(K)} \right) \\ &= h_K^{-1} C_M (C_I + 1) \|v\|^2_{L^2(K)} . \end{aligned} \tag{3.13}$$

By setting $C_{MI} = C_M(C_I + 1)$ we obtain (3.12).

■

Lemma 3.2: *There exists a constant $C_O > 0$ independent of $v,\ h,\ K$ such that*

$$\begin{gathered} h_K \|v\|^2_{L^2(\partial K)} \leq C_O \left(\|v\|^2_{L^2(K)} + k_K^2 |v|^2_{H^1(K)} \right) \\ v \in H^1(K),\ K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \end{gathered} \tag{3.14}$$

Proof: Starting from (3.4) and using the Young inequality we obtain

$$\begin{aligned} h_K \|v\|^2_{L^2(\partial K)} &\leq C_M \left(\|v\|_{L^2(K)} h_K |v|_{H^1(K)} + \|v\|^2_{L^2(K)} \right) \\ &\leq C_M \left(\|v\|^2_{L^2(K)} + h_K^2 |v|^2_{H^1(K)} + \|v\|^2_{L^2(K)} \right) \\ &\leq C_O \left(\|v\|^2_{L^2(K)} + h_K^2 |v|^2_{H^1(K)} \right), \end{aligned}$$

where $C_O = 2C_M$.

■

Lemma 3.3: *There exists a constant $C_N > 0$ independent on $v,\ h,\ K$ such that*

$$\begin{gathered} h_K |\nabla v|^2_{L^2(\partial K)} \leq C_N \left(|v|^2_{H^1(K)} + h_K^2 |v|^2_{H^2(K)} \right) \\ v \in H^2(K),\ K \in \mathcal{T}_{h,m},\ h \in (0, h_0),\ m = 1, \ldots, M. \end{gathered} \tag{3.15}$$

Proof: The proof follows from Lemma 3.2.

■

3.3 Derivation of estimates for ξ

Let us substitute $\varphi := \xi$ in (2.30) and analyze individual terms. A calculation yields

$$\begin{aligned} &2 \int_{I_m} (\xi', \xi) \, \mathrm{d}t + 2 \left(\{\xi\}_{m-1}, \xi^+_{m-1} \right) = \int_{I_m} \frac{\mathrm{d}}{\mathrm{d}t} \|\xi\|^2 \, \mathrm{d}t + 2 \left(\{\xi\}_{m-1}, \xi^+_{m-1} \right) \\ &= \left\| \xi^-_m \right\|^2 - \left\| \xi^+_{m-1} \right\|^2 + 2 \left(\{\xi\}_{m-1}, \xi^+_{m-1} \right) \end{aligned} \tag{3.16}$$

and

$$
\begin{aligned}
&2\left(\{\xi\}_{m-1},\xi_{m-1}^{+}\right)\\
&\quad=\left(\{\xi\}_{m-1},\xi_{m-1}^{+}\right)+\left(\{\xi\}_{m-1},\xi_{m-1}^{+}\right)\\
&\quad=\left\|\xi_{m-1}^{+}\right\|^{2}-\left(\xi_{m-1}^{-},\xi_{m-1}^{+}\right)+\left(\xi_{m-1}^{+}-\xi_{m-1}^{-},\xi_{m-1}^{+}-\xi_{m-1}^{-}\right)+\left(\xi_{m-1}^{+}-\xi_{m-1}^{-},\xi_{m-1}^{-}\right)\\
&\quad=\left\|\xi_{m-1}^{+}\right\|^{2}+\left\|\{\xi\}_{m-1}\right\|^{2}-\left\|\xi_{m-1}^{-}\right\|^{2}-\left(\xi_{m-1}^{-},\xi_{m-1}^{+}\right)+\left(\xi_{m-1}^{+},\xi_{m-1}^{-}\right)\\
&\quad=\left\|\xi_{m-1}^{+}\right\|^{2}+\left\|\{\xi\}_{m-1}\right\|^{2}-\left\|\xi_{m-1}^{-}\right\|^{2}. \qquad (3.17)
\end{aligned}
$$

Hence,

$$
2\int_{I_m}(\xi',\xi)\,\mathrm{d}t+2\left(\{\xi\}_{m-1},\xi_{m-1}^{+}\right)=\left\|\xi_{m}^{-}\right\|^{2}-\left\|\xi_{m-1}^{-}\right\|^{2}+\left\|\{\xi\}_{m-1}\right\|^{2}. \qquad (3.18)
$$

Further, we shall be concerned with estimates of the right-hand side of (2.30).

Lemma 3.4: *Let us have arbitrary functions $u,v\in H^1(\Omega,\mathcal{T}_{h,m})$. Then*

$$
J_{h,m}(u,v)\le\left(J_{h,m}(u,u)\right)^{1/2}\left(J_{h,m}(v,v)\right)^{1/2}. \qquad (3.19)
$$

Proof: Using the definition of the form $J_{h,m}$ and both, integral and discret, Cauchy inequalities, we obtain

$$
\begin{aligned}
J_{h,m}(u,v)= &\sum_{\Gamma\in\mathcal{F}_{h,m}^{I}}h(\Gamma)^{-1}\int_{\Gamma}[u]\,[\varphi]\,\mathrm{d}S+\sum_{\Gamma\in\mathcal{F}_{h,m}^{B}}h(\Gamma)^{-1}\int_{\Gamma}u\varphi\mathrm{d}S\\
\le &\sum_{\Gamma\in\mathcal{F}_{h,m}^{I}}h(\Gamma)^{-1}\left(\int_{\Gamma}[u]^{2}\,\mathrm{d}S\right)^{1/2}\left(\int_{\Gamma}[v]^{2}\,\mathrm{d}S\right)^{1/2}\\
&+\sum_{\Gamma\in\mathcal{F}_{h,m}^{B}}h(\Gamma)^{-1}\left(\int_{\Gamma}u^{2}\mathrm{d}S\right)^{1/2}\left(\int_{\Gamma}v^{2}\mathrm{d}S\right)^{1/2}\\
\le &\left(\sum_{\Gamma\in\mathcal{F}_{h,m}^{I}}h(\Gamma)^{-1}\int_{\Gamma}[u]^{2}\,\mathrm{d}S+\sum_{\Gamma\in\mathcal{F}_{h,m}^{B}}h(\Gamma)^{-1}\int_{\Gamma}u^{2}\mathrm{d}S\right)^{1/2}\\
&\times\left(\sum_{\Gamma\in\mathcal{F}_{h,m}^{I}}h(\Gamma)^{-1}\int_{\Gamma}[v]^{2}\,\mathrm{d}S+\sum_{\Gamma\in\mathcal{F}_{h,m}^{B}}h(\Gamma)^{-1}\int_{\Gamma}v^{2}\mathrm{d}S\right)^{1/2}\\
= &\left(J_{h,m}(u,u)\right)^{1/2}\left(J_{h,m}(v,v)\right)^{1/2}. \qquad (3.20)
\end{aligned}
$$

■

Lemma 3.5: *For an arbitrary $\tilde{k}>0$ there exists a constant $\tilde{C}>0$ independent of U,ξ,φ,h such that for an arbitrary $\varphi\in S_{h,m}^{p}$ we get an estimate*

$$
a_{h,m}(\eta,\varphi)\le\varepsilon\left(\frac{1}{\tilde{k}}\left\|\varphi\right\|_{DG,m}^{2}+\tilde{C}\sigma_{m}^{2}(\eta)\right), \qquad (3.21)
$$

where

$$\sigma_m^2(\eta) = \|\eta\|_{DG,m}^2 + \sum_{K\in\mathcal{T}_{h,m}} h_K^2 \,|\eta|_{H^2(K)}^2 \,. \tag{3.22}$$

Proof: The definition of the form a_h implies that

$$\begin{aligned}
&a_{h,m}(\eta,\varphi)\\
&= \varepsilon \sum_{K\in\mathcal{T}_{h,m}} \int_K \nabla\eta\cdot\nabla\varphi \mathrm{d}x \\
&\quad -\varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \left(\langle\nabla\eta\rangle\cdot\boldsymbol{n}_\Gamma\,[\varphi] + \theta\,\langle\nabla\varphi\rangle\cdot\boldsymbol{n}_\Gamma\,[\eta]\right)\mathrm{d}S \\
&\quad -\varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma \left(\nabla\eta\cdot\boldsymbol{n}_\Gamma\varphi + \theta\nabla\varphi\cdot\boldsymbol{n}_\Gamma\eta\right)\mathrm{d}S, \\
&\leq \varepsilon \sum_{K\in\mathcal{T}_{h,m}} \int_K |\nabla\eta\cdot\nabla\varphi|\,\mathrm{d}x \\
&\quad +\varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \frac{\left|\nabla\eta|_\Gamma^L\right| + \left|\nabla\eta|_\Gamma^R\right|}{2}\,|[\varphi]|\,\mathrm{d}S \\
&\quad +\varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \frac{\left|\nabla\varphi|_\Gamma^L\right| + \left|\nabla\varphi|_\Gamma^R\right|}{2}\,|[\eta]|\,\mathrm{d}S \\
&\quad +\varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma |\nabla\eta|\,|\varphi|\,\mathrm{d}S + \varepsilon \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma |\nabla\varphi|\,|\eta|\,\mathrm{d}S.
\end{aligned} \tag{3.23}$$

If we choose arbitrarily $\delta_1, \delta_2 > 0$, then by the application of the Young inequality and inequality (3.3) we obtain

$$\begin{aligned}
&\frac{a_{h,m}(\eta,\varphi)}{\varepsilon} \\
&\leq \sum_{K\in\mathcal{T}_{h,m}} \int_K \left(\frac{|\nabla\eta|^2}{\delta_1} + \delta_1\,|\nabla\varphi|^2\right)\mathrm{d}x \\
&\quad + \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \left(\frac{h(\Gamma)}{\delta_1}\left(\left|\nabla\eta|_\Gamma^L\right|^2 + \left|\nabla\eta|_\Gamma^R\right|^2\right) + \frac{\delta_1}{h(\Gamma)}\,|[\varphi]|^2\right)\mathrm{d}S \\
&\quad + \sum_{\Gamma\in\mathcal{F}_{h,m}^I} \int_\Gamma \left(h(\Gamma)\delta_2\left(\left|\nabla\varphi|_\Gamma^L\right|^2 + \left|\nabla\varphi|_\Gamma^R\right|^2\right) + \frac{1}{h(\Gamma)\delta_2}\,|[\eta]|^2\right)\mathrm{d}S \\
&\quad + \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma \left(\frac{h(\Gamma)}{\delta_1}\,|\nabla\eta|^2 + \frac{\delta_1}{h(\Gamma)}\,|\varphi|^2\right)\mathrm{d}S \\
&\quad + \sum_{\Gamma\in\mathcal{F}_{h,m}^B} \int_\Gamma \left(h(\Gamma)\delta_2\,|\nabla\varphi|^2 + \frac{1}{h(\Gamma)\delta_2}\,|\eta|^2\right)\mathrm{d}S.
\end{aligned}$$

Using Lemmas 3.1 and 3.3 we have

$$\begin{aligned}
&\frac{a_{h,m}(\eta,\varphi)}{\varepsilon} \\
&\leq \delta_1 \|\varphi\|^2_{DG,m} + C_{MI}\delta_2 \sum_{K\in\mathcal{T}_{h,m}} |\varphi|^2_{H^1(K)} + \frac{1}{\delta_2} J_{h,m}(\eta,\eta) \\
&\quad + \frac{1}{\delta_1} \sum_{K\in\mathcal{T}_{h,m}} |\eta|^2_{H^1(K)} + \frac{C_N}{\delta_1} \sum_{K\in\mathcal{T}_{h,m}} \left(|\eta|^2_{H^1(K)} + h_K^2 |\eta|^2_{H^2(K)} \right) \\
&\leq \frac{1}{\tilde{k}} \|\varphi\|^2_{DG,m} + \tilde{C}\sigma^2_m(\eta),
\end{aligned}$$

where

$$\begin{aligned}
\delta_1 = \tfrac{1}{2\tilde{k}},\ \delta_2 = \tfrac{1}{2\tilde{k}C_{MI}}, \\
\tilde{C} = \max\left\{ \tfrac{1}{\delta_2}, \tfrac{1}{\delta_1} + \tfrac{C_N}{\delta_1} \right\}.
\end{aligned}$$

■

With the aid of Lemmas 3.4 and 3.5 we can also show that for $\varphi \in S^{p,q}_{h,\tau}$ and $k > 0$ we have

$$|A_{h,m}(\eta,\varphi)| \leq \frac{\varepsilon}{k} \|\varphi\|^2_{DG,m} + C\varepsilon\sigma^2_m(\eta), \tag{3.24}$$

where

$$\sigma^2_m(\eta) = \|\eta\|^2_{DG,m} + \sum_{K\in\mathcal{T}_{h,m}} h_K^2 |\eta|^2_{H^2(K)}. \tag{3.25}$$

Now (2.30), where we set $\varphi := \xi$, relation (3.18) and estimates (3.8), (3.10), (3.24) imply that

$$\begin{aligned}
&\left\|\xi^-_m\right\|^2 - \left\|\xi^-_{m-1}\right\|^2 + \left\|\{\xi\}_{m-1}\right\|^2 + \varepsilon \int_{I_m} \|\xi\|^2_{DG,m}\,\mathrm{d}t \\
&\leq -2\int_{I_m} (\eta',\xi)\,\mathrm{d}t - 2\left(\{\eta\}_{m-1}, \xi^+_{m-1}\right) + \frac{2\varepsilon}{k}\int_{I_m} \|\xi\|^2_{DG,m}\,\mathrm{d}t \\
&\quad + \frac{C}{\varepsilon}\int_{I_m} \|\xi\|^2\,\mathrm{d}t + C\int_{I_m} \left(\varepsilon\sigma^2_m(\eta) + \frac{1}{\varepsilon}\tilde{\sigma}^2_m(\eta) \right) \mathrm{d}t.
\end{aligned} \tag{3.26}$$

Further, we shall be concerned with the expression

$$\int_{I_m} (\eta',\xi)\,\mathrm{d}t + \left(\{\eta\}_{m-1}, \xi^+_{m-1}\right). \tag{3.27}$$

Integration by parts yields

$$\int_{I_m} (\eta',\xi)\,\mathrm{d}t = \left(\eta^-_m, \xi^-_m\right) - \left(\eta^+_{m-1}, \xi^+_{m-1}\right) - \int_{I_m} (\eta,\xi')\,\mathrm{d}t. \tag{3.28}$$

Since $\eta = \pi u - u$ and $\xi' \in S^{p,q-1}_{h,\tau}$, by the definition of π, we have

$$\int_{I_m} (\eta,\xi')\,\mathrm{d}t = 0. \tag{3.29}$$

Thus,

$$\int_{I_m} (\eta', \xi)\,\mathrm{d}t + \left(\{\eta\}_{m-1}, \xi_{m-1}^+\right) \tag{3.30}$$
$$= \left(\eta_m^-, \xi_m^-\right) - \left(\eta_{m-1}^+, \xi_{m-1}^+\right) + \left(\eta_{m-1}^+, \xi_{m-1}^+\right) - \left(\eta_{m-1}^-, \xi_{m-1}^+\right).$$

Further, since $\xi_m^- \in S_{h,m}^p$ and

$$\eta_m^- = (\pi u)(t_m^-) - u(t_m) = \Pi_m u(t_m) - u(t_m), \tag{3.31}$$

in view of the definition of Π_m and (2.26b),

$$\left(\eta_m^-, \xi_m^-\right) = 0. \tag{3.32}$$

Similarly, $\left(\eta_{m-1}^-, \xi_{m-1}^-\right) = 0$ and, hence, using also Young's inequality, we have

$$\begin{aligned} \left|\left(\eta_{m-1}^-, \xi_{m-1}^+\right)\right| &= \left|\left(\eta_{m-1}^-, \xi_{m-1}^+ - \xi_{m-1}^-\right)\right| = \left|\left(\eta_{m-1}^-, \{\xi\}_{m-1}\right)\right| \\ &\le \frac{1}{2}\left(\left\|\{\xi\}_{m-1}\right\|^2 + \left\|\eta_{m-1}^-\right\|^2\right). \end{aligned} \tag{3.33}$$

From (3.31)-(3.33) we find that

$$\left|\int_{I_m} (\eta', \xi)\,\mathrm{d}t + \left(\{\eta\}_{m-1}, \xi_{m-1}^+\right)\right| \le \frac{1}{2}\left\|\{\xi\}_{m-1}\right\|^2 + \frac{1}{2}\left\|\eta_{m-1}^-\right\|^2. \tag{3.34}$$

This and (3.26) imply that

$$\left\|\xi_m^-\right\|^2 - \left\|\xi_{m-1}^-\right\|^2 + \varepsilon\left(1 - \frac{2}{k}\right)\int_{I_m} \|\xi\|_{DG,m}^2\,\mathrm{d}t \tag{3.35}$$
$$\le \frac{C}{\varepsilon}\int_{I_m} \|\xi\|^2\,\mathrm{d}t + 2\left\|\eta_{m-1}^-\right\|^2 + C\int_{I_m} R_m(\eta)\,\mathrm{d}t,$$

where

$$R_m(\eta) = \varepsilon\sigma_m^2(\eta) + \frac{1}{\varepsilon}\tilde{\sigma}_m^2(\eta). \tag{3.36}$$

In what follows, it will be necessary to estimate the terms with η and $\int_{I_m} \|\xi\|^2\,\mathrm{d}t$.

3.4 Estimate of $\int_{I_m} \|\xi\|^2\,\mathrm{d}t$

For the estimation of the expression $\int_{I_m} \|\xi\|^2\,\mathrm{d}t$ we shall use the technique from [34], which was applied to the analysis of the time DG methods combined with the conforming finite element method.

By $\mathcal{P}^q$ we shall denote the set of all polynomials in $t \in I\!R$ of degree $\le q$. Then $\mathcal{P}^q(0,1)$ is the set of all polynomials in $t \in (0,1)$ of degree $\le q$. In the interval $(0,1]$ we shall consider the Gauss-Radau quadrature formula

$$\int_0^1 \varphi(t)\mathrm{d}t \approx \sum_{i=1}^{q+1} w_i \varphi(\vartheta_i), \tag{3.37}$$

where $0 < \vartheta_1 < \ldots < \vartheta_{q+1} = 1$ are the Radau integration points and $w_i > 0$ are the Radau weights. We can refer, for example, to formulas from [58] on page 131. The formulas need to be transformed from interval [-1,1) to (0,1]. This can be done by following considerations. First, we seek the function

$$\tilde{t} = \alpha(t) \in [-1, 1],$$

for which holds

$$\alpha(0) = 1 \text{ and } \alpha(1) = -1.$$

This gives us the function

$$\tilde{t} = -2t + 1.$$

Using the substitution theorem and the relation (3.37) we obtain

$$\int_0^1 \varphi(t)\mathrm{d}t = -\frac{1}{2}\int_1^{-1} \varphi(\frac{1-\tilde{t}}{2})\mathrm{d}\tilde{t} = \frac{1}{2}\int_{-1}^1 \varphi(\frac{1-\tilde{t}}{2})\mathrm{d}\tilde{t} \approx \frac{1}{2}\sum_{i=1}^{q+1} \tilde{w}_i\varphi(\frac{1-\tilde{\vartheta}_i}{2}) = \sum_{i=1}^{q+1} w_i\varphi(\vartheta_i).$$

For $\vartheta_i \in (0, 1]$ and $\tilde{\vartheta}_i \in [-1, 1)$ we have

$$\vartheta_i = \frac{1-\tilde{\vartheta}_i}{2},$$
$$w_i = \frac{\tilde{w}_i}{2}.$$

Then the formula (3.37) is transformed to the interval $(t_{m-1}, t_m]$, which yields

$$\int_{I_m} \varphi\left(t\right) \mathrm{d}t \approx \tau_m \sum_{i=1}^{q+1} w_i \varphi\left(t^{m,i}\right), \tag{3.38}$$

where $t^{m,i} = t_{m-1} + \tau_m\vartheta_i$. Formulas (3.37) and (3.38) are exact for polynomials of degree $\leq 2q$.

Lemma 3.6: *Let $p \in \mathcal{P}^q$ and let $\tilde{p} \in \mathcal{P}^q$ be the Lagrange interpolation of the function $p(t)/t$ at the points $\vartheta_i,\ i = 1, \ldots, q+1$. Then*

$$\int_0^1 p'\tilde{p}\mathrm{d}t + p(0)\tilde{p}(0) = \frac{1}{2}\left(p^2(1) + \sum_{i=1}^{q+1} w_i\vartheta_i^{-2}p^2(\vartheta_i)\right). \tag{3.39}$$

Proof: Let $v \in \mathcal{P}^{q-1}(0, 1)$ be given by

$$p(t) = p(0) + tv(t). \tag{3.40}$$

According to the definition of $\tilde{p}$, we can write

$$\tilde{p}(t) = v(t) + p(0)\Lambda(t), \tag{3.41}$$

where $\Lambda \in \mathcal{P}^q(0, 1)$ is the interpolant of $1/t$ at the Radau points $\vartheta_i,\ i = 1, \ldots, q+1$. Obviously the degree of $\tilde{p}$ given by (3.41) is $\leq q$. Moreover, by (3.40) and (3.41) we have

$$\tilde{p}(\vartheta_i) = v(\vartheta_i) + p(0)\vartheta_i^{-1} \text{ and } v(\vartheta_i) = \frac{p(\vartheta_i) - p(0)}{\vartheta_i}.$$

Hence, we see that $\tilde{p}(\vartheta_i) = p(\vartheta_i)/\vartheta_i$ for $i = 1, \dots, q+1$. This means that $\tilde{p}$ given by (3.41) is the Lagrange interpolation of $p(t)/t$. Hence, $\Lambda(1) = 1$.

Now we get

$$\begin{aligned}\int_0^1 p'\tilde{p}\mathrm{d}t &= \int_0^1 \left(v(t) + tv'(t)\right)\left(v(t) + p(0)\Lambda(t)\right)\mathrm{d}t \qquad (3.42)\\ &= \int_0^1 v^2(t)\mathrm{d}t + p(0)\int_0^1 v(t)\Lambda(t)\mathrm{d}t + \int_0^1 tv'(t)v(t)\mathrm{d}t + p(0)\int_0^1 tv'(t)\Lambda(t)\mathrm{d}t.\end{aligned}$$

Integration by parts yields

$$\int_0^1 tv'(t)v(t)\mathrm{d}t = \frac{1}{2}\int_0^1 t\frac{\mathrm{d}}{\mathrm{d}t}\left(v^2(t)\right)\mathrm{d}t = \frac{1}{2}v^2(1) - \frac{1}{2}\int_2^1 v^2(t)\mathrm{d}t \qquad (3.43)$$

and

$$\int_0^1 t\Lambda'(t)\Lambda(t)\mathrm{d}t = \frac{1}{2}\int_0^1 t\frac{\mathrm{d}}{\mathrm{d}t}\left(\Lambda^2(t)\right)\mathrm{d}t = \frac{1}{2}\Lambda^2(1) - \frac{1}{2}\int_0^1 \Lambda^2(t)\mathrm{d}t = \frac{1}{2} - \frac{1}{2}\int_0^1 \Lambda^2(t)\mathrm{d}t. \qquad (3.44)$$

Hence, in view of (3.42) and (3.43),

$$\int_0^1 p'\tilde{p}\mathrm{d}t = \frac{1}{2}\int_0^1 v^2(t)\mathrm{d}t + \frac{1}{2}v^2(1) + p(0)\left(\int_0^1 tv'(t)\Lambda(t)\mathrm{d}t + \int_0^1 v(t)\Lambda(t)\mathrm{d}t\right). \qquad (3.45)$$

Let $s \in \mathcal{P}^q(0,1)$. Taking into account that $\Lambda(t), ts'(t) \in \mathcal{P}^q(0,1)$, $\Lambda(\vartheta_i) = \vartheta_i^{-1}$, $i = 1, \dots, q+1$, and the integration formula (3.37) in exact for polynomials of degree $\leq 2q$, we get

$$\begin{aligned}\int_0^1 ts'(t)\Lambda(t)\mathrm{d}t &= \sum_{i=1}^{q+1} w_i\vartheta_i s'(\vartheta_i)\vartheta_i^{-1} = \sum_{i=1}^{q+1} w_i s'(\vartheta_i)\\ &= \int_0^1 s'(t)\mathrm{d}t = s(1) - s(0). \qquad (3.46)\end{aligned}$$

Thus,

$$\int_0^1 tv'(t)\Lambda(t)\mathrm{d}t = v(1) - v(0) \qquad (3.47)$$

and

$$\int_0^1 t\Lambda'(t)\Lambda(t)\mathrm{d}t = 1 - \Lambda(0). \qquad (3.48)$$

Further, we find that

$$\int_0^1 v(t)\Lambda(t)\mathrm{d}t = \sum_{i=1}^{q+1} w_i\vartheta_i^{-1}v(\vartheta_i), \qquad (3.49)$$

$$\int_0^1 v^2(t)\mathrm{d}t = \sum_{i=1}^{q+1} w_i v^2(\vartheta_i) \qquad (3.50)$$

and

$$\int_0^1 \Lambda^2(t)\mathrm{d}t = \sum_{i=1}^{q+1} w_i \vartheta_i^{-2}. \tag{3.51}$$

Taking into account (3.44) and (3.48), we get

$$\Lambda(0) = 1 - \int_0^1 t\Lambda'(t)\Lambda(t)\mathrm{d}t = \frac{1}{2} + \frac{1}{2}\int_0^1 \Lambda^2(t)\mathrm{d}t = \frac{1}{2}\left(1 + \sum_{i=1}^{q+1} w_i \vartheta_i^{-2}\right). \tag{3.52}$$

Further, we have

$$\begin{aligned} v(1) &= p(1) - p(0), \\ v(\vartheta_i) + p(0)\vartheta_i^{-1} &= p(\vartheta_i)\vartheta_i^{-1}, \\ p(0)\tilde{p}(0) &= p(0)v(0) + p^2(0)\Lambda(0), \end{aligned} \tag{3.53}$$

as follows from (3.40) and (3.41).

Now, (3.43), (3.45), (3.47), (3.49), (3.50), (3.52), (3.53) and tedious calculation yield

$$\begin{aligned} \int_0^1 p'\tilde{p}\mathrm{d}t + p(0)\tilde{p}(0) &= \frac{1}{2}p^2(1) + \frac{1}{2}\sum_{i=1}^{q+1} w_i \left(v(\vartheta_i) + p(0)\vartheta_i^{-1}\right)^2 \\ &= \frac{1}{2}\left(p^2(1) + \sum_{i=1}^{q+1} w_i \vartheta_i^{-2} p^2(\vartheta_i)\right), \end{aligned} \tag{3.54}$$

which we wanted to prove.

■

Lemma 3.7: *Let $p \in \mathcal{P}^q$ and let $\tilde{p} \in \mathcal{P}^q$ be the Lagrange interpolation of the function $\tau_m p(t)/(t - t_{m-1})$ at the Radau points $t^{m,i},\ i = 1, \ldots, q+1$:*

$$\tilde{p}(t^{m,i}) = \tau_m p(t^{m,i})/(t^{m,i} - t_{m-1}) = p(t^{m,i})\vartheta_i^{-1}. \tag{3.55}$$

Then

$$\int_{I_m} p'\tilde{p}\mathrm{d}t + p(t_{m-1})\tilde{p}(t_{m-1}) = \frac{1}{2}\left(p^2(t_m) + \sum_{i=1}^{q+1} w_i \vartheta_i^{-2} p^2(t^{m,i})\right). \tag{3.56}$$

Proof follows by the transformation of (3.39) to the interval I_m.

Now, by $\tilde{\xi}$ we shall denote the Lagrange interpolation of $\tau_m \xi(t)/(t - t_{m-1})$ at the Radau points $t^{m,i},\ i = 1, \ldots, q+1$. This means that for each $x \in \Omega$, the function $\tilde{\xi}(\cdot, x)$ is a polynomial in t of degree $\leq q$, which means that $\tilde{\xi} \in S_{h,\tau}^{p,q}$. In what follows we shall denote

$$\|\xi\|_m^2 = \tau_m \sum_{i=1}^{q+1} w_i \vartheta_i^{-1} \left\|\xi(t^{m,i})\right\|^2. \tag{3.57}$$

Let us set $\varphi := \tilde{\xi}$ in (2.30). Then, using the relation $\left(\{\xi_{m-1}\}, \tilde{\xi}^+_{m-1}\right) = \left(\xi^+_{m-1}, \tilde{\xi}^+_{m-1}\right) - \left(\xi^-_{m-1}, \tilde{\xi}^+_{m-1}\right)$ and transferring the second term to the right-hand side, we get

$$\begin{aligned}
&\underbrace{\int_{I_m} (\xi', \tilde{\xi})\mathrm{d}t + (\xi^+_{m-1}, \tilde{\xi}^+_{m-1})}_{(a)} + \underbrace{\int_{I_m} A_{h,m}(\xi, \tilde{\xi})\mathrm{d}t}_{(b)} \\
&\quad = \underbrace{(\xi^-_{m-1}, \tilde{\xi}^+_{m-1})}_{(c)} - \underbrace{\int_{I_m} (\eta', \tilde{\xi})\mathrm{d}t - \left(\{\eta\}_{m-1}, \tilde{\xi}^+_{m-1}\right)}_{(d)} \\
&\quad - \underbrace{\int_{I_m} A_{h,m}(\eta, \tilde{\xi})\mathrm{d}t}_{(e)} + \underbrace{\int_{I_m} \left(b_{h,m}\left(u, \tilde{\xi}\right) - b_{h,m}\left(U, \tilde{\xi}\right)\right)\mathrm{d}t}_{(f)}.
\end{aligned} \tag{3.58}$$

In what follows, we shall analyze individual terms (a)-(f).

(a) By Fubini's theorem and (3.56)

$$\begin{aligned}
\int_{I_m} \left(\xi', \tilde{\xi}\right)\mathrm{d}t + \left(\xi^+_{m-1}, \tilde{\xi}^+_{m-1}\right) &= \int_\Omega \left(\int_{t_{m-1}}^{t_m} \xi'\tilde{\xi}\mathrm{d}t + \xi^+_{m-1}\tilde{\xi}^+_{m-1}\right)\mathrm{d}x \\
&= \int_\Omega \frac{1}{2}\left(\left(\xi^-_m\right)^2 + \sum_{i=1}^{q+1} w_i \vartheta_i^{-2} \left(\xi\left(t^{m,i}\right)\right)^2\right)\mathrm{d}t \\
&= \frac{1}{2}\left(\left\|\xi^-_m\right\|^2 + \sum_{i=1}^{q+1} w_i \vartheta_i^{-2} \left\|\xi\left(t^{m,i}\right)\right\|^2\right).
\end{aligned} \tag{3.59}$$

Hence, since $\vartheta_i^{-1} \geq 1$, in view of the notation (3.57), we get the inequality

$$\int_{I_m} \left(\xi', \tilde{\xi}\right)\mathrm{d}t + \left(\xi^+_{m-1}, \tilde{\xi}^+_{m-1}\right) \geq \frac{1}{2}\left(\left\|\xi^-_m\right\|^2 + \frac{1}{\tau_m}\|\xi\|^2_m\right). \tag{3.60}$$

(b) We use the following lemma:

Lemma 3.8: *Under assumptions (3.9) we have*

$$\int_{I_m} A_{h,m}\left(\xi, \tilde{\xi}\right)\mathrm{d}t \geq \frac{\varepsilon}{2}\int_{I_m} \|\xi\|^2_{DG,m}\,\mathrm{d}t. \tag{3.61}$$

Proof: In view of (2.15) and (2.17),

$$\begin{aligned}
\int_{I_m} A_{h,m}\left(\xi, \tilde{\xi}\right)\mathrm{d}t &= \varepsilon\int_{I_m} \sum_{K\in\mathcal{T}_{h,m}} \int_K \nabla\xi\cdot\nabla\tilde{\xi}\mathrm{d}x\mathrm{d}t \\
&-\varepsilon\int_{I_m} \sum_{\Gamma\in\mathcal{F}^I_{h,m}} \int_\Gamma \left(\langle\nabla\xi\rangle\cdot\boldsymbol{n}_\Gamma\left[\tilde{\xi}\right] - \theta\left\langle\nabla\tilde{\xi}\right\rangle\cdot\boldsymbol{n}_\Gamma[\xi]\right)\mathrm{d}S\mathrm{d}t \\
&-\varepsilon\int_{I_m} \sum_{\Gamma\in\mathcal{F}^B_{h,m}} \int_\Gamma \left(\nabla\xi\cdot\boldsymbol{n}_\Gamma\tilde{\xi} - \theta\nabla\tilde{\xi}\cdot\boldsymbol{n}_\Gamma\xi\right)\mathrm{d}S\mathrm{d}t + \varepsilon\int_{I_m} J_{h,m}\left(\xi, \tilde{\xi}\right)\mathrm{d}t.
\end{aligned}$$

The expressions $\xi|_\Gamma$, $[\xi]_\Gamma$, $\tilde{\xi}|_\Gamma$, $\left[\tilde{\xi}\right]_\Gamma$, $\nabla\xi$ and $\nabla\tilde{\xi}$ are polynomials in t of degree $\leq q$. Hence, $\int_K \nabla\xi\cdot\nabla\tilde{\xi}\mathrm{d}x$, $\int_\Gamma [\xi]_\Gamma\cdot\left[\tilde{\xi}\right]_\Gamma \mathrm{d}S$, $\int_\Gamma \langle\nabla\xi\rangle\cdot\boldsymbol{n}\left[\tilde{\xi}\right]\mathrm{d}S$, $J_{h,m}\left(\xi,\tilde{\xi}\right)$, etc. are polynomials in t of degree $\leq 2q$. Therefore, we can express the integrals $\int_{I_m}\ldots\mathrm{d}t$ with the aid of the integration formula (3.38). We also use the relations $\tilde{\xi}(t^{m,i}) = \xi(t^{m,i})\vartheta_i^{-1}$, $\nabla\tilde{\xi}(t^{m,i}) = \nabla\xi(t^{m,i})\vartheta_i^{-1}$, $\left[\tilde{\xi}(t^{m,i})\right] = \left[\xi(t^{m,i})\vartheta_i^{-1}\right]$. Then, by (2.15)-(2.17)we get

$$\begin{aligned}
&\int_{I_m} A_{h,m}(\xi,\tilde{\xi})\mathrm{d}t \qquad (3.62)\\
&= \varepsilon\tau_m\sum_{i=1}^{q+1} w_i\left(\sum_{K\in\mathcal{T}_{h,m}}\int_K\left(\nabla\xi(t^{m,i})\cdot\nabla\tilde{\xi}(t^{m,i}) + J_{h,m}\left(\xi(t^{m,i}),\tilde{\xi}(t^{m,i})\right)\right)\mathrm{d}x\right.\\
&\quad - \sum_{\Gamma\in\mathcal{F}_{h,m}^I}\int_\Gamma\left(\left\langle\nabla\xi(t^{m,i})\right\rangle\cdot\boldsymbol{n}_\Gamma\left[\tilde{\xi}(t^{m,i})\right] - \theta\left\langle\nabla\tilde{\xi}(t^{m,i})\right\rangle\cdot\boldsymbol{n}_\Gamma\left[\xi(t^{m,i})\right]\right)\mathrm{d}S\\
&\quad \left. - \sum_{\Gamma\in\mathcal{F}_{h,m}^B}\int_\Gamma\left(\nabla\xi(t^{m,i})\cdot\boldsymbol{n}_\Gamma\tilde{\xi}(t^{m,i}) - \theta\nabla\tilde{\xi}(t^{m,i})\cdot\boldsymbol{n}_\Gamma\xi(t^{m,i})\right)\mathrm{d}S\right)\\
&= \tau_m\sum_{i=1}^{q+1}\vartheta_i^{-1}w_i\left(a_{h,m}\left(\xi(t^{m,i}),\xi(t^{m,i})\right) + \varepsilon J_{h,m}\left(\xi(t^{m,i}),\xi(t^{m,i})\right)\right)
\end{aligned}$$

In virtue of (3.8), under the assumptions (3.9), we have

$$a_{h,m}\left(\xi(t^{m,i}),\xi(t^{m,i})\right) + \varepsilon J_{h,m}\left(\xi(t^{m,i}),\xi(t^{m,i})\right) \geq \frac{\varepsilon}{2}\left\|\xi(t^{m,i})\right\|_{DG,m}^2, \quad i=1,\ldots,q+1. \qquad (3.63)$$

If we use (3.62), (3.63), inequality $\vartheta_i^{-1}\geq 1$ and take into account that $\|\xi\|_{DG,m}^2$ is a polynomial in t of degree $\leq 2q$, we find that

$$\begin{aligned}
\int_{I_m} A_{h,m}\left(\xi,\tilde{\xi}\right)\mathrm{d}t &\geq \frac{\varepsilon}{2}\tau_m\sum_{i=1}^{q+1}\vartheta_i^{-1}w_i\left\|\xi\right\|_{DG,m}^2\\
&\geq \frac{\varepsilon}{2}\tau_m\sum_{i=1}^{q+1} w_i\left\|\xi\right\|_{DG,m}^2 = \frac{\varepsilon}{2}\int_{I_m}\left\|\xi\right\|_{DG,m}^2\mathrm{d}t,
\end{aligned}$$

what we wanted to prove.

■

Remark 3.1: *If X is a space with a scalar product $((\cdot,\cdot))$ inducing the norm $\|\cdot\|$ and the mapping $\xi: I_m\to X$ is a polynomial in t of degree $\leq q$, i.e., $\xi=\sum_{j=0}^q\alpha_j t^j$, $\alpha_j\in X$, then $\|\xi\|^2$ is a polynomial of degree $\leq 2q$, as follows from the relation*

$$\|\xi\|^2 = ((\xi,\xi)) = \left(\left(\sum_{j=0}^q\alpha_j t^j,\sum_{j=0}^q\alpha_j t^j\right)\right) = \sum_{r=0}^{2q}\left(\sum_{\substack{i,j\\ i+j=r}}((\alpha_i,\alpha_j))\right)t^r. \qquad (3.64)$$

(c) By the Cauchy inequality,

$$\left|\left(\xi_{m-1}^{-},\tilde{\xi}_{m-1}^{+}\right)\right| \leq \left\|\xi_{m-1}^{-}\right\| \left\|\tilde{\xi}_{m-1}^{+}\right\|. \tag{3.65}$$

Lemma 3.9: *There exists a constant c_1 independent of $h_K,\ \tau_m,\ \xi$ such that*

$$\left\|\tilde{\xi}_{m-1}^{+}\right\|^2 \leq \frac{c_1}{\tau_m} \left\|\xi\right\|_m^2. \tag{3.66}$$

Proof: The function $\tilde{\xi}$ is the Lagrange interpolant to $\tau_m\xi(t)/(t-t_{m-1})$ at the points $t^{m,i} = t_{m-1} + \tau_m\vartheta_i,\ i = 1,\ldots,q+1$. This means that

$$\tilde{\xi}(t) = \tau_m \sum_{i=1}^{q+1} \frac{\xi(t^{m,i})}{t^{m,i}-t_{m-1}} \prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{t-t^{m,j}}{t^{m,i}-t^{m,j}} = \tau_m \sum_{i=1}^{q+1} \frac{\xi(t^{m,i})}{\tau_m\vartheta_i} \prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{t-t_{m-1}-\tau_m\vartheta_j}{\tau_m\left(\vartheta_i-\vartheta_j\right)}.$$

Setting $t = t_{m-1}$, we get

$$\tilde{\xi}_{m-1}^{+} = \sum_{i=1}^{q+1} \xi\left(t^{m,i}\right) \vartheta_i^{-1} \prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{-\vartheta_j}{\vartheta_i-\vartheta_j}$$

and, thus, since $\vartheta_i^{-1} \leq \vartheta_1^{-1}$,

$$\begin{aligned}\left\|\tilde{\xi}_{m-1}^{+}\right\|^2 &\leq C(q) \sum_{i=1}^{q+1} \vartheta_i^{-1}\vartheta_1^{-1} \left\|\xi(t^{m,i})\right\|^2 \left(\prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{\vartheta_j}{\vartheta_i-\vartheta_j}\right)^2 \\ &\leq \tilde{C}(q) \sum_{i=1}^{q+1} \vartheta_i^{-1} \left\|\xi(t^{m,i})\right\|^2 \left(\prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{\vartheta_j}{\vartheta_i-\vartheta_j}\right)^2. \end{aligned} \tag{3.67}$$

By [58], the Radau weights are defined as

$$w_i = \int_0^1 \prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{z-\vartheta_j}{\vartheta_i-\vartheta_j} \mathrm{d}z$$

and $w^* := \min_{i=1,\ldots,q+1} w_i > 0$. Moreover, let us set

$$w^{**} := \max_{i=1,\ldots,q+1} \left(\prod_{\substack{j=1\\ j\neq i}}^{q+1} \frac{\vartheta_j}{\vartheta_i-\vartheta_j}\right)^2.$$

Hence, since $w_i \geq w^*$, we get

$$\left\|\tilde{\xi}_{m-1}^{+}\right\|^2 \leq \tilde{C}(q) \sum_{i=1}^{q+1} \vartheta_i^{-1} \left\|\xi(t^{m,i})\right\|^2 \frac{w^{**}w^*}{w^*} \leq c_1 \sum_{i=1}^{q+1} \vartheta_i^{-1} \left\|\xi(t^{m,i})\right\|^2 w_i = \frac{c_1}{\tau_m} \left\|\xi\right\|_m^2,$$

with $c_1 = \tilde{C}(q)w^{**}/w^*$.

■

(d) Integration by parts implies that

$$\begin{aligned}\int_{I_m} \; & (\eta', \tilde{\xi})\mathrm{d}t + (\{\eta\}_{m-1}, \tilde{\xi}^+_{m-1}) \\ & = -\int_{I_m} (\eta, \tilde{\xi}')\mathrm{d}t + (\eta^-_m, \tilde{\xi}^-_m) - (\eta^+_{m-1}, \tilde{\xi}^+_{m-1}) + (\eta^+_{m-1}, \tilde{\xi}^+_{m-1}) - (\eta^-_{m-1}, \tilde{\xi}^+_{m-1}).\end{aligned} \tag{3.68}$$

It holds that $\tilde{\xi}' \in S^{p,q-1}_{h,\tau}$ and thus, in view of (2.26), c),

$$\int_{I_m} (\eta, \tilde{\xi}')\mathrm{d}t = 0. \tag{3.69}$$

Further, $\eta^-_m = (\pi u)(t^-_m) - u(t_m) = \Pi_m u(t_m) - u(t_m)$ and $\tilde{\xi}^-_m \in S^p_{h,m}$. This and (2.22) imply that

$$(\eta^-_m, \tilde{\xi}^-_m) = 0. \tag{3.70}$$

Hence, in virtue of (3.68)-(3.70) we find that

$$\int_{I_m} (\eta', \tilde{\xi})\mathrm{d}t + (\{\eta\}_{m-1}, \tilde{\xi}^+_{m-1}) = -(\eta^-_{m-1}, \tilde{\xi}^+_{m-1}) \le \left\|\eta^-_{m-1}\right\| \left\|\tilde{\xi}^+_{m-1}\right\|. \tag{3.71}$$

(e) We use the following lemma:

Lemma 3.10: *If $k > 0$, then there exists a constant $C > 0$ such that*

$$\left|\int_{I_m} A_{h,m}\left(\eta, \tilde{\xi}\right) \mathrm{d}t\right| \le \frac{\varepsilon}{k} \int_{I_m} \|\xi\|^2_{DG,m} \,\mathrm{d}t + C\varepsilon \int_{I_m} \sigma^2_m(\eta) \,\mathrm{d}t. \tag{3.72}$$

Proof: Let $\hat{k} > 0$. Using (3.24) with $\varphi := \tilde{\xi}$, we get

$$\left|\int_{I_m} A_{h,m}\left(\eta, \tilde{\xi}\right) \mathrm{d}t\right| \le \frac{\varepsilon}{\hat{k}} \int_{I_m} \left\|\tilde{\xi}\right\|^2_{DG,m} \mathrm{d}t + C\varepsilon \int_{I_m} \sigma^2_m(\eta) \,\mathrm{d}t. \tag{3.73}$$

Now we shall estimate $\int_{I_m} \left\|\tilde{\xi}\right\|^2_{DG,m} \mathrm{d}t$. The function $\tilde{\xi}(t) = \sum_{j=0}^q \alpha_j t^j$, where $\alpha_j \in S^p_{h,m}$ is the Radau interpolation of the function $\tau_m \xi(t)/(t - t_{m-1})$. Hence,

$$\left\|\tilde{\xi}\left(t^{m,i}\right)\right\|^2_{DG,m} = \left\|\xi\left(t^{m,i}\right)\right\|^2_{DG,m} \vartheta_i^{-2}, \quad i = 1, \ldots, q+1,$$

and $\left\|\tilde{\xi}(t)\right\|^2_{DG,m}$ is a polynomial in t of degree $\le 2q$. Thus,we get

$$\begin{aligned}\int_{I_m} \left\|\tilde{\xi}(t)\right\|^2_{DG,m} \mathrm{d}t \; &= \; \tau_m \sum_{i=1}^{q+1} w_i \left\|\tilde{\xi}\left(t^{m,i}\right)\right\|^2_{DG,m} = \tau_m \sum_{i=1}^{q+1} w_i \vartheta_i^{-2} \left\|\xi\left(t^{m,i}\right)\right\|^2_{DG,m} \\ &\le \; \vartheta_1^{-2} \tau_m \sum_{i=1}^{q+1} w_i \left\|\xi\left(t^{m,i}\right)\right\|^2_{DG,m} = \vartheta_1^{-2} \int_{I_m} \|\xi\|^2_{DG,m} \,\mathrm{d}t.\end{aligned}$$

Hence,

$$\int_{I_m} \left\|\tilde{\xi}\right\|^2_{DG,m} \mathrm{d}t \le \hat{C} \int_{I_m} \|\xi\|^2_{DG,m} \,\mathrm{d}t. \tag{3.74}$$

From (3.73) with $\hat{k} := \hat{C}k$ and (3.74) we get the estimate (3.72), which we wanted to prove.

■

(f) Now, by (3.10), (3.74) and Young's inequality,

$$\left|\int_{I_m} b_{h,m}(u,\tilde{\xi}) - b_{h,m}(U,\tilde{\xi})\mathrm{d}t\right| \le \frac{\varepsilon}{k}\int_{I_m}\|\xi\|_{DG,m}^2\,\mathrm{d}t + \frac{C}{\varepsilon}\left(\int_{I_m}\|\xi\|^2\,\mathrm{d}t + \int_{I_m}\tilde{\sigma}_m^2\,(\eta)\,\mathrm{d}t\right). \tag{3.75}$$

Lemma 3.11: *There exist constants $C,\ C^*>0$ such that*

$$\int_{I_m}\|\xi\|^2\,\mathrm{d}t \le C\tau_m\left(\left\|\xi_{m-1}^-\right\|^2 + \left\|\eta_{m-1}^-\right\|^2 + \int_{I_m} R_m(\eta)\mathrm{d}t\right), \tag{3.76}$$

provided

$$0<\tau_m\le C^*\varepsilon, \tag{3.77}$$

where $R_m(\eta)$ is defined by (3.36).

Proof: If we proceed similarly as in the proof of (3.74), using (3.57) and the inequalities $1\le\vartheta_i^{-1}\le\vartheta_1^{-1}$, we get

$$\begin{aligned}\textstyle\int_{I_m}\|\xi\|^2\,\mathrm{d}t = \tau_m\sum_{i=1}^{q+1} w_i\,\|\xi\,(t^{m,i})\|^2 \le \|\xi\|_m^2\,,\\ \|\xi\|_m^2 \le \vartheta_1^{-1}\tau_m\textstyle\sum_{i=1}^{q+1} w_i\,\|\xi\,(t^{m,i})\|^2 = \vartheta_1^{-1}\int_{I_m}\|\xi\|^2\,\mathrm{d}t.\end{aligned} \tag{3.78}$$

Now, estimates (3.58), (3.60), (3.61), (3.65), (3.66), (3.71), (3.72) and (3.75) yield

$$\begin{aligned}&\frac{1}{2}\left\|\xi_m^-\right\|^2 + \frac{1}{2}\frac{1}{\tau_m}\|\xi\|_m^2 + \frac{\varepsilon}{2}\int_{I_m}\|\xi\|_{DG,m}^2\,\mathrm{d}t\\ &\le \left\|\xi_{m-1}^-\right\|\|\xi\|_m\sqrt{\frac{c_1}{\tau_m}} + \left\|\eta_{m-1}^-\right\|\|\xi\|_m\sqrt{\frac{c_1}{\tau_m}} + \frac{2\varepsilon}{k}\int_{I_m}\|\xi\|_{DG,m}^2\,\mathrm{d}t\\ &\quad + \frac{C}{\varepsilon}\int_{I_m}\|\xi\|^2\,\mathrm{d}t + C\varepsilon\int_{I_m}\sigma_m^2(\eta)\mathrm{d}t + \frac{C}{\varepsilon}\int_{I_m}\tilde{\sigma}_m^2(\eta)\mathrm{d}t.\end{aligned}$$

This, (3.36), (3.78), Young's inequality and the choice $k:=8$ imply that

$$\begin{aligned}&\left\|\xi_m^-\right\|^2 + \frac{\varepsilon}{2}\int_{I_m}\|\xi\|_{DG,m}^2\,\mathrm{d}t + \left(\frac{1}{2\tau_m} - \frac{\tilde{C}}{\varepsilon}\right)\int_{I_m}\|\xi\|^2\,\mathrm{d}t\\ &\le C\left(\left\|\xi_{m-1}^-\right\|^2 + \left\|\eta_{m-1}^-\right\|^2 + \int_{I_m} R_m(\eta)\mathrm{d}t\right).\end{aligned} \tag{3.79}$$

Let us put $C^*=1/(4\tilde{C})$, where $\tilde{C}$ is the constant from (3.79), and assume that (3.77) holds. Then $\frac{1}{2\tau_m}-\frac{\tilde{C}}{\varepsilon}\ge\frac{1}{4\tau_m}$ and (3.79) implies (3.76).

■

Summarizing estimates (3.35) with $k:=8$ and (3.76), we find that for $m=1,\dots,M$,

$$\left\|\xi_m^-\right\|^2 + \frac{\varepsilon}{2}\int_{I_m}\|\xi\|_{DG,m}^2\,\mathrm{d}t \le \left(1+\frac{c}{\varepsilon}\tau_m\right)\left\|\xi_{m-1}^-\right\|^2 + C\left\|\eta_{m-1}^-\right\|^2 + C\int_{I_m} R_m(\eta)\mathrm{d}t, \tag{3.80}$$

with constants $c,\ C>0$.

Lemma 3.12: *(Discrete Gronwall's lemma) Let $x_n,\ b_n$ and $c_n,\ n \in \mathbb{N}$, be non-negative sequences and let $a_n,\ n \in \mathbb{N}$, be positive nondecreasing sequence. If*

$$\begin{aligned} &x_0 + c_0 \le a_0, \\ &x_n + c_n \le a_n + \sum_{j=0}^{n-1} b_j x_j \quad \text{for } n \ge 1, \end{aligned} \tag{3.81}$$

then

$$x_n + c_n \le \left(1 + \frac{b_0 x_0}{a_0}\right) a_n \prod_{j=0}^{n-1}(1 + b_j) \quad \text{for } n \ge 0. \tag{3.82}$$

Proof: We start from relation (3.81) divided by a_n. Using the assumption of non-decreasing sequence a_n we obtain

$$\frac{x_n}{a_n} + \frac{c_n}{a_n} \le 1 + \sum_{j=0}^{n-1} b_j \frac{x_j}{a_n} \le 1 + \sum_{j=0}^{n-1} b_j \frac{x_j}{a_j}. \tag{3.83}$$

If we set $v_n := 1 + \sum_{j=0}^{n-1} b_j \frac{x_j}{a_j}$ for $n > 0$, we have

$$v_n - v_{n-1} = b_{n-1}\frac{x_{n-1}}{a_{n-1}} \le b_{n-1}\left(\frac{x_{n-1}}{a_{n-1}} + \frac{c_{n-1}}{a_{n-1}}\right) \le b_{n-1} v_{n-1}. \tag{3.84}$$

It gives us

$$\begin{aligned} v_n &\le (1 + b_{n-1}) v_{n-1} \le (1 + b_{n-1})(1 + b_{n-2}) v_{n-2} \\ &\le v_1 \prod_{j=1}^{n-1}(1 + b_j) = \left(1 + b_0 \frac{x_0}{a_0}\right) \prod_{j=1}^{n-1}(1 + b_j). \end{aligned} \tag{3.85}$$

The relation (3.85) together with (3.83) and the definition of v_n leads to (3.82).

■

Finally, we come to the *abstract error estimate.*

Theorem 3.13: *Let (3.77) hold. Then there exist constants $C,\ c > 0$ such that the error $e = U - u$ satisfies the estimate for all $m = 1, \dots, M$:*

$$\begin{aligned} &\left\|e_m^-\right\|^2 + \frac{\varepsilon}{2} \sum_{j=1}^{m} \int_{I_m} \|e\|_{DG,j}^2 \,\mathrm{d}t \\ &\le C \exp(c t_m/\varepsilon) \left(\sum_{j=1}^{m} \left\|\eta_j^-\right\|^2 + \sum_{j=1}^{m} \int_{I_j} R_j(\eta) \mathrm{d}x \right) \\ &\quad + 2 \left(\left\|\eta_m^-\right\|^2 + \varepsilon \sum_{j=1}^{m} \int_{I_j} \|\eta\|_{DG,j}^2 \,\mathrm{d}t \right). \end{aligned} \tag{3.86}$$

Proof: Summing the relation (3.80) over all $j = 1, \ldots, m$ and applying the discrete Gronwall's lemma 3.12, where we set

$$\begin{aligned} x_m &= \left\|\xi_m^-\right\|^2, \\ c_m &= \frac{\varepsilon}{2}\sum_{j=1}^{m}\int_{I_j}\|\xi\|_{DG,m}^2\,\mathrm{d}t, \\ b_m &= \frac{c}{\varepsilon}\tau_{m+1}, \\ a_m &= \left\|\xi_0^-\right\|^2 + C\sum_{j=1}^{m}\left\|\eta_{j-1}^-\right\|^2 + C\sum_{j=1}^{m}\int_{I_j} R_j(\eta)\mathrm{d}t, \end{aligned} \tag{3.87}$$

and the inequality $1 + x \le e^x$ for $x \in \mathbb{R}$ gives us the estimate

$$\begin{aligned} &\left\|\xi_m^-\right\|^2 + \frac{\varepsilon}{2}\sum_{j=1}^{m}\int_{I_j}\|\xi\|_{DG,j}^2\,\mathrm{d}t \\ &\le C\exp(ct_m/\varepsilon)\left(\left\|\xi_0^-\right\|^2 + \sum_{j=1}^{m}\left\|\eta_j^-\right\|^2 + \sum_{j=1}^{m}\int_{I_j} R_j\left(\eta\right)\mathrm{d}t\right), \end{aligned} \tag{3.88}$$

where $t_m = \sum_{j=1}^{m}\tau_j$ and $m = 1, \ldots, M$. In view of the definition of U_0^- in Definition 2.1 we have $\xi_0^- = 0$. If we use the relation $e = \xi + \eta$, the triangle inequality and the Young's inequality we obtain

$$\|e\|^2 = \|\xi+\eta\|^2 \le \|\xi\|^2 + 2\|\xi\|\,\|\eta\| + \|\eta\|^2 \le 2\left(\|\xi\|^2 + \|\eta\|^2\right), \tag{3.89}$$

$$\begin{aligned} \|e\|_{DG,j}^2 &= \|\xi+\eta\|_{DG,j}^2 \le \|\xi\|_{DG,j}^2 + 2\|\xi\|_{DG,j}\,\|\eta\|_{DG,j} + \|\eta\|_{DG,j}^2 \\ &\le 2\left(\|\xi\|_{DG,j}^2 + \|\eta\|_{DG,j}^2\right). \end{aligned} \tag{3.90}$$

Using these considerations in (3.88) we immediately get (3.86).

■

Chapter 4

Error estimation in terms of h and τ

This section will be devoted to obtaining error estimates in dependence on the mesh sizes τ and h. They will be obtained on the basis of estimate (3.88), the relations

$$\begin{aligned} e = U - u &= \xi + \eta, \\ \pi u|_{I_m} &= \pi\left(\Pi_m u\right)|_{I_m}, \\ \eta|_{I_m} = \left(\pi u - u\right)|_{I_m} &= \eta^{(1)} + \eta^{(2)}, \\ \text{with } \eta^{(1)} = \left(\Pi_m u - u\right)|_{I_m},\ \eta^{(2)} &= \left(\pi\left(\Pi_m u\right) - \Pi_m u\right)|_{I_m}, \end{aligned} \tag{4.1}$$

and estimates of individual terms on the right-hand side of (3.86) containing η, which will be proven in the sequel. To this end, we assume that the exact solution satisfies the regularity condition

$$u \in H^{q+1}\left(0,T;H^1\left(\Omega\right)\right) \cap C\left(\left[0,T\right];H^{p+1}\left(\Omega\right)\right) \tag{4.2}$$

and that the meshes satisfy shape regularity conditions (3.1), (3.2), (3.8) and (3.77). Obviously, $C\left(\left[0,T\right];H^{p+1}\left(\Omega\right)\right) \subset L^2\left(0,T;H^{p+1}\left(\Omega\right)\right)$. Moreover, let

$$\tau_m \geq Ch_m^2, \quad m = 1,\ldots,M. \tag{4.3}$$

Let us note that this assumption is not necessary, if the space meshes do not depend on time, i.e. all meshes $\mathcal{T}_{h,m},\ m = 1,\ldots,M$, are identical - see Section 4.4.

As shown in [16] if $r \geq 1$ is integer and $\mu = \min(r,p)$, then for $m = 1,\ldots,M$ and any $v \in H^{r+1}(\Omega)$ we have the standard estimates

$$\begin{aligned} \left\|\Pi_m v - v\right\|_{L^2(K)} &\leq Ch_K^{\mu+1}\left|v\right|_{H^{\mu+1}(K)}, \\ \left|\Pi_m v - v\right|_{H^1(K)} &\leq Ch_K^{\mu}\left|v\right|_{H^{\mu+1}(K)}, \\ \left|\Pi_m v - v\right|_{H^2(K)} &\leq Ch_K^{\mu-1}\left|v\right|_{H^{\mu+1}(K)} \end{aligned} \tag{4.4}$$

for $K \in \mathcal{T}_{h,m},\ h \in (0,h_0)$ and

$$\begin{aligned} &\text{(a)} \quad \left\|\Pi_m v\right\|_{L^2(K)} \leq \left\|v\right\|_{L^2(K)} \quad \text{for } v \in L^2(K),\ K \in \mathcal{T}_{h,m},\ h \in (0,h_0), \\ &\text{(b)} \quad \left|\Pi_m v\right|_{H^1(K)} \leq C\left|v\right|_{H^1(K)} \quad \text{for } v \in H^1(K),\ K \in \mathcal{T}_{h,m},\ h \in (0,h_0) \end{aligned} \tag{4.5}$$

It is possible to find that

$$D^{q+1}(\Pi_m u) = \Pi_m(D^{q+1}u). \tag{4.6}$$

Actually, by (2.22), $\Pi_m u(\cdot,t) \in S_{h,m}^p$ and for all $t \in I_m$,

$$\int_\Omega \left(\Pi_m u\left(x,t\right) - u\left(x,t\right)\right)\varphi\left(x\right)\mathrm{d}x = 0, \quad \forall \varphi \in S_{h,m}^p. \tag{4.7}$$

The differentiation with respect to t yields

$$\int_\Omega \left(D^{q+1}\Pi_m u\left(x,t\right) - D^{q+1}u\left(x,t\right)\right)\varphi\left(x\right)\mathrm{d}x = 0, \quad \forall \varphi \in S_{h,m}^p. \tag{4.8}$$

Moreover, obviously $D^{q+1}\left(\Pi_m u(t)\right) \in S_{h,m}^p$ and thus, (4.6) holds.

If we take in the mind that $\Pi_m u(\cdot,t) \in S_{h,m}^p$ then we have $\nabla\Pi_m u(\cdot,t) \in S_{h,m}^{p-1} \subset S_{h,m}^p$. So we can write

$$\int_\Omega \left(\nabla(\Pi_m u\left(x,t\right) - u\left(x,t\right))\right)\varphi\left(x\right)\mathrm{d}x = 0, \quad \forall \varphi \in S_{h,m}^p. \tag{4.9}$$

Hence,

$$\int_\Omega \left(\nabla\Pi_m u\left(x,t\right) - \nabla u\left(x,t\right)\right)\varphi\left(x\right)\mathrm{d}x = 0, \quad \forall \varphi \in S_{h,m}^p. \tag{4.10}$$

By the differentiation with respect to t we get

$$\int_\Omega \left(D^{q+1}\nabla\Pi_m u\left(x,t\right) - D^{q+1}\nabla u\left(x,t\right)\right)\varphi\left(x\right)\mathrm{d}x = 0, \quad \forall \varphi \in S_{h,m}^p. \tag{4.11}$$

Again $D^{q+1}\left(\nabla\Pi_m u(t)\right) \in S_{h,m}^p$ and we obtain

$$D^{q+1}\left(\nabla\Pi_m u\right) = \nabla\Pi_m\left(D^{q+1}u\right). \tag{4.12}$$

4.1 Time interpolation

Lemma 4.1: *Let $\varphi \in C((t_{m-1},t_m];S_{h,m}^p)$, $m = 1,\ldots,M$. Then for each $x \in K$, $K \in \mathcal{T}_{h,m}$, $t \in I_m$, $m = 1,\ldots,M$ we have*

$$\pi\varphi(x,t) = \tilde{P}_m\varphi(x,t) \tag{4.13}$$

where $\tilde{P}_m$ is defined in the following way: For $\omega \in C((t_{m-1},t_m])$,

$$\begin{aligned} &\text{(a)} \quad \tilde{P}_m\omega \in \mathcal{P}^q(I_m), \\ &\text{(b)} \quad \int_{I_m}\left(\tilde{P}_m\omega(t) - \omega(t)\right)t^j\mathrm{d}t = 0, \quad \forall j = 0,\ldots,q-1, \\ &\text{(c)} \quad \tilde{P}_m\omega(t_m-) = \omega(t_m-). \end{aligned} \tag{4.14}$$

Proof: Let $m \in \{1,\ldots,M\}$. From the definition of the operators π and $\tilde{P}_m$ it follows that for each $K \in \mathcal{T}_{h,m}$ the function $\pi\varphi$ and $\tilde{P}_m\varphi$ are on $K \times I_m$ polynomials of degree

$\leq q$ in $t \in I_m$ and of degree $\leq p$ in $x \in K$. Moreover, $\pi\varphi(x,t_m-) = \varphi(x,t_m-) = \tilde{P}_m\varphi(x,t_m-)$ for all $x \in K$. Obviously, condition (2.26),(c) is equivalent to

$$\int_{I_m}\left(\int_K (\pi\varphi(x,t)-\varphi(x,t))\,\sigma(x)\,\mathrm{d}x\right)t^j\mathrm{d}t = 0, \quad \forall j=0,\ldots,q-1, \quad \forall\sigma\in P^p(K), \quad \forall K\in\mathcal{T}_{h,m}. \tag{4.15}$$

Further, by (4.14), for any $K \in \mathcal{T}_{h,m}$,

$$\int_{I_m}\left(\tilde{P}_m\varphi(x,t)-\varphi(x,t)\right)t^j\mathrm{d}t = 0, \quad \forall j=0,\ldots,q-1, \quad \forall x\in K. \tag{4.16}$$

Let $\sigma \in P^p(K)$. Then (4.16) and Fubini's theorem imply that

$$\begin{aligned} 0 &= \int_K\left(\int_{I_m}\left(\tilde{P}_m\varphi(x,t)-\varphi(x,t)\right)t^j\mathrm{d}t\right)\sigma(x)\,\mathrm{d}x \\ &= \int_{I_m}\left(\int_K\left(\tilde{P}_m\varphi(x,t)-\varphi(x,t)\right)\sigma(x)\,\mathrm{d}x\right)t^j\mathrm{d}t, \\ &\quad \forall j=0,\ldots,q-1, \quad \forall\sigma\in P^p(K), \quad \forall K\in\mathcal{T}_{h,m}. \end{aligned} \tag{4.17}$$

Comparing (4.17) with (4.15) and taking in to account the fact that the operator π is uniquely determined by conditions (2.26), we immediately get (4.13).

■

Lemma 4.2: *If $\omega \in H^{q+1}(I_m)$, then*

$$\left\|\tilde{P}_m\omega-\omega\right\|^2_{L^2(I_m)} \leq C\tau_m^{2q+2}\left\|D^{q+1}\omega\right\|^2_{L^2(I_m)}, \tag{4.18}$$

where $C>0$ is a constant independent of ω, m and t.

Proof: We proceed in several steps.

1. We transform the interval $[t_{m-1},t_m]$ into the reference interval $[0,1]$ by the transformation

$$t = t_m - \vartheta\tau_m, \quad \vartheta\in[0,1]. \tag{4.19}$$

If $\omega \in H^{q+1}(I_m)$ and $s(\vartheta) = \omega(t_m-\vartheta\tau_m)$, then $s \in H^{q+1}(0,1)$ and

$$\tilde{P}_m(\omega)(t_m-\vartheta\tau_m) = (Ps)(\vartheta), \tag{4.20}$$

where the operator P is defined by

$$\begin{aligned} &\text{(a)} \quad Ps\in\mathcal{P}^q(0,1), \\ &\text{(b)} \quad \int_0^1 (Ps(\vartheta)-s(\vartheta))\,\vartheta^j\mathrm{d}\theta = 0 \quad \forall j=0,\ldots,q-1, \\ &\text{(c)} \quad Ps(0+) = s(0+). \end{aligned} \tag{4.21}$$

Moreover, if we set $Z_m(t) = \tilde{P}_m\omega(t)-\omega(t)$, $t\in(t_{m-1},t_m)$, $z(\vartheta) = Ps(\vartheta) - s(\vartheta)$, $\vartheta\in(0,1)$, we have $z(\vartheta) = Z_m(t_m-\vartheta\tau_m)$ and

$$D^{q+1}z(\vartheta) = (-1)^{q+1}\tau_m^{q+1}D^{q+1}Z_m(t_m-\vartheta\tau_m), \quad \vartheta\in(0,1). \tag{4.22}$$

By the substitution theorem,

$$\begin{aligned}
|z|^2_{L^2(0,1)} &= \int_0^1 |z(\vartheta)|^2 \,\mathrm{d}\vartheta \\
&= \int_0^1 |Z_m(t_m - \vartheta\tau_m)|^2 \,\mathrm{d}\vartheta = \frac{1}{\tau_m}\int_{I_m} |Z_m(t)|^2 \,\mathrm{d}t \\
&= \frac{1}{\tau_m}\,\|Z_m\|^2_{L^2(I_m)}\,, \qquad (4.23)\\
\left|D^{q+1}z\right|^2_{L^2(0,1)} &= \int_0^1 \left|D^{q+1}z(\vartheta)\right|^2 \mathrm{d}\vartheta \\
&= \int_0^1 \left|(-1)^{q+1}\tau_m^{q+1}D^{q+1}Z_m(t_m - \vartheta\tau_m)\right|^2 \mathrm{d}\vartheta \\
&= \int_0^1 \tau_m^{2q+2}\left|\tau_m^{q+1}D^{q+1}Z_m(t_m - \vartheta\tau_m)\right|^2 \mathrm{d}\vartheta \\
&= \tau^{2q+1}\int_{I_m} |Z_m(t)|^2 \,\mathrm{d}t \\
&= \tau_m^{2q+1}\left\|D^{q+1}Z_m\right\|^2_{L^2(I_m)}.
\end{aligned}$$

2. Since conditions (4.21)a)-c) determine the values of the operator P uniquely, it is clear that

$$Pr = r \quad \text{for } r \in \mathcal{P}(0,1). \qquad (4.24)$$

Now we prove that the operator P is a continuous mapping of the space $H^{q+1}(0,1)$ into $L^2(0,1)$. Let $u_n \in H^{q+1}(0,1)$, $n = 1, 2, \ldots$, $u_n \to 0$ in $H^{q+1}(0,1)$ for $n \to \infty$. The continuous imbendding $H^{q+1}(0,1) \hookrightarrow C([0,1])$ implies that

$$u_n \rightrightarrows 0 \quad \text{in } [0,1] \qquad (4.25)$$

and, hence

$$Pu_n(0) \to 0. \qquad (4.26)$$

For $j = 1, \ldots, q-1$ we have

$$\int_0^1 (Pu_n - u_n)(\vartheta)\vartheta^j \mathrm{d}\vartheta = 0. \qquad (4.27)$$

This and (4.25) imply that

$$\int_0^1 Pu_n(\vartheta)\vartheta^j \mathrm{d}\vartheta = \int_0^1 u_n(\vartheta)\vartheta^j \mathrm{d}\vartheta \to 0, \quad j = 0, \ldots, q-1. \qquad (4.28)$$

Since $Pu_n \in \mathcal{P}^q(0,1)$, we can write

$$Pu_n(\vartheta) = \sum_{i=1}^{q} c_i^{(n)}\vartheta^i + Pu_n(0), \quad \vartheta \in [0,1]. \qquad (4.29)$$

Integration yields

$$\int_0^1 Pu_n(\vartheta)\vartheta^j \mathrm{d}\vartheta = \int_0^1 \sum_{i=1}^{q} c_i^{(n)} \vartheta^{i+j} \mathrm{d}\vartheta + (Pu_n)(0) \int_0^1 \vartheta^j \mathrm{d}\vartheta \tag{4.30}$$
$$= \sum_{i=1}^{q} c_i^{(n)} \frac{1}{i+j+1} + (Pu_n)(0)\frac{1}{j+1}, \quad j = 0, \ldots, q-1.$$

Using (4.26), (4.28), (4.30) and the fact that the matrix $(\frac{1}{i+j+1})_{i,j=1,\ldots,q}$ is regular (see [39]), we find that

$$c_i^{(n)} \to 0 \quad \text{for } i = 1, \ldots, q \text{ as } n \to \infty$$

and, thus, $Pu_n \rightrightarrows 0$ in $[0,1]$ and $Pu_n \to 0$ in $L^2(0,1)$.

3. The above results allow us to apply Theorem 3.1.4 from [16] and get the estimate

$$\|z\|_{L^2(0,1)} \le C \left\|D^{q+1} z\right\|_{L^2(0,1)} \tag{4.31}$$

with a constant $C > 0$ independent of $z \in H^{q+1}(0,1)$. This and (4.23) imply that

$$\|Z_m\|_{L^2(I_m)} \le C\tau_m^{2q+2} \left\|D^{q+1} z\right\|_{L^2(I_m)}. \tag{4.32}$$

Taking into account that $D^{q+1}\tilde{P}_m\omega = 0$, we immediately get (4.18). ∎

Lemmas 4.1 and 4.2 imply that for $\varphi \in H^{q+1}(I_m, S_{h,m}^p)$ we have

$$\|\pi\varphi(x,\cdot) - \varphi(x,\cdot)\|_{L^2(I_m)}^2 = \left\|\tilde{P}_m\varphi(x,\cdot) - \varphi(x,\cdot)\right\|_{L^2(I_m)}^2 \tag{4.33}$$
$$\le C\tau_m^{2q+2} \left\|D^{q+1}\varphi(x,\cdot)\right\|_{L^2(I_m)}^2, \quad x \in K,\ K \in \mathcal{T}_{h,m}.$$

4.2 Estimates of terms with η

Our further goal is to estimate the expressions

$$\left\|\eta_m^-\right\|^2,\ \int_{I_m} \|\eta\|_{L^2(K)}^2 \,\mathrm{d}t,\ \int_{I_m} |\eta|_{H^1(K)}^2 \,\mathrm{d}t,\ \int_{I_m} |\eta|_{H^2(K)}^2 \,\mathrm{d}t,\ J_{h,m}(\eta,\eta).$$

By (4.1),

$$\|\eta\|_{L^2(K)}^2 \le 2\left\|\eta^{(1)}\right\|_{L^2(K)}^2 + 2\left\|\eta^{(2)}\right\|_{L^2(K)}^2, \tag{4.34}$$
$$|\eta|_{H^s(K)}^2 \le 2\left|\eta^{(1)}\right|_{H^s(K)}^2 + 2\left|\eta^{(2)}\right|_{H^s(K)}^2, \quad s = 1, 2.$$

Lemma 4.3: *The following estimates hold for $K \in \mathcal{T}_{h,m}$ and $m = 1, \dots, M$:*

$$\left\|\eta_m^-\right\| \leq Ch^{p+1} \left|u(t_m)\right|_{H^{p+1}(\Omega)}, \tag{4.35}$$

$$\int_{I_m} \left\|\eta^{(1)}\right\|^2_{L^2(K)} \mathrm{d}t \leq Ch_K^{2(p+1)} \left|u\right|^2_{L^2(I_m,H^{p+1}(K))}, \tag{4.36}$$

$$\int_{I_m} \left|\eta^{(1)}\right|^2_{H^1(K)} \mathrm{d}t \leq Ch_K^{2p} \left|u\right|^2_{L^2(I_m,H^{p+1}(K))}, \tag{4.37}$$

$$h_K^2 \int_{I_m} \left|\eta^{(1)}\right|^2_{H^2(K)} \mathrm{d}t \leq Ch_K^{2p} \left|u\right|^2_{L^2(I_m,H^{p+1}(K))}. \tag{4.38}$$

Proof: It is enough to use (4.4).

■

The derivation of the estimates of terms with $\eta^{(2)}$ is more complicated.

Lemma 4.4: *For $K \in \mathcal{T}_{h,m},\ m = 1, \dots, M$, we have*

$$\int_{I_m} \left\|\eta^{(2)}\right\|^2_{L^2(K)} \mathrm{d}t \leq C\tau_m^{2(q+1)} \left|u\right|^2_{H^{q+1}(I_m;L^2(K))}, \tag{4.39}$$

$$\int_{I_m} \left|\eta^{(2)}\right|^2_{H^1(K)} \mathrm{d}t \leq C\tau_m^{2(q+1)} \left|u\right|^2_{H^{q+1}(I_m;H^1(K))}, \tag{4.40}$$

$$h_K^2 \int_{I_m} \left|\eta^{(1)}\right|^2_{H^2(K)} \mathrm{d}t \leq C\tau_m^{2(q+1)} \left|u\right|^2_{H^{q+1}(I_m;H^1(K))}. \tag{4.41}$$

Proof: (a) The use of Fubini's theorem and (4.13), (4.6), (4.33), (4.5)(a) and (4.18) yields the relations

$$\begin{aligned}
\int_{I_m} \left\|\eta^{(2)}\right\|^2_{L^2(K)} \mathrm{d}t &= \int_{I_m} \left(\int_K \left|\eta^{(2)}\right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= \int_K \left(\int_{I_m} \left|\eta^{(2)}\right|^2 \mathrm{d}t \right) \mathrm{d}x = \int_K \left\| \tilde{P}_m(\Pi_m u) - \Pi_m u \right\|^2_{L^2(I_m)} \mathrm{d}x \\
&\leq C\tau_m^{2q+2} \int_K \left\| D^{q+1}(\Pi_m u) \right\|^2_{L^2(I_m)} \mathrm{d}x \\
&= C\tau_m^{2q+2} \int_{I_m} \left(\int_K \left| D^{q+1}(\Pi_m u) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= C\tau_m^{2q+2} \int_{I_m} \left(\int_K \left| \Pi_m(D^{q+1} u) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&\leq C\tau_m^{2q+2} \int_{I_m} \left(\int_K \left| D^{q+1} u \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= C\tau_m^{2q+2} \left|u\right|^2_{H^{q+1}(I_m;L^2(K))}.
\end{aligned} \tag{4.42}$$

(b) Further, due to Fubini's theorem, (4.13), (4.33), (4.12) and (4.5)(b), we find

that

$$\begin{aligned}
\int_{I_m} \left|\eta^{(2)}\right|^2_{H^1(K)} \mathrm{d}t &= \int_{I_m} \left(\int_K \left| \nabla \left(\Pi_m u - \tilde{P}_m \left(\Pi_m u \right) \right) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= \int_K \left(\int_{I_m} \sum_{j=1}^{d} \left(\frac{\partial}{\partial x_j} \left(\Pi_m u \right) - \tilde{P}_m \left(\frac{\partial}{\partial x_j} \left(\Pi_m u \right) \right) \right)^2 \mathrm{d}t \right) \mathrm{d}x \\
&\leq C\tau_m^{2q+2} \int_K \left| \nabla \left(\Pi_m u \right) \right|^2_{H^{q+1}(I_m)} \mathrm{d}x \\
&= C\tau_m^{2q+2} \int_K \left(\int_{I_m} \left| D^{q+1} \nabla \left(\Pi_m u \right) \right|^2 \mathrm{d}t \right) \mathrm{d}x \\
&= C\tau_m^{2q+2} \int_{I_m} \left(\int_K \left| \nabla \left(\Pi_m D^{q+1} u \right) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= C\tau_m^{2q+2} \int_{I_m} \left| \Pi_m \left(D^{q+1} u \right) \right|^2_{H^1(K)} \mathrm{d}t \\
&\leq C\tau_m^{2q+2} \int_{I_m} \left| D^{q+1} u \right|^2_{H^1(K)} \mathrm{d}t = C\tau_m^{2q+2} \left| u \right|^2_{H^{q+1}(I_m;H^1(K))} .
\end{aligned}$$

(c) Using a similar process as in (b) and (3.5), we find that

$$\begin{aligned}
\int_{I_m} \left|\eta^{(2)}\right|^2_{H^2(K)} \mathrm{d}t &= \int_{I_m} \left(\int_K \left| \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m u - \tilde{P}_m \left(\Pi_m u \right) \right) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= \int_K \left(\int_{I_m} \sum_{i,j=1}^{d} \left(\frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m u \right) - \tilde{P}_m \left(\frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m u \right) \right) \right)^2 \mathrm{d}t \right) \mathrm{d}x \\
&\leq C\tau_m^{2q+2} \int_K \left| \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m u \right) \right|^2_{H^{q+1}(I_m)} \mathrm{d}x \\
&= C\tau_m^{2q+2} \int_K \left(\int_{I_m} \left| D^{q+1} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m u \right) \right|^2 \mathrm{d}t \right) \mathrm{d}x \\
&= C\tau_m^{2q+2} \int_{I_m} \left(\int_K \left| \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(\Pi_m D^{q+1} u \right) \right|^2 \mathrm{d}x \right) \mathrm{d}t \\
&= C\tau_m^{2q+2} \int_{I_m} \left| \Pi_m \left(D^{q+1} u \right) \right|^2_{H^2(K)} \mathrm{d}t \\
&\leq C\tau_m^{2q+2} \int_{I_m} \left| \Pi_m \left(D^{q+1} u \right) \right|^2_{H^2(K)} \mathrm{d}t \\
&\leq C\tau_m^{2q+2} \int_{I_m} C_I^2 h_K^{-2} \left| \Pi_m \left(D^{q+1} u \right) \right|^2_{H^1(K)} \mathrm{d}t \\
&\leq \tilde{C}\tau_m^{2q+2} h_K^{-2} \int_{I_m} \left| D^{q+1} u \right|^2_{H^1(K)} \mathrm{d}t \\
&= \tilde{C}\tau_m^{2q+2} h_K^{-2} \left| u \right|^2_{H^{q+1}(I_m;H^1(K))} .
\end{aligned}$$

■

This yields (4.41).

Finally, we shall be concerned with the estimation of $\int_{I_m} J_{h,m}(\eta,\eta)\mathrm{d}t$. It holds

$$\begin{aligned} J_{h,m}(\eta,\eta) &= \sum_{\Gamma\in\mathcal{F}^I_{h,m}} h(\Gamma)^{-1}\int_\Gamma [\eta]^2\,\mathrm{d}S + \sum_{\Gamma\in\mathcal{F}^B_{h,m}} h(\Gamma)^{-1}\int_\Gamma \eta^2\mathrm{d}S \qquad (4.43)\\ &\le C\sum_{\Gamma\in\mathcal{F}^I_{h,m}} h(\Gamma)^{-1}\left(\int_\Gamma \left[\Pi_m u-u\right]^2\mathrm{d}S + \int_\Gamma \left[\pi\left(\Pi_m u\right)-\Pi_m u\right]^2\mathrm{d}S\right)\\ &\quad +C\sum_{\Gamma\in\mathcal{F}^B_{h,m}} h(\Gamma)^{-1}\left(\int_\Gamma \left(\Pi_m u-u\right)^2\mathrm{d}S + \int_\Gamma \left(\pi\left(\Pi_m u-u\right)-\Pi_m u\right)^2\mathrm{d}S\right)\\ &= C\left(J_{h,m}\left(\Pi_m u-u,\Pi_m u-u\right)+J_{h,m}\left(\pi\left(\Pi_m u\right)-\Pi_m u,\pi\left(\Pi_m u\right)-\Pi_m u\right)\right). \end{aligned}$$

Lemma 4.5: *There exists a constant $C_J>0$ such that it holds*

$$\int_{I_m} J_{h,m}(\Pi_m u-u,\Pi_m u-u)\mathrm{d}t \le C_J h^{2p}\left|u\right|^2_{L^2(I_m;H^{p+1}(\Omega))}. \qquad (4.44)$$

Proof: The definition of the form $J_{h,m}$ together with Young's inequality gives us

$$\begin{aligned} &J_{h,m}(\Pi_m u-u,\Pi_m u-u)\\ &=\sum_{\Gamma\in\mathcal{F}^I_{h,m}} h^{-1}(\Gamma)\int_\Gamma \left[\Pi_m u-u\right]^2\mathrm{d}S + \sum_{\Gamma\in\mathcal{F}^B_{h,m}} h^{-1}(\Gamma)\int_\Gamma (\Pi_m u-u)^2\mathrm{d}S\\ &=\sum_{\Gamma\in\mathcal{F}^I_{h,m}} h^{-1}(\Gamma)\int_\Gamma \left((\Pi_m u-u)|^L_\Gamma-(\Pi_m u-u)|^R_\Gamma\right)^2\mathrm{d}S\\ &\quad+\sum_{\Gamma\in\mathcal{F}^B_{h,m}} h^{-1}(\Gamma)\int_\Gamma (\Pi_m u-u)^2\mathrm{d}S\\ &\le\sum_{\Gamma\in\mathcal{F}^I_{h,m}} 2h^{-1}(\Gamma)\int_\Gamma \left\{\left((\Pi_m u-u)|^L_\Gamma\right)^2-\left((\Pi_m u-u)|^R_\Gamma\right)^2\right\}\mathrm{d}S\\ &\quad+\sum_{\Gamma\in\mathcal{F}^B_{h,m}} h^{-1}(\Gamma)\int_\Gamma (\Pi_m u-u)^2\mathrm{d}S. \end{aligned}$$

Using (3.3) and Lemma 3.2 we get

$$\begin{aligned} J_{h,m}(\Pi_m u-u,\Pi_m u-u) &\le \frac{2}{C_-}\sum_{\Gamma\in\mathcal{F}_{h,m}}\int_\Gamma h^{-1}(\Gamma)(\Pi_m u-u)^2\mathrm{d}S\\ &\le \frac{2C_O}{C_-}\sum_{K\in\mathcal{T}_{h,m}}\left(\frac{\left\|\Pi_m u-u\right\|^2_{L^2(K)}}{h_K^2}+\left|\Pi_m u-u\right|^2_{H^1(K)}\right). \end{aligned}$$

Finally, using estimations (4.4) with $\mu=p$ we obtain

$$\int_{I_m} J_{h,m}(\Pi_m u-u,\Pi_m u-u)\mathrm{d}t \le \int_{I_m}\frac{2C_O}{C_-}Ch^{2p}\sum_{K\in\mathcal{T}_{h,m}}\left|u\right|^2_{H^{p+1}(K)}\mathrm{d}t.$$

It leads to

$$\int_{I_m} J_{h,m}(\Pi_m u - u, \Pi_m u - u)\mathrm{d}t \leq C_J h^{2p} \left|u\right|^2_{L^2(I_m;H^{p+1}(\Omega))}.$$

■

Further, we shall estimate the expression

$$\int_{I_m} J_{h,m}(\pi(\Pi_m u) - \Pi_m u, \pi(\Pi_m u) - \Pi_m u)\mathrm{d}t.$$

Lemma 4.6: *Let Dirichlet data $u_D = u_D(x,t)$ have the behavior in t as a polynomial of degree $\leq q$:*

$$u_D(x,t) = \sum_{j=0}^{q} \psi_j(x) t^j, \tag{4.45}$$

where $\psi_j \in H^{p+1/2}(\partial\Omega)$ for $j = 0, \ldots, q$. Then

$$\int_{I_m} J_{h,m}\left(\pi\left(\Pi_m u\right) - \Pi_m u, \pi\left(\Pi_m u\right) - \Pi_m u\right)\mathrm{d}t \leq C\tau_m^{2q+2} \left|u\right|^2_{H^{q+1}(I_m,H^1(\Omega))}, \quad m = 1, \ldots, M. \tag{4.46}$$

For general data u_D, if there exists a constant $\bar{C} > 0$ such that $\tau_m \leq \bar{C} h_K$ for all $K \in \mathcal{T}_{h,m}$, $h \in (0, h_0)$ and $m = 1, \ldots, M$, then

$$\int_{I_m} J_{h,m}\left(\pi\left(\Pi_m u\right) - \Pi_m u, \pi\left(\Pi_m u\right) - \Pi_m u\right)\mathrm{d}t \leq C\tau_m^{2q} \left|u\right|^2_{H^{q+1}(I_m,H^1(\Omega))}, \quad m = 1, \ldots, M. \tag{4.47}$$

Proof: We proceed in two steps.

(I) Let $\Gamma \in \mathcal{F}^I_{h,m}$, i.e. $\Gamma \subset \Omega$. If we set $\varphi := \Pi_m u$, we can write

$$\begin{aligned}
&\int_{I_m} \left(\int_\Gamma \left[\pi\left(\Pi_m u\right) - \Pi_m u\right]^2 \mathrm{d}S \right) \mathrm{d}t \\
&= \int_{I_m} \left(\int_\Gamma \left[\pi\varphi - \varphi\right]^2 \mathrm{d}S \right) \mathrm{d}t \\
&= \int_\Gamma \left(\int_{I_m} \left[\pi\varphi\left(x,\cdot\right) - \varphi\left(x,\cdot\right)\right]^2 \mathrm{d}t \right) \mathrm{d}S \\
&= \int_\Gamma \left\| \left[\pi\varphi\left(x,\cdot\right) - \varphi\left(x,\cdot\right)\right] \right\|^2_{L^2(I_m)} \mathrm{d}S \\
&= \int_\Gamma \left\| \left[\tilde{P}_m\varphi\left(x,\cdot\right) - \varphi\left(x,\cdot\right)\right] \right\|^2_{L^2(I_m)} \mathrm{d}S.
\end{aligned}$$

Using the relation

$$\left[\tilde{P}_m\varphi - \varphi\right] = \tilde{P}_m\left[\varphi\right] - \left[\varphi\right], \tag{4.48}$$

and the estimate (4.33), we find that

$$\begin{aligned}
&\int_{I_m}\left(\int_\Gamma \left[\pi\left(\Pi_m u\right)-\Pi_m u\right]^2 \mathrm{d}S\right)\mathrm{d}t \\
&\quad=\int_\Gamma \left\|\left[\tilde{P}_m\varphi\left(x,\cdot\right)-\varphi\left(x,\cdot\right)\right]\right\|^2_{L^2(I_m)}\mathrm{d}S \\
&\quad\le C\tau_m^{2q+2}\int_\Gamma \left\|D^{q+1}\left[\varphi\left(x,\cdot\right)\right]\right\|^2_{L^2(I_m)}\mathrm{d}S. \qquad (4.49)
\end{aligned}$$

If we take into account that

$$\begin{aligned}
D^{q+1}\left[\varphi(x,\cdot)\right] &= \left[D^{q+1}\varphi(x,\cdot)\right], \qquad (4.50)\\
\left[D^{q+1}u\right] &= 0,
\end{aligned}$$

and use Fubini's theorem, we obtain

$$\begin{aligned}
&\int_{I_m}\left(\int_\Gamma \left[\pi\left(\Pi_m u\right)-\Pi_m u\right]^2 \mathrm{d}S\right)\mathrm{d}t \qquad (4.51)\\
&\quad=\int_\Gamma\left(\int_{I_m} \left[\pi\left(\Pi_m u\right)-\Pi_m u\right]^2 \mathrm{d}t\right)\mathrm{d}S \\
&\quad\le C\tau_m^{2q+2}\int_\Gamma\left(\int_{I_m} \left|D^{q+1}\left[\varphi\left(x,t\right)\right]\right|^2 \mathrm{d}t\right)\mathrm{d}S \\
&\quad= C\tau_m^{2q+2}\int_{I_m}\left(\int_\Gamma \left[D^{q+1}\left(\Pi_m u-u\right)\right]^2 \mathrm{d}S\right)\mathrm{d}t.
\end{aligned}$$

The application of the multiplicative trace inequality implies that

$$\begin{aligned}
&\sum_{\Gamma\in\mathcal{F}^I_{h,m}}\int_\Gamma \left[D^{q+1}\left(\Pi_m u-u\right)\right]^2\mathrm{d}S \qquad (4.52)\\
&\le C\sum_{K\in\mathcal{T}_{h,m}}\int_{\partial K}\left[D^{q+1}\left(\Pi_m u-u\right)\right]^2\mathrm{d}S = C\sum_{K\in\mathcal{T}_{h,m}}\left\|D^{q+1}\left(\Pi_m u-u\right)\right\|^2_{L^2(\partial K)} \\
&\le C\sum_{K\in\mathcal{T}_{h,m}}\Big(\left\|D^{q+1}\left(\Pi_m u-u\right)\right\|_{L^2(K)}\left|D^{q+1}\left(\Pi_m u-u\right)\right|_{H^1(K)} \\
&\quad+h_K^{-1}\left\|D^{q+1}\left(\Pi_m u-u\right)\right\|^2_{L^2(K)}\Big).
\end{aligned}$$

By (4.6),

$$D^{q+1}\left(\Pi_m u-u\right)=\Pi_m(D^{q+1}u)-D^{q+1}u. \qquad (4.53)$$

In view of (4.2), we have $D^{q+1}u\in L^2(I_m,H^1(\Omega))$. This and the approximation properties (4.4) of Π_m imply that

$$\begin{aligned}
\left\|\Pi_m\left(D^{q+1}u\right)-D^{q+1}u\right\|_{L^2(K)} &\le Ch_K\left|D^{q+1}u\right|_{H^1(K)}, \qquad (4.54)\\
\left|\Pi_m\left(D^{q+1}u\right)-D^{q+1}u\right|_{H^1(K)} &\le C\left|D^{q+1}u\right|_{H^1(K)}.
\end{aligned}$$

Summarizing (3.3), (4.51), (4.52), (4.53) and (4.54), we get

$$\int_{I_m} \left(\sum_{\Gamma \in \mathcal{F}^I_{h,m}} h(\Gamma)^{-1} \int_\Gamma \left[\pi \left(\Pi_m u \right) - \Pi_m u \right]^2 \mathrm{d}S \right) \mathrm{d}t \tag{4.55}$$
$$\leq C \tau_m^{2q+2} \int_{I_m} \sum_{K \in \mathcal{T}_{h,m}} \left| D^{q+1} u \right|^2_{H^1(K)} \mathrm{d}t = C \tau_m^{2q+2} \left| u \right|^2_{H^{q+1}(I_m; H^1(\Omega))}.$$

(II) In what follows, we shall assume that $\Gamma \in \mathcal{F}^B_{h,m}$, i.e. $\Gamma \subset \partial\Omega \cap \partial K$ for some $K \in \mathcal{T}_{h,m}$, and estimate the expression

$$(*) := \int_{I_m} \left(h(\Gamma)^{-1} \int_\Gamma \left| \pi \left(\Pi_m u \right) - \Pi_m u \right|^2 \mathrm{d}S \right) \mathrm{d}t. \tag{4.56}$$

Proceeding in a similar way as above, we find that

$$\begin{aligned} (*) &\leq C \tau_m^{2q+2} h(\Gamma)^{-1} \int_\Gamma \left\| D^{q+1} \left(\Pi_m u \right) \right\|^2_{L^2(I_m)} \mathrm{d}S \qquad (4.57) \\ &= C \tau_m^{2q+2} h(\Gamma)^{-1} \int_{I_m} \left(\int_\Gamma \left| D^{q+1} \left(\Pi_m u \right) \right|^2 \mathrm{d}S \right) \mathrm{d}t \\ &= C \tau_m^{2q+2} h(\Gamma)^{-1} \int_{I_m} \left(\int_\Gamma \left| \Pi_m \left(D^{q+1} u \right) \right|^2 \mathrm{d}S \right) \mathrm{d}t. \end{aligned}$$

If we apply the multiplicative trace inequality and use the assumption that $\tau_m \leq \bar{C} h_K$ for all $K \in \mathcal{T}_{h,m}$, we get

$$\int_{I_m} \left(\sum_{\Gamma \in \mathcal{F}^I_{h,m}} h(\Gamma)^{-1} \int_\Gamma \left| \pi \left(\Pi_m u \right) - \Pi_m u \right|^2 \mathrm{d}S \right) \mathrm{d}t \leq C \tau_m^{2q} \left| u \right|^2_{H^{q+1}(I_m, H^1(\Omega))}. \tag{4.58}$$

Now let us assume that the Dirichlet data $u_D = u_D(x,t)$ satisfy (4.45). Then $D^{q+1} u|_{\partial\Omega} = D^{q+1} u_D = 0$. This and (4.57) imply that

$$(*) \leq C \tau_m^{2q+2} \int_{I_m} \left(h(\Gamma)^{-1} \int_\Gamma \left| \Pi_m (D^{q+1} u) - D^{q+1} u \right|^2 \mathrm{d}S \right) \mathrm{d}t. \tag{4.59}$$

Now, we use (4.59), Lemma 3.2 and estimates (4.54) and get the estimate

$$\begin{aligned} &\int_{I_m} \left(\sum_{\Gamma \in \mathcal{F}^B_{h,m}} h(\Gamma)^{-1} \int_\Gamma \left| \pi \left(\Pi_m u \right) - \Pi_m u \right|^2 \mathrm{d}S \right) \mathrm{d}t \\ &\leq C C_O \tau_m^{2q+2} \sum_{K \in \mathcal{T}_{h,m}} \int_{I_m} \frac{\left\| \Pi_m (D^{q+1} u) - D^{q+1} u \right\|^2_{L^2(K)}}{h^2(K)} \mathrm{d}t \\ &\quad + C C_O \tau_m^{2q+2} \sum_{K \in \mathcal{T}_{h,m}} \int_{I_m} \left| \Pi_m (D^{q+1} u) - D^{q+1} u \right|^2_{H^1(K)} \mathrm{d}t \qquad (4.60) \\ &\leq C_{J'} \tau_m^{2q+2} \int_{I_m} \sum_{K \in \mathcal{T}_{h,m}} \left| D^{q+1} u \right|^2_{H^1(K)} \mathrm{d}t = C_{J'} \tau_m^{2q+2} \left| u \right|^2_{H^{q+1}(I_m; H^1(\Omega))}. \end{aligned}$$

Finally, summarizing estimates (4.55), (4.58) and (4.60), we get (4.46) and (4.47). ∎

4.3 Main result

In this section we shall conclude the analysis of the error estimate.

Theorem 4.7: *Let u be the exact solution satisfying the regularity condition (4.2) of problem (1.1)-(1.3) with the Dirichlet data u_D defined by (4.45). Let U be the approximate solution to problem (1.1)-(1.3) obtained by scheme (2.24) over spatial meshes $\mathcal{T}_{h,m}$ and time partition I_m, $m = 1, \ldots, M$, satisfying conditions (3.1), (3.2), (3.77) and (4.3). Then there exist constants C, $c > 0$ independent of h, τ, m, ε, u such that*

$$\begin{aligned}
&\left\|e_m^-\right\|^2 + \frac{\varepsilon}{2}\sum_{j=1}^{m}\int_{I_m}\|e\|_{DG,j}^2\,\mathrm{d}t \qquad (4.61)\\
&\leq C\exp\left(ct_m/\varepsilon\right)\left(\sum_{j=1}^{m}\left(h_j^{2p}\left|u\right|^2_{L^2(I_j,H^{p+1}(\Omega))} + \tau_j^{2q+1}\left|u\right|^2_{H^{q+1}(I_j,H^1(\Omega))}\right)\left(\varepsilon+\frac{1}{\varepsilon}\right)\right.\\
&\left. + h^{2p}\left|u\right|^2_{C([0,T],H^{p+1}(\Omega))}\right) + Ch^{2p+p}\left|u\right|^2_{C([0,T],H^{p+1}(\Omega))}\\
&+C\varepsilon\sum_{j=1}^{m}\left(h_j^{2p}\left|u\right|^2_{L^2(I_j,H^{p+1}(\Omega))} + \tau_j^{2q+2}\left|u\right|_{H^{q+1}(I_j,H^1(\Omega))}\right),\\
&m = 1,\ldots,M, \quad h \in (0,h_0),
\end{aligned}$$

or simply,

$$\begin{aligned}
&\left\|e_m^-\right\| + \frac{\varepsilon}{2}\sum_{j=1}^{m}\int_{I_m}\|e\|_{DG,j}^2\,\mathrm{d}t \qquad (4.62)\\
&\leq C\exp\left(ct_m/\varepsilon\right)\left(\left(h^{2p}\left|u\right|^2_{L^2(0,T;H^{p+1}(\Omega))} + \tau^{2q+2}\left|u\right|_{H^{q+1}(0,T;H^1(\Omega))}\right)\left(\varepsilon+\frac{1}{\varepsilon}\right)\right.\\
&\left. +h^{2p}\left|u\right|^2_{C([0,T],H^{p+1}(\Omega))}\right), \quad m = 1,\ldots,M, \quad h \in (0,h_0).
\end{aligned}$$

Proof: In order to prove (4.61), we start from (3.86) and estimate the terms containing η. In virtue of (3.36), (3.22), (3.11),

$$\begin{aligned}
R_j(\eta) &= \varepsilon\sigma_j^2(\eta) + \frac{1}{\varepsilon}\tilde{\sigma}_j^2(\eta) \qquad (4.63)\\
&= \varepsilon\left(\sum_{K\in\mathcal{T}_{h,j}}\left(\left|\eta\right|^2_{H^1(K)} + h_K^2\left|\eta\right|^2_{H^2(K)}\right) + J_{h,j}\left(\eta,\eta\right)\right)\\
&\quad \frac{1}{\varepsilon}\sum_{K\in\mathcal{T}_{h,j}}\left(\left\|\eta\right\|^2_{L^2(K)} + h_K^2\left|\eta\right|^2_{H^1(K)}\right).
\end{aligned}$$

Now, (4.63) together with (4.34) and Lemmas 4.3 and 4.4 yield the estimate

$$\int_{I_j} R_j(\eta)\mathrm{d}t \leq C\left(\varepsilon+\frac{1}{\varepsilon}\right)\sum_{K\in\mathcal{T}_{h,j}}\left(h_K^{2p}\left|u\right|^2_{L^2(I_j;H^{p+1}(K))} + \tau_j^{2q+2}\left|u\right|^2_{H^{q+1}(I_j;H^1(K))}\right). \qquad (4.64)$$

This and the inequalities $\tau_j \le \tau,\ h_K \le h_j \le h$ lead to

$$\int_{I_j} R_j(\eta)\mathrm{d}t \le C\left(\varepsilon + \frac{1}{\varepsilon}\right)\left(h^{2p}\,|u|^2_{L^2(I_j,H^{p+1}(\Omega))} + \tau^{2q+2}\,|u|^2_{H^{q+1}(I_j,H^1(\Omega))}\right) \tag{4.65}$$

Similarly, we get

$$\int_{I_j} \|\eta\|_{DG,j}\,\mathrm{d}t \le C\varepsilon \sum_{K\in\mathcal{T}_{h,j}} h_K^{2p}\,|u|^2_{L^2(I_j;H^{p+1}(K))} \le C\varepsilon h_j^{2p}\,|u|_{L^2(I_j;H^{p+1}(\Omega))}\,. \tag{4.66}$$

Further, by (4.35) and (4.3),

$$\sum_{j=1}^{m} \left\|\eta_j^-\right\|^2 \le C\sum_{j=1}^{M} \tau_j h_j^{2p}\,|u(t_j)|^2_{H^{p+1}(\Omega)} \le CTh^{2p}\,|u|^2_{C([0,T],H^{p+1}(\Omega))}\,. \tag{4.67}$$

Finally, using (3.88) and (4.65)-(4.67), we arrive at estimates (4.61) and (4.62), which we wanted to prove.

■

Remark 4.1: *As we see, estimate (4.61) is not uniform with respect to $\varepsilon \to 0$. Just on the contrary, the constant in this estimate behaves as $C exp(cT/\varepsilon)$, which blows up to ∞ as $\varepsilon \to 0$. This is consequence of the application of Young's inequality used for the treatment of nonlinear terms and Gronwall's lemma. The question, how to avoid this bad behavior of the error estimate, remains open.*

4.4 The case of the identical meshes on all time levels

If all meshes $\mathcal{T}_{h,m},\ m = 1,\dots,M$ are identical, which means that $\mathcal{T}_{h,m} = \mathcal{T}_h$ for all $m = 1,\dots,M$, then all spaces $S^p_{h,m}$ and forms $a_{h,m},\ b_{h,m},\dots$ are the same: $S^p_{h,m} = S^p_h,\ a_{h,m} = a_h,\ b_{h,m} = b_h,\dots$ for all $m = 1,\dots,M$. This implies that $\{\xi\}_{m-1} \in S^p_h$ and by (2.29), (2.26)(a) and (2.22), we have $(\eta^-_{m-1}, \{\xi\}_{m-1}) = 0$. Hence,

$$\int_{I_m} (\eta', \xi)\,\mathrm{d}t + \left(\{\eta\}_{m-1}, \xi^+_{m-1}\right) = 0. \tag{4.68}$$

Moreover, it is possible to show that the expression $\sum_{j=1}^m \left\|\eta_j^-\right\|^2$ does not appear in estimate (3.88) and instead of estimate (3.86) we get the estimate

$$\begin{aligned} &\left\|e_m^-\right\|^2 + \frac{\varepsilon}{2}\sum_{j=1}^{m}\int_{I_m} \|e\|^2_{DG,j}\,\mathrm{d}t \\ &\le C\exp(ct_m/\varepsilon)\left(\sum_{j=1}^{m}\int_{I_j} R_j(\eta)\mathrm{d}t\right) + 2\left\|\eta_m^-\right\|^2 + 2\varepsilon\sum_{j=1}^{m}\int_{I_j}\|\eta\|^2_{DG,j}\,\mathrm{d}t, \\ &m = 1,\dots,M. \end{aligned} \tag{4.69}$$

Due to fact that $\sum_{j=1}^m \left\|\eta_j^-\right\|^2$ does not appear in the abstract error estimate (3.86), assumption (4.3) can be omitted in the process of the derivation of the error estimate (4.61) and (4.62). This leads us to the following result.

Theorem 4.8: *Let u be the exact solution satisfying the regularity condition (4.2) of the problem (1.1)-(1.3) with the Dirichlet data u_D defined by (4.45). Let U be the approximate solution to the problem (1.1)-(1.3) obtained by scheme (2.24) over spatial meshes $\mathcal{T}_{h,m} = \mathcal{T}_h$ for all $m = 1, \dots, M$, and time partition I_m, $m = 1, \dots, M$, satisfying condition (3.1), (3.2) and (3.77). Then there exist constant $C, c > 0$ independent of h, τ, m, ε, u such that error estimates (4.61) and (4.62) hold.*

4.5 $L^2(Q_T)$-error estimate

Finally, we shall be concerned with the $L^2(L^2)$-error estimate, i.e. the error estimate in the norm of the space $L^2(Q_T)$.

Theorem 4.9: *Let u be the exact solution satisfying the regularity condition (4.2) of the problem (1.1)-(1.3) with the Dirichlet data u_D defined by (4.45). Let U be the approximate solution to the problem (1.1)-(1.3) obtained by scheme (2.24) over spatial meshes $\mathcal{T}_{h,m}$ and time partition I_m, $m = 1, \dots, M$, satisfying conditions (3.1), (3.2), (3.77) and (4.3). Then there exist constants $C, c > 0$ independent of h, τ, m, ε, u such that*

$$\begin{aligned} \|e\|^2_{L^2(Q_T)} \;\le\; & C\left(h^{2p+2} + e^{cT/\varepsilon} h^{2p}\right) |u|^2_{C([0,T],H^{p+1}(\Omega))} \\ & + C\left(\varepsilon + \frac{1}{\varepsilon}\right)\left(1 + e^{cT/\varepsilon}\right)\left(h^{2p} |u|^2_{L^2(0,T;H^{p+1}(\Omega))} + \tau^{2q+2} |u|^2_{H^{q+1}(0,T;H^1(\Omega))}\right) \\ & + C\left(h^{2p+2} |u|^2_{L^2(0,T;H^{p+1}(\Omega))} + |u|^2_{H^{q+1}(0,T;L^2(\Omega))}\right). \end{aligned} \tag{4.70}$$

Proof: It follows from (3.76) that

$$\int_0^T \|\xi\|^2 \,\mathrm{d}t \le C \sum_{m=1}^{M} \tau_m \left(\left\|\xi^-_{m-1}\right\|^2 + \left\|\eta^-_{m-1}\right\|^2 + \int_{I_m} R_m(\eta) \mathrm{d}t \right). \tag{4.71}$$

This and (3.89) yield

$$\int_0^T \|e\|^2 \,\mathrm{d}t \le C \sum_{m=1}^{M} \tau_m \left(\left\|\xi^-_{m-1}\right\|^2 + \left\|\eta^-_{m-1}\right\|^2 + \int_{I_m} R_m(\eta) \mathrm{d}t \right) + 2 \int_0^T \|\eta\|^2 \,\mathrm{d}t. \tag{4.72}$$

Now we use (3.88) with $m := m - 1 < M$, $\xi_0 = 0$, $\eta_0^- = \Pi_1 u^0 - u^0$ and get

$$\begin{aligned} \|e\|^2_{L^2(Q_T)} \;=\; & \int_0^T \|e\|^2 \,\mathrm{d}t \le C \sum_{m=1}^{M} \tau_m \left(\left\|\eta^-_{m-1}\right\|^2 + \int_{I_m} R_m(\eta) \mathrm{d}t \right. \\ & \left. + C e^{cT/\varepsilon} \left(\sum_{j=1}^{M} \left\|\eta_j^-\right\|^2 + \sum_{j=1}^{M} \int_{I_j} R_j(\eta) \mathrm{d}t \right) \right) + 2 \|\eta\|^2_{L^2(Q_T)}. \end{aligned} \tag{4.73}$$

Further, by (4.64), (4.67) and (4.35),

$$\sum_{j=1}^{M}\int_{I_j} R_j(\eta)\mathrm{d}t \le C\left(\varepsilon+\tfrac{1}{\varepsilon}\right)\left(h^{2p}\,|u|^2_{L^2(0,T;H^{p+1}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(0,T;H^1(\Omega))}\right) \tag{4.74}$$

$$\sum_{j=1}^{M}\left\|\eta_j^-\right\|^2 \le Ch^{2p}\,|u|^2_{C([0,T];H^{p+1}(\Omega))}\,, \tag{4.75}$$

$$\left\|\eta_{m-1}^-\right\|^2 \le Ch^{2p+2}\,|u|^2_{C([0,T],;H^{p+1}(\Omega))}\,, \tag{4.76}$$

$$\int_{I_m} R_m(\eta)\mathrm{d}t \le C\left(\varepsilon+\tfrac{1}{\varepsilon}\right)\left(h^{2p}\,|u|^2_{L^2(I_m;H^{p+1}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(I_m;H^1(\Omega))}\right). \tag{4.77}$$

Moreover, (4.34), (4.36) and (4.39) imply that

$$\begin{aligned}
\|\eta\|^2_{L^2(Q_T)} &= \sum_{m=1}^{M}\int_{I_m}\|\eta\|^2\,\mathrm{d}t \qquad (4.78)\\
&\le C\sum_{m=1}^{M}\left(h^{2p+2}\,|u|^2_{L^2(I_m;H^{p+2}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(I_m;L^2(\Omega))}\right)\\
&= C\left(h^{2p+2}\,|u|^2_{L^2(0,T;H^{p+1}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(0,T;L^2(\Omega))}\right).
\end{aligned}$$

From estimates (4.73)-(4.78) we get

$$\begin{aligned}
\|e\|^2_{L^2(Q_T)} \le\ & C\sum_{m=1}^{M}\tau_m\left(h^{2p+2}\,|u|^2_{C([0,T];H^{p+1}(\Omega))}\right. \qquad (4.79)\\
&+\left(\varepsilon+\frac{1}{\varepsilon}\right)\left(h_m^{2p}\,|u|^2_{L^2(I_m;H^{p+1}(\Omega))}+\tau_m^{2q+2}\,|u|^2_{H^{q+1}(I_m;H^1(\Omega))}\right)\Big)\\
&+e^{cT/\varepsilon}\left(h^{2p}\,|u|^2_{C([0,T];H^{p+1}(\Omega))}\right.\\
&+\left(\varepsilon+\frac{1}{\varepsilon}\right)\left(h^{2p}\,|u|^2_{L^2(0,T;H^{p+1}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(0,T;H^1(\Omega))}\right)\Big)\\
&+C\left(h^{2p+2}\,|u|^2_{L^2(0,T;H^{p+1}(\Omega))}+\tau^{2q+2}\,|u|^2_{H^{q+1}(0,T;L^2(\Omega))}\right).
\end{aligned}$$

This and the relation $\sum_{m=1}^{M}\tau_m = T$ yield the final estimate (4.70).

■

Remark 4.2: *Similarly as in Section 4.4, it is possible to formulate the $L^2(L^2)$-error estimate in the case of identical space meshes on all time levels.*

Theorem 4.10: *Let u be the exact solution satisfying the regularity condition (4.2) of the problem (1.1)-(1.3) with the Dirichlet data u_D defined by (4.45). Let U be the approximate solution to the problem (1.1)-(1.3) obtained by scheme (2.24) over spatial meshes $\mathcal{T}_{h,m}=\mathcal{T}_h$ for all $m=1,\dots,M$, and time partition I_m, $m=1,\dots,M$, satisfying condition (3.1), (3.2) and (3.77). Then there exist constant $C,c>0$ independent of h, τ, m, ε, u such that error estimate (4.70) holds.*

Part II

Numerical simulation of flow-induced vibrations

Chapter 5

Flow problem

This chapter will be devoted to the description of 2D compressible flow. Mostly we will be focused on the compressible viscous flow. Governing equations and their dimensionless form will be presented. In the second part of this chapter our aim will be concerned with the Arbitrary Lagrangian-Eulerian (ALE) method that plays crucial role in the treatment of time-dependence of the domain occupied by the fluid. Based on this the governing equations in the ALE formulation will be described.

5.1 Navier-Stokes equations for compressible viscous flow and their possible simplifications

We consider compressible flow in a bounded domain $\Omega_t \subset I\!R^2$ depending on time $t \in [0, T]$. We use the following notation: ρ – density, p – pressure, E – total energy, $\boldsymbol{v} = (v_1, v_2)$ – velocity vector, θ – absolute temperature, $c_v > 0$ – specific heat at constant volume, $\gamma > 1$ – Poisson adiabatic constant, $\mu > 0, \lambda = -2\mu/3$ – viscosity coefficients, $k > 0$ – heat conduction coefficient.

The compressible viscous flow is described by the Navier-Stokes equations. The detailed derivation of these equations can be found in [31].

The system consisting of the continuity equation, the Navier-Stokes equations and the energy equation is simply called the *compressible Navier-Stokes equations* and reads

$$\frac{\partial \rho}{\partial t} + \sum_{j=1}^{2} \frac{\partial (\rho v_j)}{\partial x_j} = 0, \tag{5.1}$$

$$\frac{\partial (\rho v_i)}{\partial t} + \sum_{j=1}^{2} \frac{\partial (\rho v_i v_j)}{\partial x_j} = \sum_{j=1}^{2} \frac{\partial \tau_{ij}}{\partial x_j}, \quad i = 1, 2, \tag{5.2}$$

$$\frac{\partial E}{\partial t} + \sum_{j=1}^{2} \frac{\partial (E v_j)}{\partial x_j} = \sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{2} \tau_{ij} v_j + k \frac{\partial \theta}{\partial x_i} \right). \tag{5.3}$$

This system needs to be completed by the thermodynamical relations

$$p = (\gamma - 1)\left(E - \frac{\rho\,|\boldsymbol{v}|^2}{2}\right), \tag{5.4}$$

$$\theta = \frac{1}{c_v}\left(\frac{E}{\rho} - \frac{|\boldsymbol{v}^2|}{2}\right). \tag{5.5}$$

Because of the modeling of the airflow we neglect the outer volume force and use the model of the *Newtonian* fluid. It means that the stress tensor $\mathcal{T} = \{\tau_{ij}\}$ is linearly dependent on the velocity deformation tensor $\mathbb{D}(\boldsymbol{v}) = \{d_{ij}\}$:

$$\mathcal{T} = (-p + \lambda \mathrm{div}\boldsymbol{v})\mathbb{I} + 2\mu\mathbb{D}(\boldsymbol{v}). \tag{5.6}$$

The term $\mathcal{T}^V = \left\{\tau_{ij}^V\right\} = \lambda \mathrm{div}\boldsymbol{v}\mathbb{I} + 2\mu\mathbb{D}(\boldsymbol{v})$ is the viscous part of the stress tensor. Thus,

$$\tau_{ij} = -p\delta_{ij} + \tau_{ij}^V, \tag{5.7}$$

$$\tau_{ij}^V = \lambda \mathrm{div}\boldsymbol{v}\delta_{ij} + 2\mu d_{ij}(\boldsymbol{v}), \quad d_{ij}(\boldsymbol{v}) = \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right), \tag{5.8}$$

$\mathbb{I}$ is the unit tensor and μ and λ are called the first and second *viscous coefficients*. In the kinetic theory of gases the conditions

$$\mu \geq 0, \qquad 3\lambda + 2\mu \geq 0, \tag{5.9}$$

are derived. The condition $3\lambda + 2\mu = 0$ holds for monoatomic gases, but this is usually used even in the case of more complicated gases. We shall assume that μ and λ are constants.

The above system (5.1)-(5.3) is equipped with the initial conditions prescribing the state of the flow in time $t = 0$

$$\begin{aligned} \boldsymbol{v}(\boldsymbol{x}, 0) &= \boldsymbol{v}^0(\boldsymbol{x}), \\ \rho(\boldsymbol{x}, 0) &= \rho^0(\boldsymbol{x}), \\ p(\boldsymbol{x}, 0) &= p^0(\boldsymbol{x}), \end{aligned} \tag{5.10}$$

where the initial data $\boldsymbol{v}^0,\ \rho^0,\ p^0$ are given.

The behavior of the fluid flow on the boundary $\partial\Omega_t$ of the domain Ω_t is described by the boundary conditions. We assume that $\partial\Omega_t$ is formed by mutually disjoint parts, where Γ_I is the inlet, Γ_O is the outlet and Γ_{W_t} denotes impermeable walls that may move in dependence on time. Hence, $\partial\Omega_t = \Gamma_I \cup \Gamma_O \cup \Gamma_{W_t}$. We consider the following boundary conditions:

$$\begin{aligned} &\text{Inlet } \Gamma_I: \quad \rho|_{\Gamma_I\times(0,T)} = \rho_D, \quad \boldsymbol{v}|_{\Gamma_I\times(0,T)} = \boldsymbol{v}_D = (v_{D1}, v_{D2}), \\ &\sum_{j=1}^{2}\left(\sum_{i=1}^{2}\tau_{ij}^V n_i\right)v_j + k\frac{\partial\theta}{\partial\boldsymbol{n}} = 0 \quad \text{on } \Gamma_I \times (0,T); \end{aligned} \tag{5.11}$$

$$\text{Moving wall } \Gamma_{W_t}: \quad \boldsymbol{v} = \boldsymbol{z}(t), \quad \frac{\partial\theta}{\partial\boldsymbol{n}} = 0 \quad \text{on } \{(\boldsymbol{x}, t);\ \boldsymbol{x} \in \Gamma_{W_t},\ t \in (0,T)\};$$

$$\text{Outlet } \Gamma_O: \quad \sum_{i=1}^{2}\tau_{ij}^V n_i = 0, \quad \frac{\partial\theta}{\partial\boldsymbol{n}} = 0, \quad j = 1, 2, \quad \text{on } \Gamma_O \times (0,T)$$

with given data $\rho_D,\ \boldsymbol{v}_D,\ \boldsymbol{z}_D$, where $\boldsymbol{z}_D$ represents the velocity of a moving wall. By $\boldsymbol{n}$ we denote the unit outer normal to the boundary $\partial\Omega_t$.

If we set $\mu = \lambda = k = 0$ in (5.1-5.5), we obtain the model of inviscid compressible flow, described by the continuity equation, the Euler equations, the energy equation and thermodynamical relations:

$$\frac{\partial \rho}{\partial t} + \sum_{j=1}^{2} \frac{\partial(\rho v_j)}{\partial x_j} = 0, \tag{5.12}$$

$$\frac{\partial(\rho v_i)}{\partial t} + \sum_{j=1}^{2} \frac{\partial(\rho v_i v_j + \delta_{ij} p)}{\partial x_j} = 0, \quad i = 1, 2, \tag{5.13}$$

$$\frac{\partial E}{\partial t} + \sum_{j=1}^{2} \frac{\partial((E+p)v_j)}{\partial x_j} = 0, \tag{5.14}$$

$$p = (\gamma - 1)(E - \rho\,|\boldsymbol{v}|^2/2). \tag{5.15}$$

This system is simply called the *compressible Euler equations.*

System (5.1)-(5.3) can be written in the form

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{i=1}^{2} \frac{\partial \boldsymbol{f}_i(\boldsymbol{w})}{\partial x_i} = \sum_{i=1}^{2} \frac{\partial \boldsymbol{R}_i(\boldsymbol{w}, \nabla \boldsymbol{w})}{\partial x_i}, \tag{5.16}$$

where

$$\boldsymbol{w} = (w_1, w_2, w_3, w_4)^T = (\rho, \rho v_1, \rho v_2, E)^T = I\!R^4 \tag{5.17}$$

is the so-called *state vector.* The functions ρ, v_1, v_2, p are called *primitive* or *physical variables*, whereas $w_1 = \rho,\ w_2 = \rho v_1,\ w_3 = \rho v_2,\ w_4 = E$ are *conservative variables.* Further,

$$\boldsymbol{f}_i(\boldsymbol{w}) = (f_{i1}(\boldsymbol{w}), \ldots, f_{i4}(\boldsymbol{w}))^T = (\rho v_i, \rho v_1 v_i + \delta_{1i} p, \rho v_2 v_i + \delta_{2i} p, (E+p)v_i)^T,\ i = 1, 2, \tag{5.18}$$

is the *flux* of the quantity $\boldsymbol{w}$ in the direction x_i. Often, $\boldsymbol{f}_i, i = 1, 2$, are called *inviscid fluxes.* If we express inviscid fluxes with the aid of the conservative variables, we obtain

$$\boldsymbol{f}_1(\boldsymbol{w}) = \begin{pmatrix} w_2 \\ w_2^2/w_1 + (\gamma-1)\left[w_4 - (w_2^2 + w_3^2)/(2w_1)\right] \\ w_2 w_3 / w_1 \\ \frac{w_2}{w_1}\left[\gamma w_4 - (\gamma-1)(w_2^2 + w_3^2)/(2w_1)\right] \end{pmatrix}, \tag{5.19}$$

$$\boldsymbol{f}_2(\boldsymbol{w}) = \begin{pmatrix} w_2 \\ w_2 w_3 / w_1 \\ w_3^2/w_1 + (\gamma-1)\left[w_4 - (w_2^2 + w_3^2)/(2w_1)\right] \\ \frac{w_3}{w_1}\left[\gamma w_4 - (\gamma-1)(w_2^2 + w_3^2)/(2w_1)\right] \end{pmatrix}. \tag{5.20}$$

The domain of definition of the vector-valued functions $\boldsymbol{f}_i$ is the open set $D \subset I\!R^4$ of vectors $\boldsymbol{w} = (w_1, \ldots, w_4)^T$ such that the corresponding density and pressure are

positive:

$$D = \left\{ \boldsymbol{w} \in I\!R^4;\ w_1 = \rho > 0,\ w_2 = \rho v_1 \in I\!R,\ w_3 = \rho v_2 \in I\!R, \right. \\ \left. w_m - \sum_{i=2}^{m-1} w_i^2/(2w_1) = p/(\gamma - 1) > 0 \right\}. \tag{5.21}$$

Obviously, $\boldsymbol{f}_s \in \mathcal{C}^1(D)^m$. By $\boldsymbol{R}_i(\boldsymbol{w}, \nabla \boldsymbol{w})$ we denote the so-called *viscous fluxes*

$$\boldsymbol{R}_i(\boldsymbol{w}, \nabla \boldsymbol{w}) = (R_{i1}, \ldots, R_{i4})^T = (0, \tau_{i1}^V, \tau_{i2}^V, \tau_{i1}^V v_1 + \tau_{i2}^V v_2 + k\partial\theta/\partial x_i)^T,\ i = 1, 2. \tag{5.22}$$

Now, let us have a short insight into properties of the inviscid fluxes $\boldsymbol{f}_s,\ s = 1, 2$. From (5.19) and (5.20) we can see that the inviscid fluxes are homogeneous, which means that they fulfill the relation

$$\boldsymbol{f}_s(\alpha \boldsymbol{w}) = \alpha \boldsymbol{f}_s(\boldsymbol{w}), \quad \alpha > 0,\ s = 1, 2. \tag{5.23}$$

Based on this it is possible to show that

$$\boldsymbol{f}_s(\boldsymbol{w}) = \mathbb{A}_s(\boldsymbol{w})\boldsymbol{w}, \quad s = 1, 2, \tag{5.24}$$

where $\mathbb{A}_s(w)$ are the 4×4 Jacobi matrices of the mapping $\boldsymbol{f}_s$ defined for $\boldsymbol{w} \in D$ by

$$\mathbb{A}_s(w) = \frac{D\boldsymbol{f}_s(w)}{D\boldsymbol{w}} = \left(\frac{\partial f_{si}(\boldsymbol{w})}{\partial w_j} \right)_{i,j=1}^m. \tag{5.25}$$

For $\boldsymbol{w} \in D$ and $\boldsymbol{n} = (n_1, n_2)^T \in I\!R^2,\ |\boldsymbol{n}| = 1$, we set

$$\mathcal{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) n_s, \tag{5.26}$$

which is the flux of the quantity $\boldsymbol{w}$ in the direction $\boldsymbol{n}$. The Jacobi matrix $D\mathcal{P}(\boldsymbol{w}, \boldsymbol{n})/D\boldsymbol{w}$ can be expressed in the form

$$\frac{D\mathcal{P}(\boldsymbol{w}, \boldsymbol{n})}{D\boldsymbol{w}} = \mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \mathbb{A}_s(\boldsymbol{w}) n_s. \tag{5.27}$$

The viscous fluxes $\boldsymbol{R}_s(\boldsymbol{w}, \nabla \boldsymbol{w})$ have a property similar to the homogeneity of the inviscid fluxes (5.23). The term $\boldsymbol{R}_s(\boldsymbol{w}, \nabla \boldsymbol{w})$ can be expressed in the form

$$\boldsymbol{R}_s(\boldsymbol{w}, \nabla \boldsymbol{w}) = \sum_{j=1}^{2} \mathbb{K}_{sj}(\boldsymbol{w}) \frac{\partial \boldsymbol{w}}{\partial x_j}, s = 1, 2, \tag{5.28}$$

where $\mathbb{K}_{sj}$ are 4×4 matrices dependent on $\boldsymbol{w}$ and independent of $\nabla \boldsymbol{w}$. Explicit formulae for $\mathbb{K}_{sj}$ read

$$\mathbb{K}_{11}(\boldsymbol{w}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -(2\mu + \lambda)\frac{w_2}{w_1^2} & (2\mu + \lambda)\frac{1}{w_1} & 0 & 0 \\ -\mu\frac{w_3}{w_1^2} & 0 & \frac{\mu}{w_1} & 0 \\ \{\mathbb{K}_{11}\}_{41} & \left(2\mu + \lambda - \frac{k}{c_v}\right)\frac{w_2}{w_1^2} & \left(\mu - \frac{k}{c_v}\frac{w_3}{w_1^2}\right) & \frac{k}{c_v w_1} \end{pmatrix},$$

$$\mathbb{K}_{12}(\boldsymbol{w}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\lambda\frac{w_3}{w_1^2} & 0 & \frac{\lambda}{w_1} & 0 \\ -\mu\frac{w_2}{w_1^2} & \frac{\mu}{w_1} & 0 & 0 \\ -(\lambda+\mu)\frac{w_2 w_3}{w_1^3} & \mu\frac{w_3}{w_1^2} & \lambda\frac{w_2}{w_1^2} & 0 \end{pmatrix},$$

$$\mathbb{K}_{21}(\boldsymbol{w}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\mu\frac{w_3}{w_1^2} & 0 & \frac{\mu}{w_1} & 0 \\ -\lambda\frac{w_2}{w_1^2} & \frac{\lambda}{w_1} & 0 & 0 \\ -(\lambda+\mu)\frac{w_2 w_3}{w_1^3} & \lambda\frac{w_3}{w_1^2} & \mu\frac{w_2}{w_1^2} & 0 \end{pmatrix},$$

$$\mathbb{K}_{22}(\boldsymbol{w}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\mu\frac{w_2}{w_1^2} & \frac{\mu}{w_1} & 0 & 0 \\ -(2\mu+\lambda)\frac{w_3}{w_1^2} & 0 & (2\mu+\lambda)\frac{1}{w-1} & 0 \\ \{\mathbb{K}_{22}\}_{41} & \left(\mu-\frac{k}{c_v}\right)\frac{w_2}{w_1^2} & \left(2\mu+\lambda-\frac{k}{c_v}\right)\frac{w_3}{w_1^2} & \frac{k}{c_v w_1} \end{pmatrix},$$

where

$$\begin{aligned} \{\mathbb{K}_{11}\}_{41} &= -(2\mu+\lambda)\frac{w_2^2}{w_1^3} - \mu\frac{w_3^2}{w_1^3} + \frac{k}{c_v}\left(-\frac{w_4}{w_1^2} + \frac{w_2^2+w_3^2}{w_1^3}\right), \\ \{\mathbb{K}_{22}\}_{41} &= -\mu\frac{w_2^2}{w_1^3} - (2\mu+\lambda)\frac{w_3^2}{w_1^3} + \frac{k}{c_v}\left(-\frac{w_4}{w_1^2} + \frac{w_2^2+w_3^2}{w_1^3}\right). \end{aligned}$$

5.2 Dimensionless form of the Navier-Stokes equations

In order to be able to carry out experiments on small models and to transfer the results to the original real flow, the *dimensionless form* of the Navier-Stokes equations is used.

Let us introduce the following positive *reference quantities*: a reference length L^*, a reference velocity U^* (scalar quantity), a reference density ρ^*, a reference viscosity μ^* and a reference heat conduction coefficient k^*. All other reference quantities can be derived from these basic ones: we choose L^*/U^* for t, $\rho^* U^{*2}$ for both p and E, U^{*2}/c_v for θ. We denote by primes the *dimensionless quantities*

$$\begin{aligned} x_i' &= \frac{x_i}{L^*}, \quad v_i' = \frac{v_i}{U^*}, \quad \boldsymbol{v}' = \frac{\boldsymbol{v}}{U^*}, \quad \rho' = \frac{\rho}{\rho^*}, \\ p' &= \frac{p}{\rho^* U^{*2}}, \quad E' = \frac{E}{\rho^* U^{*2}}, \quad \theta' = \frac{c_v\theta}{U^{*2}}, \quad t' = \frac{tU^*}{L^*}, \\ \mu' &= \frac{\mu}{\mu^*}, \quad \lambda' = \frac{\lambda}{\lambda^*}, \quad k' = \frac{k}{k^*}. \end{aligned} \tag{5.29}$$

Now, we transform equation (5.2) for $i = 1$. The reader will see that in the similar way it is possible to transform also other equations of system (5.1)-(5.3).

$$\begin{aligned} \frac{\partial \rho v_1}{\partial t}(\boldsymbol{x}, t) &= \rho^* U^* \frac{\partial \rho' v_1'}{\partial t}(\boldsymbol{x}', t') = \rho^* U^* \frac{\partial \rho' v_1'}{\partial t'}(\boldsymbol{x}', t')\frac{\partial t'}{\partial t} \\ &= \frac{\rho^* U^{*2}}{L^*}\frac{\partial \rho' v_1'}{\partial t'}(\boldsymbol{x}', t'), \end{aligned} \tag{5.30}$$

$$\begin{aligned}
\frac{\partial \rho v_1^2}{\partial x_1}(\boldsymbol{x},t) &= \rho^* U^{*2}\frac{\partial \rho' v_1'^2}{\partial x_1}(\boldsymbol{x}',t') = \rho^* U^{*2}\frac{\partial \rho' v_1'^2}{\partial x_1'}(\boldsymbol{x}',t')\frac{\partial x_1'}{\partial x_1} \\
&= \frac{\rho^* U^{*2}}{L^*}\frac{\partial \rho' v_1'^2}{\partial x_1}(\boldsymbol{x}',t') \qquad (5.31) \\
\frac{\partial p}{\partial x_1}(\boldsymbol{x},t) &= \rho^* U^{*2}\frac{\partial p'}{\partial x_1}(\boldsymbol{x}',t') = \rho^* U^{*2}\frac{\partial p'}{\partial x_1}(\boldsymbol{x}',t')\frac{\partial x_1'}{\partial x_1} \\
&= \frac{\rho^* U^{*2}}{L^*}\frac{\partial p'}{\partial x_1'}, \qquad (5.32) \\
\frac{\partial^2 v_1}{\partial x_1^2}(\boldsymbol{x},t) &= \frac{\partial}{\partial x_1}\left(\frac{\partial v_1}{\partial x_1}(\boldsymbol{x},t)\right) = U^*\frac{\partial}{\partial x_1}\left(\frac{\partial v_1'}{\partial x_1}(\boldsymbol{x}',t')\right) \\
&= U^*\frac{\partial}{\partial x_1}\left(\frac{\partial v_1'}{\partial x_1'}(\boldsymbol{x}',t')\frac{\partial x_1'}{\partial x_1}\right) = \frac{U^*}{L^*}\frac{\partial}{\partial x_1}\left(\frac{\partial v_1'}{\partial x_1'}(\boldsymbol{x}',t')\right) \\
&= \frac{U^*}{L^*}\frac{\partial^2 v_1'}{\partial x_1'^2}(\boldsymbol{x}',t')\frac{\partial x_1'}{\partial x_1} = \frac{U^*}{L^{*2}}\frac{\partial^2 v_1'}{\partial x_1'^2}(\boldsymbol{x}',t'). \qquad (5.33)
\end{aligned}$$

Remaining terms are treated in the same way:

$$\begin{aligned}
\frac{\partial \rho v_1 v_2}{\partial x_2}(\boldsymbol{x},t) &= \frac{\rho^* U^{*2}}{L^*}\frac{\partial \rho' v_1' v_2'}{\partial x_2'}(\boldsymbol{x}',t'), \qquad (5.34) \\
\frac{\partial^2 v_1}{\partial x_2^2}(\boldsymbol{x},t) &= \frac{U^*}{L^{*2}}\frac{\partial^2 v_1'}{\partial x_1'^2}(\boldsymbol{x}',t'), \qquad (5.35) \\
\frac{\partial^2 v_2}{\partial x_1 \partial x_2}(\boldsymbol{x},t) &= \frac{U^*}{L^{*2}}\frac{\partial^2 v_2'}{\partial x_1' \partial x_2'}(\boldsymbol{x}',t'). \qquad (5.36)
\end{aligned}$$

Using relations (5.30)-(5.36) we obtain the dimensionless form of equation (5.2) for $i = 1$:

$$\frac{\partial(\rho' v_1')}{\partial t'} + \sum_{j=1}^{2}\frac{\partial(\rho' v_1' v_j')}{\partial x_j'} = -\frac{\partial p'}{\partial x_1'} + \frac{\mu^*}{\rho^* U^* L^*}\sum_{j=1}^{2}\frac{\partial \tau_{1j}'^V}{\partial x_j'}. \qquad (5.37)$$

Now, if we introduce the Reynolds and Prandtl numbers by

$$Re = \frac{\rho^* U^* L^*}{\mu^*}, \quad Pr = \frac{c_p \mu^*}{k^*}, \qquad (5.38)$$

system (5.1)-(5.3) can be written in the dimensionless form

$$\begin{aligned}
\frac{\partial \rho'}{\partial t'} + \sum_{j=1}^{2}\frac{\partial(\rho' v_j')}{\partial x_j'} &= 0, \qquad (5.39) \\
\frac{\partial(\rho' v_i')}{\partial t'} + \sum_{j=1}^{2}\frac{\partial(\rho' v_i' v_j')}{\partial x_j'} &= -\frac{\partial p'}{\partial x_i'} + \frac{1}{Re}\sum_{j=1}^{2}\frac{\partial \tau_{ij}'^V}{\partial x_j'}, \quad i = 1,2, \qquad (5.40) \\
\frac{\partial E'}{\partial t'} + \sum_{j=1}^{2}\frac{\partial(E' v_j')}{\partial x_j'} &= -\sum_{j=1}^{2}\frac{\partial(p' v_j')}{\partial x_j'} + \frac{1}{Re}\left(\sum_{j=1}^{2}\frac{\partial}{\partial x_j'}\left(\sum_{i=1}^{2}\tau_{ji}'^V v_i'\right)\right) \qquad (5.41) \\
&\quad + \frac{\gamma k'}{RePr}\nabla\theta'.
\end{aligned}$$

Here the viscous part of the stress tensor has the dimensionless form

$$\tau_{ij}'^{V} = \lambda' \mathrm{div}\boldsymbol{v}'\delta_{ij} + 2\mu' d'_{ij}(\boldsymbol{v}'), \quad d'_{ij}(\boldsymbol{v}') = \left(\frac{\partial v'_i}{\partial x'_j} + \frac{\partial v'_j}{\partial x'_i}\right). \tag{5.42}$$

Finally we add the thermodynamical relations (5.4) and (5.5) expressed in the dimensionless form

$$p' = (\gamma - 1)\left(E' - \frac{\rho' \left|\boldsymbol{v}'\right|^2}{2}\right), \tag{5.43}$$

$$\theta' = \left(\frac{E'}{\rho'} - \frac{|\boldsymbol{v}'^2|}{2}\right). \tag{5.44}$$

Based on (5.39)-(5.41), the system of governing equations can be formulated in the dimensionless conservative form

$$\frac{\partial \boldsymbol{w}'}{\partial t'} + \sum_{i=1}^{2} \frac{\partial \boldsymbol{f}_i(\boldsymbol{w}')}{\partial x'_i} = \sum_{i=1}^{2} \frac{\partial \boldsymbol{R}'_i(\boldsymbol{w}', \nabla \boldsymbol{w}')}{\partial x'_i} \quad \text{in } Q_{T'}, \tag{5.45}$$

where $Q_{T'} = \Omega' \times (0, T')$, $\Omega' = \{(x'_1, x'_2);\ x'_1 = x_1/L^*,\ x'_2 = x_2/L^*,\ (x_1, x_2) \in \Omega\}$, $T' = TU^*/L^*$ and

$$\begin{aligned}
\boldsymbol{w}' &= (\rho', \rho' v'_1, \rho' v'_2, E')^T \in I\!R^4 \qquad (5.46)\\
\boldsymbol{w}' &= \boldsymbol{w}'(\boldsymbol{x}', t'),\ \boldsymbol{x}' \in \Omega',\ t' \in (0, T'),\\
\boldsymbol{f}_i(\boldsymbol{w}') &= (\rho' v'_i, \rho' v'_1 v'_i + \delta_{1i} p', \rho' v'_2 v'_i + \delta_{2i} p', (E' + p') v'_i)^T\\
\boldsymbol{R}'_i(\boldsymbol{w}', \nabla \boldsymbol{w}') &= \frac{1}{Re}(0, \tau'^{V}_{i1}, \tau'^{V}_{i2}, \tau'^{V}_{i1} v'_1 + \tau'^{V}_{i2} v'_2 + \frac{\gamma k'}{Pr}\frac{\partial \theta'}{\partial x'_i})^T.
\end{aligned}$$

Formally, system (5.45) has the same form and properties as system (5.16) and therefore the primes are omitted, if we consider the equations in the dimensionlles form.

5.3 ALE method

In order to simulate flow in a time-dependent domain, we employ *the Arbitrary Eulerian-Lagrangian (ALE) method*. Let us denote by $\Omega_{ref} = \Omega_0$ the computational domain at the initial time. It is also called the *reference or original configuration.* By $\mathcal{A}_t$ we denote a smooth, one-to-one mapping of the reference configuration onto the computational domain Ω_t at time t (the so-called *current configuration*), i.e.

$$\begin{aligned}
\mathcal{A}_t : \bar{\Omega}_{ref} &\longrightarrow \bar{\Omega}_t, \qquad (5.47)\\
\boldsymbol{X} &\longmapsto \boldsymbol{x}(\boldsymbol{X}, t) = \mathcal{A}_t(\boldsymbol{X}).
\end{aligned}$$

We call $\mathcal{A}_t$ the ALE mapping.

Based on this mapping we define the *domain velocity* $\tilde{\boldsymbol{z}}$ at all points $\boldsymbol{X}$ of the reference configuration Ω_{ref} for each time level:

$$\begin{aligned}
\tilde{\boldsymbol{z}} : \bar{\Omega}_{ref} \times (0, T) &\longrightarrow I\!R^2, \qquad (5.48)\\
\tilde{\boldsymbol{z}}(\boldsymbol{X}, t) = \tfrac{\partial}{\partial t}\boldsymbol{x}(\boldsymbol{X}, t) &= \tfrac{\partial}{\partial t}\mathcal{A}_t(\boldsymbol{X}),
\end{aligned}$$

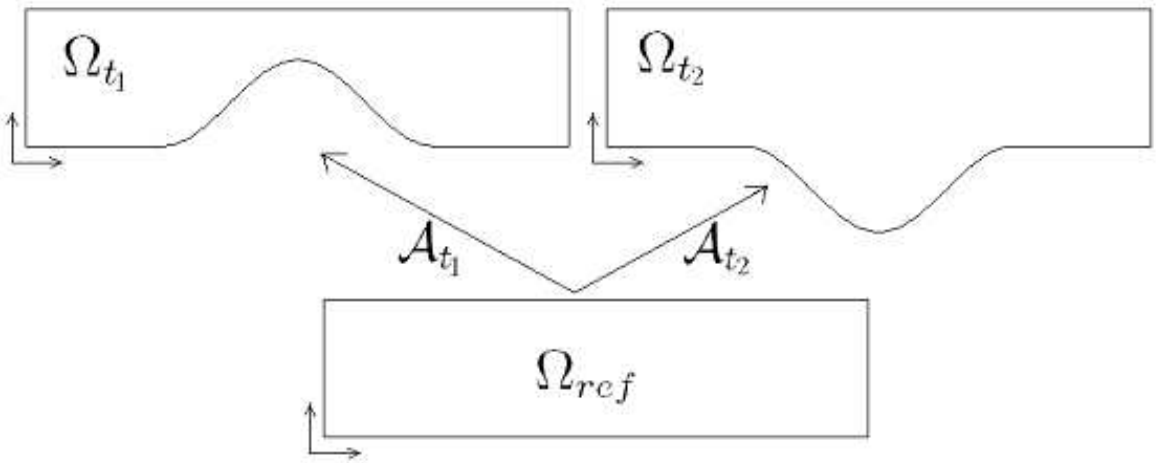


Figure 5.1: ALE mapping

which can be transformed to the space coordinates $\boldsymbol{x}$ by the relation

$$\boldsymbol{z} = \tilde{\boldsymbol{z}}(\mathcal{A}_t^{-1}(\boldsymbol{x}), t),\ t \in (0,T),\ \boldsymbol{x} \in \bar{\Omega}_t. \tag{5.49}$$

With the aid of the ALE mapping we introduce the *ALE derivative* $\frac{D^{\mathcal{A}}}{Dt}$ of a smooth function $f = f(\boldsymbol{x}, t),\ \boldsymbol{x} \in \Omega_t,\ t \in (0,T)$. We set

$$\frac{D^{\mathcal{A}}}{Dt} f(\boldsymbol{x}, t) = \frac{\partial \tilde{f}}{\partial t}(\boldsymbol{X}, t), \quad \boldsymbol{X} = \mathcal{A}_t^{-1}(\boldsymbol{x}), \tag{5.50}$$

where

$$\tilde{f}(\boldsymbol{X}, t) = f(\mathcal{A}_t(\boldsymbol{X}), t), \quad \boldsymbol{X} \in \Omega_{ref},\ t \in (0,T). \tag{5.51}$$

Using the chain rule, we find that

$$\frac{D^{\mathcal{A}}}{Dt} f = \frac{\partial f}{\partial t} + (\boldsymbol{z} \cdot \nabla) f, \tag{5.52}$$

$$\frac{D^{\mathcal{A}}}{Dt} f = \frac{\partial f}{\partial t} + \mathrm{div}(f\boldsymbol{z}) - f\mathrm{div}(\boldsymbol{z}). \tag{5.53}$$

It follows from (5.52)-(5.53) that the time derivative of a function f can be expressed in the form

$$\frac{\partial f}{\partial t} = \frac{D^{\mathcal{A}}}{Dt} f - (\boldsymbol{z} \cdot \nabla) f, \tag{5.54}$$

$$\frac{\partial f}{\partial t} = \frac{D^{\mathcal{A}}}{Dt} f + f\mathrm{div}\boldsymbol{z} - \mathrm{div}(f\boldsymbol{z}). \tag{5.55}$$

The application of relations (5.54) and (5.55) to the statevector $\boldsymbol{w}$ leads to the ALE formulations of the governing system (5.16).

If we use the relation (5.54), we will get the first possible ALE formulation

$$\frac{D^{\mathcal{A}}\boldsymbol{w}}{Dt} + \sum_{i=1}^{2} \frac{\partial \boldsymbol{f}_i(\boldsymbol{w})}{\partial x_i} - \sum_{i=1}^{2} z_i \frac{\partial \boldsymbol{w}}{\partial x_i} = \sum_{i=1}^{2} \frac{\partial \boldsymbol{R}_i(\boldsymbol{w}, \nabla \boldsymbol{w})}{\partial x_i}. \tag{5.56}$$

On the basis of (5.55) we obtain

$$\frac{D^{\mathcal{A}}\boldsymbol{w}}{Dt} + \sum_{i=1}^{2} \frac{\partial \boldsymbol{g}_i(\boldsymbol{w})}{\partial x_i} + \boldsymbol{w}\mathrm{div}\boldsymbol{z} = \sum_{i=1}^{2} \frac{\partial \boldsymbol{R}_i(\boldsymbol{w}, \nabla \boldsymbol{w})}{\partial x_i}. \tag{5.57}$$

Here $\boldsymbol{g}_i,\ i = 1, 2$, is the ALE flux of $\boldsymbol{w}$ in the direction x_i defined as

$$\boldsymbol{g}_i(\boldsymbol{w}) = \boldsymbol{f}_i(\boldsymbol{w}) - z_i\boldsymbol{w}. \tag{5.58}$$

Numerical experiments carried out in [56] proved better applicability of the ALE form (5.57). For this reason we shall be concerned just with the discretization of (5.57).

Chapter 6

Problem of an elastic structure

In this chapter we shall pay attention to the mathematical description of the deformation of elastic bodies. For the good explication of all relations we start from the presentation of terms like *stress tensor* and *strain tensor.* This allows us to derive the static elasticity problem based on which we pass to the dynamical equations of elastic body with the aid of *d'Alembert's principle.* The complex overview of the problem of elastic structure can be found, for example, in [53].

We can meet the stress tensor and other quantities also in the description of flow problem. For this reason we shall denote these quantities with the index b, if necessary.

6.1 The stress tensor

The *solid body* is a domain $\Omega^b \subset I\!R^3$. An open subset $\mathcal{V}$ of the solid body Ω^b such that $\bar{\mathcal{V}} \subset \Omega^b$ fulfilling the condition of the Lipschitz-continuous boundary $\partial\mathcal{V}$, will be called a control volume. If $\boldsymbol{x} \in \partial\mathcal{V}$ then we denote by $\boldsymbol{n}$ the unit outer normal to $\partial\mathcal{V}$ in $\boldsymbol{x}$. The *volume force* $\boldsymbol{f} = (f_1, f_2, f_3)^T$ is the density of the force acting on any particle contained in a control volume of the solid body Ω^b. We assume $\boldsymbol{f} \in C(\bar{\Omega})^3$.

The *stress vector* $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{n})$ represents the density of the inner (surface) forces in the body acting from a part $\Omega^b \setminus \mathcal{V}$ on a part $\bar{\mathcal{V}}$ at the point $\boldsymbol{x}$ and depend on the position $\boldsymbol{x}$ and on the unit outer normal $\boldsymbol{n}$. We assume that $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{n}) \in C(\bar{\Omega} \times \mathcal{S})^3$, where $\mathcal{S}$ is the surface of the unit sphere.

As a consequence of the third Newton's law we obtain $\boldsymbol{T}(\boldsymbol{x}, \boldsymbol{n}) = -\boldsymbol{T}(\boldsymbol{x}, -\boldsymbol{n})$.

The stress vector can be determined by its values for the normals parallel to the axis of the coordinates. Let us set

$$\tau_{ji}^b = T_i(\boldsymbol{x}, \boldsymbol{e}_j), \quad i, j = 1, 2, 3, \tag{6.1}$$

where $\boldsymbol{e}_1 = (1, 0, 0),\ \boldsymbol{e}_2 = (0, 1, 0),\ \boldsymbol{e}_3 = (0, 0, 1)$. Quantities $\tau_{ji}^b(\boldsymbol{x}),\ i, j = 1, 2, 3$, are called *components of the stress tensor.* Especially $\tau_{ii}^b,\ i = 1, 2, 3$, are denoted *normal stresses* and $\tau_{ji}^b,\ i \neq j,\ i, j = 1, 2, 3$, are *shear stresses.* Similarly to fluid mechanics it is possible to show that

$$T_i(\boldsymbol{x}, \boldsymbol{n}) = \sum_{j=1}^{3} n_j \tau_{ji}^b(\boldsymbol{x}), \quad i = 1, 2, 3. \tag{6.2}$$

The matrix $\left\{\tau_{ij}^b(\boldsymbol{x})\right\}_{i,j=1,2,3}$ represents the *stress tensor* at the point $\boldsymbol{x}$.

Supposing that $\tau_{ij}^b \in C^1(\Omega^b)$ and $f_i \in C(\Omega^b),\ i,j = 1,2,3$, we can express the equilibrium of the forces by the equations

$$\sum_{j=1}^{3} \frac{\partial \tau_{ji}^b}{\partial x_j}(\boldsymbol{x}) + f_i(\boldsymbol{x}) = 0, \quad i = 1,2,3,\ \boldsymbol{x} \in \Omega^b, \tag{6.3}$$

For the derivation see e.g. [53]. Further, it is possible to prove the symmetry of the stress tensor

$$\tau_{ij}^b(\boldsymbol{x}) = \tau_{ji}^b(\boldsymbol{x}), \quad i,j = 1,2,3,\ \boldsymbol{x} \in \Omega^b \tag{6.4}$$

by the equilibrium of the angular momentum.

6.2 The strain tensor

Now, we shall define the *tensor of finite strain* characterizing variation of the distance of two points of the body Ω^b. Let have a point $\boldsymbol{x} \in \Omega^b$. A deformation changes the body Ω^b to $\tilde{\Omega}^b$ and the point $\boldsymbol{x}$ change the position to point $\boldsymbol{y}$. We suppose that there exists a function of points describing the deformation

$$\boldsymbol{x} \longrightarrow \boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}), \tag{6.5}$$

where $\boldsymbol{u}(\boldsymbol{x})$ is a *vector of displacement.* We assume that the transformation $\boldsymbol{y}(\boldsymbol{x})$ is the diffeomorphism and $u_i \in C^3(\Omega^b)$.

Let $\boldsymbol{v} \in I\!R^3$ be an arbitrary vector and the point $\boldsymbol{x} + t\boldsymbol{v} \in \Omega^b$, where $t \in I\!R$ (sufficiently small). Now, we shall explore the function $\varphi(t)$ representing square difference of the length of the line segments determined by points $\boldsymbol{x}$ and $\boldsymbol{x} + t\boldsymbol{v}$ before and after the deformation

$$\begin{aligned}
\varphi(t) &= \left|[\boldsymbol{x} + t\boldsymbol{v} + \boldsymbol{u}(\boldsymbol{x} + t\boldsymbol{v})] - [\boldsymbol{x} + \boldsymbol{x}(\boldsymbol{x})]\right|^2 - \left|\boldsymbol{x} + t\boldsymbol{v} - \boldsymbol{x}\right|^2 \\
&= \left|t\boldsymbol{v} + \boldsymbol{u}(\boldsymbol{x} + t\boldsymbol{v}) - \boldsymbol{u}(\boldsymbol{x})\right|^2 - t^2\left|\boldsymbol{v}\right|^2 \\
&= \sum_{i=1}^{3}(tv_i + u_i(\boldsymbol{x} + t\boldsymbol{v}) - u_i(\boldsymbol{x}))^2 - t^2\left|\boldsymbol{v}\right|^2 \\
&= \sum_{i=1}^{3}(u_i(\boldsymbol{x} + t\boldsymbol{v}) - u_i(\boldsymbol{x}))^2 + 2t\sum_{i=1}^{3} v_i(u_i(\boldsymbol{x} + t\boldsymbol{v}) - u_i(\boldsymbol{x})).
\end{aligned} \tag{6.6}$$

Further we set $\psi_i(\tau) = u_i(\boldsymbol{x} + \tau t\boldsymbol{v}),\ i = 1,2,3$. Based on the assumption of the smoothness of $\boldsymbol{u}$ we have

$$\psi_i'(\tau) = \sum_{j=1}^{3} \frac{\partial u_i}{\partial x_j}(\boldsymbol{x} + \tau t\boldsymbol{v})tv_j, \quad i = 1,2,3. \tag{6.7}$$

Then

$$\psi_i(1) - \psi_i(0) = \int_0^1 \psi_i'(\tau)\mathrm{d}\tau, \quad i = 1,2,3. \tag{6.8}$$

Now, we can express the function φ with the aid of $\psi_i'(\tau)$:

$$\begin{aligned}\varphi(t) &= 2t\sum_{i=1}^{3}(\psi_i(1)-\psi_i(0))v_i+\sum_{i=1}^{3}(\psi_i(1)-\psi_i(0))^2 \qquad (6.9)\\ &= 2t^2\sum_{i,j=1}^{3}\int_0^1\frac{\partial u_i}{\partial x_j}(\boldsymbol{x}+\tau t\boldsymbol{v})v_jv_i\mathrm{d}\tau+t^2\sum_{i=1}^{3}\left(\sum_{j=1}^{3}\int_0^1\frac{\partial u_i}{\partial x_j}(\boldsymbol{x}+\tau t\boldsymbol{v})v_j\mathrm{d}\tau\right)^2.\end{aligned}$$

The function φ has at the point 0 the smooth second order derivative. It holds that $\varphi(0)=\varphi'(0)=0$ and only $\varphi''(0)$ can be nonzero. For this reason φ'' is the lowest order derivative of φ, which can represent the deformation

$$\begin{aligned}\frac{1}{2}\varphi''(0) &= 2\sum_{i,j=1}^{3}\frac{\partial u_i}{\partial x_j}(\boldsymbol{x})v_jv_i+\sum_{i=1}^{3}\left(\sum_{j=1}^{3}\frac{\partial u_i}{\partial x_j}(\boldsymbol{x})v_j\right)^2\\ &= \sum_{i,j=1}^{3}\left(\frac{\partial u_i}{\partial x_j}(\boldsymbol{x})+\frac{\partial u_j}{\partial x_i}(\boldsymbol{x})\right)v_iv_j+\sum_{i,j,k=1}^{3}\frac{\partial u_k}{\partial x_i}(\boldsymbol{x})\frac{\partial u_k}{\partial x_j}(\boldsymbol{x})v_iv_j. \qquad (6.10)\end{aligned}$$

Using the notation

$$2\varepsilon_{ij}=\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}+\sum_{k=1}^{3}\frac{\partial u_k}{\partial x_i}\frac{\partial u_k}{\partial x_j},\quad i,j=1,2,3 \qquad (6.11)$$

we write

$$\frac{1}{2}\varphi''(0)=2\sum_{i,j=1}^{3}\varepsilon_{ij}v_iv_j. \qquad (6.12)$$

The matrix $\{\varepsilon_{ij}\}_{i,j=1}^{3}$ is called the *tensor of finite strain*.

The linear part (with respect to the gradient of displacement) of the tensor of finite strain (6.11) $\varepsilon_{i,j}=\varepsilon_{i,j}\left(\frac{\partial u_1}{\partial x_1},\frac{\partial u_1}{\partial x_2},\ldots,\frac{\partial u_3}{\partial x_3}\right)$ is called *small strain tensor* and denoted by e_{ij} :

$$e_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right),\quad i,j=1,2,3. \qquad (6.13)$$

6.3 Generalized Hooke's law

In this section we shall describe the relation between the strain and stress tensor. For a wide range of materials the *linear generalized Hooke's law* represents a convenient characterization of their properties. It is expressed as the linear relation between the stress tensor and small strain tensor at point $\boldsymbol{x}\in\Omega^b$, i.e.

$$\tau_{ij}^b=\sum_{k,l=1}^{3}c_{ijkl}e_{kl},\quad i,j=1,2,3. \qquad (6.14)$$

We can notice that there is no absolute term in (6.14). It corresponds with the assumption that in the case of zero stress tensor also the small strain tensor is equal to zero.

If the constants $c_{ijkl}(\boldsymbol{x})$ do not depend on the choice of coordinate system, the material of the body is said to be *isotropic* at the point $\boldsymbol{x} \in \Omega^b$.

The generalized Hooke's law for isotropic material can be written in a form:

$$\tau_{ij}^b(\boldsymbol{x}) = \lambda^b(\boldsymbol{x})\mathrm{div}\boldsymbol{u}(\boldsymbol{x})\delta_{ij} + 2\mu^b(\boldsymbol{x})e_{ij}(\boldsymbol{x}), \quad i,j = 1,2,3. \tag{6.15}$$

The coefficients λ^b, μ^b are so-called *Lamé coefficients* and δ_{ij} is the Kronecker delta. The proof of (6.15) can be found in [53].

Now our aim turns to properties of the Lamé coefficients. Let suppose that there exists an inverse law to the generalized Hooke's law. So we are able to express components of the small strain tensor by the stress tensor components. Then the system (6.14) can be solved clearly with respect to e_{ij}. It is useful to take in a mind that $\mathrm{div}\boldsymbol{u} = \sum_{i=1}^{3} e_{ii}$. The determinant of the matrix of system (6.15) is equal to

$$(2\mu^b)^5(3\lambda^b + 2\mu^b). \tag{6.16}$$

Because we suppose this determinant is nonzero, it results in

$$\mu^b \neq 0, \quad 3\lambda^b + 2\mu^b \neq 0. \tag{6.17}$$

Further it is possible to show

$$\mu^b > 0, \quad \lambda^b > 0 \tag{6.18}$$

(for derivation see [53]).

In a technical practice are mostly instead of the Lamé coefficients λ^b, μ^b used the *Young modulus* E and the *Poisson ratio* σ defined by

$$\frac{1}{E} = \frac{\lambda^b + \mu^b}{\mu^b(3\lambda^b + 2\mu^b)}, \quad \sigma = \frac{\lambda^b}{2(\lambda^b + \mu^b)}. \tag{6.19}$$

In the opposite way the Lamé coefficients can be expressed by the Young modulus E and the Poisson ratio σ

$$\mu^b = \frac{E}{2(1+\sigma)}, \quad \lambda^b = \frac{E\sigma}{(1+\sigma)(1-2\sigma)}. \tag{6.20}$$

It is seen from (6.19) that

$$(\lambda^b > 0 \wedge \mu^b > 0) \Longleftrightarrow (E > 0 \wedge 0 < \sigma < \frac{1}{2}). \tag{6.21}$$

6.4 Dynamical equations of an isotropic elastic body

Till now just the conditions of equilibrium of forces were taken into account. In this section we shall derive the equations of motion of an isotropic elastic body. We assume that u_i, τ_{ij}^b and f_i, $i,j = 1,2,3$ are in general functions of the space coordinates $x_i, i = 1,2,3$ and time t and have the continuous derivatives of the required order.

Dynamical equations can be derived from the equilibrium equations (6.3) by the use of *d'Alembert's principle* (see [64]). It allows us to add the term representing the

acceleration of the body to equations (6.3) and we obtain the equations of motion in the form

$$\sum_{j=1}^{3} \frac{\partial \tau_{ij}^b}{\partial x_j} + f_i = \rho^b \frac{\partial^2 u_i}{\partial t^2}, \quad i = 1, 2, 3. \tag{6.22}$$

Further, to equations (6.22) we add the term

$$C\rho^b \frac{\partial u_i}{\partial t}, \quad i = 1, 2, \tag{6.23}$$

where $C \geq 0$ is a real constant. This term represents the dissipative structural damping, which is natural for real bodies. Then equations (6.22) read

$$\rho^b \frac{\partial^2 u_i}{\partial t^2} + C\rho^b \frac{\partial u_i}{\partial t} - \sum_{j=1}^{2} \frac{\partial \tau_{ij}^b}{\partial x_j} = f_i \quad i = 1, 2, 3. \tag{6.24}$$

6.5 Formulation of 2D problem of the motion of an isotropic elastic body

In the previous section of this chapter we suppose the three dimensional model of an elastic body. The aim of this work is the complet description of the two dimensional model of the fluid-structure interaction. For this reason we need to simplify the dynamical equations (6.24) of an elastic body for the two dimensional case. In the two-dimensional case the third component of the displacement is equal to zero:

$$u_3 = 0. \tag{6.25}$$

It means that $\boldsymbol{u} = (u_1(\boldsymbol{x}, t), u_2(\boldsymbol{x}, t))$. Similarly, we obtain

$$\tau_{i3}^b = \tau_{3i}^b = 0, \quad i = 1, 2,$$
$$f_3 = 0.$$

Let us have a time interval $(0, T)$ and a domain Ω^b with the Lipschitz-continuous boundary $\partial\Omega^b$ consisting from two disjoint parts Γ_W^b and Γ_D^b such that $\partial\Omega^b = \Gamma_W^b \cup \Gamma_D^b$. Further, we assume that $\boldsymbol{f} = 0$. Then the complete system of the two dimensional dynamical equations of an elastic body can be written in the form

$$\rho^b \frac{\partial^2 u_i}{\partial t^2} + C\rho^b \frac{\partial u_i}{\partial t} - \sum_{j=1}^{2} \frac{\partial \tau_{ij}^b}{\partial x_j} = 0 \qquad \text{in } \Omega^b \times (0, T), \quad i = 1, 2, \tag{6.26}$$

$$\boldsymbol{u}(0, \cdot) = \boldsymbol{u}^0 \qquad \text{in } \Omega^b, \tag{6.27}$$

$$\frac{\partial \boldsymbol{u}}{\partial t}(0, \cdot) = \boldsymbol{r}^0 \qquad \text{in } \Omega^b, \tag{6.28}$$

$$\boldsymbol{u} = \boldsymbol{u}^d \qquad \text{in } \Gamma_D^b \times (0, T), \tag{6.29}$$

$$\sum_{j=1}^{2} \tau_{ij}^b n_j = T_i^{\boldsymbol{n}} \qquad \text{in } \Gamma_W^b \times (0, T), \tag{6.30}$$

where equations (6.26) were completed by the initial conditions (6.27)-(6.28) and the boundary conditions (6.29)-(6.30). We assume that all presented functions are sufficiently smooth.

Chapter 7

Coupled problem

Till now we were concerned with the separate flow or structure problem as described in Chapters 5 and 6. The aim of this work is the fluid-structure interaction problem. For this reason this chapter will be devoted to the formulation of the continuous coupled problem together with the coupling conditions.

As mentioned in Chapter 5 the flow problem is represented by the Navier-Stokes equations (5.1)-(5.3) completed by the thermodynamical relations (5.4)-(5.5) and the initial conditions (5.10) and the boundary conditions (5.11) in the bounded domain $\Omega_t \subset I\!R^2$ depending on time $t \in [0, T]$.

The structure problem is defined by system (6.26)-(6.30) in the bounded domain $\Omega^b \subset I\!R^2$.

The interaction between the flow and the structure takes place on their common boundary $\tilde{\Gamma}_{W_t}$ at the time t. It is given by

$$\tilde{\Gamma}_{W_t} = \left\{ \boldsymbol{x} \in I\!R^2;\ \boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X}, t),\ \boldsymbol{X} \in \Gamma_W^b \right\}. \tag{7.1}$$

Thus, the domain Ω_t is determined by the displacement $\boldsymbol{u}$ of the part Γ_W^b at time t. It gives us the possibility how to construct a convenient ALE mapping $\mathcal{A}_t$. This procedure will be described in Section 10.1.

If the domain Ω_t occupied by the fluid at time t is known, we can solve the flow problem and compute the surface force acting on the body on the part $\tilde{\Gamma}_{W_t}$, which can be transformed to the reference configuration, i.e. to the interface Γ_W^b. In the case of the linear elasticity model, when only small deformations are considered, we get the transmission condition for the force balance between the aerodynamic forces and the forces on the structure surface

$$\sum_{j=1}^{2} \tau_{ij}^b(\boldsymbol{X}) n_j(\boldsymbol{X}) = -\sum_{j=1}^{2} \tau_{ij}^f(\boldsymbol{x}) n_j(\boldsymbol{X}), \quad i = 1, 2, \tag{7.2}$$

where τ_{ij}^f are the components of the stress tensor of the fluid:

$$\tau_{ij}^f = -p\delta_{ij} + \tau_{ij}^V, \quad i, j = 1, 2, \tag{7.3}$$

the points $\boldsymbol{x}$ and $\boldsymbol{X}$ satisfy the relation

$$\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X}, t). \tag{7.4}$$

and $\boldsymbol{n}(\boldsymbol{X}) = (n_1(\boldsymbol{X}), n_2(\boldsymbol{X}))$ denotes the unit outer normal to the body Ω^b on Γ^b_W at the point $\boldsymbol{X}$. Because the fluid problem is solved as a dimensionless one in contrast to the structural problem, we need to transform the dimensionless stress tensor of the fluid in the following way:

$$\tau^f_{ij}(\boldsymbol{x}, t) = \rho^* U^{*2} p'(\boldsymbol{x}', t') + \frac{\mu^* U^*}{L^*} \tau^V_{ij}(\boldsymbol{x}', t'). \tag{7.5}$$

This can be shown in a similar way as in (5.30)-(5.36).

Further, the fluid velocity is defined on the moving part of the boundary $\tilde{\Gamma}_{W_t}$ by the second transmission condition on the velocity equality of the fluid and structure particle on the FSI boundary

$$\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{z}_D(\boldsymbol{x}, t) = \frac{\partial \boldsymbol{u}(\boldsymbol{X}, \boldsymbol{t})}{\partial t}. \tag{7.6}$$

The obtained velocity can be written in the dimensionless form as

$$\boldsymbol{v}'(\boldsymbol{x}', t') = \frac{\boldsymbol{v}(\boldsymbol{x}, t)}{U^*}. \tag{7.7}$$

Finally, we formulate the *continuous fluid-structure interaction (FSI) problem*: We want to determine the domain $\Omega_t,\ t \in (0, T]$ and functions $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{x}, t),\ \boldsymbol{x} \in \bar{\Omega}_t,\ t \in [0, T]$ and $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{X}, t),\ \boldsymbol{X} \in \bar{\Omega}^b,\ t \in [0, T]$ satisfying equations (5.57), (6.26), the initial conditions (5.10), (6.27), (6.28), the boundary conditions (5.11), (6.29), (6.30) and the transmission conditions (7.2), (7.6).

This FSI problem represents a strongly nonlinear dynamical system. Theoretical analysis of qualitative properties of this problem, as the existence, uniqueness and regularity of its solution, is open. Therefore, in the sequel we shall be concerned with its numerical solution.

Chapter 8

Discretization of the flow problem

This chapter is devoted to the discretization of the system of equations (5.57) with the initial condition (5.10) and the boundary condition (5.11). The space semidiscretization will be carried out by the *discontinuous Galerkin finite element method* (DGFEM). The time discretization will be realized by the *backward difference formula* (BDF) method.

8.1 Space semidiscretization

We construct a polygonal approximation Ω_{ht} of the domain Ω_t. By $\mathcal{T}_{ht}$ we denote a partition of the closure $\bar{\Omega}_{ht}$ of the domain Ω_{ht} into a finite number of closed triangles K with mutually disjoint interiors such that $\bar{\Omega}_{ht} = \bigcup_{K\in\mathcal{T}_{ht}} K$.

By $\mathcal{F}_{ht}$ we denote the system of all faces of all elements $K \in \mathcal{T}_{ht}$. Further, we introduce the set of all boundary faces $\mathcal{F}_{ht}^B = \{\Gamma \in \mathcal{F}_{ht};\ \Gamma \subset \partial\Omega_{ht}\}$. In the $\mathcal{F}_{ht}^B$ we distinguish the set $\mathcal{F}_{ht}^{IO} = \left\{\Gamma \in \mathcal{F}_{ht}^B;\ \Gamma \subset \Gamma_I \cup \Gamma_O \subset \partial\Omega_{ht}\right\}$ of all boundary faces lying on the input and output, the set $\mathcal{F}_{ht}^W = \left\{\Gamma \in \mathcal{F}_{ht}^B;\ \Gamma \subset \Gamma_{W_t} \subset \partial\Omega_{ht}\right\}$ of all boundary faces lying on the impermeable wall and the set $\mathcal{F}_{ht}^D = \left\{\Gamma \in \mathcal{F}_{ht}^B;\ \text{a Dirichlet condition is prescribed on } \Gamma\right\}$ of all "Dirichlet" boundary faces. Finally, $\mathcal{F}_{ht}^I = \mathcal{F}_{ht} \setminus \mathcal{F}_{ht}^B$ denotes the set of all inner faces.

Each $\Gamma \in \mathcal{F}_{ht}$ is associated with a unit normal vector $\boldsymbol{n}_\Gamma$ to Γ. For $\Gamma \in \mathcal{F}_{ht}^B$ the normal $\boldsymbol{n}_\Gamma$ has the same orientation as the outer normal to $\partial\Omega_{ht}$. We set $d(\Gamma) =$ length of $\Gamma \in \mathcal{F}_{ht}$.

For each $\Gamma \in \mathcal{F}_{ht}^I$ there exist two neighbouring elements $K_\Gamma^{(L)}$, $K_\Gamma^{(R)} \in \mathcal{T}_{ht}$ such that $\Gamma \subset \partial K_\Gamma^{(R)} \cap \partial K_\Gamma^{(L)}$. We use the convention that $K_\Gamma^{(R)}$ lies in the direction of $\boldsymbol{n}_\Gamma$ and $K_\Gamma^{(L)}$ lies in the opposite direction to $\boldsymbol{n}_\Gamma$. See Figure 8.1. The elements $K_\Gamma^{(L)}$, $K_\Gamma^{(R)}$ are called neighbours. If $\Gamma \in \mathcal{F}_{ht}^B$, then the element adjacent to Γ will be denoted by $K_\Gamma^{(L)}$.

The approximate solution will be sought in the space of piecewise polynomial functions

$$\boldsymbol{S}_{ht} = [S_{ht}]^4, \quad \text{with } S_{ht} = \{v; v|_K \in P_r(K)\ \forall\, K \in \mathcal{T}_{ht}\}, \tag{8.1}$$

where $r \geq 1$ is an integer and $P_r(K)$ denotes the space of all polynomials on K of degree $\leq r$. A function $\boldsymbol{\varphi} \in \boldsymbol{S}_{ht}$ is, in general, discontinuous on interfaces $\Gamma \in \mathcal{F}_{ht}^I$. By $\boldsymbol{\varphi}_\Gamma^{(L)}$ and $\boldsymbol{\varphi}_\Gamma^{(R)}$ we denote the values of $\boldsymbol{\varphi}$ on Γ considered from the interior and

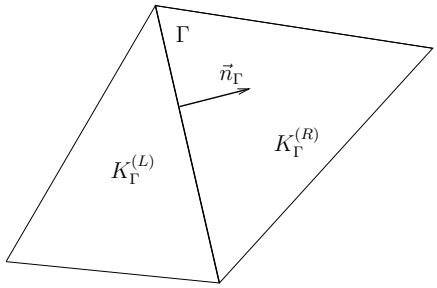


Figure 8.1: Neighbouring elements

the exterior of $K_\Gamma^{(L)}$, respectively, and set

$$\langle \boldsymbol{\varphi} \rangle_\Gamma = \frac{\boldsymbol{\varphi}_\Gamma^{(L)} + \boldsymbol{\varphi}_\Gamma^{(R)}}{2}, \tag{8.2}$$

$$[\boldsymbol{\varphi}]_\Gamma = \boldsymbol{\varphi}_\Gamma^{(L)} - \boldsymbol{\varphi}_\Gamma^{(R)}. \tag{8.3}$$

These functions represent the average and the jump of a function $\boldsymbol{\varphi} \in \boldsymbol{S}_{ht}$ on $\Gamma \in \mathcal{F}_{ht}$.

In order to derive the discrete problem, we assume that $\boldsymbol{w}$ is a sufficiently regular classical solution of (5.57). Therefore, $\boldsymbol{w}|_\Gamma^{(L)} = \boldsymbol{w}|_\Gamma^{(R)} = \boldsymbol{w}|_\Gamma$. In what follows, it will be suitable to use these symbols instead of $\boldsymbol{w}|_\Gamma$ because of the definition of the discretization forms in terms of an approximate solution. If we multiply system (5.57) by a test function $\boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht}$, integrate over $K \in \mathcal{T}_{ht}$, apply Green's theorem and sum over all elements $K \in \mathcal{T}_{ht}$, we obtain

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_{ht}} \int_K \frac{D^{\mathcal{A}}\boldsymbol{w}}{Dt} \cdot \boldsymbol{\varphi}_h \mathrm{d}\boldsymbol{x} + \sum_{K\in\mathcal{T}_{ht}} \sum_{\Gamma\subset\partial K} \int_\Gamma \sum_1^2 \boldsymbol{g}_s(\boldsymbol{w})(n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}\boldsymbol{S} \\
&\quad - \sum_{K\in\mathcal{T}_{ht}} \int_K \sum_{s=1}^2 \boldsymbol{g}_s(\boldsymbol{w}) \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x} + \sum_{K\in\mathcal{T}_{ht}} \int_K \boldsymbol{w}\mathrm{div}\boldsymbol{z} \cdot \boldsymbol{\varphi}_h \mathrm{d}\boldsymbol{x} \\
&= \sum_{K\in\mathcal{T}_{ht}} \sum_{\Gamma\subset\partial K} \int_\Gamma \sum_{s=1}^2 \boldsymbol{R}_s(\boldsymbol{w}|_\Gamma^{(L)}, \nabla\boldsymbol{w}|_\Gamma^{(L)})(n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}\boldsymbol{S} \\
&\quad - \sum_{K\in\mathcal{T}_{ht}} \int_K \sum_{s=1}^2 \boldsymbol{R}_s(\boldsymbol{w}, \nabla\boldsymbol{w}) \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{8.4}$$

Let us have a look on some terms of (8.4). First we take in mind the inviscid flux $\int_\Gamma \sum_1^2 g_s(\boldsymbol{w})(n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}\boldsymbol{S}$ through the edge Γ, $\Gamma \subset \partial K$, $K \in \mathcal{T}_{ht}$. It will be approximated with the aid of the *numerical flux* $\boldsymbol{H_g}$:

$$\sum_{s=1}^2 g_s(\boldsymbol{w})(n_\Gamma)_s \approx \boldsymbol{H_g}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma). \tag{8.5}$$

We assume the following conditions for a general numerical flux $\boldsymbol{H}$:

1. $\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n})$ is defined in $D \times D \times B_1$, where $B_1 = \{\boldsymbol{n} \in I\!R^2; |\boldsymbol{n}| = 1\}$. $\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n})$ is *locally Lipschitz-continuous*, which means that for each $a > 0$ there exists a constant $L_{\boldsymbol{H}}(a)$ such that

$$|\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n}) - \boldsymbol{H}(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \boldsymbol{n})| \leq L_{\boldsymbol{H}}(a) \left(|\boldsymbol{w}_1 - \boldsymbol{w}_1^*| + |\boldsymbol{w}_2 - \boldsymbol{w}_2^*|\right), \quad (8.6)$$
$$\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_1^*, \boldsymbol{w}_2^* \in I\!R^4, \ |\boldsymbol{w}_1|, |\boldsymbol{w}_2|, |\boldsymbol{w}_1^*|, |\boldsymbol{w}_2^*| \leq a, \ \boldsymbol{n} \in B_1.$$

2. $\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n})$ is *consistent*:

$$\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_1, \boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{g}_s(\boldsymbol{w}_1) n_s, \quad \boldsymbol{w}_1 \in I\!R^4, \ \boldsymbol{n} = (n_1, n_2) \in B_1. \quad (8.7)$$

3. $\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n})$ is *conservative*:

$$\boldsymbol{H}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{n}) = -\boldsymbol{H}(\boldsymbol{w}_2, \boldsymbol{w}_1, -\boldsymbol{n}), \quad \boldsymbol{w}_1, \boldsymbol{w}_2 \in I\!R^4, \ \boldsymbol{n} \in B_1. \quad (8.8)$$

Using (8.5), (8.8) and

$$\boldsymbol{n}|_\Gamma^{(L)} = -\boldsymbol{n}|_\Gamma^{(R)}, \quad \Gamma \in \mathcal{F}_{ht}, \quad (8.9)$$

we obtain

$$\begin{aligned}
&\sum_{K \in \mathcal{T}_{ht}} \sum_{\Gamma \subset \partial K} \int_\Gamma \sum_{1}^{2} \boldsymbol{g}_s(\boldsymbol{w})(n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}\boldsymbol{S} \quad (8.10)\\
&\approx \sum_{K \in \mathcal{T}_{ht}} \sum_{\Gamma \subset \partial K} \int_\Gamma \boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S\\
&= \sum_{\Gamma \in \mathcal{F}_{ht}^I} \int_\Gamma \left(\boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} + \boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{n}_\Gamma) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(R)} \right) \mathrm{d}S\\
&\quad + \sum_{\Gamma \in \mathcal{F}_{ht}^B} \int_\Gamma \boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S\\
&= \sum_{\Gamma \in \mathcal{F}_{ht}^I} \int_\Gamma \boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma) \cdot [\boldsymbol{\varphi}_h]_\Gamma \, \mathrm{d}S\\
&\quad + \sum_{\Gamma \in \mathcal{F}_{ht}^B} \int_\Gamma \boldsymbol{H}_{\boldsymbol{g}}(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S.
\end{aligned}$$

We shall choose the numerical flux in a form convenient for a semi-implicit linearization with respect to time. Namely, we shall seek the numerical flux in the form

$$\boldsymbol{H}(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) = \mathbb{A}_L(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) \boldsymbol{w}_L + \mathbb{A}_R(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) \boldsymbol{w}_R \quad (8.11)$$

with some matrices $\mathbb{A}_L, \ \mathbb{A}_R : \ D \times D \times I\!R^2 \longrightarrow I\!R^{4\times 4}$. As an example we can use the Vijaysundaram numerical flux $\boldsymbol{H}_{VS}$ defined in the following way. Using relations (5.24), (5.26), (5.27), we obtain

$$\begin{aligned}
\sum_{s=1}^{2} \boldsymbol{g}_s(\boldsymbol{w}) n_s &= \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) n_s - z_s n_s \boldsymbol{w} = \sum_{s=1}^{2} \left(\mathbb{A}_s(\boldsymbol{w}) n_s - z_s n_s \mathbb{I} \right) \boldsymbol{w}\\
&= \left(\mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) - (\boldsymbol{z} \cdot \boldsymbol{n}) \, \mathbb{I} \right) \boldsymbol{w} = \mathbb{P}_g(\boldsymbol{w}, \boldsymbol{n}) \boldsymbol{w} \quad (8.12)
\end{aligned}$$

We take in the mind that the matrix $\mathbb{P}(\boldsymbol{w},\boldsymbol{n})$ can be diagonalized as shown in [31]. Hence, there exists a nonsingular matrix $\mathbb{T}=\mathbb{T}(\boldsymbol{w},\boldsymbol{n})$ such that

$$\mathbb{P}(\boldsymbol{w},\boldsymbol{n})=\mathbb{T}\mathit{\Lambda}\mathbb{T}^{-1}, \tag{8.13}$$

where $\mathit{\Lambda}=\mathit{\Lambda}(\boldsymbol{w},\boldsymbol{n})=\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is the diagonal matrix with diagonal elements $\lambda_i=\lambda_i(\boldsymbol{w},\boldsymbol{n})$ equal to eigenvalues of the matrix $\mathbb{P}(\boldsymbol{w},\boldsymbol{n})$. The eigenvalues λ_i have the form

$$\begin{aligned}\lambda_1&=\boldsymbol{v}\cdot\boldsymbol{n}-a,\\ \lambda_2&=\lambda_3=\boldsymbol{n}\cdot\boldsymbol{v},\\ \lambda_4&=\boldsymbol{n}\cdot\boldsymbol{v}+a,\end{aligned} \tag{8.14}$$

where $a=\sqrt{\gamma p/\rho}$ is the speed of sound. Applying (8.13) to (8.12) we see that also matrix $\mathbb{P}_g(\boldsymbol{w},\boldsymbol{n})$ can be diagonalized

$$\mathbb{P}_g(\boldsymbol{w},\boldsymbol{n})=\mathbb{T}\mathit{\Lambda}\mathbb{T}^{-1}-(\boldsymbol{z}\cdot\boldsymbol{n})\mathbb{I}=\mathbb{T}\mathit{\Lambda}_g\mathbb{T}^{-1}, \tag{8.15}$$

where

$$\mathit{\Lambda}_g=\mathit{\Lambda}-(\boldsymbol{z}\cdot\boldsymbol{n})\mathbb{I}=\mathrm{diag}(\lambda_{g1},\lambda_{g2},\lambda_{g3},\lambda_{g4}), \tag{8.16}$$

$$\lambda_{gi}=\lambda_i-\boldsymbol{z}\cdot\boldsymbol{n}\quad\text{for } i=1,\dots,4. \tag{8.17}$$

We define the "positive" and "negative" parts of the matrix $\mathbb{P}_g$:

$$\mathbb{P}_g^{\pm}(\boldsymbol{w},\boldsymbol{n})=\mathbb{T}\mathit{\Lambda}_g^{\pm}\mathbb{T}^{-1},\quad \mathit{\Lambda}_g^{\pm}=\mathrm{diag}(\lambda_{g1}^{\pm},\dots,\lambda_{g4}^{\pm}), \tag{8.18}$$

where $\lambda_g^+=\max\{\lambda_g,0\}$ and $\lambda_g^-=\min\{\lambda_g,0\}$. Then we define the Vijayasundaram numerical flux $\boldsymbol{H}_{VS}$ as

$$\boldsymbol{H}_{VS}(\boldsymbol{w}_L,\boldsymbol{w}_R,\boldsymbol{n})=\mathbb{P}_g^{+}\left(\frac{\boldsymbol{w}_L+\boldsymbol{w}_R}{2},\boldsymbol{n}\right)\boldsymbol{w}_L+\mathbb{P}_g^{-}\left(\frac{\boldsymbol{w}_L+\boldsymbol{w}_R}{2},\boldsymbol{n}\right)\boldsymbol{w}_R. \tag{8.19}$$

The presented considerations lead us to the definition of an inviscid form

$$\begin{aligned}&\hat{b}_h(\boldsymbol{w},\boldsymbol{\varphi}_h)\\ &=-\sum_{K\in\mathcal{T}_{ht}}\int_K\sum_{s=1}^{2}(\mathbb{A}_s(\boldsymbol{w})-z_s\mathbb{I})\boldsymbol{w}\cdot\frac{\partial\boldsymbol{\varphi}_h}{\partial x_s}\mathrm{d}\boldsymbol{x}\\ &\quad+\sum_{\Gamma\in\mathcal{F}_{ht}^I}\int_\Gamma\left(\mathbb{P}_g^{+}(\langle\boldsymbol{w}\rangle_\Gamma,\boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(L)}+\mathbb{P}_g^{-}(\langle\boldsymbol{w}\rangle_\Gamma,\boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(R)}\right)\cdot[\boldsymbol{\varphi}_h]_\Gamma\,\mathrm{d}S\\ &\quad+\sum_{\Gamma\in\mathcal{F}_{ht}^B}\int_\Gamma\left(\mathbb{P}_g^{+}(\langle\boldsymbol{w}\rangle_\Gamma,\boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(L)}+\mathbb{P}_g^{-}(\langle\boldsymbol{w}\rangle_\Gamma,\boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(R)}\right)\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}S.\end{aligned} \tag{8.20}$$

Now we shall pay attention to the discretization of the viscous terms. Using relation (5.28) and the zero natural Neumann boundary condition from (5.11), the term

$\sum_{K\in\mathcal{T}_{ht}}\sum_{\Gamma\subset\partial K}\int_\Gamma\sum_{s=1}^2 \boldsymbol{R}_s(\boldsymbol{w},\nabla\boldsymbol{w})(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}\boldsymbol{S}$ can be written in the form

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_{ht}}\sum_{\Gamma\subset\partial K}\int_\Gamma\sum_{s=1}^2 \boldsymbol{R}_s(\boldsymbol{w}|_\Gamma^{(L)},\nabla\boldsymbol{w}|_\Gamma^{(L)})(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}\boldsymbol{S}\\
&=\sum_{K\in\mathcal{T}_{ht}}\sum_{\Gamma\subset\partial K}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2 \mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}S\\
&=\sum_{\Gamma\in\mathcal{F}_{ht}^I}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2 \mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}(n_\Gamma^{(L)})_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}S\\
&\quad+\sum_{\Gamma\in\mathcal{F}_{ht}^I}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2 \mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(R)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(R)}(n_\Gamma^{(R)})_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(R)}\mathrm{d}S\\
&\quad+\sum_{\Gamma\in\mathcal{F}_{ht}^D}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2 \mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}S.
\end{aligned}$$

In virtue of the sufficient regularity of $\boldsymbol{w}$, the following relations are valid:

$$\begin{aligned}
\boldsymbol{w}|_\Gamma^{(L)}&=\boldsymbol{w}|_\Gamma^{(R)},\\
\boldsymbol{n}_\Gamma^{(L)}&=-\boldsymbol{n}_\Gamma^{(R)},\\
\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}&=\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(R)}\quad\text{for } k=1,2.
\end{aligned}$$

Thus

$$\mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}=\mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(R)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(R)}=\left\langle\mathbb{K}_{sk}(\boldsymbol{w})\frac{\partial\boldsymbol{w}}{\partial x_k}\right\rangle_\Gamma. \tag{8.21}$$

This implies that

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_{ht}}\sum_{\Gamma\subset\partial K}\int_\Gamma\sum_{s=1}^2 \boldsymbol{R}_s(\boldsymbol{w}|_{\Gamma^{(L)}},\nabla\boldsymbol{w}|_{\Gamma^{(L)}})(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}\boldsymbol{S} \qquad (8.22)\\
&=\sum_{\Gamma\in\mathcal{F}_{ht}^I}\int_\Gamma\sum_{s=1}^2\left\langle\sum_{k=1}^2\mathbb{K}_{sk}(\boldsymbol{w})\frac{\partial\boldsymbol{w}}{\partial x_k}\right\rangle_\Gamma(n_\Gamma)_s\cdot[\boldsymbol{\varphi}_h]_\Gamma\,\mathrm{d}S\\
&\quad+\sum_{\Gamma\in\mathcal{F}_{ht}^D}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2 \mathbb{K}_{sk}(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{w}}{\partial x_k}|_\Gamma^{(L)}(n_\Gamma)_s\cdot\boldsymbol{\varphi}_h|_\Gamma^{(L)}\mathrm{d}S.
\end{aligned}$$

Now, we introduce *stabilization terms*, which are equal to zero for a sufficiently regular solution due to the relation $[\boldsymbol{w}]_\Gamma=0$:

$$\Theta\sum_{\Gamma\in\mathcal{F}_{ht}^I}\int_\Gamma\sum_{s=1}^2\left\langle\sum_{k=1}^2\mathbb{K}_{sk}^T(\boldsymbol{w})\frac{\partial\boldsymbol{\varphi}_h}{\partial x_k}\right\rangle_\Gamma(n_\Gamma)_s\cdot[\boldsymbol{w}]_\Gamma\,\mathrm{d}S. \tag{8.23}$$

On the part of the boundary where the Dirichlet conditions are prescribed, we add the term

$$\Theta\sum_{\Gamma\in\mathcal{F}_{ht}^D}\int_\Gamma\sum_{s=1}^2\sum_{k=1}^2\mathbb{K}_{sk}^T(\boldsymbol{w}|_\Gamma^{(L)})\frac{\partial\boldsymbol{\varphi}_h}{\partial x_k}|_\Gamma^{(L)}(n_\Gamma)_s\cdot\boldsymbol{w}|_\Gamma^{(L)}\mathrm{d}S. \tag{8.24}$$

This expression will be compensated by

$$\Theta \sum_{\Gamma\in\mathcal{F}_{ht}^D} \int_\Gamma \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}_{sk}^T(\boldsymbol{w}|_\Gamma^{(L)}) \frac{\partial \boldsymbol{\varphi}}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{w}_B \mathrm{d}S, \tag{8.25}$$

where $\boldsymbol{w}_B$ is the boundary state defined on the basis of the Dirichlet boundary conditions and extrapolation:

$$\boldsymbol{w}_B = (\rho_D, \rho_D v_{D1}, \rho_D v_{D2}, c_v \rho_D \theta_\Gamma^{(L)} + \frac{1}{2}\rho_D |\boldsymbol{v}_D|^2) \quad \text{on } \Gamma_I, \tag{8.26}$$

$$\boldsymbol{w}_B = \boldsymbol{w}_\Gamma^{(L)} \text{on } \Gamma_O, \tag{8.27}$$

$$\boldsymbol{w}_B = (\rho_\Gamma^{(L)}, \rho_\Gamma^{(L)} z_{D1}, \rho_\Gamma^{(L)} z_{D2}, c_v \rho_\Gamma^{(L)} \theta_\Gamma^{(L)} + \frac{1}{2}\rho_\Gamma^{(L)} |\boldsymbol{z}_D|^2) \quad \text{on } \Gamma_{W_t}. \tag{8.28}$$

The presented considerations lead to the definition of the viscous form

$$\begin{aligned}
\hat{a}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) = & \sum_{k\in\mathcal{T}_{ht}} \int_K \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}_{sk}(\boldsymbol{w}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_k} \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x} \\
& - \sum_{\Gamma\in\mathcal{F}_{ht}^I} \int_\Gamma \sum_{s=1}^{2} \left\langle \sum_{k=1}^{2} \mathbb{K}_{sk}(\boldsymbol{w}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_k} \right\rangle_\Gamma (n_\Gamma)_s \cdot [\boldsymbol{\varphi}_h]_\Gamma \, \mathrm{d}S \\
& - \sum_{\Gamma\in\mathcal{F}_{ht}^D} \int_\Gamma \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}_{sk}(\boldsymbol{w}_h|_\Gamma^{(L)}) \frac{\partial \boldsymbol{w}_h}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S \\
& - \Theta \sum_{\Gamma\in\mathcal{F}_{ht}^I} \int_\Gamma \sum_{s=1}^{2} \left\langle \sum_{k=1}^{2} \mathbb{K}_{sk}^T(\boldsymbol{w}_h) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k} \right\rangle_\Gamma (n_\Gamma)_s \cdot [\boldsymbol{w}]_\Gamma \, \mathrm{d}S \\
& - \Theta \sum_{\Gamma\in\mathcal{F}_{ht}^D} \int_\Gamma \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}_{sk}^T(\boldsymbol{w}_h|_\Gamma^{(L)}) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{w}|_\Gamma^{(L)} \mathrm{d}S, \\
& \boldsymbol{w}_h, \ \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht}.
\end{aligned} \tag{8.29}$$

The constant Θ is chosen as $\Theta = -1$ or $\Theta = 0$ or $\Theta = 1$, which leads to the nonsymmetric version or to the incomplete version or to the symmetric version, respectively, of the viscous form.

Now we introduce the *interior and boundary penalty form*

$$J_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) = \sum_{\Gamma\in\mathcal{F}_{ht}^I} \int_\Gamma \sigma \, [\boldsymbol{w}_h]_\Gamma \cdot [\boldsymbol{\varphi}_h]_\Gamma \, \mathrm{d}S + \sum_{\Gamma\in\mathcal{F}_{ht}^D} \int_\Gamma \sigma \boldsymbol{w}_h|_\Gamma^{(L)} \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S, \quad \boldsymbol{w}_h, \ \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht}, \tag{8.30}$$

where the weight σ is defined by

$$\sigma|_\Gamma = \frac{C_W}{d(\Gamma) Re}. \tag{8.31}$$

Here $d(\Gamma)$ denotes the diameter of $\Gamma \in \mathcal{F}_{ht}$ and $C_W > 0$ is a suitable constant. The first term on the right-hand side of (8.30) vanishes, when $\boldsymbol{w}_h$ is replaced by the exact

regular solution. The term containing the integrals over $\Gamma \in \mathcal{F}_{ht}^D$ are compensated by

$$\sum_{\Gamma \in \mathcal{F}_{ht}^D} \int_\Gamma \sigma \boldsymbol{w}_B \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S. \tag{8.32}$$

The boundary state vector $\boldsymbol{w}_B$ is again defined by (8.26)-(8.28).

The reaction form reads

$$d_h(\boldsymbol{w}, \boldsymbol{\varphi}_h) = \sum_{K \in \mathcal{T}_{ht}} \int_K \boldsymbol{w} \mathrm{div} \boldsymbol{z} \cdot \boldsymbol{\varphi}_h \mathrm{d}\boldsymbol{x}. \tag{8.33}$$

Finally, we introduce the right-hand side form containing the compensation terms (8.25) and (8.32):

$$\begin{aligned} \hat{l}_h(\boldsymbol{w}, \boldsymbol{\varphi}_h) = &\sum_{\Gamma \in \mathcal{F}_{ht}^D} \int_\Gamma \sigma \boldsymbol{w}_B \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S \\ &-\Theta \sum_{\Gamma \in \mathcal{F}_{ht}^D} \int_\Gamma \sum_{s=1}^{2} \sum_{k=1}^{2} \mathbb{K}_{sk}^T(\boldsymbol{w}|_\Gamma^{(L)}) \frac{\partial \boldsymbol{\varphi}}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{w}_B \mathrm{d}S. \end{aligned} \tag{8.34}$$

Now, the semidiscrete solution of problem (5.57) is defined as a function $\boldsymbol{w}_h \in C^1((0,T), \boldsymbol{S}_{ht})$ fulfilling the conditions

$$\begin{aligned} &\left(\frac{D^{\mathcal{A}} \boldsymbol{w}_h}{Dt}(t), \boldsymbol{\varphi}_h \right) + d_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) + \hat{b}_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) \\ &+\hat{a}_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) + J_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) = \hat{l}_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht},\ \forall t \in (0,T), \end{aligned} \tag{8.35}$$

$$\boldsymbol{w}_h(0) = \boldsymbol{w}_h^0, \tag{8.36}$$

where $\boldsymbol{w}_h^0$ is $L^2(\Omega_{h0})$-projection of $\boldsymbol{w}^0$ on $\boldsymbol{S}_{h0}$. It means that

$$\left(\boldsymbol{w}_h^0, \boldsymbol{\varphi}_h\right) = \left(\boldsymbol{w}^0, \boldsymbol{\varphi}_h\right) \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{h0}. \tag{8.37}$$

Let us mention that in our numerical experiments we use the incomplete formulation ($\Theta = 0$).

8.2 Application of the boundary conditions in the inviscid terms

For the numerical flux $\boldsymbol{H}_g(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{w}|_\Gamma^{(R)}, \boldsymbol{n}_\Gamma)$ on $\Gamma \in \mathcal{F}_{ht}$ appearing in the definition of the inviscid form $\hat{b}_h$ it is necessary to specify the boundary state $\boldsymbol{w}|_\Gamma^{(R)}$.

First we are interested in a situation on the moving impermeable wall, where the condition

$$\boldsymbol{v} \cdot \boldsymbol{n} = \boldsymbol{z} \cdot \boldsymbol{n} \tag{8.38}$$

is prescribed. We use this condition in the numerical flux $\boldsymbol{H}_g$. Using (8.38), we can write

$$\begin{aligned}
\boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n}) &:= \boldsymbol{H}_g(\boldsymbol{w},\boldsymbol{w},\boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) n_s - (\boldsymbol{z}\cdot\boldsymbol{n})\boldsymbol{w} \\
&= \boldsymbol{f}_1(\boldsymbol{w}) n_1 + \boldsymbol{f}_2(\boldsymbol{w}) n_2 - (\boldsymbol{z}\cdot\boldsymbol{n})\boldsymbol{w} \\
&= \begin{pmatrix} \rho v_1 n_1 + \rho v_2 n_2 \\ (\rho v_1^2 + p) n_1 + \rho v_1 v_2 n_2 \\ \rho v_1 v_2 n_1 + (\rho v_2^2 + p) n_2 \\ (E+p) v_1 n_1 + (E+p) v_2 n_2 \end{pmatrix} - \boldsymbol{z}\cdot\boldsymbol{n} \begin{pmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ E \end{pmatrix} \\
&= p \begin{pmatrix} 0 \\ n_1 \\ n_2 \\ \boldsymbol{z}\cdot\boldsymbol{n} \end{pmatrix} + \boldsymbol{z}\cdot\boldsymbol{n} \begin{pmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ E \end{pmatrix} - \boldsymbol{z}\cdot\boldsymbol{n} \begin{pmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ E \end{pmatrix} \\
&= (\gamma - 1) \left(w_4 - \frac{w_2^2 + w_3^2}{2 w_1} \right) \begin{pmatrix} 0 \\ n_1 \\ n_2 \\ \boldsymbol{z}\cdot\boldsymbol{n} \end{pmatrix}. \qquad (8.39)
\end{aligned}$$

We see that

$$\boldsymbol{H}_g^W(\alpha\boldsymbol{w},\boldsymbol{n}) = \alpha \boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n}), \quad \alpha > 0. \qquad (8.40)$$

In [31] it is shown that

$$\boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n}) = D_{\boldsymbol{w}} \boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n}) \boldsymbol{w}, \qquad (8.41)$$

where $D_{\boldsymbol{w}} \boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n})$ is the Jacobi matrix $D\boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n})/D\boldsymbol{w}$. The Jacobi matrix $D_{\boldsymbol{w}} \boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n})$ can be expressed in the form

$$D_{\boldsymbol{w}} \boldsymbol{H}_g^W(\boldsymbol{w},\boldsymbol{n}) = (\gamma - 1) \begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{w_2^2 + w_3^2}{2 w_1^2} n_1 & -\frac{w_2}{w_1} n_1 & -\frac{w_3}{w_1} n_1 & n_1 \\ \frac{w_2^2 + w_3^2}{2 w_1^2} n_2 & -\frac{w_2}{w_1} n_2 & -\frac{w_3}{w_1} n_2 & n_2 \\ \frac{w_2^2 + w_3^2}{2 w_1^2} (\boldsymbol{z}\cdot\boldsymbol{n}) & -\frac{w_2}{w_1} (\boldsymbol{z}\cdot\boldsymbol{n}) & -\frac{w_3}{w_1} (\boldsymbol{z}\cdot\boldsymbol{n}) & \boldsymbol{z}\cdot\boldsymbol{n} \end{pmatrix}. \qquad (8.42)$$

Further, we need to specify the value $\boldsymbol{w}|_\Gamma^{(R)}$ for $\Gamma \in \mathcal{F}_{ht}^{IO}$, when $\boldsymbol{w}|_\Gamma^{(L)}$ is known. We apply the approach using a solution of the *local Riemann problem*, which has been already described in [56].

Let us introduce a new Cartesian coordinate system $\tilde{x}_1, \tilde{x}_2$ in $\mathbb{R}^2$ with the origin at the center of gravity of the edge Γ, the coordinate $\tilde{x}_1$ is oriented in the direction of the normal $\boldsymbol{n}$ and $\tilde{x}_2$ tangent to Γ. The Euler equations transformed into this new coordinate system have the form

$$\frac{\partial \boldsymbol{q}}{\partial t} + \sum_{s=1}^{2} \frac{\partial \boldsymbol{f}_s(\boldsymbol{q})}{\partial \tilde{x}_s} = 0, \qquad (8.43)$$

as follows from the rotational invariance of the Euler equations. Here

$$\boldsymbol{q} = \mathbb{Q}(\boldsymbol{n})\boldsymbol{w}, \qquad (8.44)$$

where $\mathbb{Q}$ has form

$$\mathbb{Q}(\boldsymbol{n}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & n_1 & n_2 & 0 \\ 0 & -n_2 & n_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{8.45}$$

Now we neglect the tangential derivative $\partial/\partial\tilde{x}_2$ and get the system with one space variable $\tilde{x}_1$ in the form

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{\partial \boldsymbol{f}_1(\boldsymbol{q})}{\partial \tilde{x}_1} = 0. \tag{8.46}$$

Now we write system (8.43) in the nonconservative form

$$\frac{\partial \boldsymbol{q}}{\partial t} + \mathbb{A}_1(\boldsymbol{q})\frac{\partial \boldsymbol{q}}{\partial \tilde{x}_1} = 0. \tag{8.47}$$

Finally, we linearize this system around the state $\boldsymbol{q}_L = \mathbb{Q}(\boldsymbol{n})\boldsymbol{w}_L$ and obtain the linear system

$$\frac{\partial \boldsymbol{q}}{\partial t} + \mathbb{A}_1(\boldsymbol{q}_L)\frac{\partial \boldsymbol{q}}{\partial \tilde{x}_1} = 0, \tag{8.48}$$

which will be considered in the set $(-\infty,\ 0) \times (0,\ \infty)$ and equipped with the initial condition

$$\boldsymbol{q}(\tilde{x}_1, 0) = \boldsymbol{q}_L, \quad \tilde{x}_1 \in (-\infty,\ 0) \tag{8.49}$$

and the boundary condition

$$\boldsymbol{q}(0, t) = \boldsymbol{q}_R, \quad t > 0. \tag{8.50}$$

The goal is to choose $\boldsymbol{q}_R$ in such a way that the initial-boundary value problem (8.48)-(8.50) is well posed, i.e. has a unique solution. The solution can be written in the form

$$\boldsymbol{q}(\tilde{x}_1, t) = \sum_{s=1}^{4} \mu(\bar{x}_1, t)\boldsymbol{r}_s, \tag{8.51}$$

where $\boldsymbol{r}_s = \boldsymbol{r}_s(\boldsymbol{q}_L)$ are the eigenvectors of the matrix $\mathbb{A}_1(\boldsymbol{q}_L)$ corresponding to its eigenvalues $\tilde{\lambda}_s = \lambda_s(\boldsymbol{q}_L)$ and creating a basis in $\mathbb{R}^4$. Moreover,

$$\boldsymbol{q}_L = \sum_{s=1}^{4} \alpha_s \boldsymbol{r}_s, \quad \boldsymbol{q}_R = \sum_{s=1}^{4} \beta_s \boldsymbol{r}_s. \tag{8.52}$$

Substituting (8.51) into (8.48) and using the relation $\mathbb{A}_1(\boldsymbol{q}_L)\boldsymbol{r}_s = \tilde{\lambda}_s \boldsymbol{r}_s$, we find that problem (8.48)-(8.50) is equivalent to 4 mutually independent linear initial-boundary value scalar problems for $s = 1, \ldots, 4$:

$$\begin{aligned} &\frac{\partial \mu_s}{\partial t} + \tilde{\lambda}_s \frac{\partial \mu_s}{\partial \tilde{x}_1} = 0 \quad \text{in } (-\infty, 0) \times (0, \infty), \\ &\mu_s(\tilde{x}_1, 0) = \alpha_s, \quad \tilde{x}_1 \in (-\infty, 0), \\ &\mu_s(0, t) = \beta_s, \quad t \in (0, \infty), \end{aligned} \tag{8.53}$$

which can be solved by the method of characteristics, where the characteristics have the form

$$\tilde{x}_1 = \tilde{\lambda}_s t + \tilde{x}_1^0.$$

Using the fact that the solution μ_s is constant along characteristics, we find that

$$\mu_s(\tilde{x}_1, t) = \begin{cases} \alpha_s, & \tilde{x}_1 - \tilde{\lambda}_s t < 0, \\ \beta_s, & \tilde{x}_1 - \tilde{\lambda}_s t > 0. \end{cases} \tag{8.54}$$

The conclusion is that if

a) $\tilde{\lambda}_s > 0$, then $\beta_s = \alpha_s$ (β_s is not prescribed, but it is obtained by the extrapolation of μ_s to the boundary $\tilde{x}_1 = 0$),

b) if $\tilde{\lambda}_s = 0$, then β_s is not prescribed (but can be defined as $\beta_s = \alpha_s$ by the continuous extension of μ_s to the boundary $\tilde{x}_1 = 0$),

c) if $\tilde{\lambda}_s < 0$, then β_s must be prescribed.

Furthermore, we incorporate the fact that

$$\tilde{\lambda}_s(\boldsymbol{q}_L) = \lambda_s(\boldsymbol{w}^L, \boldsymbol{n}), \quad s = 1, \dots, 4, \tag{8.55}$$

where $\lambda_s(\boldsymbol{w}^L, \boldsymbol{n})$ are the eigenvalues of the Jacobi matrix $\mathbb{P}(\boldsymbol{w}^L, \boldsymbol{n})$ defined in (5.27). We can conclude, that we prescribe n_{pr} quantities characterizing $\boldsymbol{w}$, where n_{pr} is the number of negative eigenvalues λ_s, and extrapolate $n_{ex} = 4 - n_{pr}$ quantities. In what follows, we describe how the quantities β_s should be prescribed.

We shall take some state $\boldsymbol{q}_R^0 = \mathbb{Q}(\boldsymbol{n})\boldsymbol{w}_R^0$. The state $\boldsymbol{w}_R^0$ is the state vector of the far-field flow, or the incoming fluid at the inlet, or the initial condition, depending on the situation and interpretation. This state and above results will allow us to determine the sought boundary state $\boldsymbol{w}^R$. We express the state $\boldsymbol{q}_R^0$ in the form

$$\boldsymbol{q}_R^0 = \sum_{s=1}^{4} \gamma_s \boldsymbol{r}_s. \tag{8.56}$$

If we denote by $\mathbb{T}$ the matrix, which has $\boldsymbol{r}_s$ for its columns, we can see that for $\boldsymbol{\beta} = (\beta_1, \dots, \beta_4)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_4)^T$ we have

$$\begin{aligned} \boldsymbol{q}_L = \mathbb{T}\boldsymbol{\alpha} \;&\Leftrightarrow\; \boldsymbol{\alpha} = \mathbb{T}^{-1}\boldsymbol{q}_L, \\ \boldsymbol{q}_R^0 = \mathbb{T}\boldsymbol{\gamma} \;&\Leftrightarrow\; \boldsymbol{\gamma} = \mathbb{T}^{-1}\boldsymbol{q}_R^0. \end{aligned} \tag{8.57}$$

Now we calculate the state $\boldsymbol{q}_R$ according to the presented process:

$$\boldsymbol{q}_R := \sum_{s=1}^{4} \beta_s \boldsymbol{r}_s = \mathbb{T}\boldsymbol{\beta}, \tag{8.58}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_4)^T$ and

$$\beta_s = \begin{cases} \alpha_s, & \lambda_s \geq 0, \\ \gamma_s, & \lambda_s < 0. \end{cases} \tag{8.59}$$

Finally $\boldsymbol{w}^R = \mathbb{Q}^{-1}(\boldsymbol{n})\boldsymbol{q}_R$ and we can use this to calculate $\boldsymbol{H}(\boldsymbol{w}^L, \boldsymbol{w}^R, \boldsymbol{n})$.

In view of these results the form $\hat{b}_h$ (8.20) can be rewritten in the following way

$$\begin{aligned}
&\hat{b}_h(\boldsymbol{w}, \boldsymbol{\varphi}_h) \qquad (8.60)\\
&= -\sum_{K\in\mathcal{T}_{ht}} \int_K \sum_{s=1}^{2} (\mathbb{A}_s(\boldsymbol{w}) - z_s\mathbb{I})\boldsymbol{w} \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x}\\
&\quad + \sum_{\Gamma\in\mathcal{F}_{ht}^I} \int_\Gamma \left(\mathbb{P}_g^+(\langle \boldsymbol{w}\rangle_\Gamma, \boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(L)} + \mathbb{P}_g^-(\langle \boldsymbol{w}\rangle_\Gamma, \boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(R)} \right) \cdot [\boldsymbol{\varphi}_h]_\Gamma \,\mathrm{d}S\\
&\quad + \sum_{\Gamma\in\mathcal{F}_{ht}^{IO}} \int_\Gamma \left(\mathbb{P}_g^+(\langle \boldsymbol{w}\rangle_\Gamma, \boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(L)} + \mathbb{P}_g^-(\langle \boldsymbol{w}\rangle_\Gamma, \boldsymbol{n}_\Gamma)\boldsymbol{w}|_\Gamma^{(R)} \right) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S\\
&\quad + \sum_{\Gamma\in\mathcal{F}_{ht}^{W}} \int_\Gamma D_{\boldsymbol{w}}\boldsymbol{H}_g^W(\boldsymbol{w}|_\Gamma^{(L)}, \boldsymbol{n}|_\Gamma)\boldsymbol{w}|_\Gamma^{(L)} \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S,
\end{aligned}$$

where the state $\boldsymbol{w}|_\Gamma^{(R)}$ on $\Gamma \in \mathcal{F}_{ht}^{IO}$ is obtained by the described procedure.

8.3 Time discretization

The condition (8.35) is equivalent to a large system of ordinary differential equations. For solving this system we can apply several numerical schemes like *Runge-Kutta schemes* that are conditionally stable and the time step is strongly limited by the CFL-stability condition. This stability condition becomes very restrictive with increasing polynomial degree r of the discontinuous Galerkin space semidiscretization. Further, the fully implicit *backward Euler method* can be used. This method leads to a large system of highly nonlinear algebraic equations, whose numerical solution is rather complicated. For this reason we construct the semi-implicit scheme based on a suitable partial linearization of the form $\hat{b}_h$.

We consider a partition $0 = t_0 < t_1 < \ldots < t_M = T$ of the interval $[0, T]$ and set $\tau_m = t_m - t_{m-1},\ m = 1, \ldots, M$. We use the symbol $\boldsymbol{w}_h^m,\ \boldsymbol{w}_h^m \in \boldsymbol{S}_{ht_m}$, for the approximation of $\boldsymbol{w}_h(t_m)$. Assuming that we know $\boldsymbol{w}_h^k$ for $k \le m-1$, we are interested in finding the approximate solution $\boldsymbol{w}_h^m$ at time t_m.

First we shall approximate the ALE derivative. As an illustration we shall derive the second order scheme. We choose arbitrary, but fixed $\boldsymbol{x} \in \Omega_{ht}$. The definition of the ALE mapping gives us $\boldsymbol{x} = \mathcal{A}_t(\boldsymbol{X}),\ \boldsymbol{X} \in \Omega_{h0}$. In view of (5.51) we set $\tilde{\boldsymbol{w}}(\boldsymbol{X}, t) := \boldsymbol{w}(\mathcal{A}_t(\boldsymbol{X}), t)$. Using the Taylor expansion of the function $\tilde{\boldsymbol{w}}$ at t_m, we can write

$$\begin{aligned}
\tilde{\boldsymbol{w}}(t_{m-1}) &= \tilde{\boldsymbol{w}}(t_m - \tau_m) = \tilde{\boldsymbol{w}}(t_m) - \tau_m \frac{\partial \tilde{\boldsymbol{w}}}{\partial t} + \frac{\tau_m^2}{2}\frac{\partial^2 \tilde{\boldsymbol{w}}}{\partial t^2} + O(\tau_m^3)\\
\tilde{\boldsymbol{w}}(t_{m-2}) &= \tilde{\boldsymbol{w}}(t_m - (\tau_m + \tau_{m-1})) = \tilde{\boldsymbol{w}}(t_m) - (\tau_m + \tau_{m-1})\frac{\partial \tilde{\boldsymbol{w}}}{\partial t}\\
&\quad + \frac{(\tau_m + \tau_{m-1})^2}{2}\frac{\partial^2 \tilde{\boldsymbol{w}}}{\partial t^2} + O((\tau_m + \tau_{m-1})^3).
\end{aligned}$$

We neglect the terms $O(\tau_m^3)$ and $O((\tau_m + \tau_{m-1})^3)$ and solve the system with two

unknowns $\frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m)$ and $\frac{\partial^2 \tilde{\boldsymbol{w}}}{\partial t^2}(t_m)$, i.e.

$$\begin{pmatrix} -\tau_m & \frac{\tau_m^2}{2} \\ -(\tau_m+\tau_{m-1}) & \frac{(\tau_m+\tau_{m-1})^2}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m) \\ \frac{\partial^2 \tilde{\boldsymbol{w}}}{\partial t^2}(t_m) \end{pmatrix} \approx \begin{pmatrix} \tilde{\boldsymbol{w}}(t_{m-1}) - \tilde{\boldsymbol{w}}(t_m) \\ \tilde{\boldsymbol{w}}(t_{m-2}) - \tilde{\boldsymbol{w}}(t_m) \end{pmatrix}.$$

As a solution we obtain

$$\begin{aligned} \frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m) &\approx \frac{2\tau_m+\tau_{m+1}}{\tau_m(\tau_m+\tau_{m-1})}\tilde{\boldsymbol{w}}(t_m) - \frac{\tau_m+\tau_{m-1}}{\tau_m\tau_{m-1}}\tilde{\boldsymbol{w}}(t_{m-1}) \\ &\quad + \frac{\tau_m}{\tau_{m-1}(\tau_m+\tau_{m-1})}\tilde{\boldsymbol{w}}(t_{m-2}). \end{aligned}$$

We set

$$\hat{\boldsymbol{w}}_h^k(\boldsymbol{x}) = \boldsymbol{w}_h^k(\mathcal{A}_{t_k}(\mathcal{A}_{t_m}^{-1}(\boldsymbol{x}))), \quad \boldsymbol{x} \in \Omega_{ht_m}. \tag{8.61}$$

These assumptions lead to the second order approximation of the ALE derivative of $\boldsymbol{w}_h$ in time t_m :

$$\frac{D^{\mathcal{A}}\boldsymbol{w}_h}{Dt}(t_m) \approx \frac{2\tau_m+\tau_{m+1}}{\tau_m(\tau_m+\tau_{m-1})}\boldsymbol{w}_h^m - \frac{\tau_m+\tau_{m-1}}{\tau_m\tau_{m-1}}\hat{\boldsymbol{w}}_h^{m-1} + \frac{\tau_m}{\tau_{m-1}(\tau_m+\tau_{m-1})}\hat{\boldsymbol{w}}_h^{m-2}. \tag{8.62}$$

In a similar way we can derive an approximation of the ALE derivative of a general order q in time. It can be written in the form

$$\frac{D^{\mathcal{A}}\boldsymbol{w}_h}{Dt}(t_m) \approx \chi_0 \boldsymbol{w}_h^m + \sum_{l=1}^{q} \chi_l \hat{\boldsymbol{w}}_h^{m-l}, \tag{8.63}$$

where the coefficients $\chi_l,\ l = 0,\dots,q$ depend on $\tau_{m-l},\ l = 0,\dots,q-1$. In the beginning of our calculation, when $m < q$, we approximate the ALE derivative in time by formulas of lower order. Values of coefficients $\chi_l,\ l = 0,\dots,q$ for $q = 1,2,3$ can be found in [29].

In some terms we shall apply an extrapolation. As an example we shall derive the extrapolation of the second order. Starting from the Taylor expansion of the function $\tilde{\boldsymbol{w}}$ at the point t_m, we obtain

$$\begin{aligned} \tilde{\boldsymbol{w}}(t_{m-1}) &= \tilde{\boldsymbol{w}}(t_m-\tau_m) = \tilde{\boldsymbol{w}}(t_m) - \tau_m\frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m) + O(\tau_m^2), \\ \tilde{\boldsymbol{w}}(t_{m-2}) &= \tilde{\boldsymbol{w}}(t_m-(\tau_m+\tau_{m-1})) \\ &= \tilde{\boldsymbol{w}}(t_m) - (\tau_m+\tau_{m-1})\frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m) + O((\tau_m+\tau_{m-1})^2),. \end{aligned}$$

Again, we neglect the terms $O(\tau_m^2)$ and $O((\tau_m+\tau_{m-1})^2)$ and solve the system with two unknowns $\tilde{\boldsymbol{w}}(t_m)$ and $\frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m)$

$$\begin{pmatrix} 1 & -\tau_m \\ 1 & -(\tau_{m-1}\tau_m) \end{pmatrix} \cdot \begin{pmatrix} \tilde{\boldsymbol{w}}(t_m) \\ \frac{\partial \tilde{\boldsymbol{w}}}{\partial t}(t_m) \end{pmatrix} \approx \begin{pmatrix} \tilde{\boldsymbol{w}}(t_{m-1}) \\ \tilde{\boldsymbol{w}}(t_{m-2}) \end{pmatrix}.$$

As a solution we obtain

$$\tilde{\boldsymbol{w}}(t_m) \approx \frac{\tau_m+\tau_{m-1}}{\tau_{m-1}}\tilde{\boldsymbol{w}}(t_{m-1}) - \frac{\tau_m}{\tau_{m-1}}\tilde{\boldsymbol{w}}(t_{m-2}). \tag{8.64}$$

The definition of $\tilde{\boldsymbol{w}}(t_m)$ together with the knowledge of $\boldsymbol{w}_h^{m-1},\ \boldsymbol{w}_h^{m-2}$ and (8.61) give us the extrapolation of second order of $\boldsymbol{w}_h$ in time t_m. We shall denote by $\bar{\boldsymbol{w}}_h^m$ the extrapolation of $\boldsymbol{w}_h$ at time t_m. The extrapolation $\bar{\boldsymbol{w}}_h^m$ has the form

$$\bar{\boldsymbol{w}}_h^m = \frac{\tau_m + \tau_{m-1}}{\tau_{m-1}} \hat{\boldsymbol{w}}_h^{m-1} - \frac{\tau_m}{\tau_{m-1}} \hat{\boldsymbol{w}}_h^{m-2}. \tag{8.65}$$

In the same way we can derive the extrapolation of an arbitrary order q, which can be written in the form

$$\bar{\boldsymbol{w}}_h^m = \sum_{l=1}^{q} \iota_l \hat{\boldsymbol{w}}_h^{m-l}. \tag{8.66}$$

The constants $\iota_l,\ l = 1, \ldots, q$, depend on $\tau_{m-l},\ l = 0, \ldots, q-1$. In case of $m < q$ we use an extrapolation of order $\leq m$. Values of the coefficients $\iota_l,\ l = 0, \ldots, q$, for $q = 1, 2, 3$ can be found in [29].

In the time discretization of the inviscid terms we need to apply the extrapolation together with a linearization. This will be carried out by replacing the argument $\boldsymbol{w}_h^m$ by its extrapolation $\bar{\boldsymbol{w}}_h^m$ in nonlinearities of the form $\hat{b}_h$ defined in (8.60). This leads to the form

$$\begin{aligned}
&b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) \qquad\qquad (8.67)\\
&:= -\sum_{K \in \mathcal{T}_{ht_m}} \int_K \sum_{s=1}^{2} (\mathbb{A}_s(\bar{\boldsymbol{w}}_h^m) - z_s^m \mathbb{I}) \boldsymbol{w}_h^m \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x}\\
&\quad + \sum_{\Gamma \in \mathcal{F}_{ht_m}^I} \int_\Gamma \left(\mathbb{P}_g^+(\langle \bar{\boldsymbol{w}}_h^m \rangle_\Gamma, \boldsymbol{n}_\Gamma) \boldsymbol{w}_h^m|_\Gamma^{(L)} + \mathbb{P}_g^-(\langle \bar{\boldsymbol{w}}_h^m \rangle_\Gamma, \boldsymbol{n}_\Gamma) \boldsymbol{w}_h^m|_\Gamma^{(R)} \right) \cdot [\boldsymbol{\varphi}_h]_\Gamma \, \mathrm{d}S\\
&\quad + \sum_{\Gamma \in \mathcal{F}_{ht_m}^{IO}} \int_\Gamma \left(\mathbb{P}_g^+(\langle \bar{\boldsymbol{w}}_h^m \rangle_\Gamma, \boldsymbol{n}_\Gamma) \boldsymbol{w}_h^m|_\Gamma^{(L)} + \mathbb{P}_g^-(\langle \bar{\boldsymbol{w}}_h^m \rangle_\Gamma, \boldsymbol{n}_\Gamma) \bar{\boldsymbol{w}}_h^m|_\Gamma^{(R)} \right) \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S\\
&\quad + \sum_{\Gamma \in \mathcal{F}_{ht_m}^W} \int_\Gamma D_{\boldsymbol{w}} \boldsymbol{H}_g^W(\bar{\boldsymbol{w}}_h^m|_\Gamma^{(L)}, \boldsymbol{n}|_\Gamma) \boldsymbol{w}_h^m|_\Gamma^{(L)} \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S.
\end{aligned}$$

The same approach can be used for the viscous form (8.29) and the right-hand side form (8.34):

$$\begin{aligned}
a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) &:= \sum_{k \in \mathcal{T}_{ht_m}} \int_K \sum_{s=1}^{2} \sum_{k=1}^{2} \mathbb{K}_{sk}(\bar{\boldsymbol{w}}_h^m) \frac{\partial \boldsymbol{w}_h^m}{\partial x_k} \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} \mathrm{d}\boldsymbol{x} \qquad (8.68)\\
&- \sum_{\Gamma \in \mathcal{F}_{ht_m}^I} \int_\Gamma \sum_{s=1}^{2} \left\langle \sum_{k=1}^{2} \mathbb{K}_{sk}(\bar{\boldsymbol{w}}_h^m) \frac{\partial \boldsymbol{w}_h^m}{\partial x_k} \right\rangle_\Gamma (n_\Gamma)_s \cdot [\boldsymbol{\varphi}_h]_\Gamma \, \mathrm{d}S\\
&- \sum_{\Gamma \in \mathcal{F}_{ht_m}^D} \int_\Gamma \sum_{s=1}^{2} \sum_{k=1}^{2} \mathbb{K}_{sk}(\bar{\boldsymbol{w}}_h^m|_\Gamma^{(L)}) \frac{\partial \boldsymbol{w}_h^m}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S\\
&- \Theta \sum_{\Gamma \in \mathcal{F}_{ht_m}^I} \int_\Gamma \sum_{s=1}^{2} \left\langle \sum_{k=1}^{2} \mathbb{K}_{sk}^T(\bar{\boldsymbol{w}}_h^m) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k} \right\rangle_\Gamma (n_\Gamma)_s \cdot [\boldsymbol{w}_h^m]_\Gamma \, \mathrm{d}S
\end{aligned}$$

$$-\Theta \sum_{\Gamma\in\mathcal{F}^D_{ht_m}} \int_\Gamma \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}^T_{sk}(\bar{\boldsymbol{w}}_h^m|_\Gamma^{(L)}) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \boldsymbol{w}_h^m|_\Gamma^{(L)} \mathrm{d}S.$$

$$l_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\varphi}_h) := \sum_{\Gamma\in\mathcal{F}^D_{ht_m}} \int_\Gamma \sigma \bar{\boldsymbol{w}}_B \cdot \boldsymbol{\varphi}_h|_\Gamma^{(L)} \mathrm{d}S \tag{8.69}$$

$$-\Theta \sum_{\Gamma\in\mathcal{F}^D_{ht_m}} \int_\Gamma \sum_{s=1}^{2}\sum_{k=1}^{2} \mathbb{K}^T_{sk}(\bar{\boldsymbol{w}}_h^m|_\Gamma^{(L)}) \frac{\partial \boldsymbol{\varphi}}{\partial x_k}|_\Gamma^{(L)} (n_\Gamma)_s \cdot \bar{\boldsymbol{w}}_B \mathrm{d}S.$$

These considerations lead us to the following semi-implicit scheme: For each $m \leq 1$ we look for $\boldsymbol{w}_k^m \in \boldsymbol{S}_{ht_m}$ such that

$$\begin{aligned}
&\boldsymbol{w}_h^m \in \boldsymbol{S}_{ht_m},\\
&\left(\chi_0 \boldsymbol{w}_h^m + \sum_{l=1}^{q} \chi_l \hat{\boldsymbol{w}}_h^{m-l}, \boldsymbol{\varphi_h}\right) + d_h(\boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) + b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) \qquad (8.70)\\
&\quad + a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) + J_h(\boldsymbol{w}_h^m(t), \boldsymbol{\varphi}_h) = l_h(\bar{\boldsymbol{w}}_h^m(t), \boldsymbol{\varphi}_h)\\
&\quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht_m},\ m = 1, 2, \ldots,\\
&\boldsymbol{w}_h(0) = \boldsymbol{w}_h^0.
\end{aligned}$$

8.4 Shock capturing

In order to avoid spurious oscillations in the approximate solution in the vicinity of discontinuities or steep gradients, we apply artificial viscosity forms introduced in [33].

It is based on the *discontinuity indicator* $g^m(K)$ defined by

$$g^m(K) = \int_{\partial K} \left[\hat{\rho}_h^{m-1}\right]^2 dS/(h_K |K|^{3/4}), \quad K \in \mathcal{T}_{ht_m}, \tag{8.71}$$

where

$$\hat{\rho}_h^{m-1}(\boldsymbol{x}) = \rho_h^{m-1}(\mathcal{A}_{t_{m-1}}(\mathcal{A}_{t_m}^{-1}(\boldsymbol{x}))). \tag{8.72}$$

By $\left[\hat{\rho}_h^{m-1}\right]$ we denote the jump of the density on ∂K at time t_{m-1} and $|K|$ denotes the area of the element K. The indicator $g^m(K)$ was constructed in such a way that it takes an anisotropy of the computational mesh into account. Now we introduce the *discrete discontinuity indicator*

$$G^m(K) = 0 \quad \text{if } g^m(K) < 1,\ K \in \mathcal{T}_{ht_m} \tag{8.73}$$

$$G^m(K) = 1 \quad \text{if } g^m(K) \geq 1,\ K \in \mathcal{T}_{ht_m}, \tag{8.74}$$

and add the artificial viscosity form

$$\beta_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) = \nu_1 \sum_{K\in\mathcal{T}_{ht_m}} h_K G^m(K) \int_K \nabla \boldsymbol{w}_h^m \cdot \nabla \boldsymbol{\varphi}_h d\boldsymbol{x} \tag{8.75}$$

with $\nu_1 = \mathrm{O}(1)$ to the left-hand side of (8.70). Since the artificial viscosity form is rather local, it is proposed to augment the left-hand side of (8.70) by adding the form

$$\hat{J}_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) = \nu_2 \sum_{\Gamma \in \mathcal{F}_{ht_m}^I} \frac{1}{2}(G^m(K_\Gamma^{(L)}) + G^m(K_\Gamma^{(R)})) \int_\Gamma [\boldsymbol{w}_h^m] \cdot [\boldsymbol{\varphi}_h]\, dS, \quad (8.76)$$

where $\nu_2 = \mathrm{O}(1)$. This allows to strengthen the influence of neighbouring elements and improves the behavior of the method.

The resulting scheme has the following form: For each $m = 1,\ 2,\ \ldots$ we seek $\boldsymbol{w}_k^m \in \boldsymbol{S}_{ht_m}$ such that

$$\begin{aligned}
&\boldsymbol{w}_h^m \in \boldsymbol{S}_{ht_m},\\
&\left(\chi_0 \boldsymbol{w}_h^m + \sum_{l=1}^{q} \chi_l \hat{\boldsymbol{w}}_h^{m-l}, \boldsymbol{\varphi_h}\right) + d_h(\boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) + b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) \qquad (8.77)\\
&\quad + a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) + J_h(\boldsymbol{w}_h^m(t), \boldsymbol{\varphi}_h) + \beta_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h)\\
&\quad + \hat{J}_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{w}_h^m, \boldsymbol{\varphi}_h) = l_h(\bar{\boldsymbol{w}}_h^m(t), \boldsymbol{\varphi}_h)\\
&\quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht_m},\ m = 1, 2, \ldots,\\
&\boldsymbol{w}_h(0) = \boldsymbol{w}_h^0.
\end{aligned}$$

Chapter 9

Discretization of the structural problem

This chapter is devoted to the discretization of the structural problem

$$\rho^b \frac{\partial^2 u_i}{\partial t^2} + C\rho^b \frac{\partial u_i}{\partial t} - \sum_{j=1}^{2} \frac{\partial \tau_{ij}^b}{\partial x_j} = 0 \quad i = 1, 2, \tag{9.1}$$

with the initial conditions

$$\boldsymbol{u}(\cdot, 0) = \mathbf{0} \quad \text{in } \Omega^b, \tag{9.2}$$

$$\frac{\partial \boldsymbol{u}}{\partial t}(\cdot, 0) = \mathbf{0} \quad \text{in } \Omega^b. \tag{9.3}$$

and the boundary conditions

$$\boldsymbol{u} = 0 \quad \text{on } \Gamma_D^b \times (0, T), \tag{9.4}$$

$$\sum_{j=1}^{2} \tau_{ij}^b n_j = T_i^{\boldsymbol{n}} \quad \text{on } \Gamma_W^b \times (0, T), \quad i = 1, 2. \tag{9.5}$$

Let us remind that we apply the generalized Hooke's law for an isotropic body (6.15). Hence,

$$\tau_{ij}^b(\boldsymbol{x}) = \lambda^b(\boldsymbol{x}) \mathrm{div} \boldsymbol{u}(\boldsymbol{x}) \delta_{ij} + 2\mu^b(\boldsymbol{x}) e_{ij}(\boldsymbol{x}), \quad i, j = 1, 2.$$

The space discretization will be carried out by the *conforming finite element method* (FEM). The time discretization will be realized with the aid of the *Newmark method*.

9.1 Space discretization

By Ω_h^b we denote a polygonal approximation of the domain Ω^b. We construct a triangulation $\mathcal{T}_h^b$ of the domain Ω_h^b formed by a finite number of closed triangles with the following properties:

- $\bar{\Omega}_h^b = \bigcup_{K \in \mathcal{T}_h^b} K$.

- The intersection of two different elements $K, K' \in \mathcal{T}_h^b$ is either empty or a common edge of these elements or their common vertex.
- The vertices of elements adjacent to $\partial\Omega_h^b$, which belong to $\partial\Omega_h^b$ are elements of $\partial\Omega^b$.
- The set $\bar{\Gamma}_W \cap \bar{\Gamma}_D$ is formed by vertices of some elements $K \in \mathcal{T}_h^b$.

Further, by Γ_{Wh}^b and Γ_{Dh}^b we denote the parts of $\partial\Omega_h^b$ approximating Γ_W^b and Γ_D^b.

The approximate solution of the structural problem will be sought in the finite-dimensional space $\boldsymbol{X_h} = X_h \times X_h$, where

$$X_h = \left\{ v_h \in C(\bar{\Omega}_h^b);\ v_h|_K \in P_s(K),\ \forall K \in \mathcal{T}_h^b \right\}. \tag{9.6}$$

and $s \geq 1$ is an integer. Let us mention that in our numerical experiments we set $s = 1$. In $\boldsymbol{X}_h$ we define the subspace $\boldsymbol{V_h} = V_h \times V_h$, where

$$V_h = \left\{ y_h \in X_h;\ y_h|_{\bar{\Gamma}_{Dh}^b} = 0 \right\}. \tag{9.7}$$

The derivation of the space semidiscretization can be obtained in a standard way. If we assume that $\boldsymbol{u} = (u_1, u_2)$ is a sufficiently regular classical solution of (9.1)-(9.5), we multiply system (9.1) by any test function $y_{hi} \in V_h,\ i = 1, 2$ and integrate over Ω_h^b. This leads to

$$\int_{\Omega_h^b} \rho^b \frac{\partial^2 u_i(t)}{\partial t^2} y_i \mathrm{d}\boldsymbol{x} + \int_{\Omega_h^b} C\rho^b \frac{\partial u_i(t)}{\partial t} y_i \mathrm{d}\boldsymbol{x} - \int_{\Omega_h^b} \sum_{j=1}^{2} \frac{\partial \tau_{ij}^b(\boldsymbol{u}(t))}{\partial x_j} y_i \mathrm{d}\boldsymbol{x} = 0 \quad i = 1, 2. \tag{9.8}$$

Using the Green's theorem and applying the boundary conditions (9.5), we get

$$\int_{\Omega_h^b} \rho^b \frac{\partial^2 u_i(t)}{\partial t^2} y_i \mathrm{d}\boldsymbol{x} + \int_{\Omega_h^b} C\rho^b \frac{\partial u_i(t)}{\partial t} y_i \mathrm{d}\boldsymbol{x} + \int_{\Omega_h^b} \sum_{j=1}^{2} \tau_{ij}^b(\boldsymbol{u}(t)) \frac{\partial y_i}{\partial x_j} \mathrm{d}\boldsymbol{x} = \int_{\Gamma_W^b} T_i^{\boldsymbol{n}}(t) y_i \mathrm{d}S \quad i = 1, 2. \tag{9.9}$$

Based on this approach and the use of the generalized Hooke's law for an isotropic body (6.15), we set the forms defined for $\boldsymbol{u}_h = (u_{h1}, u_{h2})$, $\boldsymbol{y}_h = (y_{h1}, y_{h2}) \in \boldsymbol{X}_h$

$$\begin{aligned}
a_h^b(\boldsymbol{u}_h, \boldsymbol{y}_h) &= \int_{\Omega_h^b} \sum_{j=1}^{2} \tau_{ij}^b(\boldsymbol{u}_h(t)) \frac{\partial y_{hi}}{\partial x_j} \mathrm{d}\boldsymbol{x} \qquad (9.10)\\
&= \int_{\Omega_h^b} \lambda^b \mathrm{div}\boldsymbol{u}_h \, \mathrm{div}\boldsymbol{y}_h \mathrm{d}\boldsymbol{x} + 2\int_{\Omega_h^b} \mu^b \sum_{i,j=1}^{2} e_{ij}^b(\boldsymbol{u}_h)\, e_{ij}^b(\boldsymbol{y}_h) \mathrm{d}\boldsymbol{x}\\
&= \int_{\Omega_h^b} \lambda^b \left(\frac{\partial u_{h1}(t)}{\partial x_1} + \frac{\partial u_{h2}(t)}{\partial x_2} \right) \left(\frac{\partial y_{h1}}{\partial x_1} + \frac{\partial y_{h2}}{\partial x_2} \right) \mathrm{d}\boldsymbol{x}\\
&\quad + 2\int_{\Omega_h^b} \mu^b \left(\frac{\partial u_{h1}(t)}{\partial x_1} \frac{\partial y_{h1}}{\partial x_1} + \frac{\partial u_{h2}(t)}{\partial x_2} \frac{\partial y_{h2}}{\partial x_2} \right) \mathrm{d}\boldsymbol{x}\\
&\quad + \int_{\Omega_h^b} \mu^b \left(\frac{\partial u_{h1}(t)}{\partial x_2} + \frac{\partial u_{h2}(t)}{\partial x_1} \right) \left(\frac{\partial y_{h1}}{\partial x_2} + \frac{\partial y_{h2}}{\partial x_1} \right) \mathrm{d}\boldsymbol{x},
\end{aligned}$$

$$(\boldsymbol{u}_h, \boldsymbol{y}_h)_{\Omega_h^b} = \int_{\Omega_h^b} \boldsymbol{u}_h \cdot \boldsymbol{y}_h \mathrm{d}\boldsymbol{x} = \sum_{i=1}^{2} \int_{\Omega_h^b} u_i y_i \mathrm{d}\boldsymbol{x}, \tag{9.11}$$

$$(\boldsymbol{u}_h, \boldsymbol{y}_h)_{\Gamma_{Wh}^b} = \int_{\Gamma_{Wh}^b} \boldsymbol{u}_h \cdot \boldsymbol{y}_h \mathrm{d}S = \sum_{i=1}^{2} \int_{\Gamma_{Wh}^b} u_i y_i \mathrm{d}S. \tag{9.12}$$

We shall use the notation $\boldsymbol{u}_h'(t) = \frac{\partial \boldsymbol{u}_h(t)}{\partial t}$ and $\boldsymbol{u}_h''(t) = \frac{\partial^2 \boldsymbol{u}_h(t)}{\partial t^2}$. Then we define the *approximate solution of the structural problem* as a function $t \in [0,T] \to \boldsymbol{u}_h(t) \in \boldsymbol{V}_h$ such that there exist the derivatives $\boldsymbol{u}_h'(t)$, $\boldsymbol{u}_h''(t)$ and the identity

$$(\rho^b \boldsymbol{u}_h''(t), \boldsymbol{y}_h)_{\Omega_h^b} + (C\rho^b \boldsymbol{u}_h'(t), \boldsymbol{y}_h)_{\Omega_h^b} + a_h^b(\boldsymbol{u}_h(t), \boldsymbol{y}_h) = (\boldsymbol{T}_h^{\boldsymbol{n}}(t), \boldsymbol{y}_h)_{\Gamma_{Wh}^b}, \quad \forall \boldsymbol{y}_h \in \boldsymbol{V}_h, \quad \forall t \in [0,T], \tag{9.13}$$

and the initial conditions

$$\boldsymbol{u}_h(\boldsymbol{X}, 0) = 0, \quad \boldsymbol{X} \in \Omega_h^b, \tag{9.14}$$

$$\boldsymbol{u}_h'(\boldsymbol{X}, 0) = 0, \quad \boldsymbol{X} \in \Omega_h^b, \tag{9.15}$$

are satisfied.

The discrete problem (9.13)-(9.15) is equivalent to the solution of a system of ordinary differential equations. First, we start from finding the basis of the space $\boldsymbol{V}_h$ for the linear elements. We assume that k is the number of all vertices of $\mathcal{T}_h^b$ and n is the number of all vertices in $\Omega_h^b \cup \Gamma_{Wh}^b$. Then we can number the vertices in the following way: $P_1, \ldots, P_n$ are the vertices in $\Omega_h^b \cup \Gamma_{Wh}^b$ and $P_{n+1}, \ldots, P_k$ are the vertices on $\bar{\Gamma}_{Dh}^b$. Now, let us define the basis of the space X_h :

$$\varphi_j \in X_h : \ \varphi_j(P_j) = \delta_{ij}, \quad i,j = 1, \ldots, k. \tag{9.16}$$

Then the system of $K = 2k$ vector-valued functions $(\varphi_1, 0), \ldots, (\varphi_k, 0), (0, \varphi_1), \ldots, (0, \varphi_k)$ forms a basis of the space $\boldsymbol{X}_h$. The basis of V_h is formed by $\varphi_j,\ j = 1, \ldots, n$. Then the system of $N = 2n$ vector-value functions $(\varphi_1, 0), \ldots, (\varphi_n, 0), (0, \varphi_1), \ldots, (0, \varphi_n)$ form a basis of the space $\boldsymbol{V}_h$. We denote by $u_{hi} \in V_h,\ i = 1, 2$, the components of the vector $\boldsymbol{u}_h$. Then the function $u_{hi},\ i = 1, 2$ in time $t \in [0,T]$ can be express by the basis functions of the space $\boldsymbol{V}_h$ in the form

$$u_{hi}(t) = \sum_{j=1}^{n} p_j^{(i)}(t)\varphi_j, \quad i = 1, 2, \text{ with } p_j^{(i)}(t) = u_{hi}(P_j, t), \quad t \in [0,T]. \tag{9.17}$$

We see that each function $\boldsymbol{u}_h \in \boldsymbol{V}_h$ can be expressed by N coefficients $p_j^{(i)}(t) \in \mathbb{R},\ i = 1, 2,\ j = 1, \ldots, n$, and define the vector-valued function $\boldsymbol{p}(t)$:

$$\boldsymbol{p}(t) = (p_1^{(1)}(t), \ldots, p_n^{(1)}(t), p_1^{(2)}(t), \ldots, p_n^{(2)}(t))^T. \tag{9.18}$$

From (9.17) we have

$$u_{hi}'(t) = \sum_{j=1}^{n} p_j'^{(i)}(t)\varphi_j, \quad i = 1, 2, \quad t \in [0,T], \tag{9.19}$$

$$u_{hi}''(t) = \sum_{j=1}^{n} p_j''^{(i)}(t)\varphi_j, \quad i = 1, 2, \quad t \in [0,T]. \tag{9.20}$$

We transform equation (9.13) using the basis functions. We shall proceed term by term and use the basis functions $(\varphi_1,0),\dots,(\varphi_n,0),(0,\varphi_1),\dots,(0,\varphi_n)$, as a test function $\boldsymbol{y}_h$. First, we have a look at the terms $(\rho^b\boldsymbol{u}_h''(t),\boldsymbol{y}_h)_{\Omega_h^b}+(C\rho^b\boldsymbol{u}_h'(t),\boldsymbol{y}_h)_{\Omega_h^b}$:

$$(\rho^b(\boldsymbol{u}_h''(t)+C\boldsymbol{u}_h'(t)),(\varphi_j,0))_{\Omega_h^b} = \sum_{i=1}^{n}(\rho^b\varphi_i,\varphi_j)_{\Omega_h^b}(p_i''^{(1)}(t)+Cp_i'^{(1)}(t)), \quad (9.21)$$

$$(\rho^b(\boldsymbol{u}_h''(t)+C\boldsymbol{u}_h'(t)),(0,\varphi_j))_{\Omega_h^b} = \sum_{i=1}^{n}(\rho^b\varphi_i,\varphi_j)_{\Omega_h^b}(p_i''^{(2)}(t)+Cp_i'^{(2)}(t)), \quad (9.22)$$

$$j=1,\dots,n.$$

We define the elements of the matrix $\{m_{ij}\}_{i,j=1}^{N}$:

$$m_{ij}=m_{(i+n)(j+n)}=(\rho^b\varphi_i,\varphi_j)_{\Omega_h^b},\quad i,j=1,\dots,n, \quad (9.23)$$

$$m_{(i+n)j}=m_{i(j+n)}=0,\quad i,j=1,\dots,n. \quad (9.24)$$

We use the same approach in the term $a_h^b(\boldsymbol{u}_h,\boldsymbol{y}_h)$ defined by (9.10):

$$\begin{aligned} a_h^b(\boldsymbol{u}_h,(\varphi_j,0)) &= \int_{\Omega_h^b}\left((\lambda^b+2\mu^b)\frac{\partial u_{h1}}{\partial x_1}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial u_{h1}}{\partial x_2}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x} \\ &\quad+\int_{\Omega_h^b}\left(\lambda^b\frac{\partial u_{h2}}{\partial x_2}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial u_{h2}}{\partial x_1}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x} \\ &= \sum_{i=1}^{n}p_i^{(1)}\int_{\Omega_h^b}\left((\lambda^b+2\mu^b)\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x} \\ &\quad+\sum_{i=1}^{n}p_i^{(2)}\int_{\Omega_h^b}\left(\lambda^b\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x}, \quad (9.25) \\ &\quad j=1,\dots,n, \end{aligned}$$

$$\begin{aligned} a_h^b(\boldsymbol{u}_h,(0,\varphi_j)) &= \int_{\Omega_h^b}\left((\lambda^b+2\mu^b)\frac{\partial u_{h2}}{\partial x_2}\frac{\partial\varphi_j}{\partial x_2}+\mu^b\frac{\partial u_{h2}}{\partial x_1}\frac{\partial\varphi_j}{\partial x_1}\right)\mathrm{d}\boldsymbol{x} \\ &\quad+\int_{\Omega_h^b}\left(\lambda^b\frac{\partial u_{h1}}{\partial x_1}\frac{\partial\varphi_j}{\partial x_2}+\mu^b\frac{\partial u_{h1}}{\partial x_2}\frac{\partial\varphi_j}{\partial x_1}\right)\mathrm{d}\boldsymbol{x} \\ &= \sum_{i=1}^{n}p_i^{(2)}\int_{\Omega_h^b}\left((\lambda^b+2\mu^b)\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_2}+\mu^b\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_1}\right)\mathrm{d}\boldsymbol{x} \\ &\quad+\sum_{i=1}^{n}p_i^{(1)}\int_{\Omega_h^b}\left(\lambda^b\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_2}+\mu^b\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_1}\right)\mathrm{d}\boldsymbol{x}, \quad (9.26) \\ &\quad j=1,\dots,n. \end{aligned}$$

Further, we define the elements of the matrix $\{k_{ij}\}_{i,j=1}^{N}$:

$$k_{ij} = \int_{\Omega_h^b}\left((\lambda^b+2\mu^b)\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x}, \quad (9.27)$$

$$k_{(i+n)j} = \int_{\Omega_h^b}\left(\lambda^b\frac{\partial\varphi_i}{\partial x_2}\frac{\partial\varphi_j}{\partial x_1}+\mu^b\frac{\partial\varphi_i}{\partial x_1}\frac{\partial\varphi_j}{\partial x_2}\right)\mathrm{d}\boldsymbol{x}, \quad (9.28)$$

$$k_{i(j+n)} = \int_{\Omega_h^b} \left(\lambda^b \frac{\partial \varphi_i}{\partial x_1} \frac{\partial \varphi_j}{\partial x_2} + \mu^b \frac{\partial \varphi_i}{\partial x_2} \frac{\partial \varphi_j}{\partial x_1} \right) \mathrm{d}\boldsymbol{x}, \tag{9.29}$$

$$k_{(i+n)(j+n)} = \int_{\Omega_h^b} \left((\lambda^b + 2\mu^b) \frac{\partial \varphi_i}{\partial x_2} \frac{\partial \varphi_j}{\partial x_2} + \mu^b \frac{\partial \varphi_i}{\partial x_1} \frac{\partial \varphi_j}{\partial x_1} \right) \mathrm{d}\boldsymbol{x}, \tag{9.30}$$
$$i, j = 1, \ldots, n.$$

For the right-hand side we define the coefficients $g_i,\ i = 1, \ldots, N$, in the following way:

$$g_i(t) = (T_{h1}^{\boldsymbol{n}}(t), \varphi_i)_{\Gamma_{Wh}^b}, \quad i = 1, \ldots, n, \tag{9.31}$$

$$g_{i+n}(t) = (T_{h2}^{\boldsymbol{n}}(t), \varphi_i)_{\Gamma_{Wh}^b}, \quad i = 1, \ldots, n. \tag{9.32}$$

The matrix $\mathbb{M} = \{m_{ij}\}_{i,j=1}^N$ is called the *mass matrix* and the matrix $\mathbb{K} = \{k_{ij}\}_{i,j=1}^N$ is the *stiffness matrix*. The vector $\boldsymbol{G}(t) = (g_1(t), \ldots, g_N(t))$ represents the aerodynamic force. The discrete problem (9.13) is equivalent to the system of ordinary differential equations

$$\mathbb{M}(\boldsymbol{p}''(t) + C\boldsymbol{p}'(t)) + \mathbb{K}\boldsymbol{p}(t) = \boldsymbol{G}(t), \tag{9.33}$$
$$\boldsymbol{p} : [0, T] \to I\!R^N, \quad \boldsymbol{G} : [0, T] \to I\!R^N.$$

The initial conditions (9.14)-(9.15) are equivalent to

$$\boldsymbol{p}(0) = 0, \quad \boldsymbol{p}'(0) = 0. \tag{9.34}$$

9.2 Time discretization

This section will be devoted to the solution of the discrete initial value problem (9.33)-(9.34). To this end we shall use the *Newmark method*([19]), which is often used in structural mechanics.

9.2.1 Newmark method

First, we derive the Newmark scheme for the general initial value problem of the second order

$$y''(t) = \psi(t, y(t), y'(t)), \tag{9.35}$$

$$y(0) = y^0, \tag{9.36}$$

$$y'(0) = r^0, \tag{9.37}$$

where $\psi : [0, T] \times I\!R^2 \to I\!R$ is a continuous function and $y^0, r^0 \in I\!R$. We consider the partition of the time interval $[0, T]$ formed by the time instants $0 = t_0 < t_1 < \ldots < t_M = T$, where $\tau_m = t_m - t_{m-1},\ m = 1, \ldots, M$. Let us assume that $y \in C^4([0, T])$. Then we express the value of $y(t_m)$ using the Taylor expansion of the third order at the point t_{m-1} :

$$y(t_m) = y(t_{m-1} + \tau_m) = y(t_{m-1}) + \tau_m y'(t_{m-1}) + \frac{1}{2}\tau_m^2 y''(t_{m-1}) + \frac{1}{6}\tau_m^3 y'''(t_{m-1}) + O(\tau_m^4). \tag{9.38}$$

On the right-hand side of (9.38) we add and subtract the term $\tau_m^2\delta(y''(t_m)-y''(t_{m-1}))$, where $\delta \in I\!R$ is a parameter:

$$\begin{aligned} y(t_m) &= y(t_{m-1}) + \tau_m y'(t_{m-1}) + \tau_m^2 \left(\delta y''(t_m) + \left(\frac{1}{2} - \delta \right) y''(t_{m-1}) \right) \\ &\quad -\tau_m^2\delta(y''(t_m) - y''(t_{m-1})) + \frac{1}{6}\tau_m^3 y'''(t_{m-1}) + O(\tau_m^4). \end{aligned} \tag{9.39}$$

Using the Taylor expansion for the second-order time derivative of the function y at the point t_{m-1}, we obtain

$$y''(t_m) - y''(t_{m-1}) = \tau_m y'''(t_{m-1}) + O(\tau_m^2). \tag{9.40}$$

Relations (9.39) and (9.40) give us

$$\begin{aligned} y(t_m) &= y(t_{m-1}) + \tau_m y'(t_{m-1}) + \tau_m^2 \left(\delta y''(t_m) + \left(\frac{1}{2} - \delta \right) y''(t_{m-1}) \right) \\ &\quad +\tau_m^3 \left(\frac{1}{6} - \delta \right) y'''(t_{m-1}) + O(\tau_m^4). \end{aligned} \tag{9.41}$$

We consider the term $\tau_m^3 \left(\frac{1}{6} - \delta\right) y'''(t_{m-1}) + O(\tau_m^4)$ as the error of the order $O(\tau_m^3)$. Assuming that y is the solution of the initial value problem (9.35)-(9.37), we can write

$$\begin{aligned} y(t_m) &= y(t_{m-1}) + \tau_m y'(t_{m-1}) + \tau_m^2 \left(\delta\psi(t_m, y(t_m), y'(t_m)) \right. \\ &\quad \left. + \left(\frac{1}{2} - \delta \right) \psi(t_{m-1}, y(t_{m-1}), y'(t_{m-1})) \right) + O(\tau_m^3). \end{aligned} \tag{9.42}$$

With the use of the Taylor expansion we can derive the value of the first derivative of y at the point t_m :

$$y'(t_m) = y'(t_{m-1} = \tau_m) = y'(t_{m-1}) + \tau_m y''(t_{m-1}) + \frac{1}{2}\tau_m^2 y'''(t_{m-1}) + O(\tau_m^3). \tag{9.43}$$

We add and subtract the term $\phi(y''(t_m) - y''(t_{m-1}))$, where $\phi \in I\!R$ is a parameter, and get

$$\begin{aligned} y'(t_m) &= y'(t_{m-1}) + \tau_m(\phi y''(t_m)) + (1-\phi)y''(t_{m-1})) \\ &\quad +\tau_m^2 \left(\frac{1}{2} - \phi \right) y'''(t_{m-1}) + O(\tau_m^3). \end{aligned} \tag{9.44}$$

Again, we consider the term $\tau_m^2 \left(\frac{1}{2} - \phi\right) y'''(t_{m-1}) + O(\tau_m^3)$ as the error of the order $O(\tau_m^2)$. Assuming that y is the solution of the initial value problem (9.35)-(9.37), we can write

$$y'(t_m) = y'(t_{m-1}) + \tau_m(\phi\psi(t_m, y(t_m), y'(t_m)) + (1-\phi)\psi(t_{m-1}, y(t_{m-1}), y'(t_{m-1}))) + O(\tau_m^2). \tag{9.45}$$

Using the approximation $y_m \approx y(t_m),\ r_m \approx y'(t_m)$, putting $\psi_m = \psi(t_m, y_m, r_m)$ and neglecting the discretization error, we obtain the Newmark scheme

$$y_m = y_{m-1} + \tau_m r_{m-1} + \tau_m^2 \left(\delta\psi_m + \left(\frac{1}{2} - \delta \right) \psi_{m-1} \right), \tag{9.46}$$

$$r_m = r_{m-1} + \tau_m(\phi\psi_m + (1-\phi)\psi_{m-1}). \tag{9.47}$$

In our numerical experiments, the parameters $\delta = \frac{1}{4}$ and $\phi = \frac{1}{2}$ were used. This choice yields the Newmark method of the second order.

9.2.2 Time discretization of the structural problem

Using the derived Newmark scheme, we shall solve the system of ordinary differential equations (9.33) with the initial conditions (9.34).

First, we transform system (9.33) to the more suitable form:

$$\boldsymbol{p}'' = \mathbb{M}^{-1}\boldsymbol{G} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p} - C\boldsymbol{p}'. \tag{9.48}$$

Let us set $\boldsymbol{p}^0 = \boldsymbol{p}(0) = 0,\ \boldsymbol{r}^0 = \boldsymbol{p}'(0) = 0,\ \boldsymbol{G}_m = \boldsymbol{G}(t_m)$, and introduce the approximations $\boldsymbol{p}_m \approx \boldsymbol{p}(t_m),\ \boldsymbol{r}_m \approx \boldsymbol{p}'(t_m)$ for $m = 1, \ldots, M$. The Newmark scheme can be written in the form

$$\begin{aligned} \boldsymbol{p}_m &= \boldsymbol{p}_{m-1} + \tau_m \boldsymbol{r}_{m-1} + \tau_m^2 \left(\delta \left(\mathbb{M}^{-1}\boldsymbol{G}_m - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_m - C\boldsymbol{r}_m\right)\right. \\ &\quad \left. + \left(\frac{1}{2} - \delta\right)\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right)\right), \end{aligned} \tag{9.49}$$

$$\begin{aligned} \boldsymbol{r}_m &= \boldsymbol{r}_{m-1} + \tau_m \left(\phi \left(\mathbb{M}^{-1}\boldsymbol{G}_m - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_m - C\boldsymbol{r}_m\right)\right. \\ &\quad \left. + (1-\phi)\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right)\right). \end{aligned} \tag{9.50}$$

From equation (9.50) we express $\boldsymbol{r}_m$:

$$\begin{aligned} \boldsymbol{r}_m &= \frac{1}{1 + C\phi\tau_m}\left(\boldsymbol{r}_{m-1} + \tau_m \left(\phi \left(\mathbb{M}^{-1}\boldsymbol{G}_m - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_m\right)\right.\right. \\ &\quad \left.\left. + (1-\phi)\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right)\right)\right). \end{aligned} \tag{9.51}$$

The substitution of (9.51) in (9.49) yields the relation

$$\begin{aligned} \boldsymbol{p}_m &= \boldsymbol{p}_{m-1} + \tau_m \boldsymbol{r}_{m-1} + \delta\tau_m^2 \left(\mathbb{M}^{-1}\boldsymbol{G}_m - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_m - \frac{C}{1 + C\phi\tau_m}\boldsymbol{r}_{m-1}\right. \\ &\quad - \frac{C\phi\tau_m}{1 + C\phi\tau_m}\left(\mathbb{M}^{-1}\boldsymbol{G}_m - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_m\right) \\ &\quad \left. - \frac{C\phi\tau_m}{1 + C\phi\tau_m}(1-\phi)\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right)\right) \\ &\quad + \left(\frac{1}{2} - \delta\right)\tau_m^2\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right). \end{aligned} \tag{9.52}$$

For the sake of simplicity we set

$$\xi_m = \delta\tau_m^2\left(1 - \frac{C\phi\tau_m}{1 + C\phi\tau_m}\right) = \frac{\delta\tau_m^2}{1 + C\phi\tau_m}. \tag{9.53}$$

The substitution of (9.53) in (9.52) yields the relation which can be written in the form

$$\begin{aligned} &\left(\mathbb{I} + \xi_m\mathbb{M}^{-1}\mathbb{K}\right)\boldsymbol{p}_m \\ &\quad = \boldsymbol{p}_{m-1} + (\tau_m - C\xi_m)\,\boldsymbol{r}_{m-1} + \xi_m\mathbb{M}^{-1}\boldsymbol{G}_m \\ &\quad\quad + \left(C\,(\phi - 1)\,\xi_m\tau_m + \left(\frac{1}{2} - \delta\right)\tau_m^2\right)\left(\mathbb{M}^{-1}\boldsymbol{G}_{m-1} - \mathbb{M}^{-1}\mathbb{K}\boldsymbol{p}_{m-1} - C\boldsymbol{r}_{m-1}\right). \end{aligned} \tag{9.54}$$

If $\boldsymbol{p}_{m-1}$ and $\boldsymbol{r}_{m-1}$ are known, then $\boldsymbol{p}_m$ is obtained from system (9.54) and afterwards $\boldsymbol{r}_m$ is computed from (9.51).

Chapter 10

Realization of the coupled fluid-structure interaction problem

The aim of this chapter is the description of the complete coupled fluid-structure interaction problem. In the first section we shall be interested in the construction of the ALE mapping. It allows us a treatment of the time dependence of the domain Ω_t. The second section will present the coupling procedures. Namely, we shall describe the strong coupling and the weak coupling.

10.1 Construction of the ALE mapping

The aim of this section is the construction of the ALE mapping $\mathcal{A}_t$. The time-dependence of the domain is caused by the deformation of the common interface between the domain Ω_t occupied by the fluid and the elastic body:

$$\tilde{\Gamma}_{W_t} = \left\{ \boldsymbol{x} \in I\!R^2;\ \boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X}, t),\ \boldsymbol{X} \in \Gamma_W^b \right\}, \tag{10.1}$$

where $u(\boldsymbol{X}, t)$ is the displacement of the part Γ_W^b at time t.

The ALE mapping is constructed with the aid of an artificial stationary elasticity problem. We seek the displacement $\boldsymbol{d} = (d_1, d_2)$ defined in Ω_0 as a solution of the elastostatic system

$$\sum_{j=1}^{2} \frac{\partial \tau_{ij}^a}{\partial x_j} = 0 \quad \text{in } \Omega_0, \quad i = 1, 2, \tag{10.2}$$

where τ_{ij}^a are the components of the artificial stress tensor,

$$\begin{aligned} \tau_{ij}^a &= \lambda^a \mathrm{div}\boldsymbol{d}\delta_{ij} + 2\mu^a e_{ij}^a, \quad i, j = 1, 2, \\ e_{ij}^a(\boldsymbol{d}) &= \frac{1}{2}\left(\frac{\partial d_i}{\partial x_j} + \frac{\partial d_j}{\partial x_i} \right), \quad i, j = 1, 2. \end{aligned} \tag{10.3}$$

The Lamé coefficients λ^a and μ^a are related to the artificial Young modulus E^a and to the artificial Poisson number σ^a as in (6.20). The boundary conditions for $\boldsymbol{d}$ are

prescribed by

$$\begin{aligned} \boldsymbol{d}|_{\Gamma_I \cup \Gamma_O} &= 0, && (10.4)\\ \boldsymbol{d}|_{\Gamma_{W_0} \setminus \Gamma_W^b} &= 0, && (10.5)\\ \boldsymbol{d}(\boldsymbol{x},t) &= \boldsymbol{u}(\boldsymbol{x},t),\ \boldsymbol{x} \in \Gamma_W^b. && (10.6) \end{aligned}$$

The solution of (10.2) gives us the ALE mapping of $\bar{\Omega}_0$ onto $\bar{\Omega}_t$ in the form

$$\mathcal{A}_t(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x},t), \quad \boldsymbol{x} \in \bar{\Omega}_0, \tag{10.7}$$

for each time t.

System (10.2) is discretized by the conforming piecewise linear finite elements on the mesh $\mathcal{T}_{h0}$ used for computing the flow field in the beginning of the computational process in the polygonal approximation Ω_{h0} of the domain Ω_0. We seek the approximate solution $\boldsymbol{d}_h$ of the artificial stationary elasticity problem (10.2) on Ω_{h0} with the discrete boundary conditions

$$\begin{aligned} \boldsymbol{d}_h|_{\Gamma_{Ih} \cup \Gamma_{Oh}} &= 0, && (10.8)\\ \boldsymbol{d}_h|_{\Gamma_{W_0h} \setminus \Gamma_{Wh}^b} &= 0, && (10.9)\\ \boldsymbol{d}_h(\boldsymbol{x},t) &= \boldsymbol{u}(\boldsymbol{x},t),\ \boldsymbol{x} \in \Gamma_{Wh}^b. && (10.10) \end{aligned}$$

We introduce the finite element spaces

$$\begin{aligned} \boldsymbol{D}_h &= \left\{\boldsymbol{d}_h = (d_{h1}, d_{h2}) \in C(\bar{\Omega}_{h0})^2; d_{hi}|_K \in P_1(K)\ \forall K \in \mathcal{T}_{h0},\ i = 1,2\right\},\\ \boldsymbol{H}_h &= \left\{\boldsymbol{\omega}_h \in \boldsymbol{D}_h;\ \boldsymbol{\omega}_h(Q) = 0 \text{ for all vertices } Q \in \partial\Omega_0\right\}, \end{aligned}$$

and the form

$$\begin{aligned} B_h^a(\boldsymbol{d}_h, \boldsymbol{\omega}_h) &= \int_{\Omega_{h0}} \sum_{i,j=1}^{2} \tau_{ij}^a(\boldsymbol{d}_h) \frac{\partial \omega_i}{\partial x_j} \mathrm{d}\boldsymbol{x} && (10.11)\\ &= \int_{\Omega_{h0}} \lambda^a \left(\frac{\partial d_{h1}}{\partial x_1} + \frac{\partial d_{h2}}{\partial x_2}\right)\left(\frac{\partial \omega_{h1}}{\partial x_1} + \frac{\partial \omega_{h2}}{\partial x_2}\right) \mathrm{d}\boldsymbol{x}\\ &\quad + 2\int_{\Omega_{h0}} \mu^a \left(\frac{\partial d_{h1}}{\partial x_1}\frac{\partial \omega_{h1}}{\partial x_1} + \frac{\partial d_{h2}}{\partial x_2}\frac{\partial \omega_{h2}}{\partial x_2}\right) \mathrm{d}\boldsymbol{x}\\ &\quad + \int_{\Omega_{h0}} \mu^a \left(\frac{\partial d_{h1}}{\partial x_2} + \frac{\partial d_{h2}}{\partial x_1}\right)\left(\frac{\partial \omega_{h1}}{\partial x_2} + \frac{\partial \omega_{h2}}{\partial x_1}\right) \mathrm{d}\boldsymbol{x}, && (10.12) \end{aligned}$$

which was obtained from the left-hand side of (10.2) by multiplying by any test function $\boldsymbol{\omega}_h \in \boldsymbol{H}_h$, integrating over Ω_{h0} and use of Green's theorem. Then the approximate solution of problem (10.2) with the boundary conditions (10.4)-(10.6) is defined as a function $\boldsymbol{d}_h \in \boldsymbol{D}_h$ satisfying the Dirichlet boundary conditions (10.8)-(10.10) and the identity

$$B_h^a(\boldsymbol{d}_h, \boldsymbol{\omega}_h) = 0 \quad \forall \boldsymbol{\omega}_h \in \boldsymbol{H}_h. \tag{10.13}$$

The use of linear finite elements is sufficient, because we need only to know the movement of the points of the mesh.

If the displacement $\boldsymbol{d}_h$ is computed at time t_m, then in view of (10.7), the approximation of the ALE mapping is obtained in the form

$$\mathcal{A}_{t_m h}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{d}_h(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega_{h0}. \tag{10.14}$$

The knowledge of the ALE mapping at the time instants $t_m,\ t_{m-1}, t_{m-2}, \ldots$ allows us to approximate the domain velocity with the aid of the backward difference formula of a general order q at time t_m derived in the same way as in (8.63). We get

$$\boldsymbol{z}_{h,m}(\boldsymbol{x}) = \chi_0 \boldsymbol{x} + \sum_{l=1}^{q} \chi_l \mathcal{A}_{t_{m-l}}(\mathcal{A}_{t_m}^{-1}(\boldsymbol{x})), \quad \boldsymbol{x} \in \Omega_{ht_m}, \tag{10.15}$$

where the coefficients $\chi_l,\ l = 0, \ldots, q$, depend on $\tau_{m-l},\ l = 0, \ldots, q-1$. Let us mention that in our computation we use the first order formula:

$$\boldsymbol{z}_{h,m}(\boldsymbol{x}) = \frac{\boldsymbol{x} - \mathcal{A}_{t_{m-1}}(\mathcal{A}_{t_m}^{-1}(\boldsymbol{x}))}{\tau_m}, \quad \boldsymbol{x} \in \Omega_{ht_m}. \tag{10.16}$$

10.2 Coupling procedure

In the solution of the complete coupled fluid-structure interaction problem it is necessary to apply a suitable coupling procedure. The general framework can be found e.g. [9]. Here we apply two different procedures.

First, we present the *weak coupling* algorithm:

1. Compute the approximate solution of the flow problem (5.57) on the time level t_m.
2. Compute the stress tensor of the fluid τ_{ij}^f and the aerodynamical force acting on the structure and transform it to the interface Γ_{Wh}^b by (7.2).
3. Solve the elasticity problem (9.1), compute the deformation $\boldsymbol{u}_{h,m}$ at time t_m and approximate the domain $\Omega_{ht_{m+1}}$.
4. Determine the ALE mapping $\mathcal{A}_{t_{m+1}h}$ by (10.7) and approximate the domain velocity $\boldsymbol{z}_{h,m+1}$ by (10.15).
5. Set $m := m + 1$, go to 1).

The *strong coupling* procedure represents a more complicated coupling algorithm. It follows this outline:

1. Assume that the approximate solution $\boldsymbol{w}_h^m$ of the flow problem and the deformation $\boldsymbol{u}_{h,m}$ of the structure are known on the time level t_m.
2. Set $\boldsymbol{u}_{h,m+1}^0 := \boldsymbol{u}_{h,m},\ k := 1$ and apply the iterative process:
   (a) Compute the stress tensor of the fluid τ_{ij}^f and the aerodynamical force acting on the structure and transform it to the interface Γ_{Wh}^b.

(b) Solve the elasticity problem, compute the approximation of the deformation $\boldsymbol{u}_{h,m+1}^k$ and construct the approximation $\Omega_{ht_{m+1}}^k$ of the flow domain at time t_{m+1}.

(c) Determine the approximations of ALE mapping $\mathcal{A}_{t_{m+1}h}^k$ and the domain velocity $\boldsymbol{z}_{h,m+1}^k$.

(d) Solve the flow problem in $\Omega_{ht_{m+1}}^k$ and obtain the approximate solution $\boldsymbol{w}_{h,m+1}^k$.

(e) If the variation of the displacement $\boldsymbol{u}_{h,m+1}^k$ and $\boldsymbol{u}_{h,m+1}^{k-1}$ is larger than the prescribed tolerance and $k < 50$, go to a) and $k := k + 1$. Else $\Omega_{ht_{m+1}} := \Omega_{ht_m}^k$, $\boldsymbol{w}_h^{m+1} := \boldsymbol{w}_{h,m+1}^k$, $\boldsymbol{u}_h^{m+1} := \boldsymbol{u}_{h,m}^k$, $m := m + 1$ and goto 2).

The difference between these two coupling algorithms will be presented on our numerical results in Chapter 13.

Chapter 11

Algorithmization

11.1 Algorithmization of the flow problem

For the algorithmization of problem (8.77) and finding the function $\boldsymbol{w}_h^m$ we need to transform system (8.77) to the system of linear algebraic equations. This subject is described in this section.

First, we start from the choice of the appropriate basis functions of the space S_{ht}. By $\hat{K}$ we denote the *reference element* with vertices $\hat{A} = (0,0),\ \hat{B} = (1,0),\ \hat{C} = (0,1)$. Further, we assume that the space S_{ht} is created by the polynomials of degree ≥ 1. For this reason we search the basis functions of the space $P_p(\hat{K})$ in the form

$$\hat{\psi}_j(\hat{x}_1, \hat{x}_2) = \sum_{l=0}^{p} \sum_{k=0}^{p-l} q_{kl}^j (\hat{x}_1)^k (\hat{x}_2)^l, \quad j = 1, \ldots, d_p, \quad q_{kl}^j \in I\!R, \tag{11.1}$$

where

$$d_p := \frac{(p+1)(p+2)}{2} \tag{11.2}$$

is the dimension of the space $P_p(\hat{K})$. In the following way we define the set $\hat{D}$ of the points of the element $\hat{K}$:

$$\hat{D} = \left\{ \left(\frac{k}{p}, \frac{l}{p} \right);\ k, l = 0, \ldots, p,\ k + l \leq p \right\}. \tag{11.3}$$

It is possible to show that $\mathrm{card}(\hat{D}) = d_p$. Using the notation $\hat{\boldsymbol{x}}_n,\ n = 1, \ldots, d_p$, for the elements of the set $\hat{D}$, there exists the basis $\hat{\psi}_1, \ldots, \hat{\psi}_{d_p}$ of the space $P_p(\hat{K})$ fulfilling the condition $\hat{\psi}_j(\hat{\boldsymbol{x}}_n) = \delta_{jn},\ j, n = 1, \ldots, d_p$.

For each element $K_t \in \mathcal{T}_{ht}$ let us define the space $P_p(K_t)$. The element K_t has the vertices $A^{K_t},\ B^{K_t},\ C^{K_t}$, for which it holds $\mathcal{A}_t(A^{K_0}) = A^{K_t} = (a_1^{K_t}, a_2^{K_t})$, $\mathcal{A}_t(B^{K_0}) = B^{K_t} = (b_1^{K_t}, b_2^{K_t})$, $\mathcal{A}_t(C^{K_0}) = C^{K_t} = (c_1^{K_t}, c_2^{K_t})$. Further, we consider the linear mapping $F^{K_t} : \hat{K} \to K_t$ with the properties $F^{K_t}(\hat{A}) = A^{K_t},\ F^{K_t}(\hat{B}) = B^{K_t},\ F^{K_t}(\hat{C}) = C^{K_t}$. This is one-to-one mapping and we can write $\boldsymbol{x} = F^{K_t}(\hat{\boldsymbol{x}}) = \mathbb{U}^{K_t}\hat{\boldsymbol{x}} + \boldsymbol{V}^{K_t}$. The

matrices $\mathbb{U}^{K_t}$ and $\boldsymbol{V}^{K_t}$ have the forms

$$\mathbb{U}^{K_t} = \begin{pmatrix} b_1^{K_t} - a_1^{K_t} & c_1^{K_t} - a_1^{K_t} \\ b_2^{K_t} - a_2^{K_t} & c_2^{K_t} - a_2^{K_t} \end{pmatrix}, \tag{11.4}$$

$$\boldsymbol{V}^{K_t} = \begin{pmatrix} a_1^{K_t} \\ a_2^{K_t} \end{pmatrix} \tag{11.5}$$

The inverse mapping can be expressed as $\hat{\boldsymbol{x}} = (F^{K_t})^{-1}(\boldsymbol{x}) = (\mathbb{U}^{K_t})^{-1}(\boldsymbol{x} - \boldsymbol{V}^{K_t})$, where the inverse matrix $(\mathbb{U}^{K_t})^{-1}$ has the form

$$(\mathbb{U}^{K_t})^{-1} = \frac{1}{\det(\mathbb{U}^{K_t})} \begin{pmatrix} c_2^{K_t} - a_2^{K_t} & a_1^{K_t} - c_1^{K_t} \\ a_2^{K_t} - b_2^{K_t} & b_1^{K_t} - a_1^{K_t} \end{pmatrix}. \tag{11.6}$$

Here, we denote by $\det(\mathbb{U}^{K_t})$ the determinant of the matrix $\mathbb{U}^{K_t}$. We define the points $\boldsymbol{x}_n^{K_t} := F^{K_t}(\hat{\boldsymbol{x}}_n)$ $K_t \in \mathcal{T}_{ht}$, and we seek the basis functions $\psi_1^{K_t}, \ldots, \psi_{d_p}^{K_t}$ on the element $K_t \in \mathcal{T}_{ht}$ fulfilling the conditions $\psi_j^{K_t}(\boldsymbol{x}_n^{K_t}) = \delta_{jn},\ j, n = 1, \ldots, d_p$. These basis functions are defined unambiguously. Because of the linearity of the function $(F^{K_t})^{-1}$, $\hat{\psi}_j((F^{K_t})^{-1}(\boldsymbol{x}))$ is the polynomial of degree $\leq p$. It holds that $\hat{\psi}_j((F^{K_t})^{-1}(\boldsymbol{x}_n^{K_t})) = \hat{\psi}_j(\hat{\boldsymbol{x}}_n) = \delta_{jn} = \psi_j^{K_t}(\boldsymbol{x}_n^{K_t}),\ j, n = 1, \ldots, d_p$. It follows from the unambiguity that $\psi_j^{K_t}(\boldsymbol{x}) = \hat{\psi}_j((F^{K_t})^{-1}(\boldsymbol{x})),\ j = 1, \ldots, d_p$.

Using the chain rule we derive the derivative of the basis functions $\psi_j^{K_t},\ j = 1, \ldots, d_p$:

$$\begin{aligned} \frac{\partial \psi_j^{K_t}}{\partial x_1}(\boldsymbol{x}) &= \frac{\partial}{\partial x_1}\hat{\psi}_j((F^{K_t})^{-1}(\boldsymbol{x})) = \sum_{i=1}^{2} \frac{\partial \hat{\psi}_j}{\partial \hat{x}_i}((F^{K_t})^{-1}(\boldsymbol{x})) \frac{\partial ((F^{K_t})^{-1})_i}{\partial x_1}(\boldsymbol{x}) \\ &= (\nabla \hat{\psi}_j)((F^{K_t})^{-1}(\boldsymbol{x})) \cdot \left((\mathbb{U}^{K_t})^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right). \end{aligned} \tag{11.7}$$

For $\hat{\boldsymbol{x}}$ fulfilling $\boldsymbol{x} = F^{K_t}(\hat{\boldsymbol{x}})$ it holds

$$\frac{\partial \psi_j^{K_t}}{\partial x_1}(\boldsymbol{x}) = \frac{\partial \psi_j^{K_t}}{\partial x_1}(F^{K_t}(\hat{\boldsymbol{x}})) = (\nabla \hat{\psi}_j)(\hat{\boldsymbol{x}}) \cdot \left((\mathbb{U}^{K_t})^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right). \tag{11.8}$$

In a similar way we can show that

$$\frac{\partial \psi_j^{K_t}}{\partial x_2}(\boldsymbol{x}) = (\nabla \hat{\psi}_j)(\hat{\boldsymbol{x}}) \cdot \left((\mathbb{U}^{K_t})^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right). \tag{11.9}$$

If we have defined the space $P_p(K_t)$ for all elements $K_t \in \mathcal{T}_{ht}$, we can define the vector-valued basis functions $\boldsymbol{\Psi}_{j,d}^{K_t} = (\psi_j^{K_t}\delta_{1d}, \ldots, \psi_j^{K_t}\delta_{4d}),\ d = 1, \ldots, 4,\ j = 1, \ldots, d_p$. These functions form the basis of the space $\boldsymbol{S}_{ht}$. From this it follows that the number of degrees of freedom dof_h of this space is $\mathrm{dof}_h = 4d_p\mathrm{card}(\mathcal{T}_{ht})$. For the sake of simplicity we introduce the set I of the indices of elements of the triangulation $\mathcal{T}_{ht}$ such that $\mathcal{T}_{ht} = \{K_i;\ i \in I = \{0, 1, \ldots, c;\ c \in I\!N\}\}$. Then we shall number the basis functions of the space $\boldsymbol{S}_{ht}$ by the index $l = i4d_p + (d-1)d_p + j$ for $i \in I,\ d \in \{1, \ldots, 4\}, j \in$

$\{1,\dots,d_p\}$. Then $\boldsymbol{\Psi}_h^{l,m}(\boldsymbol{x}) := \boldsymbol{\Psi}_{j,d}^{K_{t_m}}$ and the sought solution can be written in the form

$$\boldsymbol{w}_h^m = \sum_{l=1}^{\mathrm{dof}_h} \zeta_l^m \boldsymbol{\Psi}_h^{l,m}. \tag{11.10}$$

In view of the ideas presented above we can write the discrete problem (8.77) in the form

$$\begin{aligned}
&\sum_{l=1}^{\mathrm{dof}_h} \zeta_l^m \chi_0\left(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}\right) + \sum_{l=1}^{\mathrm{dof}_h} \zeta_l^m \Big(d_h(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) \\
&\quad + a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + J_h(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + \beta_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) \qquad (11.11) \\
&\quad + \hat{J}_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m})\Big) = l_h(\bar{\boldsymbol{w}}_h^m(t), \boldsymbol{\Psi}_h^{r,m}) - \left(\sum_{l=1}^{q} \chi_l \boldsymbol{w}_h^{m-l}, \boldsymbol{\Psi}_h^{r,m}\right), \\
&\quad r = 1,\dots,\mathrm{dof}_h. \qquad (11.12)
\end{aligned}$$

Finally, we rewrite system (11.11) in the form

$$\mathbb{A}_h^m \boldsymbol{\zeta}_h^m = \boldsymbol{L}_h^m, \tag{11.13}$$

where $\mathbb{A}_h^m$ is the matrix with elements

$$\begin{aligned}
\{\mathbb{A}_h^m\}_{rl} \;=\; & \chi_0(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + d_h(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) \\
& + a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + J_h(\boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) + \beta_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}) \\
& + \hat{J}_h(\hat{\boldsymbol{w}}_h^{m-1}, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m}), \qquad (11.14)
\end{aligned}$$

and $\boldsymbol{L}_h^m$ is the vector with components

$$\{\boldsymbol{L}_h^m\}_r = l_h(\bar{\boldsymbol{w}}_h^m(t), \boldsymbol{\Psi}_h^{r,m}) - \left(\sum_{l=1}^{q} \chi_l \boldsymbol{w}_h^{m-l}, \boldsymbol{\Psi}_h^{r,m}\right). \tag{11.15}$$

The vector

$$\boldsymbol{\zeta}_h^m = (\zeta_1^m, \dots, \zeta_{\mathrm{dof}_h}^m)^T \tag{11.16}$$

defines the approximate solution by (11.10).

11.1.1 Numerical integration

In this section we shall be concerned with the computation of the integrals determining the elements of the matrices of system (11.13). In most cases we are not able to compute these integrals exactly. For this reason we shall use 1D and 2D quadrature formulae.

In the case of computations of 1D integrals we use 1D Gaussian quadrature formulae of higher order of accuracy derived for the interval $[0,1]$. We consider formulae that are accurate for polynomials with degree ≤ 5. Let us have a function $e(\vartheta)$ defined on the interval $[0,1]$. Then the three point rule reads

$$\int_0^1 e(\vartheta)\mathrm{d}\vartheta \approx \sum_{j=1}^{3} \varpi_j e(\vartheta_j), \tag{11.17}$$

j	ϑ_j	ϖ_j
1.	$(1-\sqrt{3/5})/2$	5/18
2.	0.5	4/9
3.	$(1+\sqrt{3/5})/2$	5/18

Table 11.1: Gauss three point rule on the interval $[0,1]$.

where ϖ_j and ϑ_j are given in Table 11.1.

Before starting the computation of 1D integrals we need to use a parameterization of all edges of the elements $K \in \mathcal{T}_{ht}$. The line segments $\hat{A}\hat{B},\ \hat{B}\hat{C},\ \hat{C}\hat{A}$ of the reference element K can be written as $\hat{\boldsymbol{\eta}}_1(\vartheta) = (\vartheta, 0),\ \hat{\boldsymbol{\eta}}_2(\vartheta) = (1-\vartheta, \vartheta,),\ \hat{\boldsymbol{\eta}}_3(\vartheta) = (0, 1-\vartheta)$ for $\vartheta \in [0,1]$. Using these functions and the function F^{K_t}, the line segments $A^{K_t}B^{K_t},\ B^{K_t}C^{K_t},\ C^{K_t}A^{K_t}$ can be expressed in the form

$$\boldsymbol{\eta}_s^{K_t}(\vartheta) = F^{K_t}(\hat{\boldsymbol{\eta}}_s(\vartheta)), \quad \vartheta \in [0,1] \text{ for } s = 1,2,3. \tag{11.18}$$

For the computation of the integrals over the line segment parameterized by $\hat{\boldsymbol{\eta}}_s$ we define the function $s_s^t(\vartheta)$:

$$s_s^t(\vartheta) := \sqrt{\left(\frac{\partial(\boldsymbol{\eta}_s^{K_t}(\vartheta))_1}{\partial\vartheta}\right)^2 + \left(\frac{\partial(\boldsymbol{\eta}_s^{K_t}(\vartheta))_2}{\partial\vartheta}\right)^2}. \tag{11.19}$$

This function represents the length of the line segment parameterized by $\hat{\boldsymbol{\eta}}_s$ and we can write $\mathrm{d}S = s_s^t(\vartheta)\mathrm{d}\vartheta$. In view of these consequences we can show an example of the computation of the integral over the edge $\Gamma \in \mathcal{F}_{ht}$ with the parametric expression $\boldsymbol{\eta}_s^{K_t}(\vartheta) = F^{K_t}(\hat{\boldsymbol{\eta}}_s(\vartheta))$. This type of integrals we can find in system (11.13), especially in form $a_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m})$. Let us define $L(\boldsymbol{x})$ on $\bar{\Omega}_{ht}$ then using (11.7), (11.19) and (11.17) we obtain

$$\begin{aligned}
&\int_\Gamma L(\boldsymbol{x})\psi_k^{K_t}(\boldsymbol{x})\frac{\partial\psi_l^{K_t}}{\partial x_1}(\boldsymbol{x})\mathrm{d}S \qquad (11.20)\\
&= \int_0^1 L(\boldsymbol{\eta}_s^{K_t}(\vartheta))\psi_k^{K_t}(\boldsymbol{\eta}_s^{K_t}(\vartheta))\frac{\partial\psi_l^{K_t}}{\partial x_1}(\boldsymbol{\eta}_s^{K_t}(\vartheta))s_s^t(\vartheta)\mathrm{d}\vartheta\\
&= \int_0^1 L(\boldsymbol{\eta}_s^{K_t}(\vartheta))\hat{\psi}_k(\hat{\boldsymbol{\eta}}_s(\vartheta))(\nabla\hat{\psi}_l)(\hat{\boldsymbol{\eta}}_s(\vartheta))\cdot\left((\mathbb{U}^{K_t})^{-1}\begin{pmatrix}1\\0\end{pmatrix}\right)s_s^t(\vartheta)\mathrm{d}\vartheta\\
&\approx \sum_{j=1}^{3}\varpi_j L(\boldsymbol{\eta}_s^{K_t}(\vartheta_j))\hat{\psi}_k(\hat{\boldsymbol{\eta}}_s(\vartheta_j))(\nabla\hat{\psi}_l)(\hat{\boldsymbol{\eta}}_s(\vartheta_j))\cdot\left((\mathbb{U}^{K_t})^{-1}\begin{pmatrix}1\\0\end{pmatrix}\right)s_s^t(\vartheta_j).
\end{aligned}$$

For the computation of 2D integrals we use 2D Gaussian quadrature formulae of higher order of accuracy derived for the reference element $\hat{K}$. We use the formulae that are accurate for polynomials with degree ≤ 5. Let us have a function f defined on the reference element $\hat{K}$. Then the seven point rule reads

$$\int_{\hat{K}} f(\hat{\boldsymbol{x}})\mathrm{d}\hat{\boldsymbol{x}} \approx \sum_{j=1}^{7}\varsigma_j f(\hat{\boldsymbol{x}}_j), \tag{11.21}$$

j	$\hat{x}_j^{(1)}$-coordinate	$\hat{x}_j^{(2)}$-coordinate	ς_j
1.	0.333333333333333	0.333333333333333	0.225
2.	0.470142064105115	0.470142064105115	0.132394152788506
3.	0.470142064105115	0.05971587178977	0.132394152788506
4.	0.05971587178977	0.470142064105115	0.132394152788506
5.	0.101286507323456	0.101286507323456	0.125939180544827
6.	0.101286507323456	0.797426985353087	0.125939180544827
7.	0.797426985353087	0.101286507323456	0.125939180544827

Table 11.2: Gauss seven point rule on the reference triangle $\hat{K}$.

where ς_j and $\hat{\boldsymbol{x}}_j = (\hat{x}_j^1, \hat{x}_j^1)$ are given in Table 11.2. As an example we compute the integral of the function from (11.20) over an element $K \in \mathcal{T}_{ht}$. Again this type of integrals can be found in system (11.13), especially in form $b_h(\bar{\boldsymbol{w}}_h^m, \boldsymbol{\Psi}_h^{l,m}, \boldsymbol{\Psi}_h^{r,m})$. Applying the substitution theorem, (11.7) and (11.21) we get

$$\begin{aligned}
&\int_K L(\boldsymbol{x})\psi_k^{K_t}(\boldsymbol{x})\frac{\partial\psi_l^{K_t}}{\partial x_1}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \qquad (11.22)\\
&= \int_{\hat{K}} L(F^{K_t}(\hat{\boldsymbol{x}}))\psi_k^{K_t}(F^{K_t}(\hat{\boldsymbol{x}}))\frac{\partial\psi_l^{K_t}}{\partial x_1}(F^{K_t}(\hat{\boldsymbol{x}}))\mathrm{det}\mathbb{U}^{K_t}\mathrm{d}\hat{\boldsymbol{x}}\\
&= \int_{\hat{K}} L(F^{K_t}(\hat{\boldsymbol{x}}))\hat{\psi}_k(\hat{\boldsymbol{x}})(\nabla\hat{\psi}_l)(\hat{\boldsymbol{x}})\cdot\left((\mathbb{U}^{K_t})^{-1}\begin{pmatrix}1\\0\end{pmatrix}\right)\mathrm{det}\mathbb{U}^{K_t}\mathrm{d}\hat{\boldsymbol{x}}\\
&\approx \sum_{j=1}^{7}\varsigma_j L(F^{K_t}(\hat{\boldsymbol{x}}_j))\hat{\psi}_k(\hat{\boldsymbol{x}}_j)(\nabla\hat{\psi}_l)(\hat{\boldsymbol{x}}_j)\cdot\left((\mathbb{U}^{K_t})^{-1}\begin{pmatrix}1\\0\end{pmatrix}\right)\mathrm{det}\mathbb{U}^{K_t}.
\end{aligned}$$

11.2 Algorithmization of the structural problem

In Section 9.1 we have defined the matrices $\mathbb{M}$, $\mathbb{K}$ and the vector $\boldsymbol{G}(t)$. In order to express their elements, we use the numerical integration. The integral over any element $K \in \mathcal{T}_h^b$ will be approximated in the following way

$$\int_K \chi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{3}|K|\sum_{i=1}^{3}\chi(P_i^K), \qquad (11.23)$$

where $\chi(\boldsymbol{x})$ is a function defined on the element $K \in \mathcal{T}_h^b$, P_i^K are the vertices of the triangle $K \in \mathcal{T}_h^b$ and $|K|$ is the surface of the triangle $K \in \mathcal{T}_h^b$. The integral over any edge $\Gamma \in \bar{\Gamma}_{Wh},\ \Gamma \subset K,\ K \in \mathcal{T}_h^b$, will be approximated by

$$\int_\Gamma \chi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \approx \frac{1}{2}|\Gamma|\,(\chi(P_1^\Gamma) + \chi(P_2^\Gamma)), \qquad (11.24)$$

where $P_1^\Gamma,\ P_2^\Gamma$ are the end points of Γ and $|\Gamma|$ denotes the length of the edge Γ.

Using definitions (9.23), (9.24) of the elements $m_{ij}, i, j = 1, \ldots, N$ of the matrix $\mathbb{M}$, we obtain

$$\begin{aligned} m_{ij} &= (\rho^b \varphi_i, \varphi_j)_{\Omega_h^b} = \sum_{K \in \mathcal{T}_{ht}} \int_K \rho^b \varphi_i \varphi_j \mathrm{d}\boldsymbol{x} \approx \frac{1}{3} \sum_{K \in \mathcal{T}_{ht}} |K| \sum_{k=1}^{3} \rho^b(P_k^K) \varphi_i(P_k^K) \varphi_j(P_k^K), \\ & \quad i, j = 1, \ldots, n. \end{aligned} \tag{11.25}$$

From the definition of the basis functions $\varphi_j,\ j = 1, \ldots, n$ in (9.16) we see that the matrix $\mathbb{M}$ is diagonal and its values on the diagonal are determined by the density ρ^b in the given vertex of the triangulation and by the size of a support of a basis function.

Let us remind that in most of cases we can assume that the density ρ^b is constant. Thus,

$$m_{ij} = \delta_{ij} m_i, \quad i, j = 1, \ldots, n, \tag{11.26}$$

where

$$m_i = \frac{1}{3} \rho^b \left| \mathrm{supp}(\varphi_i) \right| = \frac{1}{3} \rho^b \sum_{K \in \mathcal{T}_h^b;\ P_i \in K} |K|, \quad i = 1, \ldots, n. \tag{11.27}$$

Since every function $\varphi \in X_h$ is linear on each element $K \in \mathcal{T}_h^b$, the derivatives $\frac{\partial \varphi}{\partial x_i}|_K,\ i = 1, 2$, are constant. Then we use the notation

$$\frac{\partial \varphi}{\partial x_i}|_K = \varphi_K^{(i)}, \quad K \in \mathcal{T}_h^b, \tag{11.28}$$

and the elements $k_{ij},\ i, j = 1, \ldots, N$, of the matrix $\mathbb{K}$ can be written in the form

$$\begin{aligned} k_{ij} &= \int_{\Omega_h^b} \left((\lambda^b + 2\mu^b) \frac{\partial \varphi_i}{\partial x_1} \frac{\partial \varphi_j}{\partial x_1} + \mu^b \frac{\partial \varphi_i}{\partial x_2} \frac{\partial \varphi_j}{\partial x_2} \right) \mathrm{d}\boldsymbol{x} \qquad (11.29) \\ &= \sum_{K \in \mathcal{T}_h^b} \int_K (\lambda^b + 2\mu^b) \varphi_{iK}^{(1)} \varphi_{jK}^{(1)} + \mu^b \varphi_{iK}^{(2)} \varphi_{jK}^{(2)} \mathrm{d}\boldsymbol{x} \\ &= \sum_{K \in \mathcal{T}_h^b} |K| \left((\lambda^b + 2\mu^b) \varphi_{iK}^{(1)} \varphi_{jK}^{(1)} + \mu^b \varphi_{iK}^{(2)} \varphi_{jK}^{(2)} \right) \\ &= \sum_{K \in \mathcal{T}_h^b;\ P_i, P_j \in K} |K| \left((\lambda^b + 2\mu^b) \varphi_{iK}^{(1)} \varphi_{jK}^{(1)} + \mu^b \varphi_{iK}^{(2)} \varphi_{jK}^{(2)} \right), \\ k_{(i+n)j} &= \sum_{K \in \mathcal{T}_h^b;\ P_i, P_j \in K} |K| \left(\lambda^b \varphi_{iK}^{(2)} \varphi_{jK}^{(1)} + \mu^b \varphi_{iK}^{(1)} \varphi_{jK}^{(2)} \right), \qquad (11.30) \\ k_{i(j+n)} &= \sum_{K \in \mathcal{T}_h^b;\ P_i, P_j \in K} |K| \left(\lambda^b \varphi_{iK}^{(1)} \varphi_{jK}^{(2)} + \mu^b \varphi_{iK}^{(2)} \varphi_{jK}^{(1)} \right), \qquad (11.31) \\ k_{(i+n)(j+n)} &= \sum_{K \in \mathcal{T}_h^b;\ P_i, P_j \in K} |K| \left((\lambda^b + 2\mu^b) \varphi_{iK}^{(2)} \varphi_{jK}^{(2)} + \mu^b \varphi_{iK}^{(1)} \varphi_{jK}^{(1)} \right), \qquad (11.32) \\ & \quad i, j = 1, \ldots, n. \end{aligned}$$

The use of the numerical integration for the right-hand side vector $\boldsymbol{G}$ leads to

$$g_i(t) = \frac{1}{2} \sum_{\Gamma \subset \Gamma^b_{Wh}} |\Gamma| \, T^{\boldsymbol{n}}_{h1}(t, P_i), \tag{11.33}$$

$$g_{i+n}(t) = \frac{1}{2} \sum_{\Gamma \subset \Gamma^b_{Wh}} |\Gamma| \, T^{\boldsymbol{n}}_{h2}(t, P_i), \tag{11.34}$$

$$i = 1, \ldots, n.$$

Chapter 12

Implementation

In this chapter we shall describe the most important part of the program used for our numerical experiments. In our simulations we use the C program created by Václav Kučera that was originally designed for the flow in the time-independent domains. Tests of this method and program are presented in [48]. In this program we modified some parts in order to allow computations in time-dependent domains using the ALE method. The results of these modifications allowing the treatment of the motion of the domain, which is prescribed as a graph of a function, are presented in [56]. As the next step we completed this program by the C module for computations of the dynamic elasticity problem and the elastostatic problem. This module was originally developed by Adam Kosík for the computations of purely elasticity problems or the problem of the interaction of an elastic body with an incompressible flow as described in [47]. Here, the tests of the code were also carried out. The modification of the both parts of the program, the flow part and the elastic part, was necessary for the cooperation of these two different modules. The program part which treats the interaction between the flow part and the elastic part, was newly developed as well as the part containing the ALE method. Now, we have two variants of the program. Each of these two variants allows the different coupling procedure as described in Section 10.2. It means that we are able to solve the problems of the flow in time-independent domains as well as the fluid-structure interaction of the flow and an elastic structure. In the following sections of this chapter we shall pay attention to the mesh generation and some details of the program and the used technique.

Let us shortly remind that the algorithmization technique was in detail described in Chapters 8, 9 and 11. Only the small modification of the final Newmark scheme (9.54) and (9.51) will be presented in order to get a favourable program implementation. The coupling ideas and the computation of the ALE mapping are described in Chapter 10.

If we solve the flow part of the problem, we need to go from the one time level to the next one. It means that on each time level we create system (11.13). The matrix $\mathbb{A}_h^m$ of this system is nonsymmetric and sparse on each time level. For this reason we use for the solution of system (11.13) the *Generalized Minimal Residual* (GMRES) method. The description of this method can be found in e.g. [56], Section 6.3. For improving of the properties of the method we apply the block diagonal preconditioning, where the block is created by all variables of one element of the triangulation.

12.1 Mesh generation

The finite element method as well as the discontinuous Galerkin finite element method are based on the construction of a triangulation in the computational domain. On this triangulation we compute the approximate solution of the problem, where the number of the elements of the triangulation and their size play the role in the accuracy of the solution. On the other hand, the growth of the number of elements causes computations more complicated and more time-consuming. Moreover, in the case of fluid-structure interaction problem we need to connect the triangulation $\mathcal{T}_h^b$ of the domain Ω_h^b of the structure to the triangulation $\mathcal{T}_{h0}$ of the domain Ω_{h0} occupied by the fluid. It has to be done in the suitable way, because these two domains have the common boundary Γ_W^b.

For simplified computational examples as in Section 13.1 the use of tensor-product meshes would be convenient. For the reason of solving mainly more complicated problems with a broken computational domain we apply more general meshes. As an example see Section 13.2.

For the generation of meshes used in our computations we apply the open source software GMSH [37]. This program gives us the possibility to easily guarantee that the vertices of the triangulations $\mathcal{T}_h^b$ lying on the common boundary Γ_W^b are the same as the vertices of the triangulations $\mathcal{T}_{h0}$ lying on the common boundary Γ_W^b. It allows us an easy handling of the information between the fluid and the structure. The further advantage of the GMSH software is the possibility of an easy creation of subdomains of the computational domain, which allows us to define, for example, different material properties in different subdomains of the structure. Also the refinement of the mesh in the desired parts of the domain is possible without bigger complications. These possibilities make the GMSH software more advantageous in comparison with the ANGENER software [1].

On the other hand, the ANGENER software makes a convenient output format of the file containing the triangulation. For this reason we use this triangulation file format as a standard input format of the triangulation in our program. We employ the properties of this format in our program. The ANGENER format of the triangulation data guarantees the regularity of the triangulation, the positive orientation of the vertices of elements of the triangulation, the arranging of vertices on the boundary, etc.

From the above mentioned reasons we combine the advantages of both softwares. We create a mesh of the computational domain by the GMSH software and then the triangulation format is transformed to the ANGENER format using as the input format of the triangulation in our program.

12.2 Implementation of the Newmark scheme

From the discretization of the structural problem we obtain system (9.33) of ordinary differential equations with the mass matrix $\mathbb{M}$, the stiffness matrix $\mathbb{K}$ and the right-hand side vector $\boldsymbol{G}$. Since we assume in our computations the the density ρ^b and coefficients characterizing material properties are constant with respect to time, it is sufficient to compute the elements of the matrices $\mathbb{M}$ and $\mathbb{K}$ only on the first time

level. On the other hand, the right-hand side vector $\boldsymbol{G}$ has to be computed again on each time level from relations (11.33) and (11.34). Applying relations (11.26) and (11.27), we obtain the elements of the mass matrix $\mathbb{M}$ and by relations (11.29)-(11.32) we get the elements of the stiffness matrix $\mathbb{K}$. Both these matrices are sparse. System (9.33) is solved with the aid of the Newmark method.

The implementation of the Newmark method is realized by relations (9.54) and (9.51), where we multiply equation (9.54) by the matrix $\mathbb{M}$. Then for the nonuniform partition $0 = t_0 < t_1 < \ldots < t_M = T$ of the time interval $[0, T]$ with $\tau_m = t_m - t_{m-1},\ m = 1, \ldots, M$, we get system (9.54) in the following form

$$\begin{aligned}(\mathbb{M} + \xi_m \mathbb{K})\boldsymbol{p}_m &= \mathbb{M}\boldsymbol{p}_{m-1} + (\tau_m - C\xi_m)\mathbb{M}\boldsymbol{r}_{m-1} + \xi_m \boldsymbol{G}_m + (C(\phi - 1)\xi_m \tau_m \\ &\quad + \left(\frac{1}{2} - \delta\right)\tau_m^2\Big)(\boldsymbol{G}_{m-1} - \mathbb{K}\boldsymbol{p}_{m-1} - C\mathbb{M}\boldsymbol{r}_{m-1}), \qquad (12.1)\end{aligned}$$

where ξ_m is given by relation (9.53). Similarly we multiply equation (9.51) by $\mathbb{M}$:

$$\begin{aligned}\mathbb{M}\boldsymbol{r}_m &= \frac{1}{1 + C\phi\tau_m}\left(\mathbb{M}\boldsymbol{r}_{m-1} + \tau_m\left(\phi(\boldsymbol{G}_m - \mathbb{K}\boldsymbol{p}_m)\right.\right. \\ &\quad \left.\left. + (1 - \phi)(\boldsymbol{G}_{m-1} - \mathbb{K}\boldsymbol{p}_{m-1} - C\mathbb{M}\boldsymbol{r}_{m-1})\right)\right). \qquad (12.2)\end{aligned}$$

In our computations we use an identical time step $\tau_m = \tau$ for each $m \in I\!N$. It means that we can also set $\xi_m = \xi$. If we set $\tilde{\boldsymbol{r}}_m = \mathbb{M}\boldsymbol{r}_m$, equations (12.1) and (12.2) have the form

$$\begin{aligned}(\mathbb{M} + \xi\mathbb{K})\boldsymbol{p}_m &= \mathbb{M}\boldsymbol{p}_{m-1} + (\tau - C\xi)\tilde{\boldsymbol{r}}_{m-1} + \xi\boldsymbol{G}_m + (C(\phi - 1)\xi\tau \\ &\quad + \left(\frac{1}{2} - \delta\right)\tau^2\Big)(\boldsymbol{G}_{m-1} - \mathbb{K}\boldsymbol{p}_{m-1} - C\tilde{\boldsymbol{r}}_{m-1}), \qquad (12.3)\\ \tilde{\boldsymbol{r}}_m &= \frac{1}{1 + C\phi\tau}\left(\tilde{\boldsymbol{r}}_{m-1} + \tau\left(\phi(\boldsymbol{G}_m - \mathbb{K}\boldsymbol{p}_m)\right.\right. \\ &\quad \left.\left. + (1 - \phi)(\boldsymbol{G}_{m-1} - \mathbb{K}\boldsymbol{p}_{m-1} - C\tilde{\boldsymbol{r}}_{m-1})\right)\right). \qquad (12.4)\end{aligned}$$

From the reason of not doing repetitive computations, in memory we save some further information. For example, the left-hand side of (12.3) does not have to be computed on each time level and we define the matrix $\mathbb{B}$

$$\mathbb{B} = \mathbb{M} + \xi\mathbb{K} \qquad (12.5)$$

and two real coefficients

$$\eta_1^b = \frac{1}{1 + C\phi\tau}, \qquad (12.6)$$

$$\eta_2^b = C(\phi - 1)\xi\tau + \left(\frac{1}{2} - \delta\right)\tau^2. \qquad (12.7)$$

On each time level we need to compute vectors $\tilde{\boldsymbol{q}}_m^1, \tilde{\boldsymbol{q}}_m^2$ defined as

$$\tilde{\boldsymbol{q}}_m^1 = \boldsymbol{G}_m - \mathbb{K}\boldsymbol{p}_m, \qquad (12.8)$$

$$\tilde{\boldsymbol{q}}_m^2 = \tilde{\boldsymbol{q}}_m^1 - C\tilde{\boldsymbol{r}}_m. \qquad (12.9)$$

After these modifications equations (12.3) and (12.4) read

$$\mathbb{B}\boldsymbol{p}_m = \mathbb{M}\boldsymbol{p}_{m-1} + (\tau - C\xi)\tilde{\boldsymbol{r}}_{m-1} + \xi\boldsymbol{G}_m + \eta_2^b\tilde{\boldsymbol{q}}_{m-1}^2 \quad (12.10)$$

$$\tilde{\boldsymbol{r}}_m = \eta_1^b(\tilde{\boldsymbol{r}}_{m-1} + \tau(\phi\tilde{\boldsymbol{q}}_m^1 + (1-\phi)\tilde{\boldsymbol{q}}_{m-1}^2)). \quad (12.11)$$

The system of linear algebraic equations (12.10) is solved by the *method of conjugate gradients*, which use the fact that the matrix $\mathbb{B}$ is symmetric and positive definite. It follows from the properties of the matrix $\mathbb{K}$, which is also symmetric and positive definite.

12.3 Description of the program

In this section we shall be concerned with the structure of our program. Especially we shall pay attention to the file `main.c`, which represents the heart of the program. Let us remind that the flow part of the problem is solved with the aid of dimensionless variables. On the other hand, the structural part of the problem uses dimensional variables. Then data handling between these two parts of the program need to be transformed according the relations presented in Section 5.2.

12.3.1 `main.c`

The main function `int main()` starts with the declaration of the constants described in Table 12.1. The initial time step `tau` is given in the dimensionless form and is applied in the flow problem. For the application in the part of the program, where the structure problem is solve, we multiply `tau` by the constant `COEFT` representing the characteristic time (`COEFT`$= L^*/U^*$). The structures and variables used in both parts of the program are defined and the function `gettriang(vert,el,ed,tri1)` reads the information about the triangulation of the flow problem saved in the file `tri1`. As mentioned in Section 12.1, all meshes used in computations are saved in the ANGENER format. The initialization of the initial conditions for the flow problem is carried out by the functions `setinitialconditions(el,0)` and `setinitialconditions(el,1)` that determines the initial conditions as constant state, read from the data set `constants.h`. Usually we begin the computation of flow in the fixed domain Ω_{h0}. After some time interval we release the ALE mapping and solve the fluid-structure interaction problem. The function `loadstate(el,act,statex, &time,0)` allows us to start the computation of an interaction from the state saved in the data set `statex`.

Type	Notation	Use
integer	`stepsave`	Frequency of results saving.
integer	`iterations`	Total number of time levels.
double	`tau`	Initial time step.
char	`tri1[]`	Name of the data set where the triangulation is saved.

Table 12.1: Initial constants.

Notation	Variable
`0`	density
`VELOC`	velocity
`VELOCX`	x-th component of velocity
`VELOCY`	y-th component of velocity
`PRESS`	pressure
`ENTROPY`	entropy
`MACH`	Mach number

Table 12.2: Variables.

For the structure part of the problem we define the time step `taureal=tau*COEFT`. By the function `init_struct(taureal)` we initialized the structure part of the problem, when the triangulation is read as well as initial conditions, etc. For the artificial elastostatic problem used to finding of the ALE mapping we apply the analogous function `init_static_elasticity(next_moving_vertex, count_moving_vertex)`. The function `inter_edge_adresse(ed,n_inter_edge,interaction_edges)` plays an important role for the connection of the structural and flow problem. This function searches the vertices of $\mathcal{T}_h^b$ and $\mathcal{T}_{h0}$, which have the same coordinates and present the vertices of the common boundary Γ_{Wh}^b, where the interaction takes place. The addresses of these vertices are saved in the structure `interaction_edges`.

Now we shall be concerned with the heart of the main function `int main()`. It means that we describe the most important parts of the loop through all time levels that directs the whole process carried out by the program. First, we shall pay attention to this loop of the weak coupled variant of the program. The function `iteration(el,ed,&mat,x,b,act,tau,20,1E-10,&error,&gmiters,btemp,0)` computes the solution of the flow problem on the given time level. As follow from the transmission condition (7.2), we need to determine the stress tensor on the edges lying on the common boundary Γ_{Wh}^b, where the interaction takes place. This is done by the function `average_stress_tensor_interaction_edge(interaction_edges,n_inter_edge,act,stress)`. The elastic problem is solved by the function `step_struct(displacement, stress, taureal)`. The ALE mapping is computed by the function `step_static_elasticity(all_displacement, displacement)` and the triangulation on the next time level is created by the function `renewtriangulation_elastic(el, ed,vert,time,tau,all_displacement)`. The saving of the chosen flow variable is carried out by the function `savesolution(el, vert,i/stepsave,act,time,X)`, where `X` is the chosen variable. For the notation of the variables see Table 12.2. The structure variables are saved by the function `save_structure(i / stepsave)`.

In the case of the strong coupled variant of the program the loop is more complicated. Before we start this loop, we solve the flow problem on the first time level, which is realized by the function `iteration(el,ed,&mat,x,b,act,tau,20,1E-10, &error,&gmiters,btemp,0,0)`. Then we enter the loop over all time levels, where we start by computing the stress tensor for transmission condition (7.2) (`average_stress_tensor_interaction_edge(interaction_edges,n_inter_edge,act,stress)`), the structural problem (`step_struct(displacement, stress, taureal, 0)`), the ALE

mapping (`step_static_elasticity(all_displacement, displacement)`) and the new triangulation (`renewtriangulation_elastic(el,ed,vert,time,tau,all_displacement,1)`). Further, we stay on the same time level and run an inner loop described in Section 10.2. When this loop is successfully finished, the results are saved in the same way as in the weak coupled variant of the program.

12.3.2 `Constants.h`

In this data set the conditions of a computation and input variables of the flow problem are defined. In Table 12.3 we describe the most important constant for setting the computation. In Tables 12.4 - 12.8 the possible setting of these constants can be found.

Let us mention that the material properties of the elastic structure are saved in the data set `nm_cfg.h`. Constants for the artificial elastostatic problem used for determinding the ALE mapping can be found in the file `elastostatics.h`.

Notation	Use
`ELASTICITY`	Choice of dealing with a FSI problem.
`COUPLING`	Choice of the type of coupling in a FSI problem.
`NUMBER_COUPLING_LOAP`	Number of inner iterations in case of the strong coupling.
`ALE`	Setting for dealing with a time-dependent domain.

Table 12.3: Constants used for the adjustment of the type of the computation.

`ELASTICITY`	Meaning
0	Flow problem without an interaction.
1	Fluid-structure interaction problem.

Table 12.4: Adjustment of the constant `ELASTICITY`.

`COUPLING`	Meaning
0	Weak coupling.
1	Strong coupling.

Table 12.5: Adjustment of the constant `COUPLING`.

`ALE`	Type of problem
0	Problem on a time-independent domain.
11	Problem on a time-dependent domain.

Table 12.6: Adjustment of the constant `ALE`.

Notation	Initial condition
`RHO0`	Initial density.
`VX0`	Initial x-th component of velocity.
`VY0`	Initial y-th component of velocity.
`P0`	Initial pressure.

Table 12.7: Notation of initial conditions of the flow problem .

Notation	Boundary condition
`RHOIN`	Density on the inlet boundary.
`VXIN`	X-th component of velocity on the inlet boundary.
`VYIN`	Y-th component of velocity on the inlet boundary.
`POUT`	Pressure on the outlet boundary.

Table 12.8: Notation of boundary conditions of the flow problem.

Chapter 13

Numerical experiments

This chapter will be devoted to our numerical results. All presented results have been inspired by the problem of an airflow in human vocal folds. The simulation of vocal folds vibrations induced by compressible viscous flow represents a complicated problem of fluid-structure interaction.

First we shall be concentrated on the experimental analysis of computational accuracy, especially on the impact of a density of a computational mesh on the solution. For this reason we shall be interested in the solution of a problem in a simple computational domain. It allows us to compare the behaviour of the solution on meshes with a different number of elements and show a convergence tendency. All these comparisons will be carried out in context of weak and strong coupling procedures (see Section 10.2).

The second example will be focused on the demonstration of the applicability of the developed scheme, when the computational domain is more realistic and better characterizes the vocal tract. The properties of the elastic bodies better approximate material properties of human vocal folds.

Let us mention that all presented computations were carried out on the cluster Sněhurka at the Faculty of Mathematics and Physics, Charles University in Prague (see [2]).

13.1 Example 1

We consider the model of flow through a channel with two bumps which represent time dependent boundaries between the flow and a simplified model of vocal folds (see Figure 13.1). The numerical experiments were carried out for the following data: magnitude of the inlet velocity $v_{in} = 4$ m/s, the viscosity $\mu = 15 \cdot 10^{-6}\,\mathrm{kg\,m^{-1}\,s^{-1}}$, the inlet density $\rho_{in} = 1.225\,\mathrm{kg\,m^{-3}}$, the outlet pressure $p_{out} = 97611$ Pa, the Reynolds number $Re = \rho_{in} v_{in} H/\mu = 5227$, heat conduction coefficient $k = 2.428 \cdot 10^{-2}\,\mathrm{kg\,m\,s^{-2}\,K^{-1}}$, the specific heat $c_v = 721.428\,\mathrm{m^2\,s^{-2}\,K^{-1}}$, the Poisson adiabatic constant $\gamma = 1.4$. The inlet Mach number is $M_{in} = 0.012$. The parameter of the computational accuracy of the GMRES solver was 10^{-10}. The Young modulus and the Poisson ratio have values $E^b = 25000$ Pa and $\sigma^b = 0.4$, respectively, the structural damping coefficient is equal to the constant $C = 100\ \mathrm{s^{-1}}$ and the material density $\rho^b = 1040\,\mathrm{kg\,m^{-3}}$. The used time step was $8 \cdot 10^{-6}$s.

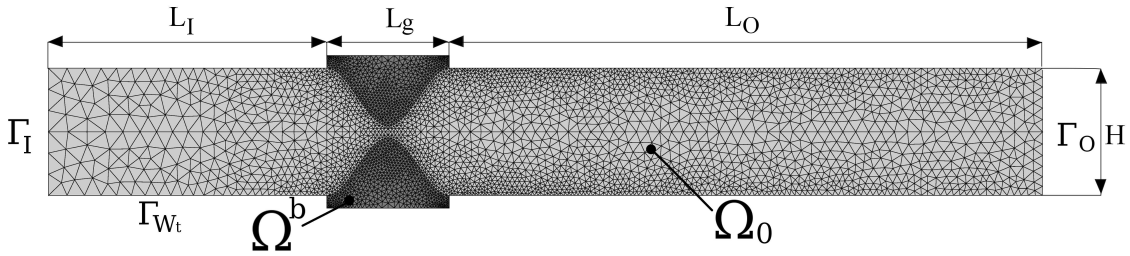


Figure 13.1: Computational domain at time $t = 0$ with a finite element mesh and the description of its size: $L_I = 50\,mm,\ L_g = 15.4\,mm,\ L_O = 94.6\,mm,\ H = 16\,mm$. The width of the channel in the narrowest part is $1.6\,mm$.

Mesh	Colour used in graphs	Flow part	Structure part
Mesh 1	red	5398	1998
Mesh 2	green	10130	2806
Mesh 3	blue	20484	4076

Table 13.1: Computational meshes.

In the numerical experiments quadratic ($r = 2$) and linear ($s = 1$) elements were used for the approximation of flow and structural problem, respectively.

In Table 13.1 we characterize the used computational meshes by the number of elements in the flow part and in the structure part of the mesh. Figure 13.1 shows the situation at the initial time $t = 0$ corresponding to the computational mesh 1. In Figure 13.2 we compare three different meshes in the flow domain and structure domain used in our computations. In the case of the flow channel we show only the narrowest part of the computational domain, which represents the most problematic part of the channel. Figure 13.3 presents the position of the point A in the flow channel, where the analysis of the mesh impact was carried out. In Figure 13.4 we see the positions of sensor points used in the analysis of flow-induced deformations.

First we tested the influence of the density of the computational meshes on the oscillations of the pressure in the point A. The corresponding Fourier analyses are carried out by the software Matlab. Figure 13.5 shows the behaviour of the pressure amplitude

$$(p - p_{average})(t) = p(A,t) - \frac{1}{T}\int_0^T p(A,t)\mathrm{d}t \tag{13.1}$$

computed with the aid of the strong coupling (on top) and the weak coupling (at the bottom). There are also presented corresponding Fourier analyses. It seems that in case of the weak coupling the mesh is already enough fine. No further improvement of the solution can be seen. See Table 13.3.

In order to compare an impact of the used coupling procedure we present the graphs of the pressure amplitude $p - p_{average}$ on the mesh 1 computed by the strong coupling (blue) and the weak coupling (red). Figure 13.6 shows that the difference between the results obtained by the strong and weak coupling is not too large. The main difference is in a higher stability of the strong coupling during the calculation on a long time interval. On the other hand, the strong coupling requires naturally

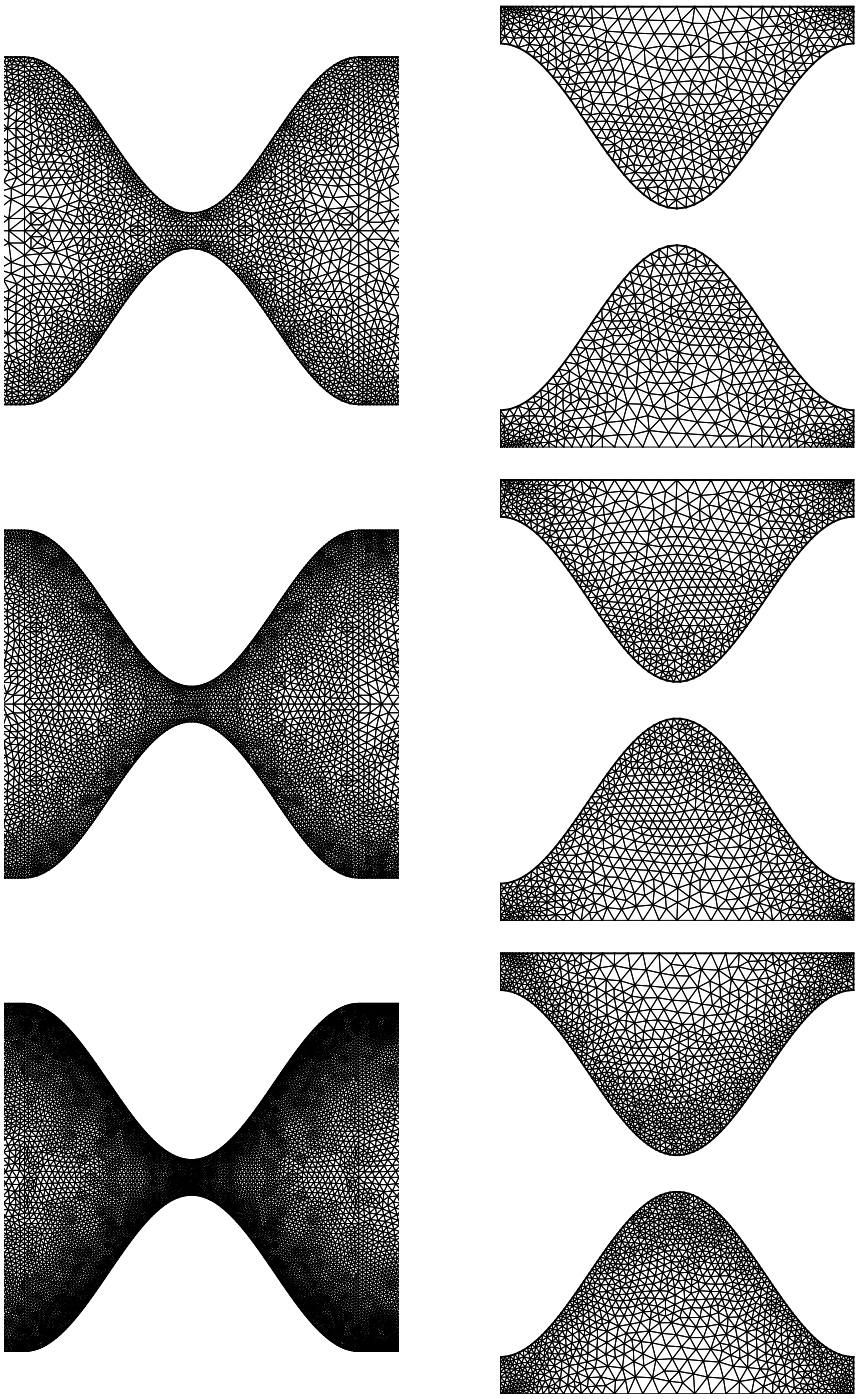


Figure 13.2: The detail of the flow meshes (left) 1, 2 and 3 in the narrowest part of the channel at time $t = 0$. The detail of the structure meshes (right) 1, 2 and 3 at time $t = 0$.

Figure 13.3: Position of the point A in the flow channel, where the analysis of the convergence tendency was carried out.

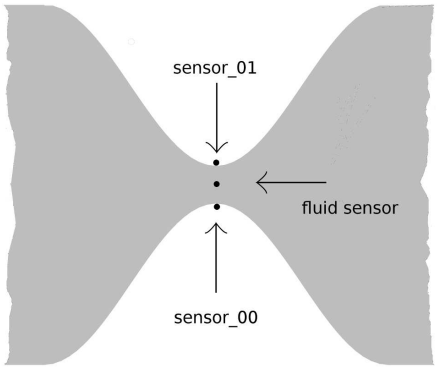


Figure 13.4: Positions of some sensors in the narrowest part of the channel used in the analysis

longer CPU time.

Now, let us deal with the flow field in the channel and the flow-induced deformations of the vocal folds model. In what follows, we shall present the results obtained by the computation on the coarse mesh (mesh 1 in Table 13.1). The coarse mesh was chosen in order to allow us the computation on the long time interval in a reasonable time. The strong coupling was used. In Figures 13.7 and 13.8 we can see the computational mesh and the velocity field near the vocal folds at several time instants. Figures 13.9 and 13.10 show the pressure isolines and the velocity isolines in the whole channel at same time instants. The maxima of the fluid velocity $v \approx 54$ ms^{-1} and the pressure 2 kPa correspond to the parameters of normal phonation. We can observe the Coanda effect represented by the attachment of the main stream (jet) successively to the upper and lower wall and formation of large scale vortices behind the glottis. The character of the vocal folds vibrations can be indicated in Figure 13.11, which shows the displacements dx and dy of the sensor points on the vocal folds surface (marked in Figure 13.4) in the horizontal and vertical directions, respectively. Moreover, the fluid pressure fluctuations in the middle of the gap as well as the Fourier analysis of the

Mesh	Colour used in graphs	Dominating frequency [Hz]
Mesh 1	red	121.3
Mesh 2	green	83.96
Mesh 3	blue	93.28

Table 13.2: Comparison of dominating frequency for the strong coupling on the different meshes.

Mesh	Colour used in graphs	Dominating frequency [Hz]
Mesh 1	red	121.3
Mesh 2	green	121.3
Mesh 3	blue	121.3

Table 13.3: Comparison of dominating frequency for the weak coupling on the different meshes.

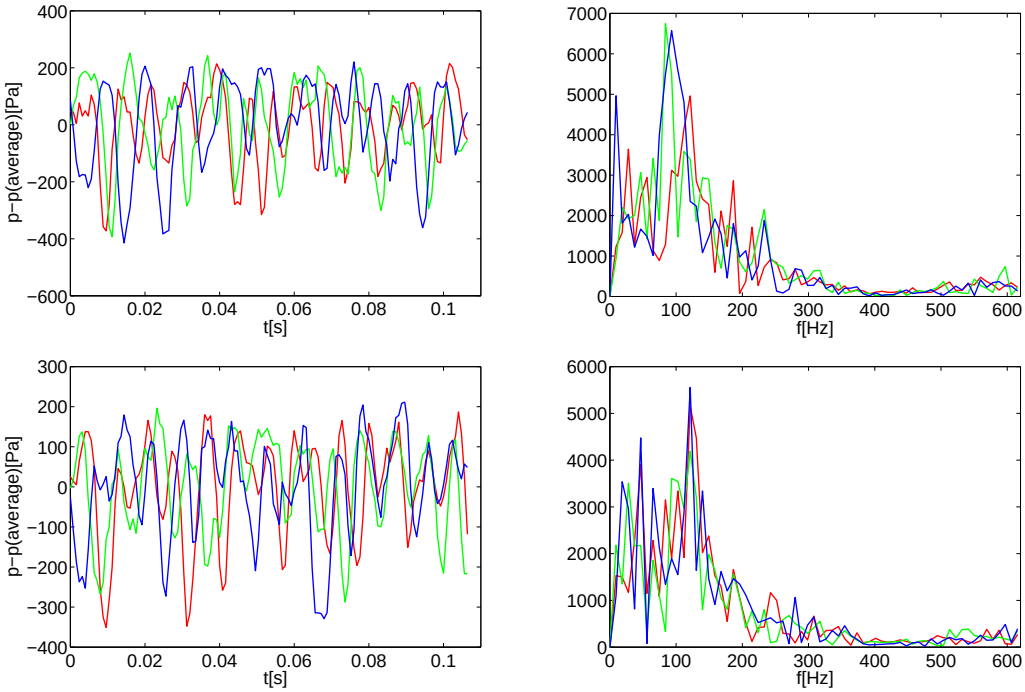


Figure 13.5: Dependence of the quantity $p - p_{average}$ and its Fourier analysis computed on three meshes: strong coupling (on top), weak coupling (at the bottom).

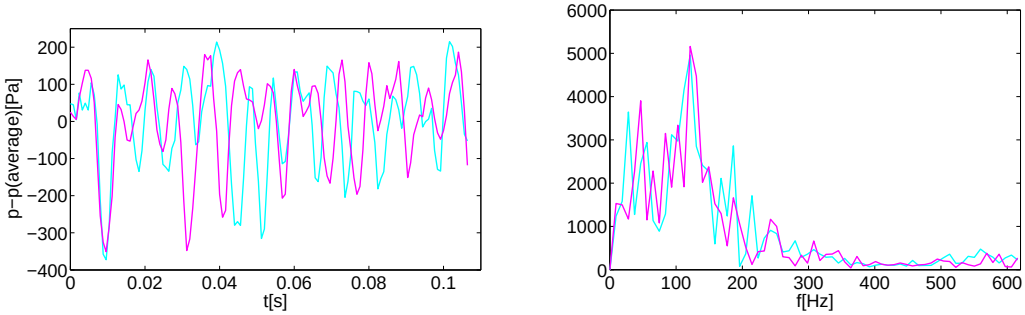


Figure 13.6: Comparison of the weak coupling (red) and the strong coupling (blue) on the mesh 1.

signals are shown here too. The vocal folds vibrations are not fully symmetric due to the Coanda effect and are composed of the fundamental horizontal mode of vibration with the corresponding frequency 113 Hz and by the higher vertical mode with the frequency 439 Hz. The increase of vertical vibrations due to the aeroelastic instability of the system results in a fast decrease of the glottal gap. At about t = 0.2 s, when the gap is nearly closed, the fluid mesh deformation in this region is too high and the numerical simulation stopped. The dominant peak at 439 Hz in the spectrum of the pressure signal corresponds well to the vertical oscillations of the glottal gap, while the influence of the lower frequency 113 Hz associated with the horizontal vocal folds motion is in the pressure fluctuations negligible. The modeled flow-induced instability of the vocal folds is called phonation onset followed in reality by a complete closing of the glottis and consequently by the vocal folds collisions producing the voice acoustic signal.

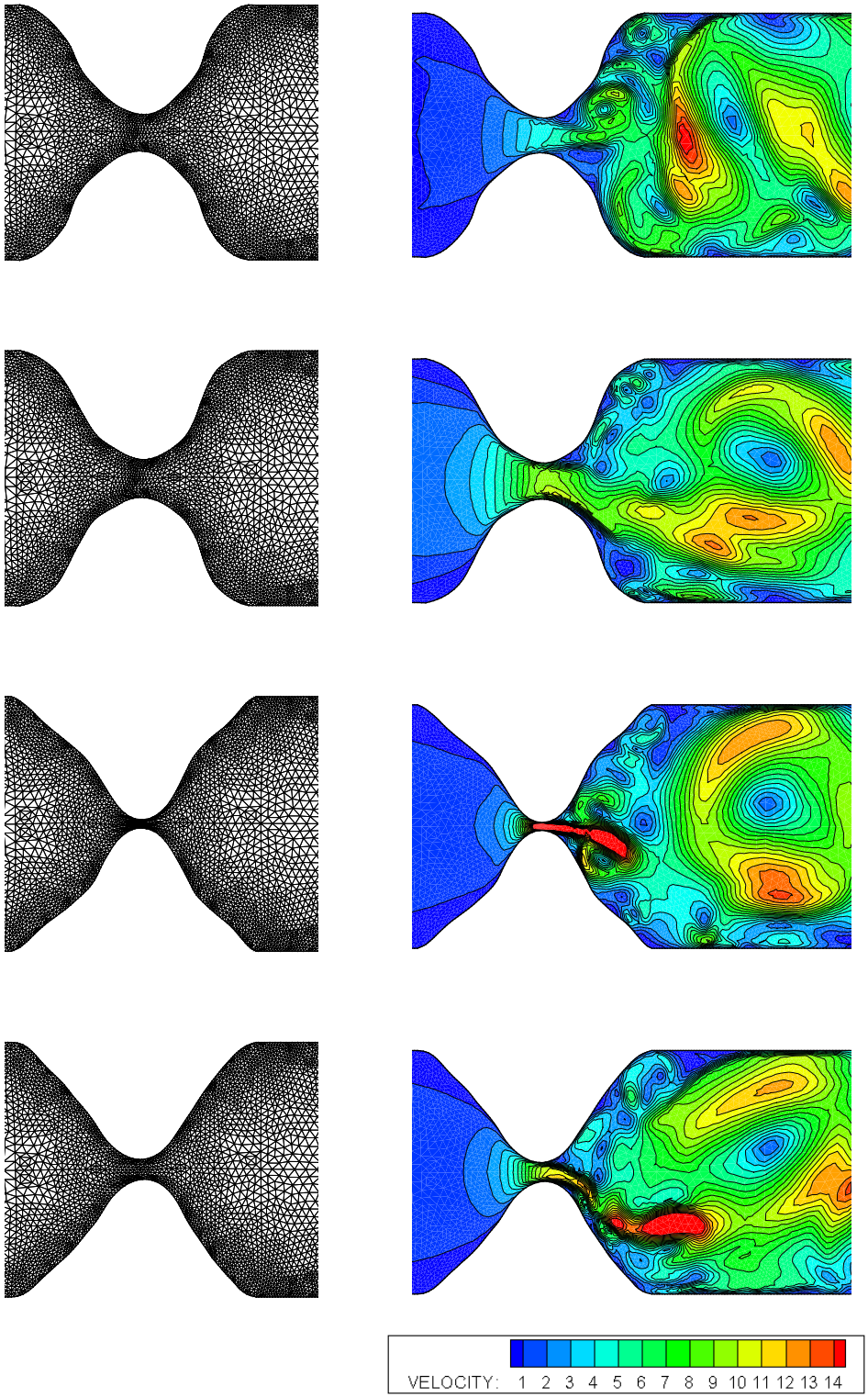


Figure 13.7: Detail of the mesh and the velocity distribution in the vicinity of the narrowest part of the channel at time instants $t = 0.1950, 0.1957, 0.1963, 0.1970$ s. The legend shows the dimensionless values of the velocity. For getting the dimensional values multiply by $U^* = 4$.

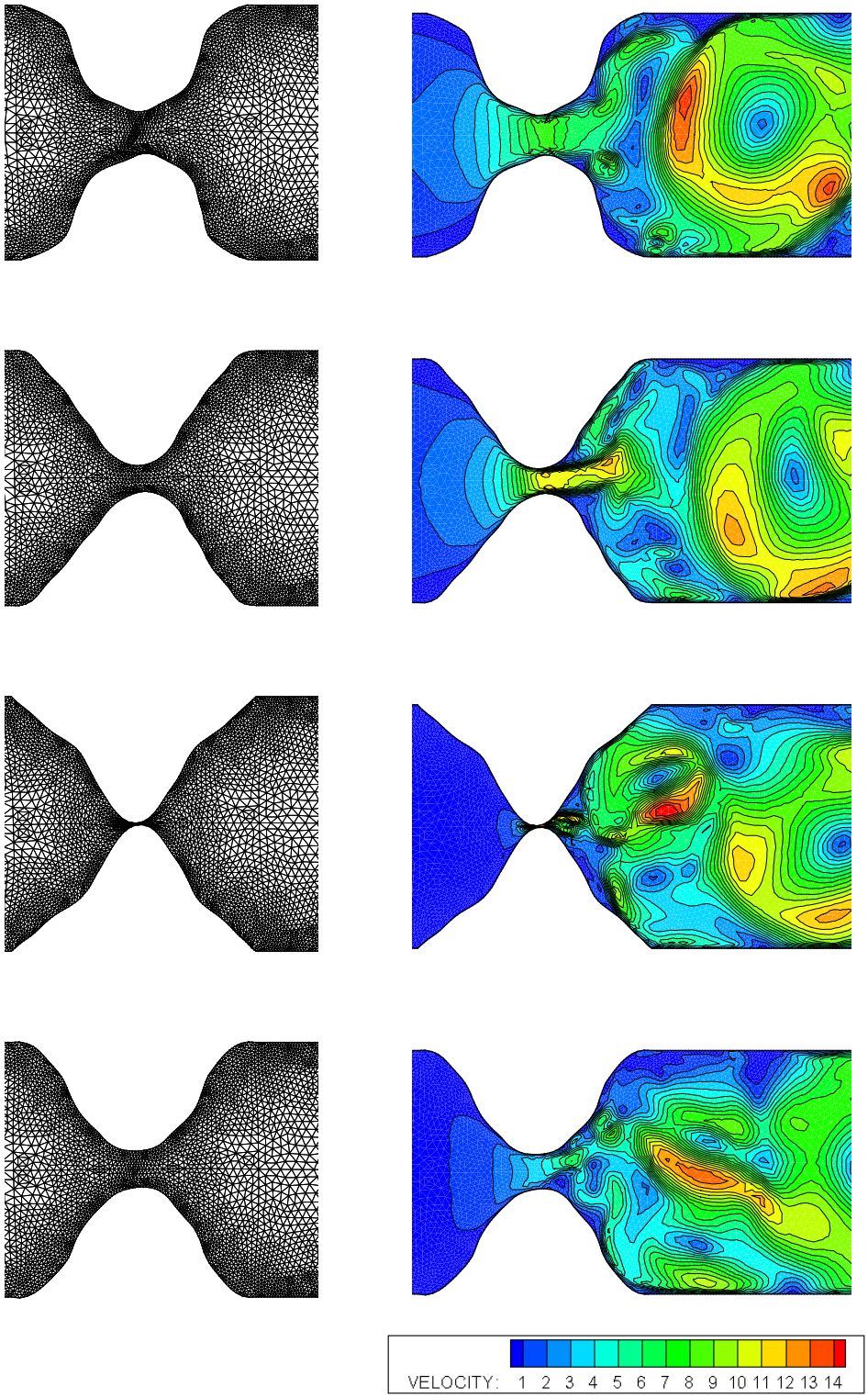


Figure 13.8: Detail of the mesh and the velocity distribution in the vicinity of the narrowest part of the channel at time instants $t = 0.1976, 0.1982, 0.1989, 0.1995$ s. The legend shows the dimensionless values of the velocity. For getting the dimensional values multiply by $U^* = 4$.

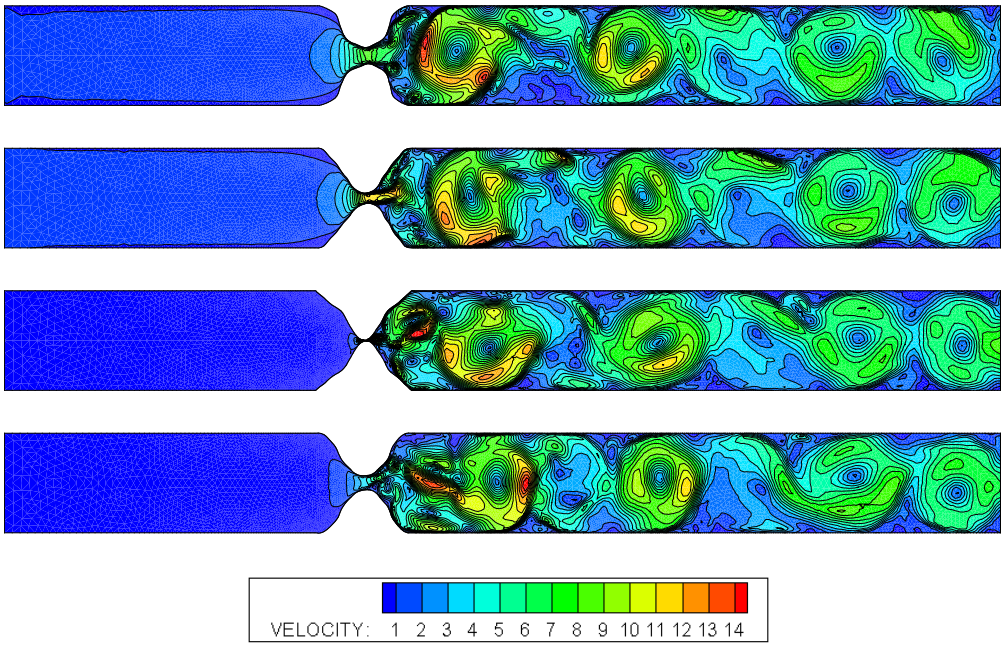


Figure 13.9: Velocity isolines at time instants $t = 0.1976, 0.1982, 0.1989, 0.1995$ s. The legend shows the dimensionless values of the velocity. For getting the dimensional values multiply by $U^* = 4$.

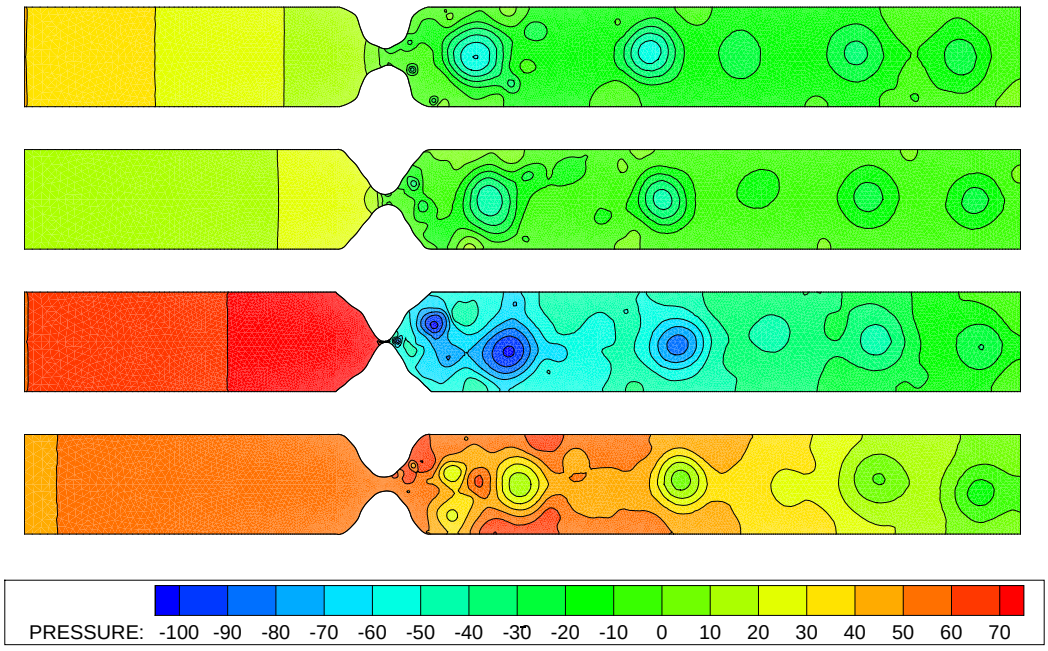


Figure 13.10: Pressure isolines at time instants $t = 0.1976, 0.1982, 0.1989, 0.1995$ s. The legend shows the dimensionless values of $p - p_{out}$. For getting the dimensional values multiply by $\rho^* U^{*2} = 19.6$.

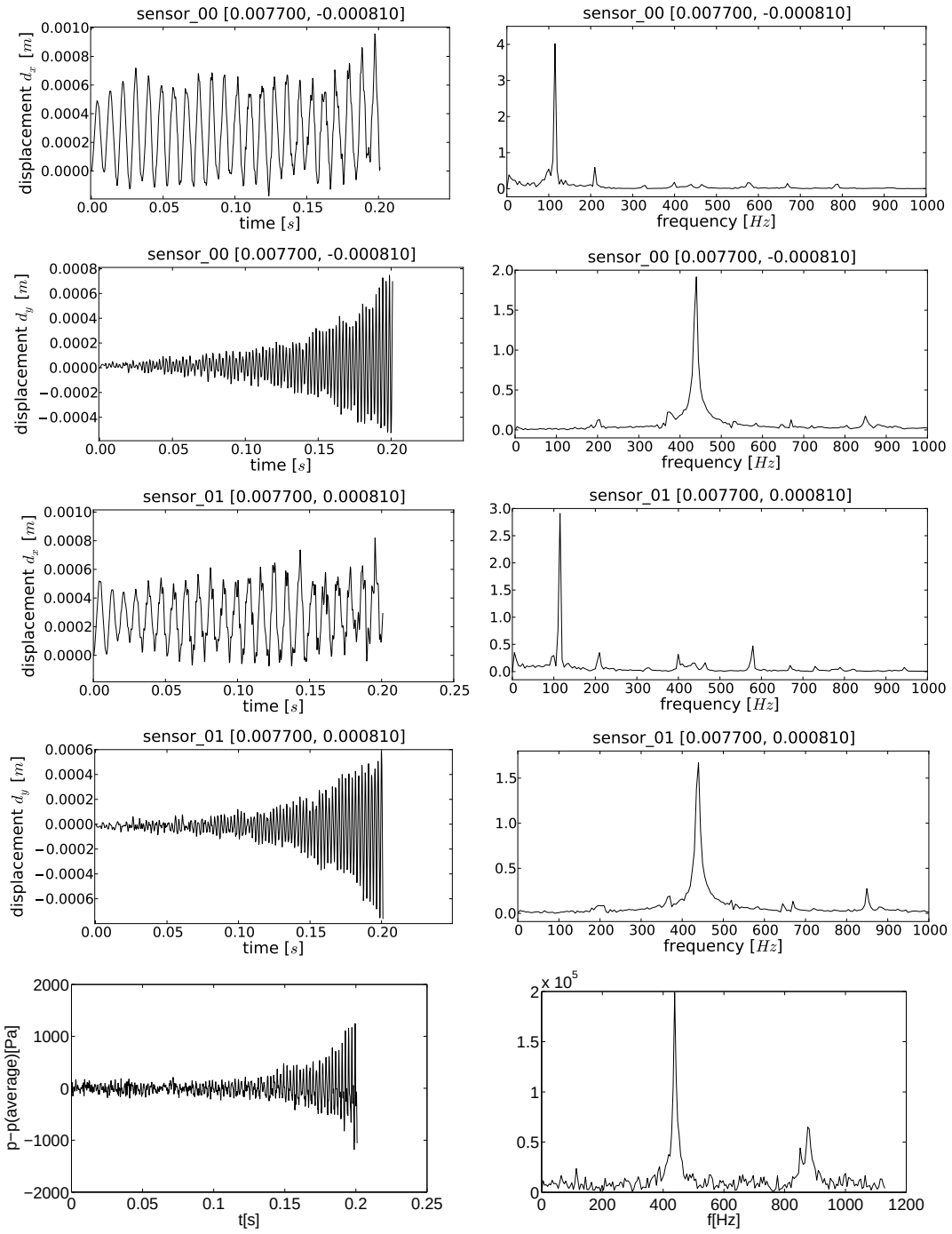


Figure 13.11: Vibrations of sensor points from the vocal folds and their Fourier analyses and the fluid pressure fluctuations in the middle of the gap and their Fourier analysis.

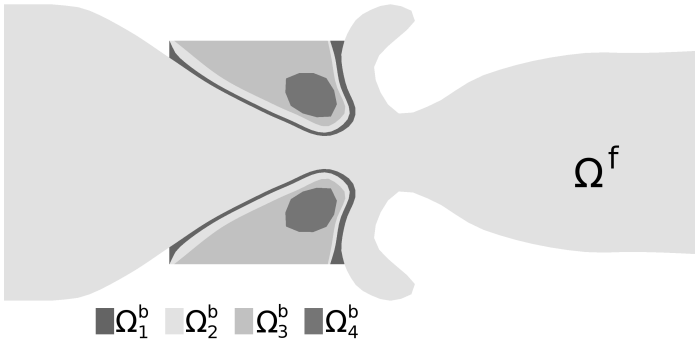


Figure 13.12: The scheme of the vocal tract.

subdomain	E^b	σ^b
Ω_1^b	100000	0.4
Ω_2^b	1000	0.495
Ω_3^b	8000	0.4
Ω_4^b	12000	0.4

Table 13.4: Material characteristics of the solid part Ω^b.

13.2 Example 2

Here we present numerical results obtained with the aid of the weak coupling technique and applied to the interaction of airflow in the domain Ω_t, representing the human vocal tract, with human vocal folds represented by the domain Ω^b.

We use the same time step $\tau = 4.35 \cdot 10^{-5}$ s for the flow problem and the structural problem. For the flow problem the following data set was applied: $\mu = 1.8375 \cdot 10^{-5}\ \mathrm{kg\,m^{-1}\,s^{-1}}$, $k = 2.428 \cdot 10^{-2}\ \mathrm{kg\,m\,s^{-3}\,K^{-1}}$, $c_v = 721.428 \mathrm{m^2\,s^{-2}\,K^{-1}}$, $Re = \rho_{in} v_{in} 2H_I/\mu = 4640$. At the inlet we prescribe the velocity vector $\boldsymbol{v}_{in} = (4, 0)\mathrm{m\,s^{-1}}$ and the density $\varrho^f_{in} = 1.225\ \mathrm{kg\,m^{-3}}$. At the outlet the pressure $p_{out} = 97611$ Pa is prescribed.

We assume that the vocal folds are isotropic bodies with the constant material density $\varrho^b = 1040\ \mathrm{kg\,m^{-3}}$. The values of the Young modulus E^b and the Poisson ratio σ^b are different in four subdomains of Ω^b. See Figures 13.12, 13.13 and Table 13.4. The damping coefficient $C = 0.1\ \mathrm{s^{-1}}$.

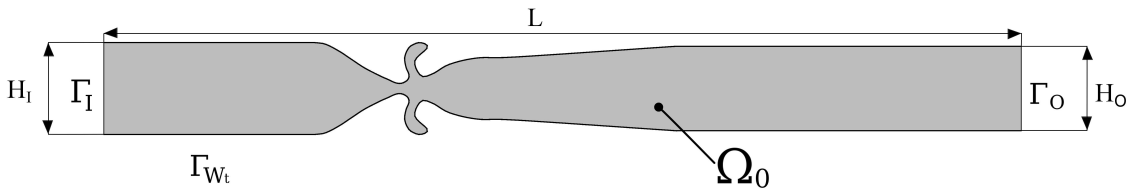


Figure 13.13: Scheme of the computational domain Ω^f at time $t = 0$ with the description of its size: $L = 87\,mm$, $H_I = 8.7\,mm$, $H_O = 8\,mm$. The width of the channel in the narrowest part is $1\,mm$.

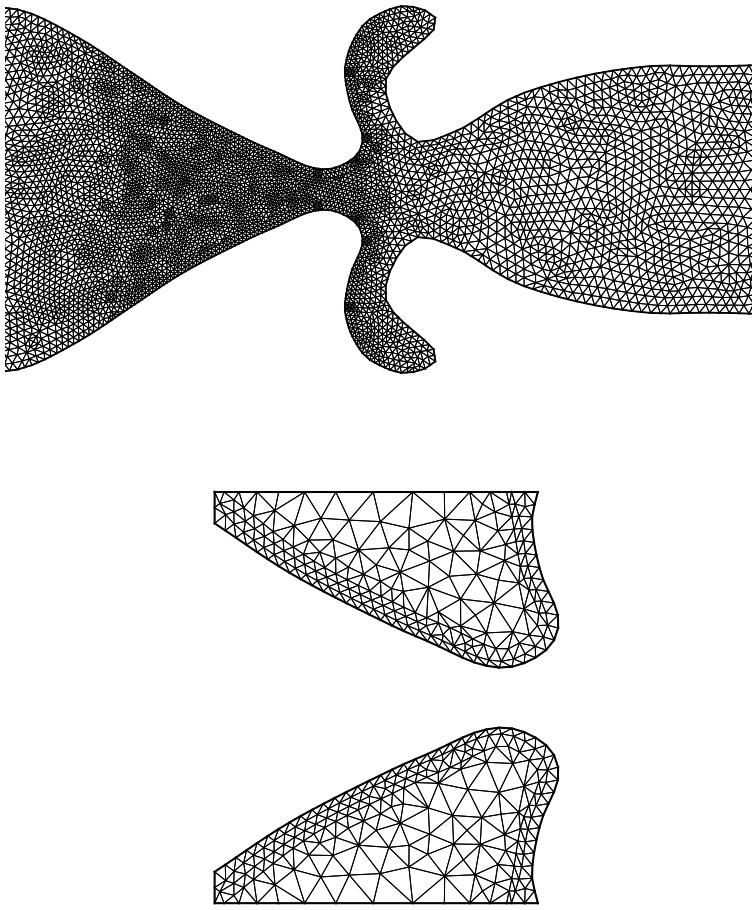


Figure 13.14: Detail of the mesh of the flow problem (on top) and the mesh of the structural problem (at the bottom).

The mesh of the domain Ω_t consists of 5002 elements. The detail of the flow domain can be seen in Figure 13.14. The mesh of the structural part of the problem has 710 elements, see Figure 13.14. Again, quadratic elements ($r = 2$) and linear elements ($s = 1$) were used for the approximation of flow and structural problem, respectively.

The character of the vocal folds vibrations can be indicated in Figure 13.15 showing horizontal and vertical displacements of the elastic body. This Figure also allows the comparison with the fluid pressure fluctuations in the middle of the gap. The Fourier analyses are provided. Figures 13.16 and 13.17 show the flow velocity and the flow pressure with the deformation of the computational domain at several time instants during the aeroelastic instability onset. The details of the flow velocity and the flow pressure in the narrowest part of the channel can be seen in Figures 13.18 and 13.19. Again we can see the Coanda effect and big vortices leaving the domain through the boundary Γ_O. As seen in Figures 13.16 and 13.17 the pressure and the velocity correspond to the parameters of normal phonation (see e.g. [63]).

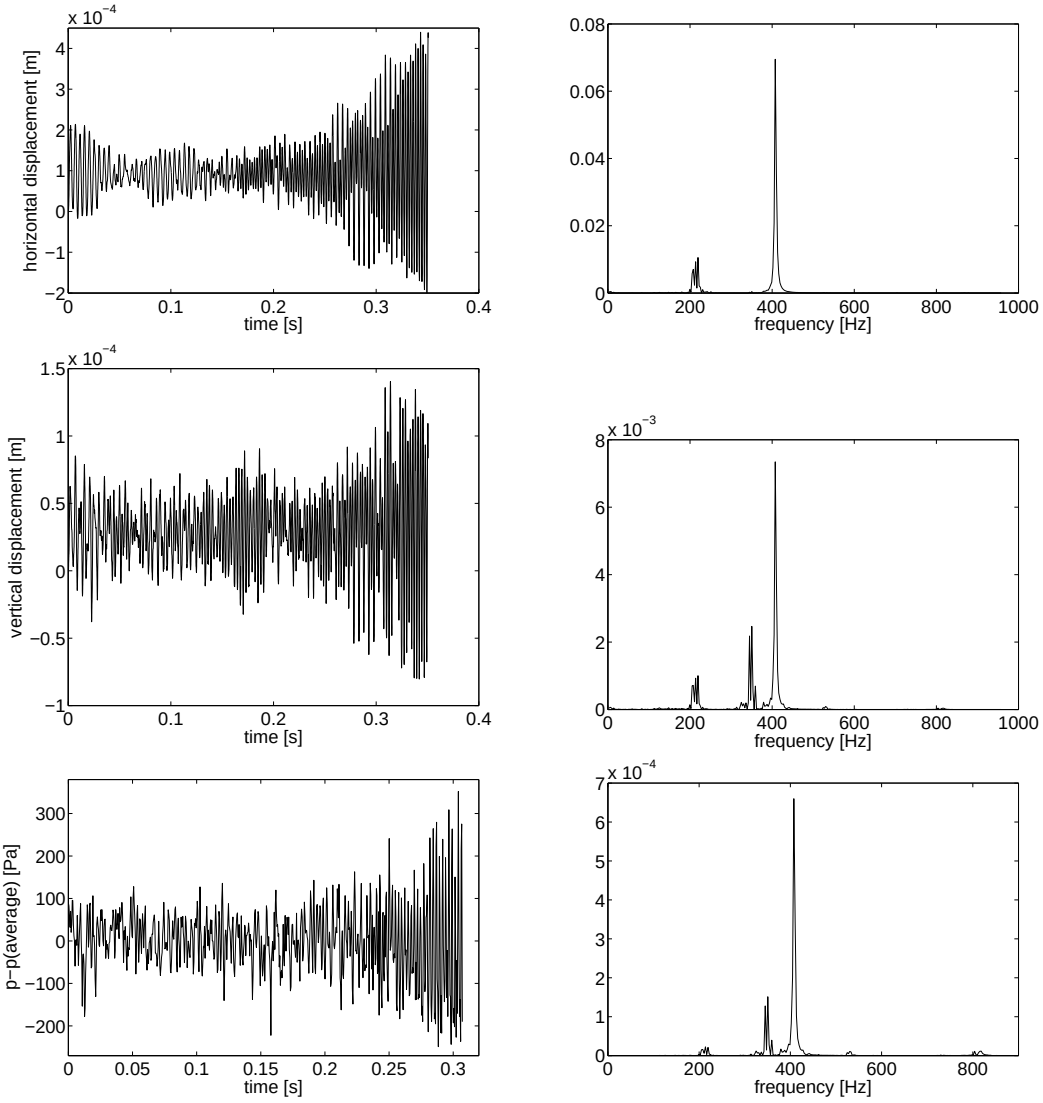


Figure 13.15: Vibrations of the sensor point lying inside the area Ω_3^b (see Figure 13.12) of the upper vocal fold and the fluid pressure fluctuations in the middle of the gap and their Fourier analyses.

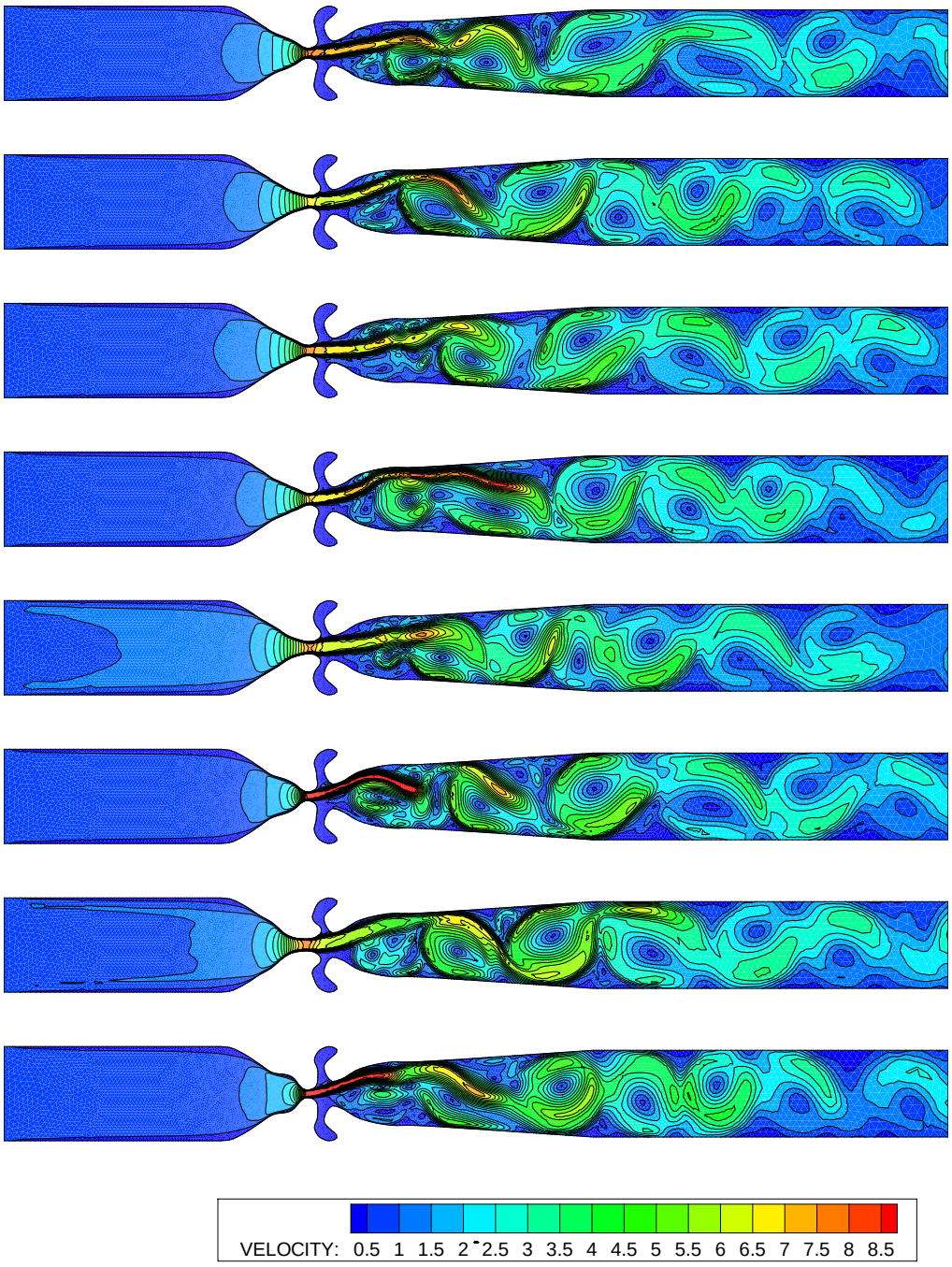


Figure 13.16: Velocity isolines at time instants $t = 0.261$, 0.272, 0.283, 0.294, 0.304, 0.315, 0.326, 0.337 s. The legend shows the dimensionless values of the velocity. For getting the dimensional values multiply by $U^* = 4$.

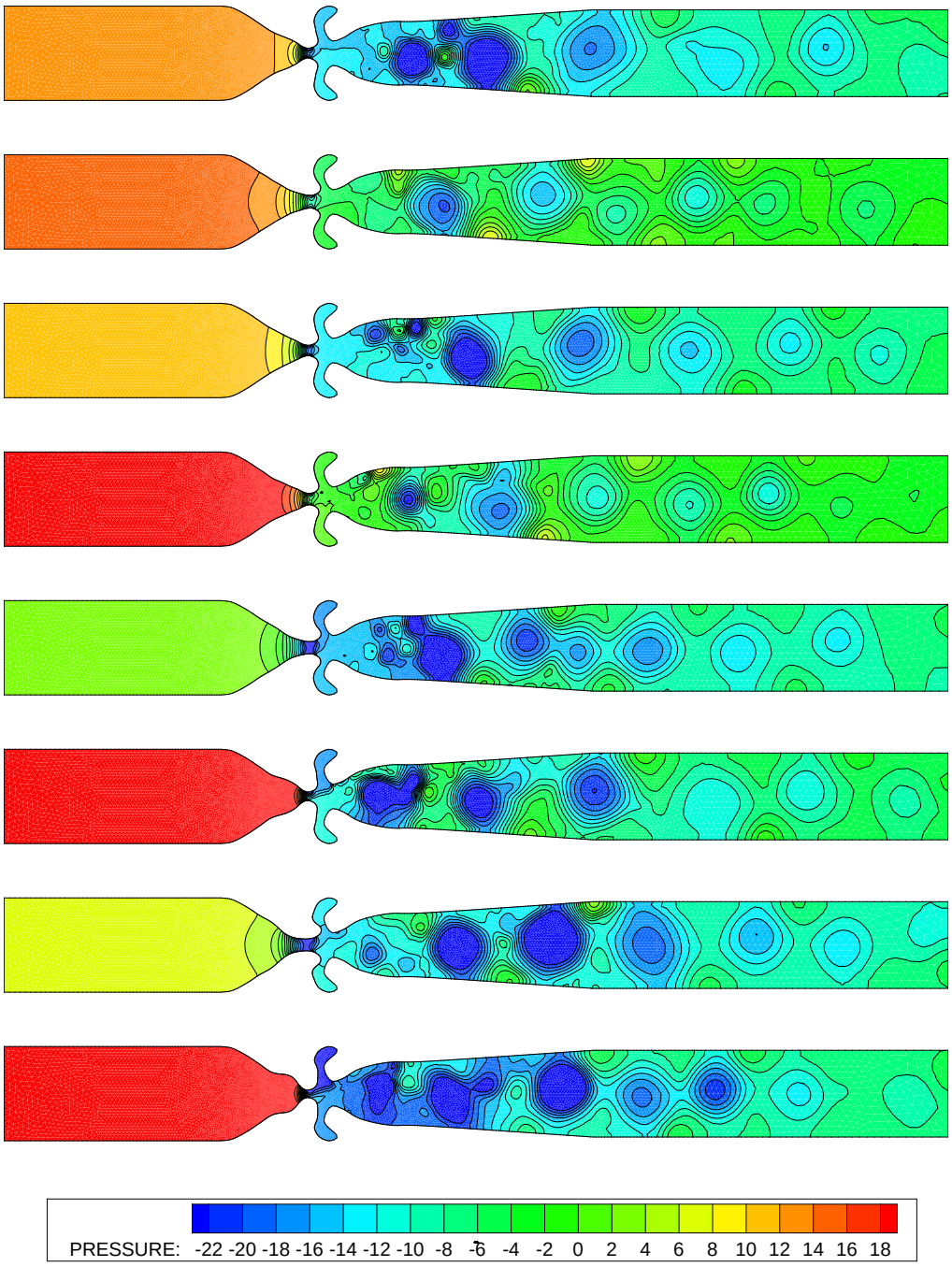


Figure 13.17: Pressure isolines at time instants $t = 0.261, 0.272, 0.283, 0.294, 0.304, 0.315, 0.326, 0.337$ s. The legend shows the dimensionless values of $p - p_{out}$. For getting the dimensional values multiply by $\rho^* U^{*2} = 19.6$.

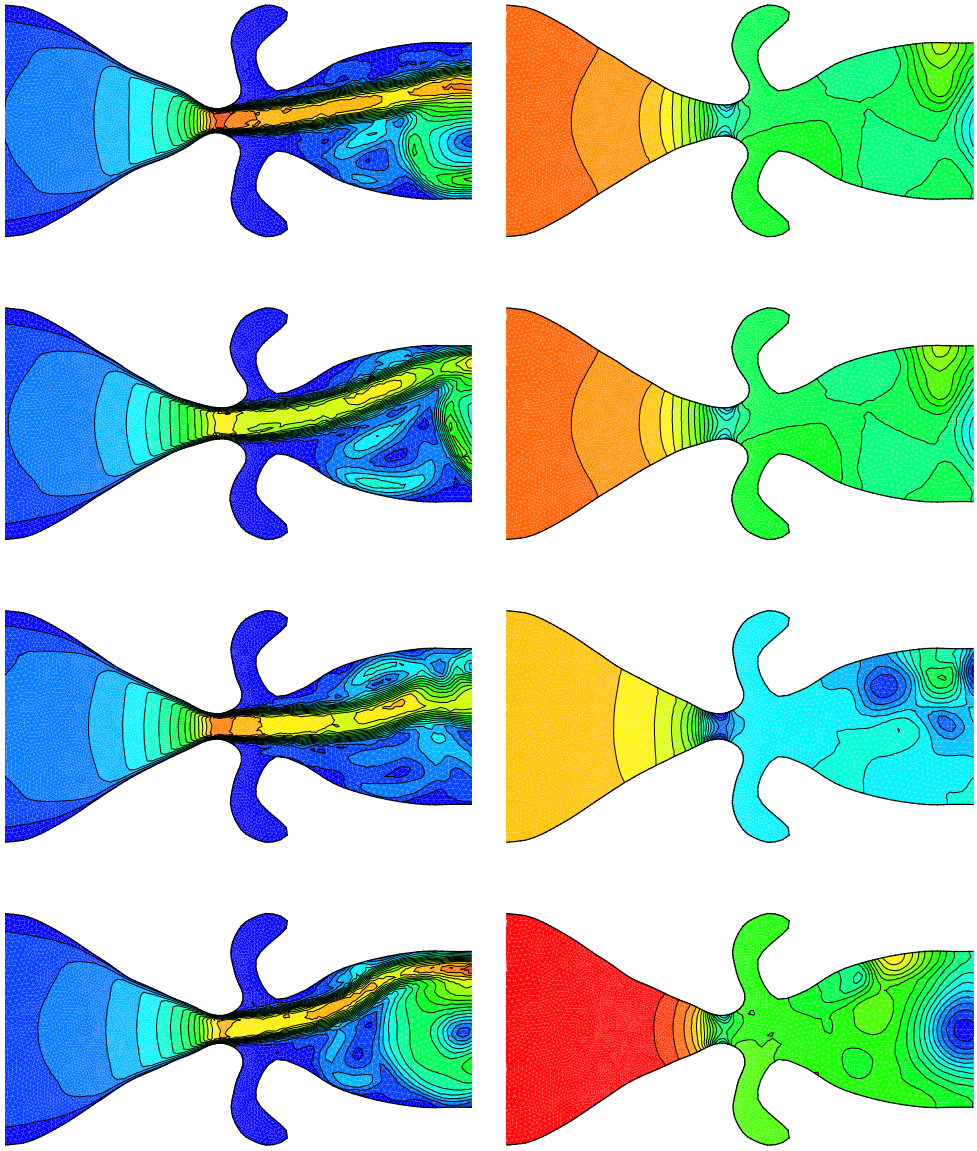


Figure 13.18: Detail of the velocity field (left) and the pressure field (right) in the neighbourhood of the narrowest part of the channel at time instants $t = 0.261, 0.272, 0.283, 0.294$ s.

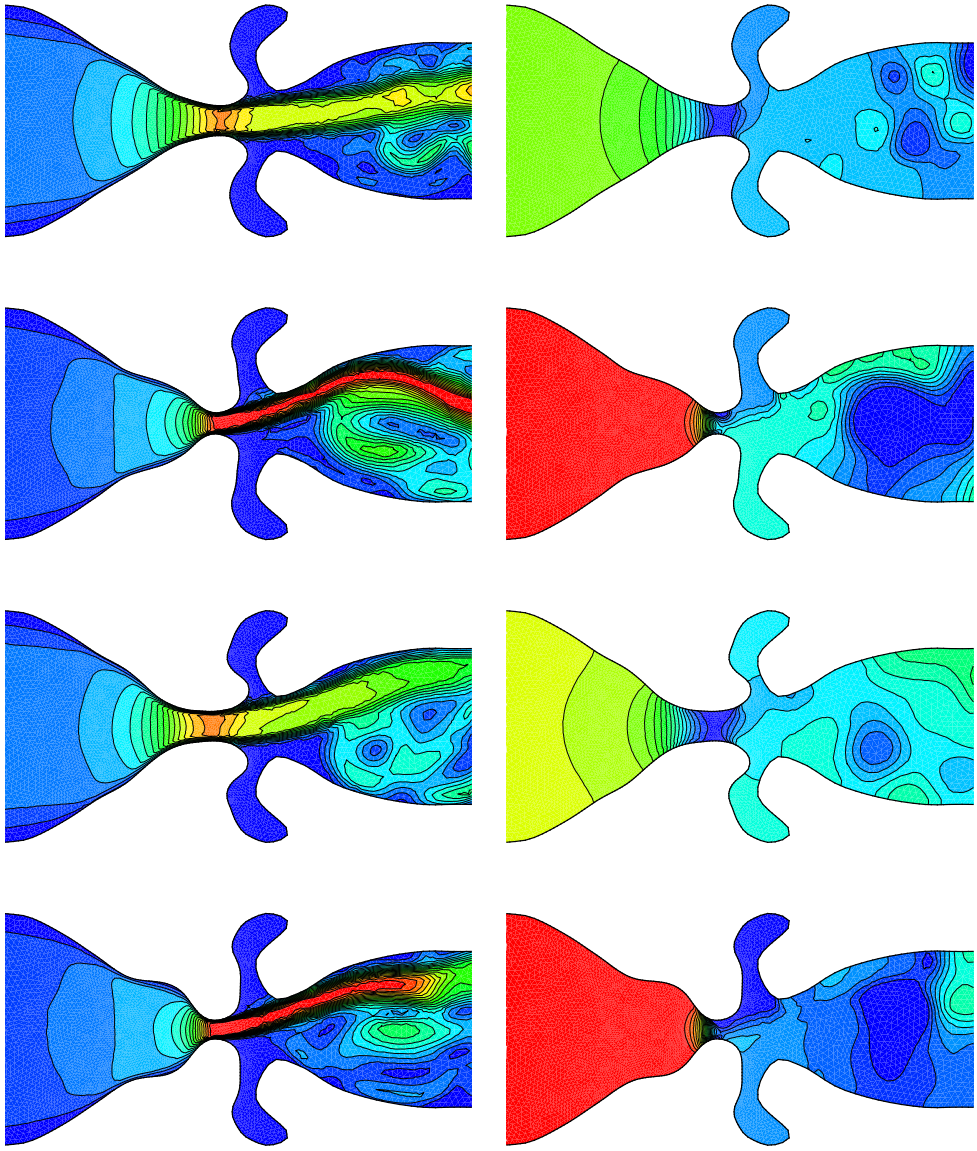


Figure 13.19: Detail of the velocity field (left) and the pressure field (right) in the neighbourhood of the narrowest part of the channel at time instants $t = 0.304, 0.315, 0.326, 0.337$ s.

Conclusion

In the first part of this thesis, we have formulated and theoretically analyzed the discontinuous Galerkin finite element method for the space-time discretization of a nonstationary convection-diffusion initial-boundary value problem with nonlinear convection, linear diffusion and a Dirichlet boundary condition. In the space discretization, we use polynomial approximations of degree $p \geq 1$, and the nonsymmetric (NIPG), incomplete (IIPG) and symmetric (SIPG) variants of the diffusion terms are considered. The discontinuous approximations of degree $q \geq 1$, in general $q \neq p$, are used in time. Under the assumption that the Dirichlet data behave in time as polynomials of degree $\leq q$, the derived estimates in $L^2(H^1)$-norm are optimal in space and time. The error estimate in $L^2(L^2)$-norm is optimal in time, but suboptimal in space. In the case of general Dirichlet data, the error estimates are suboptimal in time. The derivation of the optimal error estimates in space and time in the case of the SIPG method with the general Dirichlet data, the numerical realization of the discrete problem and the experimental demonstration of the results present the possible direction of the future work.

The second part of this thesis was devoted to the fluid-structure interaction problem motivated by the flow-induced vibrations of human vocal folds. It means that we have described the coupled problem of compressible viscous flow and the deformation of an elastic body. In the outline of this part of the thesis we can distinguish sections dedicated to the flow problem, to the structural problem and to their coupling. At the end of the work a separate chapter was devoted to examples of our numerical results.

The flow problem was described by the compressible Navier-Stokes equations and the dimensionless form of these equations was derived. Further, the governing equations were formulated in the arbitrary Lagrangian-Eulerian (ALE) form and discretized in space by the discontinuous Galerkin finite element method. The time discretization was carried out by the backward difference formula in time. The suitable algorithmization was mentioned.

The structural problem is represented by the dynamical equations of an isotropic elastic body. For its discretization we used the conforming finite element method. For the time discretization the Newmark method was applied. Again, the algorithmization was discussed.

The ALE mapping was introduced with the aid of an artificial elastostatic problem in the domain occupied by flow. We formulated two different possible coupling procedures, the strong and weak coupling. The developed method was programmed in the language C.

The last chapter of the thesis was devoted to numerical experiments carried out

for two examples. Both cases were inspired by the airflow in the human vocal tract, which induces the self oscillations of human vocal folds.

On the first simplified model we have presented both coupling procedures and compared their results. The experiments showed that the difference between the results obtained by the strong and weak coupling is not too large. The main difference is in a higher stability of the strong coupling during solving the problem on a long time interval. On the other hand, the strong coupling requires naturally longer CPU time. As follows from the description of both coupling procedures, the weak coupling is simpler than the strong coupling and thus less time consuming. Because the results of both coupling procedures perfectly match at the beginning of the computation (see Figure 13.6), the weak coupling seems to be a good tool for testing computations before starting the more complicated computations. The use of three successively refined meshes allowed the comparison of the solutions via testing quantity (13.1). These results, showed in Figure 13.5, demonstrate the convergence tendency of the method.

On the second example we have presented the more realistic problem with parameters better characterizing the properties of the tissue of human vocal folds. The elastic structure domain was split in four subdomains with the same material density, but with different Poisson ratio and Young's modulus in each of them. Also the shape of the computational channel was more realistic. Even if there were big deformations of the structure causing creation of the massive vortices in the flow domain, no difficulties in the flow part were marked and the vortices were smoothly leaving the domain.

The possible future work is the treatment of the complete closure of the channel. This effect occurs during the phonation in human vocal folds and together with the acoustic resonances of the human vocal tract causes the creation of human voice.

Bibliography

[1] *ANGENER*. http://www.karlin.mff.cuni.cz/ ˜ dolejsi/free.htm.

[2] *Cluster Sněhurka*. http://cluster.karlin.mff.cuni.cz/mw/index.php/Cluster_Sněhurka.

[3] F. Alipour, C. Brücker, D. Cook, A. Gömmel, M. Kaltenbacher, W. Mattheus, L. Mongeau, E. Nauman, R. Schwarze, I. Tokuda, and S. Zörner. Mathematical models and numerical schemes for simulation of human phonation. *Current Bioinformatics*, 6:323–343, 2011.

[4] F. Alipour and I. Titze. Combined simulation of two-dimensional airflow and vocal fold vibration. In *P. J. Davis and N. H. Fletcher (editors): Vocal fold physiology, controlling complexity and chaos*. San Diego, 1996.

[5] D. Arnold. An interionr penalty finite element method with discontinuous elements. *SIAM J.Numer. Anal.*, 19:742–760, 1982.

[6] D. Arnold, F. Brezzi, B. Cockburn, and D. Marini. Discontinuous Galerkin methods for elliptic problems. In *Discontinuous Galerkin Methods. Theory, Computation and Applications. Lecture Notes in Computational Science and Engineering, vol. 11*, pages 89–101. Springer, Berlin, 2000.

[7] D. Arnold, F. Brezzi, B. Cockburn, and D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J.Numer. Anal.*, 39:1749–1779, 2001.

[8] I. Babuška and C. Baumann. A discontinuous *hp* finite element method for diffusion problems, 1-D analysis. *Comput. Math. Appl.*, 37:103–122, 1999.

[9] S. Badia and R. Codina. On some fluid-structure iterative algorithms using pressure segregation methods. Application to aeroelasticity. *Int. J. Numer. Meth. Engng.*, 72:46–71, 2007.

[10] G. Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comput.*, 31:45–59, 1977.

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

[12] F. Bassi and S. Rebay. High-order accurate discontinuous finite element solution of the 2D Euler equations. *J. Comput. Phys.*, 138:251–285, 1997.

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

[14] R. Bisplinghoff, H. Ashley, and R. Halfman. *Aeroelaticity*. Dover, New York, 1996.

[15] P. Castillo, B. Cockburn, D. Schötzau, and C. Schwab. Optimal a priori estimates for the *hp*-version ofthe local discontinuous Galerkin method for convection-diffusion problems. *Math. Comput.*, 71:455–478, 2001.

[16] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1979.

[17] B. Cockburn and C. Shu. The local discontinuous Galerkin method for time-dependent cenvection-diffusion systems. *SIAM J.Numer. Anal.*, 35:2440–2463, 1998.

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

[19] A. Curnier. *Computational Methods in Solid Mechanics*. Kluwer Academic Publishing Group, Dodrecht, 1994.

[20] M. De Vries, H. Schutte, A. Veldman, and G. Verkerke. Glottal flow through a two-mass model: comparison of Navier-Stokes solutions with simplified models. *J. Acoust. Soc. Am.*, 111:1874–1853, 2002.

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

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

[23] V. Dolejší and M. Feistauer. Error estimates of the discontinuous Galerkin method for nonlinear nonstationary convection-diffusion problems. *Numer. Func. Anal. Optimiz.*, 26:349–383, 2005.

[24] V. Dolejší, M. Feistauer, and J. Hozman. Analysis of semi-implicit DGFEM for nonlinear convection-diffusion problems on nonconforming meshes. *Comput. Methods Appl. Mech. Engrg.*, 196:2813–2827, 2007.

[25] V. Dolejší, M. Feistauer, and C. Schwab. A finite volume discontinuous Galerkin scheme for nonlinear convection-diffusion problems. *Calcolo*, 39:1–40, 2002.

[26] V. Dolejší, M. Feistauer, and V. Sobotíková. A discontinuous Galerkin method for nonlinear convection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 194:2709–2733, 2005.

[27] E. Dowell. *Aeroelasticity of Plates and Shells*. Kluwer, Dordrecht, 1974.

[28] E. Dowell. *A Modern Course in Aeroelaticity.* Kluwer Academic Publishers, 1995.

[29] J. Česenek. *Nespojitá Galerkinova metoda pro řešení stlačitelného vazkého proudění. Disertační práce.* Matematicko-fyzikální fakulta, Univerzita Karlova v Praze, 2011.

[30] M. Feistauer. Optimal error estimates in the DGFEM for nonlinear convection-diffusion problems. In *Numerical Mathematics and Advanced Applications, ENUMATH 2007*, pages 323–330. Springer, Heidelberg, 2008.

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

[32] M. Feistauer, J. Hájek, and K. Švadlenka. Space-time discontinuous Galerkin method for solving nonstationary linear convection-diffusion-reaction problems. *Appl. Math.*, 52:197–234, 2007.

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

[34] M. Feistauer, V. Kučera, K. Najzar, and J. Prokopová. Analysis of space-time discontinuous Galerkin method for nonlinear convection-diffusion problems. *Numer. Math.*, 117:251–288, 2011.

[35] M. Feistauer and K. Švadlenka. Discontinuous Galerkin method of lines for solving nonstationary singularly perturbed linear problems. *J. Numer. Math.*, 2:97–117, 2004.

[36] Y. Fung. *An Introduction to the Theory of Aeroelaticity.* Dover Publications, New York, 1969.

[37] C. Geuzaine and J.-F. Remacle. Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79:1309–1331, 2009.

[38] C. Grandmont. Existence of a weak solutions for the unsteady interaction of a viscous fluid with an elastic plate. *SIAM J. Math. Sci.*, 40:716–737, 2008.

[39] R. Gregory and D. Karney. *A Collection of Matrices for Testing Computational Algorithms.* Wiley-Interscience, New York, 1969.

[40] M. Guidorzi, M. Padula, and P. Plotnikov. Hopf solutions to a fluid-elastic interaction model. *Math. Models Methods Appl. Sci.*, 18:215–269, 2008.

[41] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for compressible euler equations. In *Technical report 2001-42 (SFB 359)*. IWR Heidelberg.

[42] K. Hoffman and V. Starovoitov. On a motion of a solid body in a viscous fluid. Two-dimensional case. *Advanced in Mathematical Sciences and Applications*, 9:633–648, 1999.

[43] J. Horáček, P. Šidlof, and J. Švec. Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces. *J. Fluids Struct.*, 20:853–869, 2005.

[44] J. Horáček and J. Švec. Aeroelastic model of vocal-fold-shaped vibrating element for studying the phonation threshold. *J. Fluids Struct.*, 16:931–955, 2002.

[45] P. Houston, C. Schwab, and E. Süli. Discontinuous *hp*-finite element methods for advection-diffusion problems. *SIAM J. Numer. Anal.*, 39:2133–2163, 2002.

[46] C. Johnson and J. Pitkäranta. An analysis ofthe discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comput.*, 46:1–26, 1986.

[47] A. Kosík. *Interakce proudící tekutiny a elastického tělesa. Diplomová práce.* Matematicko-fyzikální fakulta, Univerzita Karlova v Praze, 2010.

[48] V. Kučera. *Higher order methods of the solution of compressible flows. Doctoral Thesis.* Faculty of mathematics and physics, Charles university in Prague, 2007.

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

[50] P. Le Saint and P. Raviart. On a finite element method for solving the neutron transport equation. In *Mathematical Aspects of FInite Elements in Partial Differential Equations*, pages 89–145. Academi Press, New York, 1974.

[51] E. Naudasher and D. Rockwell. *Flow-induced vibrations.* A.A.Balkema, Rotterdam, 1994.

[52] J. Neustupa. Existence of a weak solution to the Navier-Stokes equation in a general time-varying domain by the Rothe method. *Mathematical Methods in the Applied Sciences*, 32:653–683, 2009.

[53] J. Nečas and I. Hlaváček. *Úvod do pružných a pružně plastických těles.* SNTL, Praha, 1983.

[54] M. Paidoussis. *Fluid-Structure Interactions. Slender Structures and Axial Flow. Volume I.* Academic Press, San Diego, 1998.

[55] M. Paidoussis. *Fluid-Structure Interactions. Slender Structures and Axial Flow. Volume II.* Academic Press, London, 2004.

[56] J. Prokopová. *Numerical Solution of Compressible Flow. Master Thesis.* Faculty of mathematics and physics, Charles university in Prague, 2008.

[57] P. Punčochářová-Pořízková, K. Kozel, and J. Horáček. Simulation of unsteady compressible flow in a channel with vibrating walls - influence of frequency. *Computers and Fluids*, 46:404–410, 2011.

[58] A. Ralston. *A First Course in Numerical Analysis (czech translation)*. Academia, Praha, 1973.

[59] W. Reed and T. Hill. Triangular mesh methods for the neutron transport equation. In *Technical report LA-UR-73-479*. Los Alamos Scientific Laboratory, 1973.

[60] B. Rivière and M. Wheeler. A discontinuous Galerkin method applied to nonlinear parabolic equations. In *Discontinuous Galerkin Methods. Theory, Computation and Applications. Lecture Notes in Computational Science and Engineering, vol. 11*, pages 231–244. Springer, Berlin, 2000.

[61] T. Roubíček. *Nonlinear Partial Differential Equations with Applications*. Birkhäuser, Basel, 2005.

[62] I. Titze. *Principles of Voice Production*. National Center for Voice and Speech, Iowa City, 2000.

[63] I. Titze. *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech, Denver and Iowa City, 2006.

[64] V. Trkal. *Mechanika hmotných bodů a tuhého tělesa*. Praha, 1956.

[65] J. Van der Vegt and H. Van der Ven. Space-timediscontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flow. Part I: general formulation. *J. Comput. Phys.*, 182:546–585, 2002.

[66] M. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15:152–161, 1978.

[67] Z. Zhang, J. Neubauer, and D. Berry. Physical mechanisms of phonation onset: A linear stability analysis of an aeroelastic continuum model of phonation. *J. Acoust. Soc. Am.*, 122:2279–2295, 2007.