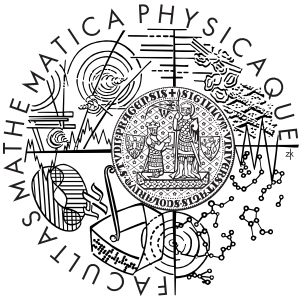
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Jaroslava Prokopová

Numerical solution of compressible flow

Department of Numerical Mathematics

Supervisor: Prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.

Study branch: Mathematical and Computer Modeling in Physics and Engineering

2008

I would like to thank all those who supported me in my master study and the work on my thesis. I very appreciate the help and guidance received from my supervisor Prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c. and I am grateful for numerous remarks, corrections and advices he gave me throughout my work. I would like to thank RNDr. Václav Kučera, Ph.D. for the help he provided me during the work on this thesis and the FEM program. I am also obliged to Doc. RNDr. Jan Zítko, CSc. for fruitful discussions about GMRES method.

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

My thanks also go to the Charles University in Prague that provided financial support for my research work. Through my master study, my work was partially supported by the Grant GAChU 48607.

I confirm having prepared the master thesis by my own, and having listed all used sources of information in the bibliography. I agree with lending the master thesis.

In Prague, April 18, 2008 .....

Jaroslava Prokopová

Contents

List of Figures

List of Tables

Název práce: Numerické řešení stlačitelného proudění
Autor: Bc. Jaroslava Prokopová
Katedra: Katedra numerické matematiky
Vedoucí práce: Prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.
e-mail vedoucího: feist@karlin.mff.cuni.cz

Abstrakt: Předkládaná práce se věnuje problematice proudění nevazké stlačitelné tekutiny v časově proměnné oblasti. Jsou zde popsány Eulerovy rovnice, jejich vlastnosti a řešení pomocí nespojité Galerkinovy metody konečných prvků (DGFEM) v časově nezávislé oblasti. Hlavní náplní práce je studium dané problematiky v časově proměnných oblastech. Za tímto účelem je zde představena tzv. ALE metoda. Pro řídící rovnice v ALE formulaci je odvozena jejich prostorová a časová diskretizace opět pomocí DGFEM metody. Krátce je zmíněna i stabilizace schématu a řešení vzniklé lineární soustavy pomocí GMRES metody. Na závěr jsou uvedeny a porovnány výsledky získané pomocí dvou rozdílných ALE formulací řídících rovnic v obdélníkové oblasti s pohyblivou částí spodní stěny.

Klíčová slova: stlačitelné Eulerovy rovnice, ALE metoda, nespojitá Galerkinova metoda konečných prvků, okrajové podmínky, numerický tok, stabilizace schématu, GMRES metoda

Title: Numerical solution of compressible flow
Author: Bc. Jaroslava Prokopová
Department: Department of Numerical Mathematics
Supervisor: Prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.
Supervisor's e-mail address: feist@karlin.mff.cuni.cz

Abstract: This work deals with the problem of inviscid, compressible flow in a time-dependent domain. We describe mathematical properties of the Euler equations and the system of governing equations is solved with the aid of the discontinuous Galerkin finite element method (DGFEM) in the time-indepentent domain. The main aim of this work is the study of this problem in time-dependent domains. For this reason the Arbitrary Lagrangian-Eulerian (ALE) method is presented. The governing equations are formulated in the ALE formulation and discretized in space and time by the DGFEM. Shortly we mention the shock capturing of the obtained scheme and the solution of the resulting linear system with the aid of Generalized Minimal Residual (GMRES) method. At the end of this work we present and compare results obtained by two different ALE formulations of the governing equations in the rectangular domain with a moving part of lower wall.

Keywords: compressible Euler equations, Arbitrary Lagrangian-Eulerian method, Discontinuous Galerkin finite element method, boundary conditions, numerical flux, shock capturing, Generalized Minimal Residual method

Introduction

At the current speed of technology progress, fluid-structure interaction affects an increasing number of technical applications - airfoil and helicopter rotor blade vibration, stability of suspension bridges, towers, smokestacks and skyscrapers, vibration of turbine blades or flow in heat exchangers and nuclear reactors. Through investigations of catastrophic disasters caused by wind-induced vibrations, such as ruptures of aircraft wings, collapse of the Tacoma Narrows bridge on November 7, 1940 or breakdown of the cooling towers in Ferrybridge on November 1, 1965 lead to development of a new scientific and technical discipline: the aeroelasticity.

The aeroelasticity calculations need to combine the methods of three classical branches of mechanics: dynamics of rigid bodies and structures, fluid dynamics and elasticity. In the majority case, the consequences of the aeroelastic effects are rather undesirable - the flow-induced vibration may affect negatively the operation of the systems, lead to material fatigue or induce excessive noise generation. However, there are processes where the fluid-structure interaction plays a crutial role; this is the case of voice production in human vocal folds.

This thesis is concerned with the numerical solution of the compressible flow. Especially we are focused on the fluid flow in the 2D time-dependent domain. This thesis can be considered as a preparation step for solving the problem of flow in human vocal folds or other aeroelastic problems.

We deal with inviscid compressible flow in a rectangular channel with moving lower wall, where the motion of this wall is prescribed by a periodic function. We choose a compressible flow, because there could be possibility of complete closure of the channel. This could be very useful in the simulation of the flow in human vocal folds, because the majority of papers dealing with this topic is concerned with an incompressible flow. This type of flow does not allow the complete closure of the testing channel although in reality the glottal region of the human vocal folds can be comlepetely closed at some moments.

The mathematical model of the fluid flow is represented by the system consisting of the 2D Euler equations, the continuity equation and the energy equation, equipped with initial conditions and mixed boundary conditions. This system is solved by the discontinuous Galerkin Finite Element method (DGFEM).

The unsteadiness of the the flow is caused by a prescribed periodic motion of a part of the channel wall with large amplitudes. Even in the case that the inlet velocity has low Mach number it may happen that the flow becomes transonic with discontinuities due to a very narrow part of the channel. It is a reason why the suitable shock capturing technique is applied.

The time-dependence of the computational domain requires to use techniques working on moving meshes. A suitable choice is to apply the arbitrary Lagrangian-Eulerian (ALE) method, which is based on the reformulation of the Euler equations using an ALE mapping of the reference configuration onto the current configuration for the time under consideration.

The use of the DGFEM leads to a large discrete system of nonlinear algebraic equations. In order to solve the problem in the shortest possible time, a suitable linearization and sufficiently fast solver has to be applied in each time step.

In this thesis, attention is paid step-by-step to the following aspects: implementation of the ALE method, semi-implicit time discretization and space discontinuous Galerkin finite element discretization of the system of governing equations for the problem in the time-independent domain as well as in the time-dependent domain, stabilization of the DGFEM method, solving of the resulting linear system by the Generalized Residual (GMRES) method and other aspects of the problem. The computational results are presented.

Chapter 1

Fundamental equations describing the compressible flow

In this chapter the basics of the gas dynamics will be presented.

We will consider a time interval $(0,T) \subset \mathbb{R}$, during which we follow the fluid motion, and the domain $\Omega_t \subset \mathbb{R}^3$ occupied by the fluid at time $t \in (0,T)$. So we define the set $\mathcal{M}$ by $\mathcal{M} = \{(\boldsymbol{x},t); \boldsymbol{x} \in \Omega_t, t \in (0,T)\}$ and assume that it is open.

The *fundamental hypotesis* will be used. It assumes that exactly one fluid particle passes through each point $\boldsymbol{x} \in \Omega_t$ at any time t.

1.1 Description of the flow

Two different ways of describing the flow can be used.

1.1.1 Lagrangian description

This approach of flow considers the motion of each individual fluid particle. The trajectories of the particles can be described by the equation

$$\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t) \tag{1.1}$$

(i.e. $x_i = \varphi_i(\boldsymbol{X}, t),\ i = 1,2,3$), where $\boldsymbol{X}$ represents the *reference* determining the particle under consideration. We can also use a more detailed description of the motion of the fluid particle in the form

$$\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t_0; t) \tag{1.2}$$

which determines, at time t, the position $\boldsymbol{x}$ of the article passing through the point (given by the reference) $\boldsymbol{X}$ at time t_0. Then

$$\boldsymbol{X} = \boldsymbol{\varphi}(\boldsymbol{X}, t_0; t_0)$$

provided the references are identical with the coordinates of particles at time t_0.

The components X_1, X_2, X_3 of the refrence $\boldsymbol{X}$ are called *Lagrangian coordinates* in contrast to the *Eulerian coordinates* x_1, x_2, x_3.

The Lagrangian description in the form (1.2) is useful for studying the flow of a piece of fluid formed by the same particles at each time instant and filling a domain $\mathcal{V}(t) \subset I\!R^3$ at time t.

The *velocity* and the *acceleration* of the fluid particles given by the reference $\boldsymbol{X}$ are defined as

$$\hat{\boldsymbol{v}}(\boldsymbol{X},t) = \frac{\partial \boldsymbol{\varphi}}{\partial t}(\boldsymbol{X},t) \quad \left(= \frac{\partial \boldsymbol{\varphi}}{\partial t}(\boldsymbol{X},t_0;t)\right) \tag{1.3}$$

$$\hat{\boldsymbol{a}}(\boldsymbol{X},t) = \frac{\partial^2 \boldsymbol{\varphi}}{\partial t^2}(\boldsymbol{X},t) \quad \left(= \frac{\partial^2 \boldsymbol{\varphi}}{\partial t^2}(\boldsymbol{X},t_0;t)\right), \tag{1.4}$$

respectively, provided the above derivatives exist.

1.1.2 Eulerian description

The second way of discribing the flow is based on the determination of the *velocity* $\boldsymbol{v}(\boldsymbol{x},t)$ of the fluid particle passing through the point $\boldsymbol{x}$ at time t. With respect to (1.1) and (1.3) we can write

$$\boldsymbol{v}(\boldsymbol{x},t) = \hat{\boldsymbol{v}}(\boldsymbol{X},t) = \frac{\partial \boldsymbol{\varphi}}{\partial t}(\boldsymbol{X},t), \text{ where } \boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X},t). \tag{1.5}$$

Under the assumption that

$$\boldsymbol{v} \in \left[\mathcal{C}^1(\mathcal{M})\right]^3, \tag{1.6}$$

the *acceleration* of the particle passing though the point $\boldsymbol{x}$ at time t is expressed as

$$\boldsymbol{a}(\boldsymbol{x},t) = \frac{\partial \boldsymbol{v}}{\partial t}(\boldsymbol{x},t) + \sum_{i=1}^{3} v_i(\boldsymbol{x},t)\frac{\partial \boldsymbol{v}}{\partial x_i}(\boldsymbol{x},t), \tag{1.7}$$

which can be also written in the form

$$\boldsymbol{a} = \frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot \mathrm{grad})\boldsymbol{v} = \frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v}, \tag{1.8}$$

where we omit the variables $\boldsymbol{x}$ and t for simplicity.

Let us introduce the symbol

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla \tag{1.9}$$

called the *material* (or total) *derivative* with respect to time. The partial derivative $\partial/\partial t$ is called the *local derivative* and the term $(\boldsymbol{v}\cdot\mathrm{grad})$ is referred to as the *convective derivative*. We see that the acceleration of a fluid particle is expressed in Eulerian coordinates as the material derivative of the velocity:

$$\boldsymbol{a} = \frac{\mathrm{D}\boldsymbol{v}}{\mathrm{D}t} := \frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}\cdot\mathrm{grad}\boldsymbol{v}. \tag{1.10}$$

1.1.3 The transition from the Eulerian description to the Lagrangian description

The problem of the transition from the Eulerian description to the Lagrangian description is equivalent to the determination of the paths of fluid particles on the basis of a given velocity field $\boldsymbol{v}(\boldsymbol{x},t)$.The trajectory of the fluid particle passing through the point $\boldsymbol{X} \in \Omega_{t_0}$ at time $t_0 \in (0,T)$ is given as the solution of the initial value problem

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{v}(\boldsymbol{x},t), \quad \boldsymbol{x}(t_0) = \boldsymbol{X}. \tag{1.11}$$

The following theorem gives us the solution of the transition problem.

Theorem 1.1: *Under assumption (1.6) the following statements hold:*

1. *For each $(\boldsymbol{X},t_0) \in \mathcal{M}$ problem (1.11) has exactly one maximal solution $\boldsymbol{\varphi}(\boldsymbol{X},t_0;t)$ (defined for t from a certain interval $(\alpha_{\boldsymbol{X},t_0},\beta_{\boldsymbol{X},t_0})$).*

2. *The mapping $\boldsymbol{\varphi}$ has continuous first order partial derivatives with respect to X_1,X_2,X_3,t_0,t and continuous derivatives $\partial^2\boldsymbol{\varphi}/\partial t\partial X_i$, $\partial^2\boldsymbol{\varphi}/\partial t_0\partial X_i$. $i=1,2,3$, in its domain of definition $\{(\boldsymbol{X},t_0,t);\ (\boldsymbol{X},t_0)\in\mathcal{M},\ t\in(\alpha_{\boldsymbol{X},t_0},\beta_{\boldsymbol{X},t_0})\}$.*

The prove of this can be found in [1].

1.2 The transport theorem

Let a function $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{x},t) : \mathcal{M} \longrightarrow I\!R$ be the Eulerian representation of some physical quantity transported by fluid particles and let us consider a system of fluid particles filling a *bounded domain* $\mathcal{V}(t) \subset \Omega_t$ at time t. The total amount of the quantity given by the function $\boldsymbol{F}$ that is contained in the volume $\mathcal{V}(t)$ at time t equals the integral

$$\mathcal{F}(t) = \int_{\mathcal{V}(t)} \boldsymbol{F}(\boldsymbol{x},t)d\boldsymbol{x}. \tag{1.12}$$

For the formulation of fundamental equations describing the flow we will need to calculate the rate of change of the quantity $\boldsymbol{F}$ bound on the system of particles considered. It means we shall be interested in the derivative

$$\frac{d\mathcal{F}(t)}{dt} = \frac{d}{dt}\int_{\mathcal{V}(t)} \boldsymbol{F}(\boldsymbol{x},t)d\boldsymbol{x}. \tag{1.13}$$

Let us suppose that $\boldsymbol{F} \in \mathcal{C}^1(\mathcal{M})$ and $\boldsymbol{v} \in [\mathcal{C}^1(\mathcal{M})]^3$ and let $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{X},t_0;t)$ be the mapping from Theorem (1.1). This mapping defines the changes of the domain $\mathcal{V}(t)$ with time. Let $t_0 \in (0,T)$ be an arbitrary fixed time instant and $\mathcal{V}(t_0) \subset \Omega_{t_0}$. Then

$$\mathcal{V}(t) = \{\boldsymbol{\varphi}(\boldsymbol{X},t_0;t);\ \boldsymbol{X} \in \mathcal{V}(t_0)\}. \tag{1.14}$$

By $J(\boldsymbol{X},t)$ we shall denote the Jacobian of the mapping '$\boldsymbol{X}\in\mathcal{V}(t_0)\longrightarrow\boldsymbol{\varphi}(\boldsymbol{X},t_0;t)\in\mathcal{V}(t)$':

$$J(\boldsymbol{X},t)=\det\frac{D\boldsymbol{\varphi}(\boldsymbol{X},t_0;t)}{D\boldsymbol{X}}=\det\begin{pmatrix}\frac{\partial\varphi_1}{\partial X_1}, & \frac{\partial\varphi_1}{\partial X_2}, & \frac{\partial\varphi_1}{\partial X_3}\\ \frac{\partial\varphi_2}{\partial X_1}, & \frac{\partial\varphi_2}{\partial X_2}, & \frac{\partial\varphi_2}{\partial X_3}\\ \frac{\partial\varphi_3}{\partial X_1}, & \frac{\partial\varphi_3}{\partial X_2}, & \frac{\partial\varphi_3}{\partial X_3}\end{pmatrix}(\boldsymbol{X},t_0;t). \tag{1.15}$$

The prove of the following technical lemma can be found in [2].

Lemma 1.2: *Let $t_0\in(0,T)$, $\mathcal{V}(t_0)$ be a bounded domain and let $\overline{\mathcal{V}(t_0)}\subset\Omega_{t_0}$. Then there exists an interval $(t_1,t_2)\ni t_0$ such that the following conditions are satisfied:*

1. *The mapping '$t\in(t_1,t_2)$, $\boldsymbol{X}\in\mathcal{V}(t_0)\longrightarrow\boldsymbol{x}=\boldsymbol{\varphi}(\boldsymbol{X},t_0;t)\in\mathcal{V}(t)$' has continuous first order derivatives with respect to t,X_1,X_2,X_3 and continuous second order derivatives $\partial^2\boldsymbol{\varphi}/\partial t\partial X_i$, $i=1,2,3$.*

2. *The mapping '$\boldsymbol{X}\in\mathcal{V}(t_0)\longrightarrow\boldsymbol{x}=\boldsymbol{\varphi}(\boldsymbol{X},t_0;t)\in\mathcal{V}(t)$' is a continuously differentiable one-to-one mapping of $\mathcal{V}(t_0)$ onto $\mathcal{V}(t)$ with the Jacobian (1.15) which is continuous and bounded and satisfies the condition*

$$J(\boldsymbol{X},t)>0\quad\forall\boldsymbol{X}\in\mathcal{V}(t_0),\ \forall t\in(t_1,t_2).$$

3. *The inclusion*

$$\left\{(\boldsymbol{x},t);\ t\in[t_1,t_2]\,,\ \boldsymbol{x}\in\overline{\mathcal{V}(t)}\right\}\subset\mathcal{M}$$

holds and thus the mapping $\boldsymbol{v}$ has continuous and bounded first order derivatives on $\{(\boldsymbol{x},t);\ t\in(t_1,t_2),\ \boldsymbol{x}\in\mathcal{V}(t)\}$.

4. $\boldsymbol{v}(\boldsymbol{\varphi}(\boldsymbol{X},t_0;t),t)=\frac{\partial\boldsymbol{\varphi}}{\partial t}(\boldsymbol{X},t_0;t)\ \forall\boldsymbol{X}\in\mathcal{V}(t_0),\ \forall t\in(t_1,t_2)$.

Next we present the lemma playing an important role in the fluid dynamics and is proved in [2].

Lemma 1.3: *Let conditions 1)-4) from Lemma 1.2 be satisfied. Then the function $J=J(\boldsymbol{X},t)$ has a continuous and bounded partial derivative $\partial J/\partial t$ for $\boldsymbol{X}\in\mathcal{V}(t_0)$, $t\in(t_1,t_2)$, and*

$$\begin{aligned}\frac{\partial J}{\partial t}(\boldsymbol{X},t) &= J(\boldsymbol{X},t)\mathrm{div}\boldsymbol{v}(\boldsymbol{x},t),\\ \boldsymbol{x} &= \boldsymbol{\varphi}(\boldsymbol{X},t_0;t).\end{aligned} \tag{1.16}$$

Now we can prove the so-called *transport theorem* using two preceding lemmas.

Theorem 1.4: *Let conditions 1)-4) from Lemma 1.2 be satisfied and let the function* $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{x}, t)$ *have continuous and bounded first order derivatives on the set* $\{(\boldsymbol{x}, t);\ t \in (t_1, t_2),\ \boldsymbol{x} \in \mathcal{V}(t)\}$. *Then for each* $t \in (t_1, t_2)$ *there exists a finite derivative*

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial t}(t) &= \frac{d}{dt}\int_{\mathcal{V}(t)} \boldsymbol{F}(\boldsymbol{x}, t)d\boldsymbol{x} \qquad (1.17)\\ &= \int_{\mathcal{V}(t)} \left[\frac{\partial \boldsymbol{F}}{\partial t}(\boldsymbol{x}, t) + \boldsymbol{v}(\boldsymbol{x}, t)\cdot \mathrm{grad}\boldsymbol{F}(\boldsymbol{x}, t) + \boldsymbol{F}(\boldsymbol{x}, t)\mathrm{div}\boldsymbol{v}(\boldsymbol{x}, t)\right] d\boldsymbol{x}\\ &= \int_{\mathcal{V}(t)} \left[\frac{\partial \boldsymbol{F}}{\partial t}(\boldsymbol{x}, t) + \mathrm{div}(\boldsymbol{F}\boldsymbol{v})(\boldsymbol{x}, t)\right] d\boldsymbol{x}. \end{aligned}$$

Proof: By the substitution theorem, the integral $\mathcal{F}(t)$ can be written in the form

$$\mathcal{F}(t) = \int_{\mathcal{V}(t_0)} \boldsymbol{F}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)J(\boldsymbol{X}, t)d\boldsymbol{X}.$$

Since t_0 is fixed and the integration domain $\mathcal{V}(t_0)$ does not depend on time t, the theorem on differentiation of an integral with respect to a parameter can be applied:

$$\frac{d\mathcal{F}}{dt}(t) = \int_{\mathcal{V}(t_0)} \left[\left(\frac{\partial \boldsymbol{F}}{\partial t}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t) + \sum_{i=1}^{3} \frac{\partial \boldsymbol{F}}{\partial x_i}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)\frac{\partial \varphi_i}{\partial t}(\boldsymbol{X}, t_0; t)\right)\cdot J(\boldsymbol{X}, t)\right] d\boldsymbol{X}$$

$$+ \int_{\mathcal{V}(t_0)} \left[\boldsymbol{F}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)\frac{\partial J}{\partial t}(\boldsymbol{X}, t)\right] d\boldsymbol{X}.$$

The assumptions considered guarantee the correctness of the differentiation under the integral sign. In view of Lemma 1.2 and relation 4) from Lemma 1.3, we get the identity

$$\frac{d\mathcal{F}}{dt}(t) = \int_{\mathcal{V}(t_0)} \left[\frac{\partial \boldsymbol{F}}{\partial t}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t) + \sum_{i=1}^{3} \frac{\partial \boldsymbol{F}}{\partial x_i}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)v_i(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)\right] J(\boldsymbol{X}, t)d\boldsymbol{X}$$

$$+ \int_{\mathcal{V}(t_0)} \left[\boldsymbol{F}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)\mathrm{div}\boldsymbol{v}(\boldsymbol{\varphi}(\boldsymbol{X}, t_0; t), t)\right] J(\boldsymbol{X}, t)d\boldsymbol{X}.$$

By using the inverse substitution, transforming the integral over $\mathcal{V}(t_0)$ onto the integral over $\mathcal{V}(t)$, we immediately obtain relation (1.17).

■

This theorem will be use in what follows. We shall introduce there the matematical formulations of fundamental physical laws: The law of conservation of mass, the law of conservation of momentum and the law of conservation of energy. From these, briefly called *conservation laws*, we will derive the fundamental differential equations of fluid dynamics: the continuity equation, the equation of motion and the energy equation.

1.3 The continuity equation

The *density of fluid* is a function

$$\rho : \mathcal{M} = \{(\boldsymbol{x}, t);\ t \in (0,T),\ \boldsymbol{x} \in \Omega_t\} \longrightarrow (0, +\infty)$$

which allows us to determine the mass $m(\mathcal{V}; t)$ of the fluid contained in any subdomain $\mathcal{V} \subset \Omega_t$:

$$m(\mathcal{V}; t) = \int_{\mathcal{V}} \rho(\boldsymbol{x}, t) d\boldsymbol{x}. \tag{1.18}$$

Let $\rho \in \mathcal{C}^1$ and $\boldsymbol{v} \in [\mathcal{C}^1(\mathcal{M})]^3$. We will consider an arbitrary time instant $t_0 \in (0,T)$ and a moving piece of fluid formed by the same particles at each instant and filling at time t_0 a bounded domain $\mathcal{V} \subset \overline{\mathcal{V}} \subset \Omega_{t_0}$, called *control volume*. The domain occupied by this piece of fluide at time $t \in (t_1, t_2)$ will be denoted by $\mathcal{V}(t)$. (t_1, t_2) is sufficiently small time interval containing t_0 with properties from Lemma 1.2. Hence, $\mathcal{V}(t_0) = \mathcal{V}$ and conditions 1)-4) from Lemma 1.2 are satisfied.

Since the domain $\mathcal{V}(t)$ is formed by the same particles at each time instant, the *concervation of mass* can be formulated in the following way:

The mass of the piece of fluid represented by the domain $\mathcal{V}(t)$ does not depend on time t.

This means that

$$\frac{dm(\mathcal{V}(t); t)}{dt} = 0, \quad t \in (t_1, t_2), \tag{1.19}$$

where with respect to (1.18) we have

$$dm(\mathcal{V}(t); t) = \int_{\mathcal{V}_t} \rho(\boldsymbol{x}, t) d\boldsymbol{x}. \tag{1.20}$$

The assumptions of the transport Theorem 1.4 are satisfied for the function $\boldsymbol{F} = \rho$. We get the identity

$$\int_{\mathcal{V}(t)} \left[\frac{\partial \rho}{\partial t}(\boldsymbol{x}, t) + \operatorname{div}(\rho \boldsymbol{v})(\boldsymbol{x}, t) \right] d\boldsymbol{x} = 0, \quad t \in (t_1, t_2). \tag{1.21}$$

If we substitute $t := t_0$ and take into account that $\mathcal{V}(t_0) = \mathcal{V}$, we conclude that

$$\int_{\mathcal{V}} \left[\frac{\partial \rho}{\partial t}(\boldsymbol{x}, t_0) + \operatorname{div}(\rho \boldsymbol{v})(\boldsymbol{x}, t_0) \right] d\boldsymbol{x} = 0 \tag{1.22}$$

for an arbitrary $t_0 \in (0,T)$ and an arbitrary control volume $\mathcal{V}$ in Ω_{t_0}.

Let us introduce following lemma which we will use to derive the differential form of the mass concervation law. The proof can be found in [3].

Lemma 1.5: *Let $\Omega \subset I\!R^N$ be an open set. Then we have:*

1. *If $f \in \mathcal{C}^0(\Omega)$, then*

$$\lim_{\operatorname{diam}(V(\boldsymbol{a})) \to 0_+} \frac{1}{|V(\boldsymbol{a})|} \int_{V(\boldsymbol{a})} f \, d\boldsymbol{x} = f(\boldsymbol{a}) \text{ for all } \boldsymbol{a} \in \Omega, \tag{1.23}$$

where $V(a) \subset \Omega$ denotes an open set containing a and

$$\operatorname{diam}(V(\boldsymbol{a})) := \sup\{|\boldsymbol{x}-\boldsymbol{y}|\,;\ \boldsymbol{x},\ \boldsymbol{y} \in V(\boldsymbol{a})\}\,.$$

2. *If $f \in \mathcal{C}^0(\Omega)$, then $f = 0$ in Ω if and only if $\int_{\mathcal{V}} f\ dx = 0$ for any open and bounded set $\mathcal{V} \subset \bar{\mathcal{V}} \subset \Omega$.*

Now, using the continuity of the integrand in (1.22) and assertion 2) of Lemma 1.5 and writing t instead of t_0, we conclude that

$$\frac{\partial \rho}{\partial t}(\boldsymbol{x},t) + \operatorname{div}(\rho(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t)) = 0, \quad t \in (0,T), \boldsymbol{x} \in \Omega_t. \tag{1.24}$$

This equation is the differential form of the law of conservation of mass and is called the *continuity equation*.

1.4 The equation of motion

We will derive basic dynamical equations describing fluid motion from the *law of conservation of momentum* which can be formulated in the following way:

The rate of change of the total momentum of a piece of fluid formed by the same particles at each time and occupying the domain $\mathcal{V}(t)$ at the instant t is equal to the force acting on $\mathcal{V}(t)$.

Let $\rho \in \mathcal{C}^1(\mathcal{M})$, $\boldsymbol{v} \in [\mathcal{C}^1(\mathcal{M})]^3$. The total momentum of particles contained in $\mathcal{V}(t)$ is given by

$$\boldsymbol{\mathcal{H}}(\mathcal{V}(t)) = \int_{\mathcal{V}(t)} \rho(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t)d\boldsymbol{x}. \tag{1.25}$$

Moreover, denoting by $\boldsymbol{\mathcal{F}}(\mathcal{V}(t))$ the force acting on the volume $\mathcal{V}(t)$, the law of conservation of momentum reads

$$\frac{d\boldsymbol{\mathcal{H}}(\mathcal{V}(t))}{dt} = \boldsymbol{\mathcal{F}}(\mathcal{V}(t)), \quad t \in (t_1,t_2). \tag{1.26}$$

Using the transport theorem, we get

$$\int_{\mathcal{V}(t)} \left[\tfrac{\partial}{\partial t}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)) + \operatorname{div}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t))\right] d\boldsymbol{x} = \mathcal{F}_i(\mathcal{V}(t)), \tag{1.27}$$
$$i = 1,2,3,\ t \in (t_1,t_2).$$

Taking into account that $t_0 \in (0,T)$ is an arbitrary time instant and $\mathcal{V}(t_0) = \mathcal{V} \subset \overline{\mathcal{V}} \subset \Omega_{t_0}$, we get the law of conservation of momentum in the form where t is written instead t_0:

$$\int_{\mathcal{V}} \left[\frac{\partial}{\partial t}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)) + \operatorname{div}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t))\right] d\boldsymbol{x} = \mathcal{F}_i(\mathcal{V};t) \tag{1.28}$$

where $i = 1, 2, 3$ for an arbitrary $t \in (0, T)$ and an arbitrary control volume $\mathcal{V}$ in Ω_t. The vector $\boldsymbol{\mathcal{F}}(\mathcal{V}; t)$ with components $\mathcal{F}_i(\mathcal{V}; t)$ denotes the force acting on the volume $\mathcal{V}$ at time t.

In the following we will characterize the vector $\boldsymbol{\mathcal{F}}(\mathcal{V}; t)$ to be able to rewrite (1.28) as a differential equation. First the distinguishing of two types of forces acting in fluids will be explained:

1. The *volume force* (also called outer or body forces) $\boldsymbol{\mathcal{F}}_v(\mathcal{V}; t)$ acting at the time t on the particles contained in a controle volume $\mathcal{V} \subset \overline{\mathcal{V}} \subset \Omega_t$ is expressed by its density (related to the unit of mass) $\boldsymbol{f} \in \mathcal{C}(\mathcal{M})^3$:

$$\boldsymbol{\mathcal{F}}_v(\mathcal{V}; t) = \int_{\mathcal{V}} \rho(\boldsymbol{x}, t) \boldsymbol{f}(\boldsymbol{x}, t) \, d\boldsymbol{x}. \tag{1.29}$$

2. The *surface force* (or inner force) $\mathbf{F}_S$, by which the fluid contained outside the domain $\mathcal{V}$ acts on a set $S \subset \partial\mathcal{V}$, is expressed with the use of the *stress vector* $\mathbf{T}(x, t, \boldsymbol{n})$ characterising the density of the surface force:

$$\mathbf{F}_S = \int_S \mathbf{T}(\boldsymbol{x}, t, \boldsymbol{n}(\boldsymbol{x})) \, dS, \tag{1.30}$$

where $\boldsymbol{n}(\boldsymbol{x})$ is the unit outer normal to $\partial\mathcal{V}$ at $\boldsymbol{x}$. We shall assume that $\mathbf{T} \in [\mathcal{C}(\mathcal{M} \times \mathcal{S}_1)]^3$, where $\mathcal{S}_1$ is surface of the unit sphere with centre at the origin. Then the total surface force acting at the time t on the control volume $\mathcal{V}$ from outside has the form

$$\boldsymbol{\mathcal{F}}_s(\mathcal{V}; t) = \int_{\partial\mathcal{V}} \mathbf{T}(\boldsymbol{x}, t, \boldsymbol{n}(\boldsymbol{x})) \, dS. \tag{1.31}$$

The stress vector $\mathbf{T}(\boldsymbol{x}, t, \boldsymbol{n})$ can be expressed with the aid of some of its values for certain normals. Let us choose the normals parallel to the coordinate axes and set

$$\tau_{ij} = T_i(\boldsymbol{x}, t, \boldsymbol{e}_j), \quad i, \ j = 1, 2, 3, \tag{1.32}$$

where $\boldsymbol{e}_j$ are unit vectors with directions of coordinate axes. The quantities $\tau_{ij} = \tau_{ij}(\boldsymbol{x}, t), \ i, \ j = 1, 2, 3$ are called the *components of the stress tensor*

$$\mathcal{T} = \begin{pmatrix} \tau_{11}, & \tau_{12}, & \tau_{13} \\ \tau_{21}, & \tau_{22}, & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix}. \tag{1.33}$$

Then

$$T_i(\boldsymbol{x}, t, \boldsymbol{n}) = \sum_{j=1}^{3} n_j \tau_{ji}(\boldsymbol{x}, t), \quad i = 1, 2, 3. \tag{1.34}$$

Assuming that $\rho,\ v_i,\ \tau_{ij} \in \mathcal{C}^1(\mathcal{M})$ and $f_i \in \mathcal{C}(\mathcal{M})$ $(i,\ j = 1,2,3)$, expressing the total force acting on the fluid contained in a control volume $\mathcal{V}$ and substituing in (1.28), we obtain

$$\int_{\mathcal{V}} \left[\frac{\partial}{\partial t}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)) + \mathrm{div}(\rho(\boldsymbol{x},t)v_i(\boldsymbol{x},t)\boldsymbol{v}(\boldsymbol{x},t))\right]\, d\boldsymbol{x} \tag{1.35}$$
$$= \int_{\mathcal{V}} \rho(\boldsymbol{x},t)f_i(\boldsymbol{x},t)\, d\boldsymbol{x} + \int_{\partial\mathcal{V}} \sum_{j=1}^{3} \tau_{ji}(\boldsymbol{x},t)n_j(\boldsymbol{x})\, dS, \quad i = 1,2,3,$$

for each $t \in (0,T)$ and an arbitrary control volume $\mathcal{V}$ in Ω_t.

Moreover, applying Green's theorem and Lemma 1.5, we get the *equation of motion of a general fluid in differential conservation form*:

$$\frac{\partial}{\partial t}(\rho v_i) + \mathrm{div}(\rho v_i \boldsymbol{v}) = \rho f_i + \sum_{j=1}^{3} \frac{\partial \tau_{ji}}{\partial x_j}, \quad i = 1,2,3. \tag{1.36}$$

This can be written as

$$\frac{\partial}{\partial t}(\rho \boldsymbol{v}) + \mathrm{div}(\rho \boldsymbol{v} \otimes \boldsymbol{v}) = \rho \boldsymbol{f} + \mathrm{div}\mathcal{T}, \quad i = 1,2,3. \tag{1.37}$$

1.5 The symmetry of the stress tensor

In this section we will demonstrate the relation between the law of conservation of the moment of momentum and the symmetry of the stress tensor.

Let us assume that $\rho,\ v_i,\ \tau_{ij} \in \mathcal{C}^1(\mathcal{M})$ and $f_i \in \mathcal{C}(\mathcal{M})$ $(i,\ j = 1,2,3)$. As above, we consider a control volume $\mathcal{V} = \mathcal{V}(t)$ formed by the same fluid particles at each time instant $t \in (t_1, t_2)$. The law of conservation of the moment of momentum can be formulated in the following way:

The rate of change of the moment of momentum of the piece of fluid occupying the volume $\mathcal{V}(t)$ at any time t is equal to the sum of the moments of the volume and surface forces acting on this volume.

Hence,

$$\frac{d}{dt}\int_{\mathcal{V}(t)} \boldsymbol{x} \times (\rho \boldsymbol{v}(\boldsymbol{x},t))\, d\boldsymbol{x} \tag{1.38}$$
$$= \int_{\mathcal{V}(t)} \boldsymbol{x} \times (\rho \boldsymbol{f}(\boldsymbol{x},t))\, d\boldsymbol{x} + \int_{\partial\mathcal{V}(t)} \boldsymbol{x} \times \mathbf{T}(\boldsymbol{x},t,\boldsymbol{n}(\boldsymbol{x}))dS.$$

Theorem 1.6: *The law of concervation of the moment of momentum (1.38) is valid if and only if the stress tensor $\mathcal{T}$ is symmetric.*

Proof: We will show that the following is valid

$$\frac{d}{dt}\int_{\mathcal{V}(t)} \boldsymbol{x} \times (\rho \boldsymbol{v})\, d\boldsymbol{x} = \int_{\mathcal{V}(t)} \boldsymbol{x} \times \boldsymbol{g}(\boldsymbol{x},t)\, d\boldsymbol{x},$$

where

$$\boldsymbol{g}_i(\boldsymbol{x},t) = \frac{\partial \rho v_i}{\partial t}(\boldsymbol{x},t) + \mathrm{div}(\rho v_i v)(\boldsymbol{x},t)$$

$$\boldsymbol{x} \times (\rho \boldsymbol{v}) = \begin{pmatrix} x_2\rho v_3 - x_3\rho v_2 \\ x_3\rho v_1 - x_1\rho v_3 \\ x_1\rho v_2 - x_2\rho v_1 \end{pmatrix}$$

Let us consider the first component and show that

$$\begin{aligned} &\frac{d}{dt}\int_{\mathcal{V}(t)} x_2\rho v_3 - x_3\rho v_2 \; d\boldsymbol{x} \\ &= \int_{\mathcal{V}(t)} \left[x_2 \left(\frac{\partial(\rho v_3)}{\partial t} + \mathrm{div}(\rho v_3 \boldsymbol{v}) \right) - x_3 \left(\frac{\partial(\rho v_2)}{\partial t} + \mathrm{div}(\rho v_2 \boldsymbol{v}) \right) \right] \; d\boldsymbol{x}. \end{aligned}$$

The use of the transport Theorem 1.4 yields

$$\begin{aligned} &\frac{d}{dt}\int_{\mathcal{V}(t)} x_2\rho v_3 - x_3\rho v_2 \; d\boldsymbol{x} \\ &= \int_{\mathcal{V}(t)} \left[\frac{\partial(x_2\rho v_3)}{\partial t} + \mathrm{div}(x_2\rho v_3 \boldsymbol{v}) - \frac{\partial(x_3\rho v_2)}{\partial t} - \mathrm{div}(x_3\rho v_2 \boldsymbol{v}) \right] \; dx \\ &= \int_{\mathcal{V}(t)} \left[\frac{\partial x_2}{\partial t}\rho v_3 + x_2\frac{\partial(\rho v_3)}{\partial t} + x_2 \mathrm{div}(\rho v_3 \boldsymbol{v}) + \nabla x_2 \cdot \rho v_2 \boldsymbol{v} \right] \; dx \\ &\quad - \int_{\mathcal{V}(t)} \left[\frac{\partial x_3}{\partial t}\rho v_2 + x_3\frac{\partial(\rho v_2)}{\partial t} + x_3 \mathrm{div}(\rho v_2 \boldsymbol{v}) + \nabla x_3 \cdot \rho v_2 \boldsymbol{v} \right] \; dx \end{aligned}$$

$$\begin{aligned} &= \int_{\mathcal{V}(t)} \left[v_2\rho v_3 + x_2\frac{\partial(\rho v_3)}{\partial t} + x_2 \mathrm{div}(\rho v_3 \boldsymbol{v}) + \rho v_3 v_2 \right] \; dx \\ &\quad - \int_{\mathcal{V}(t)} \left[v_3\rho v_2 + x_3\frac{\partial(\rho v_2)}{\partial t} + x_3 \mathrm{div}(\rho v_2 \boldsymbol{v}) + \rho v_2 v_3 \right] \; dx \\ &= \int_{\mathcal{V}(t)} \left[x_2 \left(\frac{\partial(\rho v_3)}{\partial t} + \mathrm{div}(\rho v_3 \boldsymbol{v}) \right) - x_3 \left(\frac{\partial(\rho v_2)}{\partial t} + \mathrm{div}(\rho v_2 \boldsymbol{v}) \right) \right] \; dx, \end{aligned}$$

what we wished to prove. Now, we can write the law of conservation of the moment of momentum as

$$\int_{\mathcal{V}(t)} \boldsymbol{x} \times \boldsymbol{g}(\boldsymbol{x},t) \; d\boldsymbol{x} = \int_{\mathcal{V}(t)} \boldsymbol{x} \times (\rho \boldsymbol{f})(\boldsymbol{x},t) \; d\boldsymbol{x} + \int_{\partial\mathcal{V}(t)} \boldsymbol{x} \times \mathbf{T}(\boldsymbol{x},t,\boldsymbol{n}(\boldsymbol{x})) \; dS.$$

The first component of the resulting vector is

$$\int_{\mathcal{V}(t)} (x_2 g_3 - x_3 g_2) \; d\boldsymbol{x} = \int_{\mathcal{V}(t)} (x_2\rho f_3 - x_3\rho f_2) \; d\boldsymbol{x} + \int_{\partial\mathcal{V}(t)} (x_2 \sum_{j=1}^{3} n_j\tau_{j3} - x_3 \sum_{j=1}^{3} n_j\tau_{j2}) \; dS,$$

where using Green's theorem we can replace

$$\int_{\partial\mathcal{V}(t)} (x_2 \sum_{j=1}^{3} n_j\tau_{j3} - x_3 \sum_{j=1}^{3} n_j\tau_{j2}) \; dS = \sum_{j=1}^{3} \int_{\mathcal{V}(t)} \frac{\partial}{\partial x_j}(x_2\tau_{j3} - x_3\tau_{j2}) \; d\boldsymbol{x}.$$

Hence,

$$\begin{aligned}
&\int_{\mathcal{V}(t)} [x_2(g_3 - \rho f_3) - x_3(g_2 - \rho f_2)] \; d\boldsymbol{x} \\
&= \int_{\mathcal{V}(t)} \left[x_2 \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \tau_{j3} + \tau_{j3} \sum_{j=1}^{3} \frac{\partial x_2}{\partial x_j} - x_3 \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \tau_{j2} - \tau_{j2} \sum_{j=1}^{3} \frac{\partial x_3}{\partial x_j} \right] \; d\boldsymbol{x} \\
&\int_{\mathcal{V}(t)} \left[x_2 \left(g_3 - \rho f_3 - \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \tau_{j3} \right) - x_3 \left(g_2 - \rho f_2 - \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \tau_{j2} \right) \right] \; d\boldsymbol{x} \\
&= \int_{\mathcal{V}(t)} (\tau_{23} - \tau_{32}) \; d\boldsymbol{x}.
\end{aligned}$$

Applying the law of conservation of momentum in the differential form (1.37) we get

$$0 = \int_{\mathcal{V}(t)} (\tau_{23} - \tau_{32}) \; d\boldsymbol{x}.$$

Because the control volume $\mathcal{V}$ can be chosen arbitrary, we showed that $\tau_{23} = \tau_{32}$. In the same way it can be proved for all others indices $i,\ j$. So the tensor $\mathcal{T}$ is symmetric. If we proceed in the opposite direction, we prove the law of conservation of the moment of momentum from the symmetry of tensor $\mathcal{T}$.

■

1.6 The Navier-Stokes equations

The relation between the stress tensor and other quantities describing fluid flow, particulary the velocity and its derivatives, represent the so-called *reological equations* of the fluid. The simplest rheological equation

$$\mathcal{T} = -p\mathbb{I}, \tag{1.39}$$

characterizes inviscid fluid. Here p is the pressure and $\mathbb{I}$ is the unit tensor. Besides the pressure forces, the friction shear forces also act in the real fluids as a consequence of the *viscosity*. Therefore, in the case of viscous fluid, we add a contribution $\mathcal{T}'$ characterizing the shear stress to the term $-p\mathbb{I}$:

$$\mathcal{T} = -p\mathbb{I} + \mathcal{T}'. \tag{1.40}$$

For identification of the viscous part $\mathcal{T}'$ of the stress tensor, we shall use *Stoke's postulates*:

a. $\mathcal{T} = -p\mathbb{I} + \mathcal{T}'$.

b. The tensor $\mathcal{T}'$ is a continuous function of the deformation velocity tensor,

$$\mathbb{D} = \mathbb{D}(\boldsymbol{v}) = (d_{ij})_{i,j=1}^{3}, \quad d_{ij} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right), \tag{1.41}$$

is independent of other kinematic variables and does not explicitly depend on the position in the fluid and on time either.

c. Fluid is an *isotropic* medium. This means that its properties are the same in all space directions.

d. If the deformation velocity tensor is zero, only the pressure force acts in the fluid. Hence, if $\mathbb{D} = 0$, then $\mathcal{T} = -p\mathbb{I}$.

e. The relation between $\mathcal{T}'$ and $\mathbb{D}$ is linear.

In mathematical language the above postulates can be formulated as follows:

A. $\mathcal{T} = -p\mathbb{I} + \mathcal{T}'$.

B. $\mathcal{T}' = f(\mathbb{D})$, f is continuous.

C. The form of mapping f is invariant with respect to the transformation of the Cartesian coordinate system: $\mathbb{S}\mathcal{T}'\mathbb{S}^{-1} = f(\mathbb{S}\mathbb{D}\mathbb{S}^{-1})$ for any orthonormal matrix $\mathbb{S}$.

D. $f(0) = 0$.

E. The mapping f is linear.

Then it is possible to show that the following representation holds true ([2]):

Theorem 1.7: *Under the above conditions A)-E), the stress tensor has the form*

$$\mathcal{T} = (-p + \lambda \mathrm{div}\boldsymbol{v})\mathbb{I} + 2\mu\mathbb{D}(\boldsymbol{v}), \tag{1.42}$$

where λ, μ are constants or scalar function of termodynamical quantities.

If the stress tensor depends linearly on the velocity deformation tensor as in (1.42), the fluid is called *Newtonian*, which is the case of gases.

Let us assume that $\rho \in \mathcal{C}^1(\mathcal{M})$ and $\partial\boldsymbol{v}/\partial t$ and $\partial^2\boldsymbol{v}/\partial x_i \partial x_j \in \mathcal{C}(\mathcal{M})$ $(i,\ j = 1, 2, 3)$ and substitute relation (1.42) into the general equations of motion (1.37). We get the so-called *Navier-Stokes equations*

$$\frac{\partial(\rho\boldsymbol{v})}{\partial t} + \mathrm{div}(\rho\boldsymbol{v} \otimes \boldsymbol{v}) \tag{1.43}$$

$$= \rho\boldsymbol{f} - \mathrm{grad}p + \mathrm{grad}(\lambda \mathrm{div}\boldsymbol{v}) + \mathrm{div}(2\mu\mathbb{D}(\boldsymbol{v})), \tag{1.44}$$

where λ and μ are called the first and second *viscosity coefficients*. In the kinetic theory of gases the conditions

$$\mu \geq 0, \qquad 3\lambda + 2\mu \geq 0, \tag{1.45}$$

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. Moreover, we shall assume that μ and λ are constants.

1.7 The energy equation

In this section we derive the energy equation representing the law of conservation of energy. Let us recall that the power of the force $\mathbf{F}$ acting on a particle passing through the point $\boldsymbol{x}$ at time t is

$$W(\boldsymbol{x}, t) = \mathbf{F}(\boldsymbol{x}, t) \cdot \boldsymbol{v}(\boldsymbol{x}, t). \tag{1.46}$$

We still consider a piece of fluid represented by a control volume $\mathcal{V}(t)$ satisfying assumptions from Section 1.3. The law of conservation of energy can be formulated as follows:

The rate of change of the total energy of the fluid particles occupying the domain $\mathcal{V}(t)$ at time t is equal to the sum of the volume force acting on the volume $\mathcal{V}(t)$ and the surface force acting on the surface $\partial\mathcal{V}(t)$, and of the amount of heat transmitted to $\mathcal{V}(t)$.

Let us denote by $\mathcal{E}(\mathcal{V}(t))$ the total energy of the fluid particles contained in the domain $\mathcal{V}(t)$ and by $\mathcal{Q}(\mathcal{V}(t))$ the amount of heat transmitted to $\mathcal{V}(t)$ at time t. Taking into account the character of outer and inner forces acting on the domain $\mathcal{V}(t)$, determined by the density $\boldsymbol{f}$ of the volume force and the stress vector $\mathbf{T}$, we get the identity representing the law of conservation of energy:

$$\begin{aligned} \frac{d}{dt}\mathcal{E}(\mathcal{V}(t)) \;&=\; \int_{\mathcal{V}(t)} \rho(\boldsymbol{x}, t)\boldsymbol{f}(\boldsymbol{x}, t) \cdot \boldsymbol{v}(\boldsymbol{x}, t) \\ &\quad + \int_{\partial\mathcal{V}(t)} \mathbf{T}(\boldsymbol{x}, t, \boldsymbol{n}(\boldsymbol{x})) \cdot \boldsymbol{v}(\boldsymbol{x}, t) dS + \mathcal{Q}(\mathcal{V}(t)) \end{aligned} \tag{1.47}$$

Futher, we can write

$$\begin{aligned} &\text{a) } \mathcal{E}(\mathcal{V}(t)) = \int_{\mathcal{V}(t)} E(\boldsymbol{x}, t) d\boldsymbol{x}, \\ &\text{b) } E = \rho\left(e + \frac{|\boldsymbol{v}|^2}{2}\right), \\ &\text{c) } \mathcal{Q}(\mathcal{V}(t)) = \int_{\mathcal{V}(t)} \rho(\boldsymbol{x}, t) q(\boldsymbol{x}, t) d\boldsymbol{x} - \int_{\partial\mathcal{V}(t)} \mathbf{q}(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) dS. \end{aligned} \tag{1.48}$$

Here E is the total energy, e is the specific internal energy (i.e. per unit mass) associated with molecular and atomic behaviour, $|\boldsymbol{v}|^2/2$ is the density of the kinetic energy,

q represents the density of heat source (related to unit mass) and $\mathbf{q}$ is the heat flux. By virtue of the so-called *Fourier's law,*

$$\mathbf{q} = -k \ \mathrm{grad}\theta, \tag{1.49}$$

so that

$$\int_{\partial\mathcal{V}(t)} \mathbf{q}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x})dS = \int_{\partial\mathcal{V}(t)} k(\boldsymbol{x},t)\frac{\partial\theta(\boldsymbol{x},t)}{\partial\boldsymbol{n}}dS, \tag{1.50}$$

where k is the heat conduction coefficient and θ is the absolute temperature. From the second law of termodynamics, which will be presented later, it can be proven that $k \geq 0$. We shall assume that k is constant.

Substituting (1.34) and (1.48), a-c) into (1.47), we get

$$\begin{aligned}\frac{d}{dt}\int_{\mathcal{V}(t)} E(\boldsymbol{x},t)d\boldsymbol{x} &= \int_{\mathcal{V}(t)} \rho(\boldsymbol{x},t)\boldsymbol{f}(\boldsymbol{x},t)\cdot\boldsymbol{v}(\boldsymbol{x},t)d\boldsymbol{x} \\ &+ \int_{\partial\mathcal{V}(t)} \sum_{i,j=3}^{3} \tau_{ji}(\boldsymbol{x},t)n_j(\boldsymbol{x})v_i(\boldsymbol{x},t)dS \\ &+ \int_{\mathcal{V}(t)} \rho(\boldsymbol{x},t)q(\boldsymbol{x},t)d\boldsymbol{x} - \int_{\partial\mathcal{V}(t)} \mathbf{q}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x})dS.\end{aligned} \tag{1.51}$$

We still assume some smoothness of functions describing the flow. Let $\rho,\ v_i,\ \tau_{ij},\ \mathrm{q}_i \in \mathcal{C}^1(\mathcal{M})$ and $f_i,\ q \in \mathcal{C}(\mathcal{M})$ $(i,j = 1,2,3)$. By virtue of the transport theorem 1.4, Green's theorem and Lemma 1.5, we derive from (1.51) the *energy equation* written in the differential conservative form:

$$\frac{\partial E}{\partial t} + \mathrm{div}(E\boldsymbol{v}) = \rho\boldsymbol{f}\cdot\boldsymbol{v} + \mathrm{div}\mathcal{T}\boldsymbol{v} + \rho q - \mathrm{div}\mathbf{q}. \tag{1.52}$$

For a *Newtonian fluid* we have

$$\begin{aligned}\frac{\partial E}{\partial t} + \mathrm{div}(E\boldsymbol{v}) &= \rho\boldsymbol{f}\cdot\boldsymbol{v} - \mathrm{p}\boldsymbol{v} + \mathrm{div}(\lambda\boldsymbol{v}\ \mathrm{div}\boldsymbol{v}) \\ &+ \mathrm{div}(2\mu\mathbb{D}(\boldsymbol{v})\boldsymbol{v}) + \rho q - \mathrm{div}\mathbf{q}.\end{aligned} \tag{1.53}$$

The system given by the equations (1.24), (1.43), (1.53) forms the so-called *full system of equations of a Newtonian fluid.*

1.8 Thermodynamical relations

In order to complete the conservation law system, additional equations derived in thermodynamics have to be included.

The absolute temperature θ, the density ρ and the pressure p are called the *state variables.* All these quantities are positive functions. The gas is characterized by the equation of state

$$p = p(\rho,\theta) \tag{1.54}$$

and

$$e = e(\rho,\theta). \tag{1.55}$$

On the basis of these equations it is possible to express p and θ as functions of e and ρ :

$$p = p(e, \rho), \tag{1.56}$$
$$\theta = \theta(e, \rho). \tag{1.57}$$

Very often the so-called *perfect gas* (also called ideal gas) is considered and then state variables satisfy the equation of state in the form

$$p = R\theta\rho. \tag{1.58}$$

$R > 0$ is the *gas constant*, which can be expressed in the form

$$R = c_p - c_v, \tag{1.59}$$

where c_p and c_v denote the *specific heat at constant pressure* and the *specific heat at constant volume*, respectively. From experiments we know that $c_p > c_v$, so that $R > 0$. c_p and c_v can be considered constant, which is assumed for perfect gases. This is true for a relatively large range of temperature as showed experiments. The quantity

$$\gamma = \frac{c_p}{c_v} > 1 \tag{1.60}$$

is called the *Poisson adiabatic constant.*

The internal energy of a perfect gas is given by

$$e = c_v\theta. \tag{1.61}$$

1.8.1 Entropy

One of the important thermodynamical quantities is the entropy S, defined by the relation

$$\theta dS = de + pdV, \tag{1.62}$$

where $V = 1/\rho$ is so-called specific volume. This identity is derived in thermodynamics under the assumption that the internal energy is a function of S and $V : e = e(S, V)$, which explains the meaning of the differentials in (1.62).

Theorem 1.8: *For a perfect gas we have*

$$S = c_v \ln \frac{p/p_0}{(\rho/\rho_0)^\gamma} + const \tag{1.63}$$
$$= c_v \ln \frac{\theta/\theta_0}{(\rho/\rho_0)^{\gamma-1}} + const,$$

where p_0 and ρ_0 are fixed (reference) values of pressure and density,respectively, and $\theta_0 = p_0/(R\rho_0)$.

Proof: Using (1.61) and the relation $V = 1/\rho$, we can write (1.62) in the form

$$\theta dS = c_v d\theta - \frac{p}{\rho^2 d\rho}. \tag{1.64}$$

From this and (1.58)-(1.60) we obtain

$$dS = c_v \frac{d\theta}{\theta} - \frac{p}{\rho\theta}\frac{d\rho}{\rho} = c_v \frac{d(p/\rho)}{(p/\rho)} - R\frac{d\rho}{\rho} = c_v d\ln\frac{p/p_0}{(\rho/\rho_0)^\gamma} = c_v d\ln\frac{\theta/\theta_0}{(\rho/\rho_0)^{\gamma-1}},$$

which immediately yields (1.63).

■

In the flow considered is a reversible process, which means that the system is in equilibrium with the surrounding medium at each time instant, the *first law of thermodynamics* is valid:

$$\delta\mathcal{Q} = de + pdV,^1 \tag{1.65}$$

where $\delta\mathcal{Q}$ is the elementary heat transmittion (related to unit of mass). This means that the heat transmitted to the system is equal to the sum of the energy increment and the elementary work performed on the system by the pressure force. From this and (1.63) we find

$$dS = \frac{\delta\mathcal{Q}}{\theta}. \tag{1.66}$$

1.8.2 The second law of thermodynamics

In the irreversible processes, equality (1.63) does not hold in general and is replaced by the inequality

$$dS \geq \frac{\delta\mathcal{Q}}{\theta} \tag{1.67}$$

called the *second law of thermodynamics.* For a system of fluid particles occupying a domain $\mathcal{V}(t)$ at time t we postulate the second law of thermodynamics mathematically in the form

$$\frac{d}{dt}\int_{\mathcal{V}(t)} \rho(\boldsymbol{x},t)S(\boldsymbol{x},t)d\boldsymbol{x} \geq \int_{\mathcal{V}(t)} \frac{\rho(\boldsymbol{x},t)q(\boldsymbol{x},t)}{\theta(\boldsymbol{x},t)}d\boldsymbol{x} - \int_{\partial\mathcal{V}(t)} \frac{\mathbf{q}(\boldsymbol{x},t)\cdot\boldsymbol{n}(\boldsymbol{x})}{\theta(\boldsymbol{x},t)}dS, \tag{1.68}$$

where q and $\mathbf{q}$ denote the density of heat source and the heat flux. The left-hand side of (1.68) represents the rate of the entropy contained in the volume $\mathcal{V}(t)$, and the first and second integral of the right-side are called the *entropy production* and the *entropy flux.* Let $\rho,\ \theta,\ v_i,\ \mathrm{q}_i \in \mathcal{C}^1(\mathcal{M}),\ q,\ f_i \in \mathcal{C}(\mathcal{M}),\ i = 1,2,3.$ By virtue of the transport Theorem 1.4 and the continuity equation (1.24), from (1.68) we obtain the inequality

$$\rho\frac{\mathrm{D}S}{\mathrm{D}t} \geq \frac{\rho q}{\theta} - \mathrm{div}\left(\frac{\mathbf{q}}{\theta}\right). \tag{1.69}$$

[1]We use the symbol $\delta\mathcal{Q}$, because the elementary heat transmittion depends of the transition of the system from one state to another and, therefore, it cannot be expressed as the differential $d\mathcal{Q}$.

1.9 Description of the flow of a heat-conductive gas

For description of a real heat conductive gas it is necessary to use the system consisting of the continuity equation, the Navier-Stokes equations, the energy equation and thermodynamical relations. Using equations (1.47) b), (1.56) and (1.57), the complete system for a perfect gas reads:

$$\rho_t + \mathrm{div}(\rho \boldsymbol{v}) = 0, \tag{1.70}$$

$$(\rho \boldsymbol{v})_t + \mathrm{div}(\rho \boldsymbol{v} \otimes \boldsymbol{v}) = \rho \boldsymbol{f} - \nabla p + \nabla(\lambda \mathrm{div} \boldsymbol{v}) + \mathrm{div}(2\mu \mathbb{D}(\boldsymbol{v})), \tag{1.71}$$

$$E_t + \mathrm{div}(E\boldsymbol{v}) = \rho \boldsymbol{f} \cdot \boldsymbol{v} - \mathrm{div}(p\boldsymbol{v}) + \mathrm{div}(\lambda \boldsymbol{v} \mathrm{div} \boldsymbol{v}) \\ + \mathrm{div}(2\mu \mathbb{D}(\boldsymbol{v})\boldsymbol{v}) + \rho q + \mathrm{div}(k \nabla \theta), \tag{1.72}$$

$$p = (\gamma - 1)(E - \rho |\boldsymbol{v}|^2 /2), \tag{1.73}$$

$$\theta = (E/\rho - \rho |\boldsymbol{v}|^2 /2)/c_v, \tag{1.74}$$

where the velocity deformation tensor $\mathbb{D}(\boldsymbol{v})$ is expressed in (1.41). We simply call this system the *compressible Navier-Stokes equations* for a heat conductive gas.

If we set $\mu = \lambda = k = 0$, we obtain the model of inviscid compressible flow, described by the continuity equation, the Euler equations, the energy equation and thermodynamical relations. Since gases are light, usually it is possible to neglect the effect of the volume force. Neglecting heat sources also, we get the system for the perfect inviscid gas

$$\rho_t + \mathrm{div}(\rho \boldsymbol{v}) = 0, \tag{1.75}$$

$$(\rho \boldsymbol{v})_t + \mathrm{div}(\rho \boldsymbol{v} \otimes \boldsymbol{v}) + \nabla p = 0, \tag{1.76}$$

$$E_t + \mathrm{div}((E + p)\boldsymbol{v}) = 0, \tag{1.77}$$

$$p = (\gamma - 1)(E - \rho |\boldsymbol{v}|^2 /2). \tag{1.78}$$

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

Chapter 2

The Euler equations

In this chapter we shall be concerned with some special properties of the Euler equations important for the construction of numerical schemes which will be presented it futher chapters.

2.1 Properties

Let us consider the unsteady flow of an inviscid gas in a domain $\Omega \subset I\!R^N$ $(1 \leq N \leq 3)$. It is governed by the continuity equation, the Euler equations of motion and the energy equation, to which we add closing thermodynamical relations. We neglect the heat transfer thus we assume *adiabatic* flow. Moreover, we neglect the outer volume force, because the gas is light. We shall be concerned with the flow of a perfect gas, for which the equation of state has the form

$$p = R\rho\theta, \tag{2.1}$$

where R is the gas constant. The system of governing equations presented in the previous chapter (see (1.75)-(1.78)) considered in the space-time cylinder $Q_T = \Omega \times (0,T)$ can be written in the form

$$\frac{\partial \rho}{\partial t} + \sum_{s=1}^{N} \frac{\partial(\rho v_s)}{\partial x_s} = 0, \tag{2.2}$$

$$\frac{\partial(\rho v_i)}{\partial t} + \sum_{s=1}^{N} \frac{\partial(\rho v_i v_s + \delta_{is} p)}{\partial x_s} = 0, \quad i = 1, \ldots, N, \tag{2.3}$$

$$\frac{\partial E}{\partial t} + \sum_{s=1}^{N} \frac{\partial((E+p)v_s)}{\partial x_s} = 0, \tag{2.4}$$

$$p = (\gamma - 1)(E - \rho \left|\boldsymbol{v}\right|^2 /2). \tag{2.5}$$

Here v_s are the components of the velocity vector $\boldsymbol{v} = (v_1, \ldots, v_N)^T$ in the directions x_s $(s = 1, \ldots, 3)$, ρ is density, p is the pressure, E is the total energy, i.e.

$$E = \rho(c_v\theta + \left|\boldsymbol{v}\right|^2 /2), \tag{2.6}$$

and θ is the absolute temperature. For a perfect gas we assume that the specific heat c_v at constant volume is a constant.

System (2.2)-(2.5) can be written as

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{s=1}^{N} \frac{\partial \boldsymbol{f_s}(\boldsymbol{w})}{\partial x_s} = 0, \tag{2.7}$$

where

$$\boldsymbol{w} = (w_1, \ldots, w_m)^T = (\rho, \rho v_1, \ldots, \rho_N, E)^T \in I\!R^m \quad m = N + 2, \tag{2.8}$$

is the so-called *state vector,* and

$$\begin{aligned} \boldsymbol{f}_s(\boldsymbol{w}) &= (f_{s1}(\boldsymbol{w}), \ldots, f_{sm}(\boldsymbol{w}))^T \\ &= (\rho v_s, \rho v_1 v_s + \delta_{1s} p, \ldots, \rho v_N v_s + \delta_{Ns} p, (E+p)v_s)^T \\ &= \Bigg(w_{s+1}, w_2 w_{s+1}/w_1 + \delta_{1,s}(\gamma - 1)(w_m - \sum_{i=2}^{m-1} w_i^2/(2w_1)), \ldots, \\ &\quad w_{m-1} w_{s+1}/w_1 + \delta_{m-2,s}(\gamma - 1)(w_m - \sum_{i=2}^{m-1} w_i^2/(2w_1)), \\ &\quad w_{s+1}(\gamma w_m - (\gamma - 1)\sum_{i=2}^{m-1} w_i^2/(2w_1)/w_1 \Bigg)^T \end{aligned} \tag{2.9}$$

is the *flux* of the quantity $\boldsymbol{w}$ in the direction x_s. Often, $\boldsymbol{f}_s, s = 1, \ldots, N$, are called *inviscid Euler fluxes.* Usually, system (2.2)-(2.5)(i.e. (2.7)) is called the system of the Euler equations, or simply Euler equations. The functions $\rho, v_1, \ldots, v_N, p$ are called *primitive variables*, whereas $w_1 = \rho$, $w_2 = \rho v_1, \ldots,$ $w_{m-1} = \rho v_N$, $w_m = E$ are *conservative variables.* Sometimes $\rho,\ v_1, \ldots, v_N, \theta$ are called *physical variables.* The domain of definition of the vector-valued functions $\boldsymbol{f}_s$ is the open set $D \subset I\!R^m$ of vectors $\boldsymbol{w} = (w_1, \ldots, w_m)^T$ such that the corresponding density and pressure are positive:

$$\begin{aligned} D = \{ \boldsymbol{w} \in I\!R^m;\ w_1 = \rho > 0,\ w_s = \rho v_{s-1} \in I\!R \text{ for } s = 2, \ldots, m-1, \\ w_m - \sum_{i=2}^{m-1} w_i^2/(2w_1) = p/(\gamma - 1) > 0 \Bigg\}. \end{aligned} \tag{2.10}$$

Obviously, $\boldsymbol{f}_s \in \mathcal{C}^1(D)^m$.

Differentiation in (2.7) and the use of the chain rule lead to a *first order quasilinear system* of partial diferential equations

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{s=1}^{N} \mathbb{A}_s(w) \frac{\partial \boldsymbol{w}}{\partial x_s} = 0, \tag{2.11}$$

where $\mathbb{A}_s(w)$ are $m \times m$ matrices defined for $\boldsymbol{w} \in D$ by

$$\mathbb{A}_s(w) = \frac{D\boldsymbol{f}_s(w)}{D\boldsymbol{w}} = \left(\frac{\partial f_{si}(\boldsymbol{w})}{\partial w_j} \right)_{i,j=1}^{m}. \tag{2.12}$$

So $\mathbb{A}_s(w)$ is the Jacobi matrix of the mapping $\boldsymbol{f}_s$. For $\boldsymbol{w} \in D$ and $\boldsymbol{n} = (n_1, \ldots, n_N)^T \in I\!R^N$ we denote

$$\mathcal{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{N} \boldsymbol{f}_s(\boldsymbol{w}) n_s, \tag{2.13}$$

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

$$\frac{D\mathcal{P}(\boldsymbol{w}, \boldsymbol{n})}{D\boldsymbol{w}} = \mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{N} \mathbb{A}_s(\boldsymbol{w}) n_s. \tag{2.14}$$

In the following lemma we will investigate some properties of the system (2.7) of the Euler equations.

Lemma 2.1: *The vector-valued functions $\boldsymbol{f}_s$ defined by (2.9) are homogeneous mapping of order 1:*

$$\boldsymbol{f}_s(\alpha \boldsymbol{w}) = \alpha \boldsymbol{f}_s(\boldsymbol{w}), \quad \alpha > 0. \tag{2.15}$$

Moreover, we have

$$\boldsymbol{f}_s(\boldsymbol{w}) = \mathbb{A}_s(\boldsymbol{w}) \boldsymbol{w}. \tag{2.16}$$

Similary,

$$\mathcal{P}(\alpha \boldsymbol{w}, \boldsymbol{n}) = \alpha \mathcal{P}(\boldsymbol{w}, \boldsymbol{n}), \quad \alpha > 0, \tag{2.17}$$

$$\mathcal{P}(\boldsymbol{w}, \boldsymbol{n}) = \mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) \boldsymbol{w}. \tag{2.18}$$

Proof: Relation (2.15) immediately follows from (2.9). Since $\boldsymbol{f}_s \in \mathcal{C}^1(D)^m$, the expression $(D\boldsymbol{f}_s(\boldsymbol{w})/D\boldsymbol{w})\boldsymbol{w} = \mathbb{A}_s(\boldsymbol{w})\boldsymbol{w}$ is the derivative of $\boldsymbol{f}_s$ in the direction $\boldsymbol{w}$ at the point $\boldsymbol{w}$. By the definition of the derivative and (2.15),

$$\begin{aligned} \mathbb{A}_s(\boldsymbol{w})\boldsymbol{w} &= \lim_{\alpha \to 0} \frac{\boldsymbol{f}_s(\boldsymbol{w} + \alpha \boldsymbol{w}) - \boldsymbol{f}_s(\boldsymbol{w})}{\alpha} \\ &= \lim_{\alpha \to 0} \frac{(1+\alpha)\boldsymbol{f}_s(\boldsymbol{w}) - \boldsymbol{f}_s(\boldsymbol{w})}{\alpha} = \boldsymbol{f}_s(\boldsymbol{w}). \end{aligned} \tag{2.19}$$

Relations (2.17) and (2.18) are consequences of the definitions of $\mathcal{P}$ and $\mathbb{P}$ and the above results.

■

2.2 The 2D case

The main results obtained for 2D flow are summarized in the following lemma which will not be proved. (See [4].)

Lemma 2.2: *If $N = 2$, then*

$$\begin{aligned} \boldsymbol{f}_1(\boldsymbol{w}) &= \big(w_2, w_2^2/w_1 + (\gamma-1)\left[w_4 - (w_3^2+w_3^2)/(2w_1)\right], \\ &\quad w_2 w_3/w_1, w_2\left[\gamma w_4 - (\gamma-1)(w_3^2+w_3^2)/(2w_1)\right]/w_1\big)^T \end{aligned} \tag{2.20}$$

and, with the notation $\boldsymbol{v} = (u, v)$,

$$\mathbb{A}_1(\boldsymbol{w}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{\gamma-1}{2}|\boldsymbol{v}|^2 - u^2 & (3-\gamma)u & (1-\gamma)v & 1-\gamma \\ -uv & v & u & 0 \\ u((\gamma-1)|\boldsymbol{v}|^2 - \gamma\frac{E}{\rho}) & \gamma\frac{E}{\rho} - (\gamma-1)u^2 - \frac{\gamma-1}{2}|\boldsymbol{v}|^2 & (1-\gamma)uv & \gamma u \end{pmatrix}. \tag{2.21}$$

The matrix $\mathbb{A}_1(\boldsymbol{w})$ has the eigenvalues

$$\tilde{\lambda}_1 = u - a, \quad \tilde{\lambda}_2 = \tilde{\lambda}_3 = u, \quad \tilde{\lambda}_4 = u + a \tag{2.22}$$

and the corresponding eigenvectors

$$\begin{aligned} \boldsymbol{r}_1(\boldsymbol{w}) &= (1, u-a, v, |\boldsymbol{v}|^2/2 + a^2(\gamma-1) - ua)^T, \\ \boldsymbol{r}_2(\boldsymbol{w}) &= (1, u, v, |\boldsymbol{v}|^2/2)^T, \\ \boldsymbol{r}_3(\boldsymbol{w}) &= (1, u, v-a, |\boldsymbol{v}|^2/2 - va)^T, \\ \boldsymbol{r}_4(\boldsymbol{w}) &= (1, u+a, v, |\boldsymbol{v}|^2/2 + a^2/(\gamma-1) + ua)^T. \end{aligned} \tag{2.23}$$

We have

$$\tilde{\mathbb{T}}^{-1}(\boldsymbol{w})\mathbb{A}_1(\boldsymbol{w})\tilde{\mathbb{T}}(\boldsymbol{w}) = \tilde{\mathbb{\Lambda}}(\boldsymbol{w}), \tag{2.24}$$

where

$$\tilde{\mathbb{\Lambda}}(\boldsymbol{w}) = \mathrm{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_4), \tag{2.25}$$

$$\tilde{\mathbb{T}}(\boldsymbol{w}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ u-a & u & u & u+a \\ v & v & v-a & v \\ \frac{|\boldsymbol{v}|^2}{2} + \frac{a^2}{\gamma-1} - ua & \frac{|\boldsymbol{v}|^2}{2} & \frac{|\boldsymbol{v}|^2}{2} - va & \frac{|\boldsymbol{v}|^2}{2} + \frac{a^2}{\gamma-1} + ua \end{pmatrix} \tag{2.26}$$

and

$$\tilde{\mathbb{T}}^{-1}(\boldsymbol{w}) = \begin{pmatrix} \frac{1}{2}\left(\frac{(\gamma-1)|\boldsymbol{v}|^2}{2} + ua\right) & -\frac{a+u(\gamma-1)}{2} & -\frac{v(\gamma-1)}{2} & \frac{\gamma-1}{2} \\ a^2 - va - (\gamma-1)\frac{|\boldsymbol{v}|^2}{2} & u(\gamma-1) & a + v(\gamma-1) & 1-\gamma \\ va & 0 & -a & 0 \\ \frac{1}{2}\left(\frac{(\gamma-1)|\boldsymbol{v}|^2}{2} - ua\right) & \frac{a-u(\gamma-1)}{2} & -\frac{v(\gamma-1)}{2} & \frac{\gamma-1}{2} \end{pmatrix}. \tag{2.27}$$

The *rotational invariance* of Euler equations is represented by the relations

$$\mathcal{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) n_s = \mathbb{Q}^{-1}(\boldsymbol{n}) \boldsymbol{f}_1(\mathbb{Q}(\boldsymbol{n})\boldsymbol{w}), \tag{2.28}$$

$$\mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \mathbb{A}_s(\boldsymbol{w}) n_s = \mathbb{Q}^{-1}(\boldsymbol{n}) \mathbb{A}_1(\mathbb{Q}(\boldsymbol{n})\boldsymbol{w}) \mathbb{Q}(\boldsymbol{n}),$$

$$\boldsymbol{n} = (n_1, n_2) \in I\!R, \ |\boldsymbol{n}| = 1, \ \boldsymbol{w} \in D,$$

where

$$\mathbb{Q}(\boldsymbol{n}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & n_1 & n_2 & 0 \\ 0 & -n_2 & n_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.29}$$

Definition 2.1: *The system (2.11) is called hyperbolic in the domain $D \subset \mathbb{R}^4$, if the matrix $\mathbb{P}(\boldsymbol{w}, \boldsymbol{n})$ defined by (2.14) has only real eigenvalues for each $\boldsymbol{n} = (n_1, n_2)^T \in \mathbb{R}^2$ and $\boldsymbol{w} \in D$. The system is diagonally hyperbolic if the matrix $\mathbb{P}(\boldsymbol{w}, \boldsymbol{n})$ can be diagonalized. This means that there exists a nonsingular matrix $\mathbb{T} = \mathbb{T}(\boldsymbol{w}, \boldsymbol{n})$ such that*

$$\mathbb{T}^{-1}\mathbb{P}\mathbb{T} = \mathbb{\Lambda} = \mathbb{\Lambda}(\boldsymbol{w}, \boldsymbol{n}) = \operatorname{diag}(\lambda_1, \ldots, \lambda_4) = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}. \tag{2.30}$$

Theorem 2.3: *The 2D Euler equations form a diagonally hyperbolic system in the sense of Definition 2.1.*
The proof can be found in [4].

At the end it is useful to add that for the next computations we need express the gas equation with the aid of the state vector. We can see that the velocity, density and energy can be written by the state vector: $\rho = w_1$, $v_1 = w_2/w_1$, $v_2 = w_3/w_1$ and $E = w_4$. Using the equation (2.5) we get the gas equation

$$p = (\gamma - 1)\left(w_4 - \frac{1}{2}\left(\frac{w_2^2}{w_1} + \frac{w_3^2}{w_1}\right)\right). \tag{2.31}$$

Chapter 3

The Euler equations in time-dependent domains

This chapter is concerned with modelling of flow in domains with moving boundaries. The Arbitrary Eulerian-Lagrangian (ALE) method will be presented and its two possible applications to two-dimensional Euler equations will be discussed. We assume that $(0,T)$ with $T>0$ is a time interval and by Ω_t we denote a computational domain occupied by the fluid at time t. The set $\bar{\Omega}_{ref}\times(0,T)=\{(\boldsymbol{x},t);\ \boldsymbol{x}\in\Omega_t,\ t\in(0,T)\}$ is called the space-time cylinder.

3.1 ALE method

In order to simulate flow in a time-dependent domain, we employ the Arbitrary Eulerian-Lagrangian 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.* A smooth, one-to-one mapping of the reference configuration onto the computational domain Ω_t at time t (the so-called *current configuration*) is denoted by $\mathcal{A}_t$, i.e.

$$\begin{aligned}\mathcal{A}_t : \bar{\Omega}_{ref} &\longrightarrow \bar{\Omega}_t,\\ \boldsymbol{X} &\longmapsto \boldsymbol{x}(\boldsymbol{X},t)=\mathcal{A}_t(\boldsymbol{X}).\end{aligned} \tag{3.1}$$

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:

$$\begin{aligned}\tilde{\boldsymbol{z}} : \bar{\Omega}_{ref}\times(0,T) &\longrightarrow I\!R^2,\\ \tilde{\boldsymbol{z}}(\boldsymbol{X},t)=\tfrac{\partial}{\partial t}\boldsymbol{x}(\boldsymbol{X},t)&=\tfrac{\partial}{\partial t}\mathcal{A}_t(\boldsymbol{X}),\end{aligned} \tag{3.2}$$

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

$$\boldsymbol{z}=\tilde{\boldsymbol{z}}\circ\mathcal{A}_t^{-1}, \tag{3.3}$$

$$\text{i.e.}\quad \boldsymbol{z}=\tilde{\boldsymbol{z}}(\mathcal{A}_t^{-1}(\boldsymbol{x}),t),\ t\in(0,T),\ \boldsymbol{x}\in\bar{\Omega}_t. \tag{3.4}$$

With the aid of the ALE mapping we introduce the so-called *ALE derivative* $\frac{D^{\mathcal{A}}}{Dt}$ for a function $f:\bar{\Omega}_{ref}\times(0,T)\to I\!R$. We set

$$\frac{D^{\mathcal{A}}}{Dt}f(\boldsymbol{x},t)=\frac{\partial\tilde{f}}{\partial t}(\boldsymbol{X},t),\quad \boldsymbol{X}=\mathcal{A}_t^{-1}(\boldsymbol{x}), \tag{3.5}$$

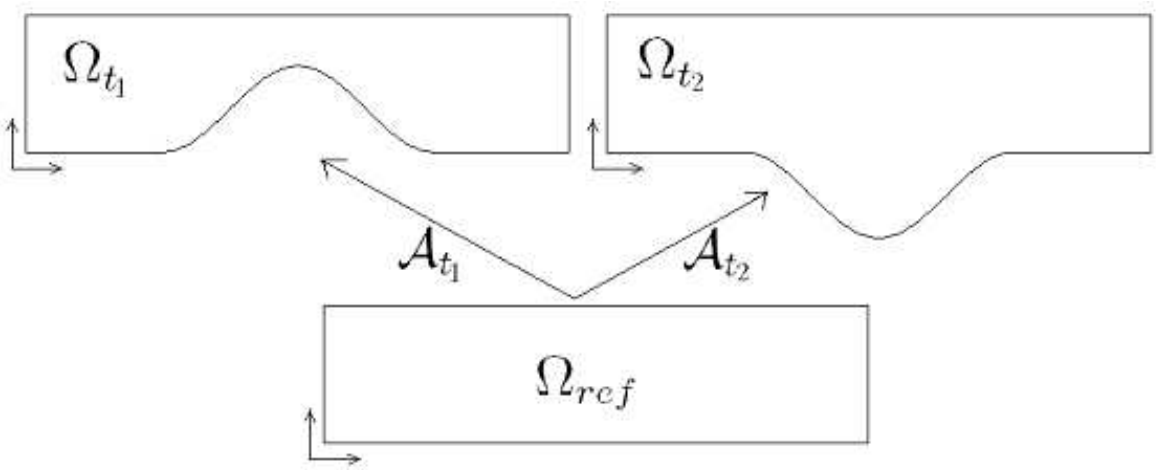


Figure 3.1: ALE mapping

where $\tilde{f} = f \circ \mathcal{A}_t$, i.e. $\tilde{f}(\boldsymbol{X}, t) = f(\mathcal{A}_t(\boldsymbol{X}), t),\ \boldsymbol{X} \in \Omega_{ref},\ t \in (0, T)$. Using the chain rule and assuming smooth f, we find that

$$\frac{D^{\mathcal{A}}}{Dt} f = \frac{\partial f}{\partial t} + (\boldsymbol{z} \cdot \nabla) f, \tag{3.6}$$

$$\frac{D^{\mathcal{A}}}{Dt} f = \frac{\partial f}{\partial t} + \mathrm{div}(f\boldsymbol{z}) - f\mathrm{div}(\boldsymbol{z}). \tag{3.7}$$

It follows from (3.6)-(3.7) that the time derivative of a function f can be expressed in the form

$$\frac{\partial f}{\partial t} = \frac{D^{\mathcal{A}}}{Dt} f - (\boldsymbol{z} \cdot \nabla) f, \tag{3.8}$$

$$\frac{\partial f}{\partial t} = \frac{D^{\mathcal{A}}}{Dt} f + f\mathrm{div}(\boldsymbol{z}) - \mathrm{div}(f\boldsymbol{z}). \tag{3.9}$$

3.2 ALE description of the Euler equations

This section deals with two different formulations of the two-dimensional Euler equations implementing the ALE method. The diffence of this formulations is based on a different expressions of the ALE derivative (see (3.6) and (3.7)).

3.2.1 Formulation I

The system of governing equations for inviscide compressible flow was presented in Section 1.9 (see (1.75)-(1.78)). In Section 2.1 we rewrote it in a more convenient way (see (2.2)-(2.5)) and reformulated it using the state vector $\boldsymbol{w}$ and the flux $\boldsymbol{f}_s(\boldsymbol{w})$ of $\boldsymbol{w}$ in the direction $x_s,\ s = 1,\ 2$ (see (2.7)-(2.9)). Now, applying the ALE derivative in the first form (3.6) and using the corresponding relation (3.8), for the state vector $\boldsymbol{w}$ we get the system of Euler equations in the ALE formulation

$$\frac{D^{\mathcal{A}}\boldsymbol{w}}{Dt} + \sum_{s=1}^{2} \frac{\partial \boldsymbol{f}_s(\boldsymbol{w})}{\partial x_s} - \sum_{s=1}^{2} z_s \frac{\partial \boldsymbol{w}}{\partial x_s} = 0. \tag{3.10}$$

3.2.2 Formulation II

Using the relation (3.9), we obtain the Euler equations in the form

$$\frac{D^{\mathcal{A}}\boldsymbol{w}}{Dt} + \sum_{s=1}^{2} \frac{\partial \boldsymbol{g}_s(\boldsymbol{w})}{\partial x_s} + \boldsymbol{w}\mathrm{div}\boldsymbol{z} = 0, \tag{3.11}$$

where $\boldsymbol{g}_s,\ s = 1,\ 2$, is the ALE flux of $\boldsymbol{w}$ in the direction x_s defined as

$$g_s(\boldsymbol{w}) = f_s(\boldsymbol{w}) - z_s\boldsymbol{w}. \tag{3.12}$$

Both formulation I and II will be used in our numerical simulations of flow in domains with moving boundaries.

Chapter 4

Discretization of compressible flow problem by the discontinuous Galerkin method

In this chapter we shall be concerned with the discontinuous Galerkin finite element method (DGFEM) applied to the solution of inviscid compressible flow. The discretization of the Euler equations in both time-independent and time-dependent domains will be described.

4.1 Discretization in the time-independent domain

We shall be concerned with inviscid compressible two-dimensional flow. Let $T > 0$, $\Omega \subset \mathbb{R}^2$ and Q_T be the same as in Section 2.1. Futhermore, we define disjoint boundary components $\Gamma_I,\ \Gamma_O,\ \Gamma_W$, the *inlet, outlet* and *impermeable wall*, respectively, such that $\partial\Omega = \Gamma_I \cup \Gamma_O \cup \Gamma_W$. We set $\Gamma_{IO} = \Gamma_I \cup \Gamma_O$. For the detailed description of the system of Euler equations for inviscid compressible flow see Chapter 2.

4.1.1 Space semidiscretization

Let Ω_h be a polygonal approximation of Ω. By T_h we denote a partition of Ω_h consisting of elements $K_i \in T_h,\ i \in I$, e.g. triangles or quadrilaterals. ($I \subset \mathbb{Z}^+ = \{0,\ 1,\ 2,\ \ldots\}$ is a suitable index set.) By Γ_{ij} we denote a common edge between two neighbouring elements K_i and K_j. We set $s(i) = \{j \in I;\ K_j \text{ is a neighbour of } K_i.\}$ The boundary $\partial\Omega_h$ is formed by a finite number of faces of elements K_i adjecent to $\partial\Omega_h$. We denote all these boundary faces by S_j, where $j \in I_b \subset \mathbb{Z}^- = \{-1,\ -2,\ \ldots.\}$ Now we set $\gamma(i) = \{j \in I_b;\ S_j \text{ is a face of } K_i \in T_h\}$ and $\Gamma_{ij} = S_j$ for $K_i \in T_h$ such that $S_j \subset \partial K_i,\ j \in I_b$. For K_i not containing any boundary face S_j we set $\gamma(i) = \emptyset$. Obviously, $s(i) \cap \gamma(i) = \emptyset$ for all $i \in I$. If we write $S(i) = s(i) \cup \gamma(i)$, we have

$$\partial K_i = \bigcup_{j \in S(i)} \Gamma_{ij}, \quad \partial K_i \cap \partial\Omega_h = \bigcup_{j \in \gamma(i)} \Gamma_{ij}. \tag{4.1}$$

The symbol $\boldsymbol{n}_{ij} = \left((n_{ij})_1, (n_{ij})_2\right)$ will denote the unit outer normal to ∂K_i on the side Γ_{ij}.

The approximate solution will be sought at each time instant t as an element of the finite-dimensional space

$$\boldsymbol{S}_h = \boldsymbol{S}^{r,-1}(\Omega_h, T_h) = \{\varphi_h;\ \varphi_h|_K \in P^r(K) \quad \forall K \in T_h\}^4, \tag{4.2}$$

where $r \geq 0$ is an integer and $P^r(K)$ denotes the space of all polynomials on K of degree $\leq r$. Function $\boldsymbol{\varphi} \in \boldsymbol{S}_h$ are in general discontinuous on interfaces Γ_{ij}. By $\boldsymbol{\varphi}|_{\Gamma_{ij}}$ and $\boldsymbol{\varphi}|_{\Gamma_{ji}}$ we denote the values of $\boldsymbol{\varphi}$ on Γ_{ij} considered from the interior and the exterior of K_i, respectively. The symbols

$$\langle \boldsymbol{\varphi} \rangle_{ij} = \frac{1}{2}\left(\boldsymbol{\varphi}|_{\Gamma_{ij}} + \boldsymbol{\varphi}|_{\Gamma_{ji}}\right), \quad [\boldsymbol{\varphi}]_{ij} = \boldsymbol{\varphi}|_{\Gamma_{ij}} - \boldsymbol{\varphi}|_{\Gamma_{ji}} \tag{4.3}$$

denote the average and jump of a function $\boldsymbol{\varphi}$ on $\Gamma_{ij} = \Gamma_{ij}$.

In order to derive the discrete problem, we assume that the exact solution $\boldsymbol{w}$ is sufficiently regular (e.g. continuously differentiable in $\bar{\Omega} \times [0, T]$.) We multiply (2.7) by a test function $\boldsymbol{\varphi} \in \boldsymbol{S}_h$ and integrate over any element $K_i,\ i \in I$. Applying Green's theorem and summing over all $i \in I$, we obtain

$$\begin{aligned} \frac{d}{dt} \sum_{K_i \in T_h} \int_{K_i} \boldsymbol{w}(t) \cdot \boldsymbol{\varphi} d\boldsymbol{x} = \sum_{K_i \in T_h} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t)) \cdot \frac{\partial \boldsymbol{\varphi}}{\partial x_s} d\boldsymbol{x} \\ - \sum_{K_i \in T_h} \sum_{j \in S(i)} \int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t))\,(n_{ij})_s \cdot \boldsymbol{\varphi} dS. \end{aligned} \tag{4.4}$$

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

$$\int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w})\,(n_{ij})_s \cdot \boldsymbol{\varphi} dS \approx \int_{\Gamma_{ij}} \boldsymbol{H}(\boldsymbol{w}_h(t)|_{\Gamma_{ij}}, \boldsymbol{w}_h(t)|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi} dS. \tag{4.5}$$

If we introduce the forms

$$(\boldsymbol{w}_h, \boldsymbol{\varphi}_h)_h = \int_{\Omega_h} \boldsymbol{w}_h \cdot \boldsymbol{\varphi}_h d\boldsymbol{x}, \tag{4.6}$$

$$\begin{aligned} \tilde{b}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) = - \sum_{K_i \in T_h} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t)) \cdot \frac{\partial \boldsymbol{\varphi}}{\partial x_s} d\boldsymbol{x} \\ + \sum_{K_i \in T_h} \sum_{j \in S(i)} \int_{\Gamma_{ij}} \boldsymbol{H}(\boldsymbol{w}_h(t)|_{\Gamma_{ij}}, \boldsymbol{w}_h(t)|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi}_h dS, \end{aligned} \tag{4.7}$$

we can define an *approximate solution of* (2.7) as a function $\boldsymbol{w}_h$ satisfying the conditions

$$\begin{aligned} &\text{(a)} \quad \boldsymbol{w}_h \in C^1\left([0, T]; \boldsymbol{S}_h\right), \\ &\text{(b)} \quad \frac{d}{dt}(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h)_h + \tilde{b}_h(\boldsymbol{w}_{h(t)}, \boldsymbol{\varphi}_h) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h,\ \forall t \in (0, T), \\ &\text{(c)} \quad \boldsymbol{w}_h(0) = \Pi_h \boldsymbol{w}^0, \end{aligned} \tag{4.8}$$

where $\Pi_h \boldsymbol{w}^0$ is the L^2-projection of $\boldsymbol{w}^0$ from the initial condition

$$\boldsymbol{w}(\boldsymbol{x}, 0) = \boldsymbol{w}^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{4.9}$$

on the space $\boldsymbol{S}_h$. This means that $\Pi_h \boldsymbol{w}^0 \in \boldsymbol{S}_h$ and

$$(\Pi_h \boldsymbol{w}^0, \boldsymbol{\varphi}_h)_h = (\boldsymbol{w}^0, \boldsymbol{\varphi}_h) \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h. \tag{4.10}$$

If we set $r = 0$, then we obtain the finite volume method using piecewise constant approximations.

The numerical flux $\boldsymbol{H}$ is assumed to be (locally) Lipschitz-continuous, consistent, i.e.

$$\boldsymbol{H}(\boldsymbol{w}, \boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) n_s, \tag{4.11}$$

and conservative, i.e.

$$\boldsymbol{H}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{n}) = -\boldsymbol{H}(\boldsymbol{w}, \boldsymbol{u}, -\boldsymbol{n}). \tag{4.12}$$

We choose the numerical flux with a convenient form for the semi-implicit linearization with respect to time. Particularly, this numerical flux can be written in the form

$$\boldsymbol{H}(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) = \mathbb{A}_L(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) \boldsymbol{w}_L + \mathbb{A}_R(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) \boldsymbol{w}_R \tag{4.13}$$

with some matrices $\mathbb{A}_L,\ \mathbb{A}_R :\ I\!R^4 \times I\!R^4 \times I\!R^2 \longrightarrow I\!R^{4\times 4}$.

In particular, we shall be concerned with the Vijaysundaram numerical flux $\boldsymbol{H}_{VS}$. We use Definition 2.1 and define the "absolute value", "positive" and "negative" parts of matrix $\mathbb{P}$ defined by (2.14):

$$\begin{aligned} |\mathbb{P}|(\boldsymbol{w}, \boldsymbol{n}) &= \mathbb{T} |\mathit{\Lambda}| \mathbb{T}^{-1}, & |\mathit{\Lambda}| &= \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_4|), \\ \mathbb{P}^{\pm}(\boldsymbol{w}, \boldsymbol{n}) &= \mathbb{T} \mathit{\Lambda}^{\pm} \mathbb{T}^{-1}, & \mathit{\Lambda}^{\pm} &= \operatorname{diag}(\lambda_1^{\pm}, \ldots, \lambda_2^{\pm}), \end{aligned} \tag{4.14}$$

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

$$\boldsymbol{H}_{VS}(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) = \mathbb{P}^+ \left(\frac{\boldsymbol{w}_L + \boldsymbol{w}_R}{2}, \boldsymbol{n} \right) \boldsymbol{w}_L + \mathbb{P}^- \left(\frac{\boldsymbol{w}_L + \boldsymbol{w}_R}{2}, \boldsymbol{n} \right) \boldsymbol{w}_R. \tag{4.15}$$

For explicit formula for $\mathbb{T},\ \mathit{\Lambda},\ \mathbb{T}^{-1}$ see Section 2.2. The eigenvalues λ_i have the form

$$\begin{aligned} \lambda_1 &= \boldsymbol{v} \cdot \boldsymbol{n} - a, \\ \lambda_2 = \lambda_3 &= \boldsymbol{n} \cdot \boldsymbol{v}, \\ \lambda_4 &= \boldsymbol{n} \cdot \boldsymbol{v} + a, \end{aligned} \tag{4.16}$$

where $a = \sqrt{\gamma p / \rho}$ is the speed of sound.

4.1.2 Time discretization

The condition (4.8), b) 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 use. This method leads to a large system of highly nonlinear algebraic equations, whose numerical solution is rather complicated. For these reasons we choose the semi-implicit scheme based on a suitable partial linearization of the form $\tilde{b}_h$, which gives us a higher order unconditionally stable scheme.

We consider a partition $0 = t_0 < t_1 < t_2 \dots$ of the interval $(0, T)$ and set $\tau_k = t_{k+1} - t_k$. We use the symbol $\boldsymbol{w}_h^k$ for the approximation of $\boldsymbol{w}_h(t_k)$. This technique is based on the linearization of the form $\tilde{b}_h$ carried out with the aid of the homogeneity of the fluxes $\boldsymbol{f}_s$ and the use of the Vijayasundaram numerical flux (4.15). In this way we obtain the form

$$b_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = -\sum_{K_i \in T_h} \int_{K_i} \sum_{s=1}^{2} \mathbb{A}_s(\boldsymbol{w}_h^k(\boldsymbol{x}))\boldsymbol{w}_h^{k+1}(\boldsymbol{x}) \cdot \frac{\partial \boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s} d\boldsymbol{x} \tag{4.17}$$

$$+ \sum_{K_i \in T_h} \sum_{j \in S(i)} \int_{\Gamma_{ij}} \left[\mathbb{P}^+ \left(\left\langle \boldsymbol{w}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} + \mathbb{P}^- \left(\left\langle \boldsymbol{w}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}} \right] \cdot \boldsymbol{\varphi}_h dS,$$

which is linear with respect to the second argument $\boldsymbol{w}_h^{k+1}$ and the third argument $\boldsymbol{\varphi}_h$.

On the basis of the above considerations we obtain the following semi-implicit scheme: For each $k \geq 0$ find $\boldsymbol{w}_h^{k+1}$ such that

$$\begin{aligned}
&\text{(a)} \quad \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_h, \\
&\text{(b)} \quad \left(\frac{\boldsymbol{w}_h^{k+1} - \boldsymbol{w}_h^k}{\tau_k}, \boldsymbol{\varphi}_h \right)_h + b_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h, \ k = 0, 1, \dots, \\
&\text{(c)} \quad \boldsymbol{w}_h^0 = \Pi_h \boldsymbol{w}^0.
\end{aligned} \tag{4.18}$$

This is a first order accurate scheme in time. In the solution of nonstationary flows, it is necessary to apply a scheme, which is sufficiently accurate in space as well as in time. One possibility is to apply the following two step second order time discretization: In (4.18), we use the second order approximation $\tilde{\boldsymbol{w}}_h^{k+1}$ of $\boldsymbol{w}_h(t_{k+1})$ obtained with the aid of extrapolation,

$$\tilde{\boldsymbol{w}}_h^{k+1} = \frac{\tau_k + \tau_{k-1}}{\tau_{k-1}} \boldsymbol{w}_h^k - \frac{\tau_k}{\tau_{k-1}} \boldsymbol{w}_h^{k-1}, \tag{4.19}$$

which replaces the state $\boldsymbol{w}_h^k$ in the form b_h. Moreover the second order backward difference approximation of the time derivative of the solution at t_{k+1} is used

$$\begin{aligned}
\frac{\partial \boldsymbol{w}_h(\boldsymbol{x}, t)}{\partial t}|_{t=t_{k+1}} \approx & \frac{2\tau_k + \tau_{k-1}}{\tau_k (\tau_k + \tau_{k-1})} \boldsymbol{w}_h(\boldsymbol{x}, t_{k+1}) - \frac{\tau_k + \tau_{k-1}}{\tau_k \tau_{k-1}} \boldsymbol{w}_h(\boldsymbol{x}, t_k) \\
& + \frac{\tau_k}{\tau_{k-1} (\tau_k + \tau_{k-1})} \boldsymbol{w}_h(\boldsymbol{x}, t_{k-1}).
\end{aligned} \tag{4.20}$$

As a result we get the following *two-step second-order scheme*: For each $k \geq 1$ find $\boldsymbol{w}_h^{k+1}$ such that

$$\begin{aligned}
&\text{(a)} \quad \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_h, \\
&\text{(b)} \quad \frac{2\tau_k + \tau_{k-1}}{\tau_k(\tau_k + \tau_{k-1})} \left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right)_h + b_h(\tilde{\boldsymbol{w}}_h^{k+1}, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = \frac{\tau_k + \tau_{k-1}}{\tau_k \tau_{k-1}} \left(\boldsymbol{w}_h^{k}, \boldsymbol{\varphi}_h\right)_h \\
&\qquad -\frac{\tau_k}{\tau_{k-1}(\tau_k + \tau_{k-1})} \left(\boldsymbol{w}_h^{k-1}, \boldsymbol{\varphi}_h\right)_h \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h,\ k = 0, 1, \ldots, \qquad (4.21) \\
&\text{(c)} \quad \boldsymbol{w}_h^0 = \Pi_h \boldsymbol{w}^0,\ \boldsymbol{w}_h^1 \text{ obtained by the Runge} - \text{Kutta method.}
\end{aligned}$$

In order to guarantee the stability of the scheme, we use the CLF condition

$$\tau_k \max_{K_i \in T_h} \frac{1}{|K_i|} \left(\max_{j \in S_i} |\Gamma_{ij}| \, \lambda^{\max}_{\mathbb{P}\left(\boldsymbol{w}_h^k|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}\right)} \right) \leq CLF, \qquad (4.22)$$

where $|K_i|$ denotes the area of K_i, $|\Gamma_{ij}|$ the lenght of the edge Γ_{ij}, CFL a given constant and $\lambda^{\max}_{\mathbb{P}\left(\boldsymbol{w}_h^k|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}\right)}$ is the maximal eigenvalue of the matrix $\mathbb{P}(\boldsymbol{w}_h^k|_{\Gamma_{ij}}, \boldsymbol{n}_{ij})$, where the maximum is taken over Γ_{ij}. This condition is a heuristic extension of a similar condition applied in the finite volume method (see, e.g. [4]).

4.1.3 Boundary conditions

If $\Gamma_{ij} \subset \partial\Omega_h$, i.e. $j \in \gamma(i)$, it is necessary to specify the boundary state $\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}$ appearing in the numerical flux $\boldsymbol{H}$ in the definition of the inviscid form b_h.

On $\Gamma = \Gamma_{ij} \subset \Gamma_W$, i.e. solid impermeable wall with normal $\boldsymbol{n} = \boldsymbol{n}_{ij}$, we prescribe the so-called *no-stick* condition:

$$\boldsymbol{v} \cdot \boldsymbol{n} = 0 \text{ on } \Gamma. \qquad (4.23)$$

We use the approximation

$$\int_{\Gamma_{ij}} \boldsymbol{H}\left(\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}, \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}\right) \cdot \boldsymbol{\varphi}_h dS \approx \int_{\Gamma_{ij}} \boldsymbol{F}_W\left(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{n}_{ij}\right) \cdot \boldsymbol{\varphi}_h dS, \qquad (4.24)$$

where

$$\boldsymbol{F}_W\left(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{n}\right) = \frac{D\mathcal{P}}{D\boldsymbol{w}}(\boldsymbol{w}_{\boldsymbol{h}}^{\boldsymbol{k}}, \boldsymbol{n}) \boldsymbol{w}_h^{k+1} = \mathbb{P}(\boldsymbol{w}_h^k, \boldsymbol{n}) \boldsymbol{w}_h^{k+1}. \qquad (4.25)$$

Taking in to account the no-stick condition on the impermeable wall, the normal component of the inviscid flux has the form

$$\mathcal{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \boldsymbol{f}^s(\boldsymbol{w}) n_s = (\boldsymbol{v} \cdot \boldsymbol{n}) + p(0, n_1, n_2, \boldsymbol{v} \cdot \boldsymbol{n})^T = p(0, n_1, n_2, 0)^T. \qquad (4.26)$$

If we extrapolate the value of the pressure by $p_R = p_L$, we can define the numerical flux

$$\boldsymbol{H}(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) = p(0, n_1, n_2, 0)^T. \qquad (4.27)$$

In the case of the inlet and outlet conditions there is a problem, which quantities should be prescribed (*Dirichlet* condition) and which should be extrapolated onto Γ

from the adjacent element (*Neumann*-type condition). One possibility used often in practice is given in [4] using the method of characteristics. We shall describe here in detail the second possibility.

Let $\Gamma = \Gamma_{ij} \subset \Gamma_{IO}$ and $\boldsymbol{n} = \boldsymbol{n}_{ij}$ be the outer unit normal to K_i on Γ. In order to compute $\boldsymbol{H}(\boldsymbol{w}_i, \boldsymbol{w}_j, \boldsymbol{n})$, we need to specify the value $\boldsymbol{w}_j$, when $\boldsymbol{w}_i$ is known.

Let $\boldsymbol{n} = \boldsymbol{n}_{ij}$ be the outer normal to K_i on $\Gamma = \Gamma_{ij}$. 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

$$\frac{\partial \boldsymbol{q}}{\partial t} + \sum_{s=1}^{2} \frac{\partial \boldsymbol{f}_s(\boldsymbol{q})}{\partial \tilde{x}_s} = 0, \tag{4.28}$$

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

$$\boldsymbol{q} = \mathbb{Q}(\boldsymbol{n})\boldsymbol{w}, \tag{4.29}$$

where $\mathbb{Q}$ has form (2.29).

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

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{\partial \boldsymbol{f}_1(\boldsymbol{q})}{\partial \tilde{x}_1} = 0. \tag{4.30}$$

Now we write system (4.28) in the nonconservative form

$$\frac{\partial \boldsymbol{q}}{\partial t} + \mathbb{A}_1(\boldsymbol{q})\frac{(\boldsymbol{q})}{\partial \tilde{x}_1} = 0. \tag{4.31}$$

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

$$\frac{\partial \boldsymbol{q}}{\partial t} + \mathbb{A}_1(\boldsymbol{q}_i)\frac{(\boldsymbol{q})}{\partial \tilde{x}_1} = 0, \tag{4.32}$$

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

$$\boldsymbol{q}(\tilde{x}_1, 0) = \boldsymbol{q}_i, \quad \tilde{x}_1 \in (-\infty,\ 0) \tag{4.33}$$

and the boundary condition

$$\boldsymbol{q}(0, t) = \boldsymbol{q}_j, \quad t > 0. \tag{4.34}$$

The goal is to choose $\boldsymbol{q}_j$ in such a way that the initial-boundary value problem (4.32)-(4.34) is well posed, i.e. has a unique solution. The solution can be written in the form

$$\boldsymbol{q}(\tilde{x}_1, t) = \sum_{s=1}^{4} \mu(\bar{x}_1, t)\boldsymbol{r}_s, \tag{4.35}$$

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

$$\boldsymbol{q}_i = \sum_{s=1}^{4} \alpha_s \boldsymbol{r}_s, \quad \boldsymbol{q}_j = \sum_{s=1}^{4} \beta_s \boldsymbol{r}_s. \tag{4.36}$$

Substituting (4.35) into (4.32) and using the relation $\mathbb{A}_1(\boldsymbol{q}_i)\boldsymbol{r}_s = \tilde{\lambda}_s \boldsymbol{r}_s$, we find that problem (4.32)-(4.34) is equivalent to 4 mutually independent linear initial-boundary value scalar problems for $s = 1, \ldots, 4$:

$$\begin{aligned} &\frac{\partial \mu_s}{\partial t} + \tilde{\lambda}_s \frac{\partial \mu_s}{\partial \tilde{x}_1} = 0 \quad \text{in } (-\infty, 0) \times (0, \infty), \\ &\mu_s(\tilde{x}_1, 0) = \alpha_s, \quad \tilde{x}_1 \in (-\infty, 0), \\ &\mu_s(0, t) = \beta_s, \quad t \in (0, \infty), \end{aligned} \tag{4.37}$$

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

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

using the fact that the solution μ_s is constant along characteristics. Now the solution is

$$\mu_s(\tilde{x}_1, t) = \begin{cases} \alpha_s, & \tilde{x}_1 - \tilde{\lambda}_s t < 0, \\ \beta_s, & \tilde{x}_1 - \tilde{\lambda}_s t > 0. \end{cases} \tag{4.38}$$

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

$$\tilde{\lambda}_s(\boldsymbol{q}_i) = \lambda_s(\boldsymbol{w}_i, \boldsymbol{n}), \quad s = 1, \ldots, 4, \tag{4.39}$$

where $\lambda_s(\boldsymbol{w}_i, \boldsymbol{n})$ are the eigenvalues of the Jacobi matrix $\mathbb{P}(\boldsymbol{w}_i, \boldsymbol{n})$ defined in (2.14). In [4] the conclusion is drawn, 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. We propose to prescribed variables based on the local linearized problem.

We shall take some state $\boldsymbol{q}_j^0 = \mathbb{Q}(\boldsymbol{n})\boldsymbol{w}_j^0$. The state $\boldsymbol{w}_j^0$ 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 $\boldsymbol{w}_j$. We express the state $\boldsymbol{q}_j^0$ in the form

$$\boldsymbol{q}_j^0 = \sum_{s=1}^{4} \gamma_s \boldsymbol{r}_s. \tag{4.40}$$

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

$$\begin{aligned} \boldsymbol{q}_i = \mathbb{T}\boldsymbol{\alpha} \;&\Leftrightarrow\; \boldsymbol{\alpha} = \mathbb{T}^{-1}\boldsymbol{q}_i, \\ \boldsymbol{q}_j^0 = \mathbb{T}\boldsymbol{\gamma} \;&\Leftrightarrow\; \boldsymbol{\gamma} = \mathbb{T}^{-1}\boldsymbol{q}_j^0. \end{aligned} \tag{4.41}$$

Now we calculate the state $\boldsymbol{q}_j$ according to the presented process:

$$\boldsymbol{q}_j := \sum_{s=1}^{4} \beta_s \boldsymbol{r}_s = \mathbb{T}\boldsymbol{\beta}, \tag{4.42}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_4)^T$ and

$$\beta_s = \begin{cases} \alpha_s, & \lambda_s \geq 0, \\ \gamma_s, & \lambda_s < 0. \end{cases} \tag{4.43}$$

Finally $\boldsymbol{w}_j = \mathbb{Q}^{-1}(\boldsymbol{n})\boldsymbol{q}_j$ and we can use this to calculate $\boldsymbol{H}(\boldsymbol{w}_i, \boldsymbol{w}_j, \boldsymbol{n})$.

In the framework of the presented theory, these boundary conditions seem to give the natural choice for $\boldsymbol{w}_j$. However, we must keep in mind two simplifications that we have made during the derivation:

a) We have neglected tangential derivatives of the solution in order to get a simplified equation (4.30).

b) We have avoided the nonlinearity of problem (4.30) by local linearization.

Nonetheless, experiments show that this method applied to the approximation of inlet and outlet boundary conditions allows to pass the density and pressure waves through the boundaries without reflection.

4.1.4 Shock capturing

For high-speed flow with shock waves and contact discontinuities it is necessary to avoid the *Gibbs phenomenon* manifested by spurious overshoots and undershoots in computed quantities near discontinuities and steep gradients. In spite of dealing with the low Mach number flow in our work, where these phenomena do not occur, we have to solve the problem in the transonic case. In this case these phenomena cause instabilities in the semi-implicit solution.

One possibility for avoiding the Gibbs phenomenon is the use of the limiting of order of acuracy of the method in the vicinity of discontinuities. The limiting technique is motivated by the paper [5], on the basis of which the left-hand side of (4.18)(b) and (4.21)(b) is augmented by an artificial viscosity form. However, since this form is nonzero also in regions where the exact solution is regular, a small nonphysical entropy production can appear in these regions. Therefore, this technique is combined with the approach proposed in [6]. It is based on the discontinuity indicator $g^k(i)$ defined by

$$g^k(i) = \int_{\partial K_i} \left[\rho_h^k\right]^2 dS/(h_{K_i} |K_i|^{3/4}), \quad K_i \in T_h. \tag{4.44}$$

By $\left[\rho_h^k\right]$ we denote the jump of the density on ∂K_i at time t. The indicator $g^k(i)$ was constructed in such a way that it takes an anisotropy of the computational mesh into account. Now we introduce the discrete discontinuity indicator

$$G^k(i) = 0 \quad \text{if } g^k(i) < 1, \ K_i \in T_h \tag{4.45}$$

$$G^k(i) = 1 \quad \text{if } g^k(i) \geq 1, \ K_i \in T_h, \tag{4.46}$$

and add the artificial viscosity form

$$\beta_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}) = \nu_1 \sum_{i \in I} h_{K_i} G^k(i) \int_{K_i} \nabla \boldsymbol{w}_h^{k+1} \cdot \nabla \boldsymbol{\varphi} d\boldsymbol{x} \tag{4.47}$$

with $\nu_1 = \mathrm{O}(1)$ to the left-hand side of (4.18)(b) and (4.21)(b). Since the artificial viscosity form is rather local, it is proposed to augment the left-hand side of (4.18)(b) and (4.21)(b) by adding the form

$$J_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}) = \nu_2 \sum_{i\in I} \sum_{j\in s(i)} \frac{1}{2}(G^k(i) + G^k(j)) \int_{\Gamma_{ij}} \left[\boldsymbol{w}_h^{k+1}\right] \cdot [\boldsymbol{\varphi}]\, dS, \tag{4.48}$$

where $\nu_2 = \mathrm{O}(1)$, which allows to strengthen the influence of neighbouring elements and improves the behaviour of the method.

Thus, the resulting scheme obtained by limiting of (4.18)(b) reads:

$$\begin{aligned}
&\text{(a)} \quad \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_h, \\
&\text{(b)} \quad \left(\frac{\boldsymbol{w}_h^{k+1} - \boldsymbol{w}_h^k}{\tau_k}, \boldsymbol{\varphi}_h\right)_h + b_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) + \beta_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}) \\
&\qquad + J_h(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h,\ k = 0, 1, \ldots, \\
&\text{(c)} \quad \boldsymbol{w}_h^0 = \Pi_h \boldsymbol{w}^0.
\end{aligned} \tag{4.49}$$

Similarly, we obtained a stabilized version of the scheme (4.21). The same stabilization technique can be used easily in the problem of time-dependent domain.

4.1.5 Approximation of the boundary - Isoparametric elements

So far we have worked only with polygonal domains $\Omega \subset \mathbb{R}^2$. This is rather limiting, when we approach practical problems, in which we seldom meet completely polygonal (polyhedral) shapes. In practice, this means that we have a domain Ω with a curved boundary and have to approximate it by some Ω_h, which is polygonal. In the finite volume method this works well, since we seek piecewise constant solutions. Also in the conforming finite element method with P^1 elements applied to elliptic or parabolic problems, polygonal approximations of the boundary yield optimal error estimates. However in the case of DGFE higher-order approximations, numerical experiments show that this method does not give good results in the vicinity of curved parts of $\partial\Omega$. As stated in [4], refining the mesh locally does not help and undesired phenomena occur - for instance nonphysical entropy production. In order to get good behavior near curved segments of the boundary when using higher order discretizations, it is necessary to introduce a higher order approximation of the boundary $\partial\Omega$ and adjacent elements.

Let $\Omega \subset \mathbb{R}^2$ and T_h be its partition formed by triangles $K_i, i \in I$. Let $\hat{K}$ be the reference triangle. Let

$$\hat{P}^0 = (0;0), \quad \hat{P}^1 = (1;0), \quad \hat{P}^2 = (0;1) \tag{4.50}$$

be the vertices of $\hat{K}$ and

$$\hat{P}^{12} = (1/2;1/2). \tag{4.51}$$

Let $\{K_i, i \in I_c\}$ with $I_c \subset I$ be a set of triangles adjacent to a curved part of $\partial\Omega$. For $i \in I_c$ let $P_i^k, k = 0, 1, 2$, be the vertices of K_i such that $P_i^0 \in \Omega$, $P_i^1, P_i^2 \in \partial\Omega$.

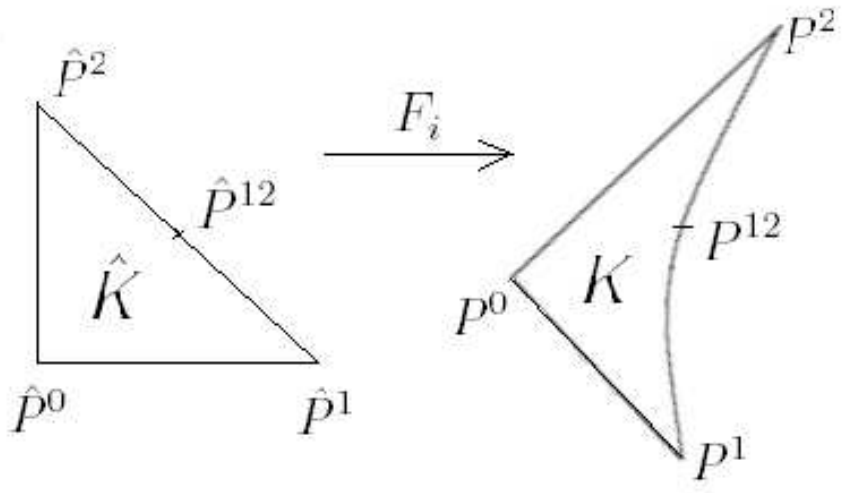


Figure 4.1: Bilinear mapping:$F_i : \hat{K}_i \to K_i$

We suppose that the the center P_i^{12} of the curved side with endpoints P_i^1, P_i^2 is close to the center of the linear segment $P_i^1 P_i^2$ - this is natural for triangulations that are dense enough. Under these assumptions we can find a unique bilinear mapping F_i defined on $\hat{K}$, $F_i = (F_i^1, F_i^2)$ such that

$$\begin{aligned} &F_i(\hat{P}^k) = P_i^k, \quad k = 0, 1, 2, \\ &F_i(\hat{P}^{12}) = P_i^{12}. \end{aligned} \tag{4.52}$$

Triangles $K_i, i \in I_c$, are replaced by the curved triangles defined by

$$\bar{K}_i := F_i(\hat{K}), \tag{4.53}$$

which have two straight sides and one curved side approximating the curved segment of $\partial\Omega$ adjacent to K_i, see Figure 4.1. If $i \notin I_c$ then F_i is a linear mapping and therefore $\bar{K}_i = K_i$.

In the described discretization we need to evaluate volume and boundary integrals over elements and their boundaries - here we describe the modification of the method for curved elements $\bar{K}_i, i \in I_c$ - the simpler case when $i \notin I_c$ is treated in the same manner, only the mapping F_i is linear. We denote by

$$J_{F_i}(\hat{\boldsymbol{x}}) := \frac{DF_i}{D\hat{\boldsymbol{x}}}(\hat{\boldsymbol{x}}), \quad \hat{\boldsymbol{x}} \in \hat{K}, \tag{4.54}$$

the Jacobi matrix of the mapping F_i. Test functions $\boldsymbol{\varphi}$ and the approximate solution $\boldsymbol{w}(\cdot, t)$ are defined on $\hat{K}_i$ as

$$\begin{aligned} &\boldsymbol{\varphi}(\boldsymbol{x}) = \hat{\boldsymbol{\varphi}}(F_i^{-1}(\boldsymbol{x})), \quad \boldsymbol{x} \in \bar{K}_i, \\ &\boldsymbol{w}_h(\boldsymbol{x}, t) = \hat{\boldsymbol{w}}_i(F_i^{-1}(\boldsymbol{x}), t), \quad \boldsymbol{x} \in \bar{K}_i,\ t \in [0, T], \end{aligned} \tag{4.55}$$

where $\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{w}}_i(\cdot, t) \in [P^p(\hat{K})]^m$.

The forms in (4.8) are evaluated in the following way: The $L^2(K_i)$-scalar product is expressed, using the substitution as

$$\int_{\bar{K}_i} \boldsymbol{w}_h(\boldsymbol{x},t)\cdot\boldsymbol{\varphi}_h(\boldsymbol{x})\,d\boldsymbol{x} = \int_{\hat{K}} \hat{\boldsymbol{w}}_i(\hat{\boldsymbol{x}},t)\cdot\hat{\boldsymbol{\varphi}}_h(\hat{\boldsymbol{x}})\det J_{F_i}(\hat{\boldsymbol{x}})\,d\hat{\boldsymbol{x}},\quad i\in I. \tag{4.56}$$

In the inviscid volume terms in $\tilde{b}_h$ we have to use the fact that

$$(\hat{\nabla}\hat{\boldsymbol{\varphi}}_h)(\hat{\boldsymbol{x}}) = J_{F_i}(\hat{\boldsymbol{x}})(\nabla\boldsymbol{\varphi}_h)(F_i(\hat{\boldsymbol{x}})), \tag{4.57}$$

thus

$$(\nabla\boldsymbol{\varphi}_h)(F_i(\hat{\boldsymbol{x}})) = [J_{F_i}(\hat{\boldsymbol{x}})]^{-1}(\hat{\nabla}\hat{\boldsymbol{\varphi}}_h)(\hat{\boldsymbol{x}}) \tag{4.58}$$

and

$$\begin{aligned}
&\int_{\bar{K}_i}\sum_{s=1}^{2}\boldsymbol{f}_s(\boldsymbol{w}_h(\boldsymbol{x},t))\cdot\frac{\partial\boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s}\,d\boldsymbol{x} \qquad (4.59)\\
&=\int_{\hat{K}}(\boldsymbol{f}_1(\hat{\boldsymbol{w}}_i),\boldsymbol{f}_2(\hat{\boldsymbol{w}}_i))\,[J_{F_i}(\hat{\boldsymbol{x}})]^{-1}\hat{\nabla}\hat{\boldsymbol{\varphi}}_h(\hat{\boldsymbol{x}})\det J_{F_i}(\hat{\boldsymbol{x}})\,d\hat{\boldsymbol{x}}\\
&=\int_{\hat{K}}\sum_{s=1}^{2}\boldsymbol{f}_s(\hat{\boldsymbol{w}}_i(\boldsymbol{x},t))\cdot\sum_{j=1}^{2}\frac{\partial\hat{\boldsymbol{\varphi}}_h(\hat{\boldsymbol{x}})}{\partial\hat{\boldsymbol{x}}_j}\frac{\partial(F_i^{-1})^j}{\partial x_s}(F_i(\hat{\boldsymbol{x}}))\,\det J_{F_i}(\hat{\boldsymbol{x}})\,d\hat{\boldsymbol{x}},\quad i\in I,
\end{aligned}$$

where $(F_i^{-1})^j$ denotes the j-th component of the inverse mapping F_i^{-1}. However, the evaluation using the inverse $[J_{F_i}(\hat{\boldsymbol{x}})]^{-1}$ is simpler then calculating the inverse F_i^{-1} and than its Jacobi matrix. One can see that these two approaches are equivalent since

$$\frac{DF_i^{-1}}{D\boldsymbol{x}}(F_i(\hat{\boldsymbol{x}})) = \left[\frac{DF_i}{D\hat{\boldsymbol{x}}}(\hat{\boldsymbol{x}})\right]^{-1} \tag{4.60}$$

following from the identity $\boldsymbol{x} = F_i(F_i^{-1}(\boldsymbol{x}))$.

Boundary integrals over a curved side $\Gamma_{ij}\subset\partial K_i$ in the boundary terms of the form $\tilde{b}_h$ are computed with the aid of a suitable parameterization of Γ_{ij} and the side $\hat{\Gamma}$ of $\hat{K}$ corresponding to Γ_{ij} in the mapping F_i:

$$\boldsymbol{x} = \boldsymbol{x}(\xi) = F_i(\hat{\boldsymbol{x}}(\xi)),\quad \xi\in[0,1]. \tag{4.61}$$

If we put

$$u(\boldsymbol{x}) := \boldsymbol{H}(\boldsymbol{w}|_{\Gamma_{ij}}(\boldsymbol{x},t),\boldsymbol{w}|_{\Gamma_{ji}}(\boldsymbol{x},t),\boldsymbol{n}_{ij})\cdot\boldsymbol{\varphi}_h(\boldsymbol{x}) \tag{4.62}$$

(for a fixed t), we get

$$\begin{aligned}
\int_{\Gamma_{ij}} u(\boldsymbol{x})\,dS &= \int_0^1 u(\boldsymbol{x}(\xi))|\boldsymbol{x}'(\xi)|\,d\xi \qquad (4.63)\\
&= \int_0^1 u(F_i(\hat{\boldsymbol{x}}(\xi)))\left\{\sum_{j=1}^{2}\left(\sum_{k=1}^{2}\frac{\partial F_i^j(\hat{\boldsymbol{x}}(\xi))}{\partial\hat{\boldsymbol{x}}_k}\hat{\boldsymbol{x}}_k'(\xi)\right)^2\right\}^{1/2}d\xi.
\end{aligned}$$

The parametrization $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\xi)$ of $\hat{\Gamma}$ is expressed in the form

$$\hat{\boldsymbol{x}}(\xi) = A + \xi(B-A), \tag{4.64}$$

j	$\hat{x}_j^{(1)}$-coordinate	$\hat{x}_j^{(2)}$-coordinate	α_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 4.1: Gauss seven point rule on the reference triangle $\hat{K}$.

j	ξ_j	β_j
1.	$(1-\sqrt{3/5})/2$	5/18
2.	0.5	4/9
3.	$(1+\sqrt{3/5})/2$	5/18

Table 4.2: Gauss three point rule on the unit interval $\hat{\Gamma}$.

where A, B are the endpoints of $\hat{\Gamma}$.

For the evaluation of integrals over $\hat{\Gamma}$ and $\hat{K}$ we use 1D and 2D Gaussian quadrature formulae of higher order of accuracy: both are accurate for polynomials with degree ≤ 5. In 2D it is the seven point rule,

$$\int_{\hat{K}} f(\hat{\boldsymbol{x}})d\hat{\boldsymbol{x}} \approx \sum_{j=1}^{7} \alpha_j f(\hat{x}_j), \tag{4.65}$$

where α_j and $\hat{x}_j = (\hat{x}_j^1, \hat{x}_j^2)$ are given in Table 4.1 and f is a function that we integrate over $\hat{K}$. In 1D we use the three point rule,

$$\int_0^1 s(\xi)d\xi \approx \sum_{j=1}^{3} \beta_j s(\xi_j), \tag{4.66}$$

where β_j and ξ_j are given in Table 4.2 and s is a function that we integrate over $\hat{\Gamma}$.

4.2 Discretization in time-dependent domain

We shall deal with inviscid compressible flow in a bounded domain $\Omega_t \subset I\!R^2$ depending on time $t \in [0, T]$. We assume that the boundary $\partial\Omega_t$ consists of three disjoint parts $\partial\Omega_t = \Gamma_I \cup \Gamma_O \cup \Gamma_{W_t}$, where Γ_I and Γ_O represent the *inlet* and *outlet* and Γ_{W_t} represents a moving *impermeable wall*.

Futher, we shall consider this notation for the discretization of the both formulations of Euler equations in the time-dependent domain: Let Ω_{ht} be a polygonal

approximation of Ω_t at time t. By T_{ht} we denote a partition of Ω_{ht} consisting of elements $K_i \in T_{ht},\ i \in I_t$, e.g. triangles or quadrilaterals. ($I_t \subset \mathbb{Z}^+ = \{0,\ 1,\ 2,\ \ldots\}$ is a suitable index set.) By Γ_{ij} we denote a common edge between two neighbouring elements K_i and K_j. We set $s_t(i) = \{j \in I_t;\ K_j \text{ is a neighbour of } K_i.\}$ The boundary $\partial\Omega_{ht}$ is formed by a finite number of faces of elements K_i adjacent to $\partial\Omega_{ht}$. We denote all these boundary faces by S_j, where $j \in I_{bt} \subset \mathbb{Z}^- = \{-1,\ -2,\ \ldots\}$. Now we set $\gamma_t(i) = \{j \in I_{bt};\ S_j \text{ is a face of } K_i \in T_{ht}\}$ and $\Gamma_{ij} = S_j$ for $K_i \in T_{ht}$ such that $S_j \subset \partial K_i,\ j \in I_{bt}$. For K_i not containing any boundary face S_j we set $\gamma_t(i) = \emptyset$. Obviously, $s_t(i) \cap \gamma_t(i) = \emptyset$ for all $i \in I_t$. If we write $S_t(i) = s_t(i) \cup \gamma_t(i)$, we have

$$\partial K_i = \bigcup_{j \in S_t(i)} \Gamma_{ij}, \quad \partial K_i \cap \partial\Omega_{ht} = \bigcup_{j \in \gamma_t(i)} \Gamma_{ij}. \tag{4.67}$$

The symbol $\boldsymbol{n}_{ij} = \left((n_{ij})_1, (n_{ij})_2\right)$ will denote the unit outer normal to ∂K_i on the side Γ_{ij}.

The approximate solution will be sought at each time instant t as an element of the finite-dimensional space

$$\boldsymbol{S}_{ht} = \boldsymbol{S}^{r,-1}(\Omega_{ht}, T_{ht}) = \{\varphi_h;\ \varphi_h|_K \in P^r(K) \quad \forall K \in T_{ht}\}^4, \tag{4.68}$$

where $r \geq 0$ is an integer and $P^r(K)$ denotes the space of all polynomials on K of degree$\leq r$. Function $\boldsymbol{\varphi} \in \boldsymbol{S}_{ht}$ are in general discontinuous on interfaces Γ_{ij}. By $\boldsymbol{\varphi}|_{\Gamma_{ij}}$ and $\boldsymbol{\varphi}|_{\Gamma_{ji}}$ we denote the values of $\boldsymbol{\varphi}$ on Γ_{ij} considered from the interior and the exterior of K_i, respectively. The symbols

$$\langle \boldsymbol{\varphi} \rangle_{ij} = \frac{1}{2}\left(\boldsymbol{\varphi}|_{\Gamma_{ij}} + \boldsymbol{\varphi}|_{\Gamma_{ji}}\right), \quad [\boldsymbol{\varphi}]_{ij} = \boldsymbol{\varphi}|_{\Gamma_{ij}} - \boldsymbol{\varphi}|_{\Gamma_{ji}} \tag{4.69}$$

denote the avarege and jump of a function $\boldsymbol{\varphi}$ on $\Gamma_{ij} = \Gamma_{ij}$.

4.2.1 ALE formulation I of the Euler equations

In this subsection we shall discretize the ALE formulation I of the Euler equations, which we derived in Section 3.2.1.

Space semidiscretization

In order to derive the discrete problem, we multiply (3.10) by a test function $\boldsymbol{\varphi} \in \boldsymbol{S}_{ht}$ and integrate over any element $K_i,\ i \in I_t$. Applying Green's theorem and summing over all $i \in I_t$, we obtain

$$\begin{aligned}\sum_{K_i \in T_{ht}} \int_{K_i} \frac{D^{\mathcal{A}}\boldsymbol{w}(t)}{Dt} \cdot \boldsymbol{\varphi} d\boldsymbol{x} &= \sum_{K_i \in T_{ht}} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t)) \cdot \frac{\partial \boldsymbol{\varphi}}{\partial x_s} d\boldsymbol{x} \\ &- \sum_{K_i \in T_{ht}} \sum_{j \in S_t(i)} \int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t))\,(n_{ij})_s \cdot \boldsymbol{\varphi} dS + \sum_{K_i \in T_{ht}} \int_{K_i} \sum_{s=1}^{2} z_s \frac{\partial \boldsymbol{w}(t)}{\partial x_s} \cdot \boldsymbol{\varphi} d\boldsymbol{x}.\end{aligned} \tag{4.70}$$

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

$$\int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}) \left(n_{ij}\right)_s \cdot \boldsymbol{\varphi} dS \approx \int_{\Gamma_{ij}} \boldsymbol{H}_f(\boldsymbol{w}_h(t)|_{\Gamma_{ij}}, \boldsymbol{w}_h(t)|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi} dS. \tag{4.71}$$

The properties of the flux $\boldsymbol{H}_f$ are the same as in the Section 4.1.1.

Now, we can introduce the forms

$$\begin{aligned}
(\frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt}, \boldsymbol{\varphi}_h)_h &= \int_{\Omega_{ht}} \frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt} \cdot \boldsymbol{\varphi}_h d\boldsymbol{x}, &(4.72)\\
\tilde{b}_h^1(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) &= -\sum_{K_i \in T_{ht}} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}(t)) \cdot \frac{\partial \boldsymbol{\varphi}}{\partial x_s} d\boldsymbol{x} &(4.73)\\
&\quad + \sum_{K_i \in T_{ht}} \sum_{j \in S_t(i)} \int_{\Gamma_{ij}} \boldsymbol{H}_f(\boldsymbol{w}_h(t)|_{\Gamma_{ij}}, \boldsymbol{w}_h(t)|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi}_h dS,\\
d_h^1(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) &= -\sum_{K_i \in T_{ht}} \int_{K_i} \sum_{s=1}^{2} z_s \frac{\partial \boldsymbol{w}_h(t)}{\partial x_s} \cdot \boldsymbol{\varphi}_h d\boldsymbol{x} &(4.74)
\end{aligned}$$

It allows us to define an *approximate solution of* (3.10) as a function $\boldsymbol{w}_h = \boldsymbol{w}_t(t)$ satisfying the conditions

$$\begin{aligned}
&\text{(a)} \quad \boldsymbol{w}_h(t) \in \boldsymbol{S}_{ht}, \ \forall t \in [0, T], &(4.75)\\
&\text{(b)} \quad (\frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt}, \boldsymbol{\varphi}_h)_h + \tilde{b}_h^1(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) + d_h^1(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht}, \ \forall t \in (0, T),\\
&\text{(c)} \quad \boldsymbol{w}_h(0) = \Pi_h \boldsymbol{w}^0,
\end{aligned}$$

where $\Pi_h \boldsymbol{w}^0$ is the L^2-projection of $\boldsymbol{w}^0$ from the initial condition

$$\boldsymbol{w}(\boldsymbol{x}, 0) = \boldsymbol{w}^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega_0 \tag{4.76}$$

on the space $\boldsymbol{S}_{h0}$.

Time discretization

Now, we introduce the partition $0 = t_0 < t_1 < \ldots$ of the time interval $[0, T]$ and set $\tau_k = t_{k+1} - t_k$. The function $\boldsymbol{w}_h(\cdot, t_k)$ will be approximated by $\boldsymbol{w}^k$, defined in Ω_{t_k}. If we set

$$\hat{\boldsymbol{w}}_h^j(\boldsymbol{x}) = \boldsymbol{w}^j\left(\mathcal{A}_{t_j}\left(\mathcal{A}_{t_{k+1}}^{-1}\right)(\boldsymbol{x})\right), \quad \boldsymbol{x} \in \Omega_{t_{k+1}}, \tag{4.77}$$

then we can approximate the ALE derivative using the Euler backward method of the first order:

$$\left(\frac{D^{\mathcal{A}}\boldsymbol{w}_h(\boldsymbol{x}, t)}{Dt}, \boldsymbol{\varphi}_h)\right)|_{t_{k+1}} \approx \left(\frac{\boldsymbol{w}^{k+1}(\boldsymbol{x}) - \hat{\boldsymbol{w}}_h^k(\boldsymbol{x})}{\tau_k}, \boldsymbol{\varphi}_h\right), \quad \boldsymbol{x} \in \Omega_{ht_{k+1}}, \tag{4.78}$$

where we take in account the relation (3.5).

For the linearization of the form $\tilde{b}_h^1$ we use the homogeneity of the fluxes $\boldsymbol{f}_s$ and the Vijayasundaram numerical flux (4.15). In this way we obtain the form

$$b_h^1(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = - \sum_{K_i \in T_{ht_{k+1}}} \int_{K_i} \sum_{s=1}^{2} \mathbb{A}_s(\hat{\boldsymbol{w}}_h^k(\boldsymbol{x})) \boldsymbol{w}_h^{k+1}(\boldsymbol{x}) \cdot \frac{\partial \boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s} d\boldsymbol{x}$$
$$+ \sum_{K_i \in T_{ht_{k+1}}} \sum_{j \in S_{t_{k+1}}(i)} \int_{\Gamma_{ij}} \left[\mathbb{P}^+ \left(\frac{\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ij}} + \hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}}{2}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \right. \tag{4.79}$$
$$\left. + \mathbb{P}^- \left(\frac{\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ij}} + \hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}}{2}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}} \right] \cdot \boldsymbol{\varphi}_h dS,$$

which is linear with respect to the second argument $\boldsymbol{w}_h^{k+1}$ and the third argument $\boldsymbol{\varphi}_h$. For the description of the matrices $\mathbb{P}^{\pm}$ and $\mathbb{D}^{\pm}$ see the Section 4.1.1. The form d_h^1 we let implicit. It means

$$d_h^1\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) = - \sum_{K_i \in T_{ht_{k+1}}} \int_{K_i} \sum_{s=1}^{2} z_s^{k+1} \frac{\partial \boldsymbol{w}_h^{k+1}(\boldsymbol{x})}{\partial x_s} \cdot \boldsymbol{\varphi}_h d\boldsymbol{x}, \tag{4.80}$$

where $\boldsymbol{z}^{k+1}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}, t_{k+1})$.

From the above considerations we obtain the following semi-implicit scheme: For each $k \geq 0$ find $\boldsymbol{w}_h^{k+1}$ such that

$$\begin{aligned}
&\text{(a)} \quad && \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_{ht_{k+1}}, \\
&\text{(b)} \quad && \left(\frac{\boldsymbol{w}_h^{k+1} - \hat{\boldsymbol{w}}_h^k}{\tau_k}, \boldsymbol{\varphi}_h \right) + b_h^1(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) + d_h^1\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) = 0 \\
& && \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht_{k+1}}, \ k = 0, 1, \ldots, \\
&\text{(c)} \quad && \boldsymbol{w}_h^0 = \Pi_h \boldsymbol{w}^0.
\end{aligned} \tag{4.81}$$

Boundary conditions

In the case of inlet and outlet conditions we use the same conditions like in Section 4.1.3.

On the moving wall we prescribe the impermeability condition

$$\boldsymbol{v} \cdot \boldsymbol{n} = \boldsymbol{z} \cdot \boldsymbol{n}, \tag{4.82}$$

where $\boldsymbol{n}$ is unit outer normal to Γ_{W_t} and $\boldsymbol{z}$ is the wall velocity. We get the normal components of the inviscid flux in the form

$$\sum_{s=1}^{2} \boldsymbol{f}^s(\boldsymbol{w}) n_s = (\boldsymbol{z} \cdot \boldsymbol{n}) + p(0, n_1, n_2, \boldsymbol{v} \cdot \boldsymbol{n})^T. \tag{4.83}$$

4.2.2 ALE formulation II of the Euler equations

The aim of this subsection is to discretize the ALE formulation II of the Euler equations, which was derived in the Section 3.2.2.

Space semidiscretization

We use a similar approach as in the previous Sections 4.1.1 and 4.2.1. We multiply (3.11) by a test function $\boldsymbol{\varphi} \in \boldsymbol{S}_{ht}$, integrate over any element $K_i,\ i \in I_t$, apply Green's theorem and sum over all $i \in I_t$. The resulting formula is

$$\sum_{K_i \in T_{ht}} \int_{K_i} \frac{D^{\mathcal{A}} \boldsymbol{w}(t)}{Dt} \cdot \boldsymbol{\varphi} d\boldsymbol{x} = \sum_{K_i \in T_{ht}} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{g}_s(\boldsymbol{w}(t)) \cdot \frac{\partial \boldsymbol{\varphi}}{\partial x_s} d\boldsymbol{x} \tag{4.84}$$

$$- \sum_{K_i \in T_{ht}} \sum_{j \in S_t(i)} \int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{g}_s(\boldsymbol{w}(t)) \left(n_{ij}\right)_s \cdot \boldsymbol{\varphi} dS - \sum_{K_i \in T_{ht}} \int_{K_i} \mathrm{div} \boldsymbol{z} \left(\boldsymbol{w} \cdot \boldsymbol{\varphi}\right) d\boldsymbol{x},$$

where we use the same relation for the flux g_s presented in the experession (3.12). We apply again the approximation of fluxes through the face Γ_{ij} by a numerical flux $\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{n})$. It means that

$$\int_{\Gamma_{ij}} \sum_{s=1}^{2} \boldsymbol{g}_s(\boldsymbol{w}) \left(n_{ij}\right)_s \cdot \boldsymbol{\varphi} dS \approx \int_{\Gamma_{ij}} \boldsymbol{H}_g(\boldsymbol{w}_h(t)|_{\Gamma_{ij}}, \boldsymbol{w}_h(t)|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi} dS. \tag{4.85}$$

Here $\boldsymbol{H}_g$ is an analogy to the Vijayasundaram numerical flux consistent with the fluxes $\boldsymbol{g}_s,\ s = 1, 2$.

Taking into account that

$$\frac{D\boldsymbol{g}_s(\boldsymbol{w})}{D\boldsymbol{w}} = \frac{D\boldsymbol{f}_s(\boldsymbol{w})}{D\boldsymbol{w}} - z_s \mathbb{I} = \mathbb{A}_s - z_s \mathbb{I}, \tag{4.86}$$

we can write

$$\tilde{\mathbb{P}}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \frac{D\boldsymbol{g}_s(\boldsymbol{w})}{D\boldsymbol{w}} n_s = \sum_{s=1}^{2} \left(\mathbb{A}_s n_s - z_s n_s \mathbb{I}\right) = \mathbb{P}(\boldsymbol{w}, \boldsymbol{n}) - (\boldsymbol{z} \cdot \boldsymbol{n}) \mathbb{I}, \tag{4.87}$$

where matrix $\mathbb{P}$ is the same one as in (2.14). In view of (2.30), there exists a nonsingular matrix $\mathbb{T}$ such that $\tilde{\mathbb{P}} = \mathbb{T} \tilde{\mathit{\Lambda}} \mathbb{T}^{-1}$, where $\mathit{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and $\lambda_1 = \boldsymbol{v} \cdot \boldsymbol{n} - a,\ \lambda_2 = \lambda_3 = \boldsymbol{v} \cdot \boldsymbol{n},\ \lambda_1 = \boldsymbol{v} \cdot \boldsymbol{n} + a$. This implies that

$$\tilde{\mathbb{P}} = \mathbb{T} \tilde{\mathit{\Lambda}} \mathbb{T}^{-1}, \quad \tilde{\mathit{\Lambda}} = \mathrm{diag}(\lambda_1 - \boldsymbol{z} \cdot \boldsymbol{n}, \lambda_2 - \boldsymbol{z} \cdot \boldsymbol{n}, \lambda_3 - \boldsymbol{z} \cdot \boldsymbol{n}, \lambda_4 - \boldsymbol{z} \cdot \boldsymbol{n}). \tag{4.88}$$

Now we define $\tilde{\mathbb{P}^+}$ and $\tilde{\mathbb{P}^-}$, "positive" and "negative" parts of the matrix $\tilde{\mathbb{P}}$, as in (4.15) and introduce the modification of the Vijayasundaram numerical flux

$$\boldsymbol{H}_g(\boldsymbol{w}_L, \boldsymbol{w}_R, \boldsymbol{n}) = \tilde{\mathbb{P}}^+ \left(\frac{\boldsymbol{w}_L + \boldsymbol{w}_R}{2}, \boldsymbol{n}\right) \boldsymbol{w}_L + \tilde{\mathbb{P}}^- \left(\frac{\boldsymbol{w}_L + \boldsymbol{w}_R}{2}, \boldsymbol{n}\right) \boldsymbol{w}_R. \tag{4.89}$$

Now, we define the forms

$$(\frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt},\boldsymbol{\varphi}_h)_h = \int_{\Omega_{ht}} \frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt}\cdot\boldsymbol{\varphi}_h d\boldsymbol{x}, \tag{4.90}$$

$$\tilde{b}_h^2(\boldsymbol{w}_h,\boldsymbol{\varphi}_h) = -\sum_{K_i\in T_{ht}}\int_{K_i}\sum_{s=1}^{2}\boldsymbol{g}_s(\boldsymbol{w}(t))\cdot\frac{\partial\boldsymbol{\varphi}}{\partial x_s}d\boldsymbol{x} + \sum_{K_i\in T_{ht}}\sum_{j\in S_t(i)}\int_{\Gamma_{ij}}\boldsymbol{H}_g(\boldsymbol{w}_h(t)|_{\Gamma_{ij}},\boldsymbol{w}_h(t)|_{\Gamma_{ji}},\boldsymbol{n}_{ij})\cdot\boldsymbol{\varphi}_h dS, \tag{4.91}$$

$$d_h^2(\boldsymbol{w}_h,\boldsymbol{\varphi}_h) = -\sum_{K_i\in T_{ht}}\int_{K_i}\mathrm{div}\boldsymbol{z}\,(\boldsymbol{w}\cdot\boldsymbol{\varphi})\,d\boldsymbol{x}. \tag{4.92}$$

We obtain an *approximate solution of* (3.11) as a function $\boldsymbol{w}_h = \boldsymbol{w}_h(t)$ satisfying the conditions

$$\text{(a)}\quad \boldsymbol{w}_h(t)\in\boldsymbol{S}_{ht},\ \forall t\in[0,T], \tag{4.93}$$

$$\text{(b)}\quad (\frac{D^{\mathcal{A}}\boldsymbol{w}_h(t)}{Dt},\boldsymbol{\varphi}_h)_h + \tilde{b}_h^2(\boldsymbol{w}_h(t),\boldsymbol{\varphi}_h) - d_h^2(\boldsymbol{w}_h(t),\boldsymbol{\varphi}_h) = 0 \quad \forall\boldsymbol{\varphi}_h\in\boldsymbol{S}_{ht},\ \forall t\in(0,T),$$

$$\text{(c)}\quad \boldsymbol{w}_h(0) = \Pi_h\boldsymbol{w}^0,$$

where $\Pi_h\boldsymbol{w}^0$ is the L^2-projection of $\boldsymbol{w}^0$ from the initial condition

$$\boldsymbol{w}(\boldsymbol{x},0) = \boldsymbol{w}^0(\boldsymbol{x}),\quad \boldsymbol{x}\in\Omega_0 \tag{4.94}$$

on the space $\boldsymbol{S}_{h0}$.

Time discretization

The process of the time discretization is carried out similarly as in Section 4.2.1. We define a partially linearized form b_h^2 to the form $\tilde{b}_h^2$:

$$b_h^2(\hat{\boldsymbol{w}}_h^k,\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h) = -\sum_{K_i\in T_{ht_{k+1}}}\int_{K_i}\sum_{s=1}^{2}(\mathbb{A}_s(\hat{\boldsymbol{w}}^k(\boldsymbol{x}))-z_s(\boldsymbol{x}))\mathbb{I})\boldsymbol{w}^{k+1}(\boldsymbol{x}))\cdot\frac{\partial\boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s}d\boldsymbol{x}$$
$$+\sum_{K_i\in T_{ht_{k+1}}}\sum_{j\in S_{t_{k+1}}(i)}\int_{\Gamma_{ij}}\left[\tilde{\mathbb{P}}^+\left(\frac{\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ij}}+\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}}{2},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}\right. \tag{4.95}$$
$$\left.+\tilde{\mathbb{P}}^-\left(\frac{\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ij}}+\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}}{2},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}\right]\cdot\boldsymbol{\varphi}_h dS.$$

The term d_h^2 will again be treated implicitly.

These considerations lead us to the following semi-implicit scheme: For each $k\geq 0$ find $\boldsymbol{w}_h^{k+1}$ such that

$$\text{(a)}\quad \boldsymbol{w}_h^{k+1}\in\boldsymbol{S}_{ht_{k+1}},$$

$$\text{(b)}\quad \left(\frac{\boldsymbol{w}_h^{k+1}-\hat{\boldsymbol{w}}_h^k}{\tau_k},\boldsymbol{\varphi}_h\right) + b_h^2(\hat{\boldsymbol{w}}_h^k,\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h) - d_h^2\left(\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h\right) = 0 \tag{4.96}$$
$$\forall\boldsymbol{\varphi}_h\in\boldsymbol{S}_{ht_{k+1}},\ k=0,1,\ldots,$$

$$\text{(c)}\quad \boldsymbol{w}_h^0 = \Pi_h\boldsymbol{w}^0.$$

Chapter 5

Flow in the channel with oscilating walls

In this chapter we shall present the choice of the ALE mapping that we use in our numerical experiments and its properties.

Let us recall the basic characteristics of the ALE method presented in Chapter 3. We choose reference (ALE) coordinates $\boldsymbol{X} = (X_1, X_2)$ like the Cartesian coordinates at time $t = 0$.

5.1 Construction of ALE mapping

We assume that the inlet of the channel is an abscissa forming a part of line $X_1 = a$ and the outlet is also an abscissa forming a part of line $X_1 = b$, where $a, b \in I\!R,\ a < b$. Next we describe the dependence of the movement of the wall on time by the following functions

$$\begin{array}{ll} \text{upper wall} & x_2 = \phi(X_1, t), \quad X_1 \in [a, b]\,,\ t \in [0, T] \\ \text{lower wall} & x_2 = \psi(X_1, t), \quad X_1 \in [a, b]\,,\ t \in [0, T]\,, \end{array}$$

where $\phi(X_1, t) > \psi(X_1, t)$ for all $X_1 \in [a, b]\,,\ t \in [0, T]$. Let us assume ϕ and ψ are smooth function, i.e. $\phi, \psi \in \mathcal{C}^1([a, b] \times [0, T])$. This means that for each $t \in [0, T]\,$, $\Omega_t = \{(x_1, x_2);\ \psi(X_1, t) < x_2 < \phi(X_1, t),\ x_1 = X_1 \in (a, b)\}$.

We choose a linear transformation of the coordinate x_2 and define the ALE mapping by

$$\begin{aligned} & x_1(\boldsymbol{X}, t) = X_1, \\ & x_2(\boldsymbol{X}, t) = \psi(X_1, t) + \frac{X_2 - \psi(X_1, 0)}{\phi(X_1, 0) - \psi(X_1, 0)} \left(\phi(X_1, t) - \psi(X_1, t)\right), \\ & (X_1, X_2) \in \Omega_0,\ \mathcal{A}_t(\boldsymbol{X}) = \left(x_1\left(\boldsymbol{X}, t\right), x_2\left(\boldsymbol{X}, t\right)\right). \end{aligned} \tag{5.1}$$

The inverse $\mathcal{A}_t^{-1} : \bar{\Omega}_t \to \bar{\Omega}_0$ to the ALE mapping has the form

$$\begin{aligned} X_1 = \left(\mathcal{A}_t^{-1}\left(x_1, x_2\right)\right)_1 = X_1\left(x_1, x_2, t\right) &= x_1, \\ X_2 = \left(\mathcal{A}_t^{-1}\left(x_1, x_2\right)\right)_2 = X_2\left(x_1, x_2, t\right) &= \left(x_2 - \psi(x_1, t)\right) \frac{\phi(x_1, 0) - \psi(x_1, 0)}{\phi(x_1, t) - \psi(x_1, t)} + \psi(x_1, 0) \end{aligned} \tag{5.2}$$

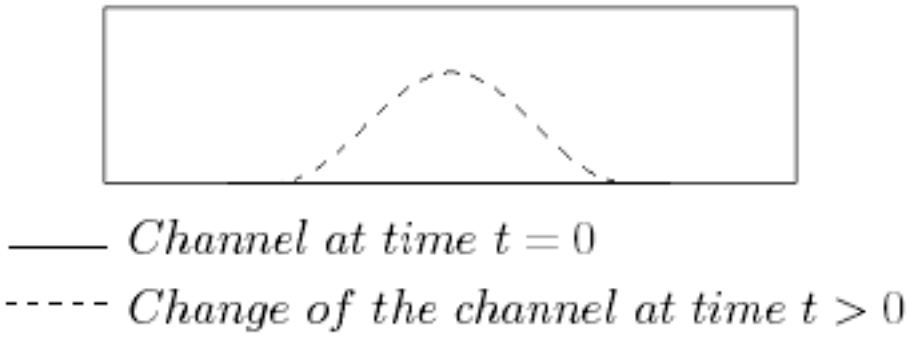


Figure 5.1: Channel with moving walls

for all $\boldsymbol{x} = (x_1, x_2) \in \bar{\Omega}_t$ and all $t \in [0, T]$.

Now, we can derive the domain velocity $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2)$ in the reference coordinates $\boldsymbol{X} \in \bar{\Omega}_0$ by the form (3.3) and transform it to the space coordinates $\boldsymbol{x} \in \bar{\Omega}_t$ by the relation (3.4). For more details see Chapter 3. Using these relations and the independence of first coordinate of our chosen ALE mapping on time, we obtain

$$\tilde{z}_1(\boldsymbol{X}, t) = \frac{\partial}{\partial t} X_1 = 0. \tag{5.3}$$

From (5.2) we get the second coordinate of domain velocity in the reference coordinates $\boldsymbol{X} \in \bar{\Omega}_0$:

$$\begin{aligned} \tilde{z}_1(\boldsymbol{X}, t) &= \frac{\partial}{\partial t}\left(\psi(X_1, t) + \frac{X_2 - \psi(X_1, 0)}{\phi(X_1, 0) - \psi(X_1, 0)}\left(\phi(X_1, t) - \psi(X_1, t)\right)\right) \\ &= \frac{\partial \psi}{\partial t}(X_1, t) + \frac{X_2 - \psi(X_1, 0)}{\phi(X_1, 0) - \psi(X_1, 0)}\left(\frac{\partial \phi}{\partial t}(X_1, t) - \frac{\partial \psi}{\partial t}(X_1, t)\right) \end{aligned} \tag{5.4}$$

With the aid of (3.4) and (5.3) we express the domain velocity $\boldsymbol{z} = (z_1, z_2)$ in the space coordinates:

$$z_1(\boldsymbol{x}, t) = 0, \tag{5.5}$$

$$z_2(\boldsymbol{x}, t) = \tfrac{\partial \psi}{\partial t}(x_1, t) + \tfrac{x_2 - \psi(x_1, t)}{\phi(x_1, 0) - \psi(x_1, 0)}\left(\tfrac{\partial \phi}{\partial t}(x_1, t) - \tfrac{\partial \psi}{\partial t}(x_1, t)\right). \tag{5.6}$$

5.2 Example

In this section we shall be concerned with an example of the ALE mapping used in our numerical experiments. We assume that

$$\begin{gathered} a = -2,\ b = 2, \\ \psi(X_1, t) = \alpha \sin t\ (\cos(\pi X_1) + 1), \quad X_1 \in [-1, 1],\ t \in [0, T], \\ \psi(X_1, t) = 0, \quad X_1 \in [-2, 1) \cup (1, 2],\ t \in [0, T], \\ \phi(X_1, t) = 1, \quad X_1 \in [-2, 2],\ t \in [0, T], \end{gathered} \tag{5.7}$$

where $\alpha \in \left[0, \frac{1}{2}\right]$ is a parameter determining the amplitude of deformation of the lower wall. The upper wall is formed by a straight segment.

From the definition of the functions ψ and ϕ we calculate their time derivative:

$$\begin{aligned} \tfrac{\partial \phi}{\partial t}(X_1,t) = 0 \quad \forall X_1 \in [-2,2]\,,\ t \in [0,T]\,, \\ \tfrac{\partial \psi}{\partial t}(X_1,t) = 0 \quad \forall X_1 \in [-2,-1)\cup(1,2]\,,\ t \in [0,T]\,, \\ \tfrac{\partial \psi}{\partial t}(X_1,t) = \alpha\,\cos t\,(\cos(\pi X_1)+1) \quad \forall X_1 \in [-1,1]\,,\ t \in [0,T]\,. \end{aligned} \tag{5.8}$$

It shows us that the function $\frac{\partial \psi}{\partial t}$ is continuous in its domain of definition. Since the functions ψ and ϕ are independent of the second reference coordinate X_2, it is sufficient for verifying the smoothness of the ALE mapping to explore the continuity of functions $\frac{\partial \phi}{\partial X_1}$, $\frac{\partial \psi}{\partial X_1}$. It is an easy consequence of the relation

$$\begin{aligned} \tfrac{\partial \phi}{\partial X_1} = 0 \quad \forall X_1 \in [-2,2]\,,\ t \in [0,T]\,, \\ \tfrac{\partial \psi}{\partial X_1}(X_1,t) = 0 \quad \forall X_1 \in [-2,-1)\cup(1,2]\,,\ t \in [0,T]\,, \\ \tfrac{\partial \psi}{\partial X_1}(X_1,t) = -\pi\alpha\,\sin t\,(\sin(\pi X_1) \quad \forall X_1 \in [-1,1]\,,\ t \in [0,T]\,. \end{aligned} \tag{5.9}$$

Then, we can see that the ALE mapping $\mathcal{A}_t \in \mathcal{C}^1(\bar{\Omega}_0 \times [0,T])$.

Chapter 6

Algorithm development

6.1 Basis functions

Here we shall present the basis functions for linear elements P^1 and quadratic elements P^2.

For linear elements P^1, such basis $\{\varphi_{in} \in S_h;\ i \in I,\ n = 1, 2, 3\}$ is used that $\varphi_{in}(P_{i'}^{n'}) = \delta_{ii'}\delta_{nn'}$, where $P_i^n,\ n = 1, 2, 3$, are vertices of element K_i and δ is the Kronecker symbol.

For quadratic elements P^2 we use such basis $\{\varphi_{in} \in S_h;\ i \in I,\ n = 1, \ldots, 6\}$ that $\varphi_{in}(P_{i'}^{n'}) = \delta_{ii'}\delta_{nn'}$, where $P_i^n,\ n = 1, 2, 3$, are vertices of element K_i and $P_i^n,\ n = 4, 5, 6$, are midpoints of edges of K_i. These are standard local basis functions as known from the Finite Element Method and they work quite well.

Experiments were done with the simple monomial basis $1, x, y, x^2, y^2, xy$ as an alternative, for which evaluation is simpler. However, the latter basis is very "non-orthogonal" and local mass matrices $\mathbb{B}_i$ are ill-conditioned, causing a great loss of accuracy.

6.2 Construction of the linear system

6.2.1 Time-independent domain

For the construction of the linear system we have to come back to the time-discretization of our problem described in Section 4.1.2. Let us shortly remind the most important point of the time discretization. We assume a partition $0 = t_0 < t_1 < \ldots$ of the time interval $[0, T]$ with a time step $\tau_k = t_{k+1} - t_k$. We seek $\boldsymbol{w}_h^k \approx \boldsymbol{w}_h(t_k)$ such that

$$
\begin{aligned}
&(a) \quad \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_h, \\
&(b) \quad \left(\frac{\boldsymbol{w}_h^{k+1} - \boldsymbol{w}_h^k}{\tau_k}, \boldsymbol{\varphi}_h\right) + \tilde{b}_h(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h,\ k = 0, 1, \ldots, \\
&(c) \quad \boldsymbol{w}_h^0 = \tilde{\boldsymbol{w}}_h^0,
\end{aligned}
\tag{6.1}
$$

where $\tilde{\boldsymbol{w}}_h^0$ is an $\boldsymbol{S}_h$ approximation of the initial condition $\boldsymbol{w}^0$. This scheme leads to a large system of highly nonlinear equations due to the nonlinearity of the form $\tilde{b}_h$

in the variable $\boldsymbol{w}_h^{k+1}$. The numerical solution of such a system is very complicated and time consuming. Therefore in [7] a simplified linearization of problem (6.2) is presented in order to obtain a large (sparse) system of linear equations rather than solving a nonlinear system.

We shall treat the interior and boundary terms in (4.7) separately:

$$b_h(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) = -\tilde{\sigma}_1 + \tilde{\sigma}_2, \tag{6.2}$$

where

$$\tilde{\sigma}_1 = \sum_{K_i \in T_h} \int_{K_i} \sum_{s=1}^{2} \boldsymbol{f}_s(\boldsymbol{w}_h^{k+1}) \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} d\boldsymbol{x}, \tag{6.3}$$

$$\tilde{\sigma}_2 = \sum_{K_i \in T_h} \sum_{j \in S(i)} \int_{\Gamma_{ij}} \boldsymbol{H}(\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}, \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \cdot \boldsymbol{\varphi}_h dS. \tag{6.4}$$

For $\tilde{\sigma}_1$ we use the property of the Euler fluxes $\boldsymbol{f}_s$ given in Section 2.1 by the form (2.16) and set

$$\sigma_1 := \sum_{K_i \in T_h} \int_{K_i} \sum_{s=1}^{2} \mathbb{A}_s(\boldsymbol{w}_h^k) \boldsymbol{w}_h^{k+1} \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_s} d\boldsymbol{x}. \tag{6.5}$$

In order to treat the term $\tilde{\sigma}_2$, we shall use the method of numerical flux described in Sections 4.1.1 and 4.1.2. It means that for interior edges the form (4.18) is used.

For edges $\Gamma_{ij} \subset \Gamma_{IO}$ we cannot simply apply this linearization since we have no information about $\boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}$ - this is caused by the fact that the Inlet and Outlet boundary conditions are not a priori given and can change roles for complex flows. A simple solution is to treat these terms explicitly, i.e. $\boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}} \approx \boldsymbol{w}_h^k|_{\Gamma_{ji}}$, where the latter state is calculated using a method from Section 4.1.3. In contrast to [7], we consider more suitable to use here $\boldsymbol{w}_{ij}^{k+1}$ (instead of $\boldsymbol{w}_{ij}^k$ from [7]). Thus, inlet and outlet terms have the form:

$$\sum_{K_i \in T_h} \sum_{j \in \gamma_{IO}(i)} \int_{\Gamma_{ij}} \left[\mathbb{P}^+ \left(\left\langle \boldsymbol{w}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} + \mathbb{P}^- \left(\left\langle \boldsymbol{w}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij} \right) \boldsymbol{w}_h^k|_{\Gamma_{ji}} \right] \cdot \boldsymbol{\varphi}_h dS, \tag{6.6}$$

where $\gamma_{IO}(i) = \{ j \in S(i);\ \Gamma_{ij} \subset \Gamma_{IO} \}$.

For $\Gamma_{ij} \subset \Gamma_W$ special treatment is needed - according to Section 4.1.3 we put

$$\begin{aligned} &\sum_{i \in I} \sum_{j \in \gamma_W(i)} \int_{\Gamma_{ij}} \boldsymbol{H}\left(\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}, \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}\right) \cdot \boldsymbol{\varphi}_h dS \\ &\quad \approx \sum_{i \in I} \sum_{j \in \gamma_W(i)} \int_{\Gamma_{ij}} \boldsymbol{F}_W\left(\boldsymbol{w}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{n}_{ij}\right) \cdot \boldsymbol{\varphi}_h dS. \end{aligned} \tag{6.7}$$

Finally we can define the linearized edge term as

$$\begin{aligned}\sigma_2 \;:=\; & \sum_{i\in I}\sum_{j\in s(i)}\int_{\Gamma_{ij}}\Big[\mathbb{P}^+\left(\left\langle \boldsymbol{w}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \\ & +\mathbb{P}^-\left(\left\langle \boldsymbol{w}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}\Big]\cdot\boldsymbol{\varphi}_h dS \\ & +\sum_{i\in I}\sum_{j\in\gamma_{IO}(i)}\int_{\Gamma_{ij}}\Big[\mathbb{P}^+\left(\left\langle \boldsymbol{w}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \\ & +\mathbb{P}^-\left(\left\langle \boldsymbol{w}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k}|_{\Gamma_{ji}}\Big]\cdot\boldsymbol{\varphi}_h dS \\ & +\sum_{i\in I}\sum_{j\in\gamma_W}\int_{\Gamma_{ij}}\mathbb{P}(\boldsymbol{w}_h^k,\boldsymbol{n}_{ij})\boldsymbol{w}_h^{k+1}\cdot\boldsymbol{\varphi}_h dS.\end{aligned} \tag{6.8}$$

Now we obtained the semi-implicit linearized form as

$$b_h^{SI}(\boldsymbol{w}_h^k,\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h)=-\sigma_1+\sigma_2, \tag{6.9}$$

where σ_1 and σ_2 are given in (6.9) and (6.5) respectively. We obtain the semi-implicit linearized scheme satisfying the conditions:

$$\begin{aligned}&(a)\quad \boldsymbol{w}_h^{k+1}\in\boldsymbol{S}_h, \\ &(b)\quad \left(\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h\right)+\tau_k b_h^{SI}(\boldsymbol{w}_h^k,\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h)=\left(\boldsymbol{w}_h^k,\boldsymbol{\varphi}_h\right)\quad \forall\boldsymbol{\varphi}_h\in\boldsymbol{S}_h,\ k=0,1,\ldots, \\ &(c)\quad \boldsymbol{w}_h^0=\tilde{\boldsymbol{w}}_h^0\end{aligned} \tag{6.10}$$

with $\tilde{\boldsymbol{w}}_h^0$ representing an $\boldsymbol{S}_h$ approximation of initial condition $\boldsymbol{w}^0$.

Let $\mathcal{B}=\{\boldsymbol{w}_\alpha\}_{\alpha=1}^n$ be a basis in the space $\boldsymbol{S}_h$ with $n=\dim\boldsymbol{S}_h$. We seek the approximate solution $\boldsymbol{w}_h\in\boldsymbol{S}_h$ in the form

$$\boldsymbol{w}_h(t)=\sum_{\alpha=1}^n\xi_\alpha(t)\boldsymbol{w}_\alpha. \tag{6.11}$$

Due to the linearity of the form b_h^{SI}in the variable $\boldsymbol{\varphi}$, we can use as test functions only elements of the basis $\mathcal{B}$.

Let $\boldsymbol{\xi}^k=(\xi_1^k,\ldots,\xi_n^k)$. Then our semi-implicit scheme (6.11) can be written in the matrix representation

$$\mathbb{A}(\boldsymbol{\xi}^k)\boldsymbol{\xi}^{k+1}=\boldsymbol{g}(\boldsymbol{\xi}^k), \tag{6.12}$$

where $\boldsymbol{g}:I\!R^n\to I\!R^n$ and $\mathbb{A}:I\!R^n\to I\!R^{n\times n}$ is an $n\times n$ nonsymmetric matrices. This matrix has the form $\mathbb{A}=\mathbb{M}+\tau_k\mathbb{B}$, where $\mathbb{M}=\{m_{ij}\}_{i,j=1}^n$ is the symmetric, positive definite mass matrix with entries $m_{ij}=\int_\Omega\boldsymbol{w}_i\cdot\boldsymbol{w}_j d\boldsymbol{x}$ and $\mathbb{B}$ represents the form b_h^{SI}.

We must take into account that the basis functions are in the space $\boldsymbol{S}_h=[S_h]^4$. We use the P^1 and P^2 basis functions for S_h as in Section 6.1 'separately' for each component and get the basis for $[S_h]^4$. Then we write (6.11) as

$$\boldsymbol{w}_h^k(\boldsymbol{x})=\sum_{i\in I}\sum_{j=1}^{n_p}\sum_{l=1}^{4}\xi_{ijl}(t_k)\boldsymbol{w}_{ijl}(\boldsymbol{x}), \tag{6.13}$$

where $\operatorname{supp} \boldsymbol{w}_{ijl} \subset K_i$, n_p = number of degrees of freedom for $P^p(K_i)$, $n_0 = 1, n_1 = 3, n_2 = 6$ and $\boldsymbol{w}_{ijl}^{k,(m)} = 0$, if $m \neq l$, where we use the following notation described with the aid of a vector $\boldsymbol{u}$: $\boldsymbol{u}^{(m)} = m$-th component of vector $\boldsymbol{u}$. Using this representation we are able to 'cluster' the basis functions with common support elements and representing the same unknown (i.e. with common nonzero component) thus achieving the block-diagonality of $\mathbb{M}$ - with $n_p \times n_p$ blocks. Let us note that in practice the order p of approximation can be chosen separetly for every component of the state vector. Thus n_p becomes $n_{p(l)}$. This option is incorporated into the implementation of the presented scheme.

Since $\mathbb{M}$ is nonsingular, we can expect that for small τ_k the matrix $\mathbb{A}$ is also nonsingular. Furthermore, for sufficiently small τ_k the matrix will be close to the block diagonal matrix. One can therefore expect better behavior of the linear solver. On the other hand, we want to avoid the limitations imposed on τ_k via the $CLF-$condition. If we choose large τ_k, then we may expect slower convergence of an iterative linear solver.

6.2.2 Time-dependent domain

Here we describe the above process for the case of a time-dependent domain.

ALE formulation I

We shall again pass throught the time-discretization described in Section 4.2.1. For a partition $0 = t_0 < t_1 < \dots$ of the time interval $[0, T]$ with a time step $\tau_k = t_{k+1} - t_k$ we obtain the semi-implicit linearized scheme satisfying the conditions:

$$
\begin{aligned}
&\text{(a)} \quad \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_{ht_{k+1}}, \\
&\text{(b)} \quad \left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) + \tau_k b_h^{1SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) + \tau_k d_h^1\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) = \left(\boldsymbol{w}_h^k, \boldsymbol{\varphi}_h\right) \\
&\qquad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht_{k+1}},\ k = 0, 1, \dots, \\
&\text{(c)} \quad \boldsymbol{w}_h^0 = \tilde{\boldsymbol{w}}_h^0
\end{aligned}
\tag{6.14}
$$

with $\tilde{\boldsymbol{w}}_h^0$ representing an $\boldsymbol{S}_{h0}$ approximation of the initial condition $\boldsymbol{w}^0$. We use the same notation as in Section 4.2.1, i.e.

$$
\hat{\boldsymbol{w}}_h^j(\boldsymbol{x}) = \boldsymbol{w}^j\left(\mathcal{A}_{t_j}\left(\mathcal{A}_{t_{k+1}}^{-1}\right)(\boldsymbol{x})\right), \quad \boldsymbol{x} \in \Omega_{t_{k+1}}.
$$

We shall describe in detail the forms $b_h^{1SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h)$ and $d_h^1\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right)$. We write these two forms in the following way:

$$
\begin{aligned}
b_h^{1SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) \;=\; & -\sum_{K_i \in T_{ht_{k+1}}} \int_{K_i} \sum_{s=1}^{2} \mathbb{A}_s(\hat{\boldsymbol{w}}_h^k(\boldsymbol{x})) \boldsymbol{w}_h^{k+1}(\boldsymbol{x}) \cdot \frac{\partial \boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s} d\boldsymbol{x} \\
& + \sum_{i \in I_{t_{k+1}}} \sum_{j \in s_{t_{k+1}}(i)} \int_{\Gamma_{ij}} \Big[\mathbb{P}^+\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \\
& + \mathbb{P}^-\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}\Big] \cdot \boldsymbol{\varphi}_h dS
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i\in I_{t_{k+1}}}\sum_{j\in\gamma_{t_{k+1}IO}(i)}\int_{\Gamma_{ij}}\Big[\mathbb{P}^{+}\left(\left\langle\hat{\boldsymbol{w}}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}}\\
&+\mathbb{P}^{-}\left(\left\langle\hat{\boldsymbol{w}}_h^k\right\rangle_{ij},\boldsymbol{n}_{ij}\right)\hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}\Big]\cdot\boldsymbol{\varphi}_h dS\\
&+\sum_{i\in I_{t_{k+1}}}\sum_{j\in\gamma_{t_{k+1}W}}\int_{\Gamma_{ij}}\mathbb{P}(\hat{\boldsymbol{w}}_h^k,\boldsymbol{n}_{ij})\boldsymbol{w}_h^{k+1}\cdot\boldsymbol{\varphi}_h dS, \qquad (6.15)
\end{aligned}
$$

$$
d_h^1\left(\boldsymbol{w}_h^{k+1},\boldsymbol{\varphi}_h\right) = -\sum_{K_i\in T_{ht_{k+1}}}\int_{K_i}\sum_{s=1}^{2}z_s^{k+1}\frac{\partial\boldsymbol{w}_h^{k+1}}{\partial x_s}\cdot\boldsymbol{\varphi}_h d\boldsymbol{x}, \qquad (6.16)
$$

where $\boldsymbol{z}^{k+1}(\boldsymbol{x})=\boldsymbol{z}(\boldsymbol{x},t_{k+1})$.

Let $\mathcal{B}=\{\boldsymbol{w}_\alpha\}_{\alpha=1}^n$ be a basis in the space $\boldsymbol{S}_{ht_{k+1}}$ with $n=\dim\boldsymbol{S}_{ht_{k+1}}$. We seek the approximate solution $\boldsymbol{w}_h^{k+1}\in\boldsymbol{S}_{ht_{k+1}}$ in the form

$$
\boldsymbol{w}_h^{k+1}=\sum_{\alpha=1}^{n}\xi_\alpha^{k+1}\boldsymbol{w}_\alpha. \qquad (6.17)
$$

Similarly we set

$$
\hat{\boldsymbol{w}}_h^{k}=\sum_{\alpha=1}^{n}\xi_\alpha^{k}\boldsymbol{w}_\alpha. \qquad (6.18)
$$

Due to the linearity of the form b_h^{1SI}in the variable $\boldsymbol{\varphi}$, we can use as test functions only elements of the basis $\mathcal{B}$.

Let $\boldsymbol{\xi}^k=(\xi_1^k,\ldots,\xi_n^k)$. Then our semi-implicit scheme (6.15) can be written in the matrix representation

$$
\mathbb{A}(\boldsymbol{\xi}^k)\boldsymbol{\xi}^{k+1}=\boldsymbol{q}(\boldsymbol{\xi}^k), \qquad (6.19)
$$

where $\boldsymbol{q}:I\!R^n\to I\!R^n$ and $\mathbb{A}:I\!R^n\to I\!R^{n\times n}$ is an $n\times n$ nonsymmetric matrices. This matrix has the form $\mathbb{A}=\mathbb{M}+\tau_k\mathbb{B}^1+\tau_k\mathbb{D}^1$, where $\mathbb{M}=\{m_{ij}\}_{i,j=1}^n$ is the symmetric, positive definite mass matrix with entries $m_{ij}=\int_\Omega\boldsymbol{w}_i\cdot\boldsymbol{w}_j d\boldsymbol{x}$, $\mathbb{B}^1$ represents the form b_h^{1SI} and $\mathbb{D}^1$ represents the form d_h^1.

Using the P^1 and P^2 basis functions for $S_{ht_{k+1}}$, respectively S_{ht_k} as in Section 6.1 'separetly' for each component we get the basis for $\boldsymbol{S}_{ht_{k+1}}=\left[S_{ht_{k+1}}\right]^4$, respectively $\boldsymbol{S}_{ht_k}=\left[S_{ht_k}\right]^4$. Then we write (6.17) and (6.18)

$$
\boldsymbol{w}_h^{k+1}(\boldsymbol{x}) = \sum_{i\in I}\sum_{j=1}^{n_p}\sum_{l=1}^{4}\xi_{ijl}(t_{k+1})\boldsymbol{w}_{ijl}(\boldsymbol{x}), \qquad (6.20)
$$

$$
\hat{\boldsymbol{w}}_h^{k}(\boldsymbol{x}) = \sum_{i\in I}\sum_{j=1}^{n_p}\sum_{l=1}^{4}\xi_{ijl}(t_{k})\boldsymbol{w}_{ijl}(\boldsymbol{x}), \qquad (6.21)
$$

where the same notation as in Section 6.2.1 is used.

ALE formulation II

For the ALE formulation II we shall apply the same process as for the ALE formulation I. We again assume a partition $0=t_0<t_1<\ldots$ of the time interval $[0,T]$ with

a time step $\tau_k = t_{k+1} - t_k$. The obtained semi-implicit linearized scheme satisfies the conditions:

$$\begin{aligned}
&\text{(a)} \quad && \boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_{ht_{k+1}}, \\
&\text{(b)} \quad && \left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) + \tau_k b_h^{2SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) - \tau_k d_h^2\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) = \left(\boldsymbol{w}_h^k, \boldsymbol{\varphi}_h\right) \\
& && \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_{ht_{k+1}},\ k = 0, 1, \ldots, \\
&\text{(c)} \quad && \boldsymbol{w}_h^0 = \tilde{\boldsymbol{w}}^0
\end{aligned} \tag{6.22}$$

with $\tilde{\boldsymbol{w}}^0$ representing an $\boldsymbol{S}_{ht_{k+1}}$ approximation of the initial condition $\boldsymbol{w}^0$ and the forms $b_h^{2SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h)$ and $d_h^2\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right)$:

$$\begin{aligned}
b_h^{2SI}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h) \;=\; & -\sum_{K_i \in T_{ht_{k+1}}} \int_{K_i} \sum_{s=1}^{2} (\mathbb{A}_s(\hat{\boldsymbol{w}}^k(\boldsymbol{x})) - z_s^{k+1}(\boldsymbol{x}))\mathbb{I})\boldsymbol{w}^{k+1}(\boldsymbol{x})) \cdot \frac{\partial \boldsymbol{\varphi}_h(\boldsymbol{x})}{\partial x_s} d\boldsymbol{x}, \\
& + \sum_{i \in I_{t_{k+1}}} \sum_{j \in s_{t_{k+1}}(i)} \int_{\Gamma_{ij}} \Big[\tilde{\mathbb{P}}^+\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \\
& + \tilde{\mathbb{P}}^-\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ji}}\Big] \cdot \boldsymbol{\varphi}_h dS \\
& + \sum_{i \in I_{t_{k+1}}} \sum_{j \in \gamma_{t_{k+1}IO}(i)} \int_{\Gamma_{ij}} \Big[\tilde{\mathbb{P}}^+\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \boldsymbol{w}_h^{k+1}|_{\Gamma_{ij}} \\
& + \tilde{\mathbb{P}}^-\left(\left\langle \hat{\boldsymbol{w}}_h^k \right\rangle_{ij}, \boldsymbol{n}_{ij}\right) \hat{\boldsymbol{w}}_h^k|_{\Gamma_{ji}}\Big] \cdot \boldsymbol{\varphi}_h dS \\
& + \sum_{i \in I_{t_{k+1}}} \sum_{j \in \gamma_{t_{k+1}W}} \int_{\Gamma_{ij}} \tilde{\mathbb{P}}(\hat{\boldsymbol{w}}_h^k, \boldsymbol{n}_{ij}) \boldsymbol{w}_h^{k+1} \cdot \boldsymbol{\varphi}_h dS,
\end{aligned} \tag{6.23}$$

$$d_h^2\left(\boldsymbol{w}_h^{k+1}, \boldsymbol{\varphi}_h\right) \;=\; -\sum_{K_i \in T_{ht_{k+1}}} \int_{K_i} \operatorname{div} \boldsymbol{z}^{k+1}\left(\hat{\boldsymbol{w}}_h^{k+1} \cdot \boldsymbol{\varphi}_h\right) d\boldsymbol{x}. \tag{6.24}$$

Let us note that we use the following notation $\tilde{\mathbb{P}}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{s=1}^{2} \frac{D\boldsymbol{g}_s(\boldsymbol{w})}{D\boldsymbol{w}} n_s$, where $\frac{D\boldsymbol{g}_s(\boldsymbol{w})}{D\boldsymbol{w}} = \frac{D\boldsymbol{f}_s(\boldsymbol{w})}{D\boldsymbol{w}} - z_s\mathbb{I} = \mathbb{A}_s - z_s\mathbb{I}$.

Let $\mathcal{B} = \{\boldsymbol{w}_\alpha\}_{\alpha=1}^n$ be a basis in the space $\boldsymbol{S}_{ht_{k+1}}$ with $n = \dim \boldsymbol{S}_{ht_{k+1}}$. We seek the approximate solution $\boldsymbol{w}_h^{k+1} \in \boldsymbol{S}_{ht_{k+1}}$ in the form

$$\boldsymbol{w}_h^{k+1} = \sum_{\alpha=1}^{n} \xi_\alpha^{k+1} \boldsymbol{w}_\alpha. \tag{6.25}$$

Similarly we set

$$\hat{\boldsymbol{w}}_h^k = \sum_{\alpha=1}^{n} \xi_\alpha^k \boldsymbol{w}_\alpha. \tag{6.26}$$

Due to the linearity of the form b_h^{1SI} in the variable $\boldsymbol{\varphi}$, we can use as test functions only elements of the basis $\mathcal{B}$.

Let $\boldsymbol{\xi}^k = (\xi_1^k, \ldots, \xi_n^k)$. Then our semi-implicit scheme (6.23) can be written in the matrix form

$$\mathbb{A}(\boldsymbol{\xi}^k)\boldsymbol{\xi}^{k+1} = \boldsymbol{q}(\boldsymbol{\xi}^k), \tag{6.27}$$

where $\boldsymbol{q} : I\!R^n \to I\!R^n$ and $\mathbb{A} : I\!R^n \to I\!R^{n\times n}$ is an $n \times n$ nonsymmetric matrices. This matrix has the form $\mathbb{A} = \mathbb{M} + \tau_k\mathbb{B}^2 + \tau_k\mathbb{D}^2$, where $\mathbb{M} = \{m_{ij}\}_{i,j=1}^n$ is the symmetric, positive definite mass matrix with entries $m_{ij} = \int_\Omega \boldsymbol{w}_i \cdot \boldsymbol{w}_j d\boldsymbol{x}$, $\mathbb{B}^2$ represents the form b_h^{2SI} and $\mathbb{D}^2$ represents the form d_h^2.

Further we proceed as for the ALE formulation I. It means that we use the P^1 and P^2 basis functions for $S_{ht_{k+1}}$ as in Section 6.1 'separately' for each component. We get the basis for $\boldsymbol{S}_{ht_{k+1}} = \left[S_{ht_{k+1}}\right]^4$, respectively $\boldsymbol{S}_{ht_k} = [S_{ht_k}]^4$. Then we write (6.25) and (6.26)

$$\boldsymbol{w}_h^{k+1}(\boldsymbol{x}) = \sum_{i\in I}\sum_{j=1}^{n_p}\sum_{l=1}^{4} \xi_{ijl}^{k+1}\boldsymbol{w}_{ijl}(\boldsymbol{x}), \tag{6.28}$$

$$\hat{\boldsymbol{w}}_h^{k}(\boldsymbol{x}) = \sum_{i\in I}\sum_{j=1}^{n_p}\sum_{l=1}^{4} \xi_{ijl}^{k}\boldsymbol{w}_{ijl}(\boldsymbol{x}), \tag{6.29}$$

where the same notation as in Section 6.2.1 is used.

6.3 The GMRES Method of solving the linear system

In this section we shall present a method for solving our nonsymmetric linear system. In our numerical experiments we use the Generalized Minimal Residual (GMRES) Method that will be derived in the following section. The GMRES is a *nonstationary iterative method,* where the computation involves information that changes at each iteration. We can classify it to the so-called *projective Krylov methods.*

6.3.1 Basic concepts

Let us have a linear system

$$\mathbb{A}\boldsymbol{x} = \boldsymbol{b}, \tag{6.30}$$

where $\mathbb{A} \in I\!R^{n\times n}$ is a nonsingular matrix, $\boldsymbol{b} \in I\!R^n$.

The preconditioning represents replacing of the system (6.30) by the system

$$\mathbb{M}_1\mathbb{A}\mathbb{M}_2\boldsymbol{y} = \mathbb{M}_1\boldsymbol{b}, \quad \boldsymbol{x} = \mathbb{M}_2\boldsymbol{y}, \tag{6.31}$$

where $\mathbb{M}_1 \in I\!R^{n\times n}$ and $\mathbb{M}_2 \in I\!R^{n\times n}$ are nonsingular matrices. These matrices are especially constructed to improve the convergence propreties of the method. The process of searching of the solution of the system is similar for both the preconditioned system (6.31) and the unpreconditioned system (6.30). Hence in the following we shall assume that the system (6.30) is already preconditioned.

Let $\boldsymbol{x}_0 \in I\!R^n$ denote the initial guess of the solution, $\boldsymbol{r}_0 = \boldsymbol{b} - \mathbb{A}\boldsymbol{x}_0$ the corresponding residual and $\boldsymbol{x}^* \in I\!R^n$ the exact solution.

Definition 6.1: *By the notation $\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$ we denote the Krylov space generated by the Krylov sequence $\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{k-1}\boldsymbol{r}_0$, i.e.*

$$\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0) = \text{span}\left\{\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{k-1}\boldsymbol{r}_0, \right\}.$$

Definition 6.2: (Condition of the minimal residual) *Let us define the m-th approximation of the solution of the system (6.30) by the relation*

$$\boldsymbol{x}_m = \arg\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_m(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{b} - \mathbb{A}\boldsymbol{x}\|, \tag{6.32}$$

where $\|\cdot\|$ is the Euclidean norm. The relation (6.32) is called the condition of the minimal residual.

Definition 6.3: (The degree of the residual $\boldsymbol{r}_0$ with respect to the matrix $\mathbb{A}$) *By the degree of the residual $\boldsymbol{r}_0$ with respect to the matrix $\mathbb{A}$ we understand the number $\vartheta(\mathbb{A}, \boldsymbol{r}_0)$ defined by the relation*

$$\vartheta(\mathbb{A}, \boldsymbol{r}_0) = \min\left\{m | \dim\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0) = \dim\mathcal{K}_{m+1}(\mathbb{A}, \boldsymbol{r}_0)\right\}. \tag{6.33}$$

In following we shall use a shorter notation $\vartheta = \vartheta(\mathbb{A}, \boldsymbol{r}_0)$.

6.3.2 Properties of the Krylov space

Theorem 6.1: *Let vectors $\boldsymbol{x}_i$, $i \in \{1, \ldots, n\}$, perform the condition of the minimal residual (6.2). Then the following holds:*

$$\begin{aligned}
&\text{(i)} && \boldsymbol{x}_\vartheta = \boldsymbol{x}^*,\\
&\text{(ii)} && \boldsymbol{x}_i \neq \boldsymbol{x}_\vartheta \ for\ i < \vartheta,\\
&\text{(iii)} && \boldsymbol{x}_i = \boldsymbol{x}^* \ for\ i > \vartheta,\\
&\text{(iv)} && \dim\mathcal{K}_i(\mathbb{A}, \boldsymbol{r}_0) = \vartheta,\ \forall i > \vartheta.
\end{aligned}$$

Proof:

ad (i)

Let us assume that $\dim\mathcal{K}_\vartheta(\mathbb{A}, \boldsymbol{r}_0) = \dim\mathcal{K}_{\vartheta+1}(\mathbb{A}, \boldsymbol{r}_0) = \vartheta$. Then we can write

$$\mathbb{A}^\vartheta\boldsymbol{r}_0 = \sum_{i=0}^{\vartheta-1}\alpha_i\mathbb{A}^i\boldsymbol{r}_0. \tag{6.34}$$

In agreement with (6.33), we have $\alpha_0 \neq 0$. If we divide (6.34) by α_0, we obtain the equation

$$\frac{1}{\alpha_0}\mathbb{A}^\vartheta\boldsymbol{r}_0 = \boldsymbol{r}_0 + \sum_{i=1}^{\vartheta-1}\frac{\alpha_i}{\alpha_0}\mathbb{A}^i\boldsymbol{r}_0. \tag{6.35}$$

Taking into account that $\boldsymbol{x}^* - \boldsymbol{x}_0 = -\mathbb{A}^{-1}\boldsymbol{r}_0$ holds, we obtain from the equation (6.35) the relation

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = \underbrace{-\sum_{i=0}^{\vartheta-2} \frac{\alpha_{i+1}}{\alpha_0}\mathbb{A}^i\boldsymbol{r}_0 + \frac{1}{\alpha_0}\mathbb{A}^{\vartheta-1}\boldsymbol{r}_0}_{\in\mathcal{K}_\vartheta(\mathbb{A},\boldsymbol{r}_0)}, \tag{6.36}$$

It shows that $\boldsymbol{x}^* \in \boldsymbol{x}_0 + \mathcal{K}_\vartheta(\mathbb{A}, \boldsymbol{r}_0)$. The regularity of the matrix and the use of the condition of the minimal residual (6.32) give us immediately the identity $\boldsymbol{x}^* = \boldsymbol{x}_0$.

ad (ii)

This part of the proof will be carried out by a contradiction. Let assume the existence of $k < \vartheta$ that satisfies $\boldsymbol{x}_k = \boldsymbol{x}^*$ and $\boldsymbol{x}_i \neq \boldsymbol{x}^*$, for all $i < k$. Because $\boldsymbol{x}^* \in \boldsymbol{x}_0 + \mathcal{K}_k(\mathbb{A}, \boldsymbol{r}_0)$, there exist $\beta_1, \ldots, \beta_k \in \mathbb{R}$ such that

$$\boldsymbol{x}^* = \boldsymbol{x}_0 + \sum_{i=0}^{k-1} \beta_{i+1}\mathbb{A}^i\boldsymbol{r}_0. \tag{6.37}$$

The condition $\boldsymbol{x}_i \neq \boldsymbol{x}^*$, for all $i < k$ gives us $\beta_k \neq 0$. When we transfer $\boldsymbol{x}_0$ to the left-hand side of the relation (6.37), multiply the resulting relation by the matrix $\mathbb{A}$, replace $\mathbb{A}(\boldsymbol{x}^* - \boldsymbol{x}_0)$ by $\boldsymbol{r}_0$ and divide the obtained relation by β_k, we get

$$\frac{1}{\beta_k}\boldsymbol{r}_0 = \mathbb{A}^k\boldsymbol{r}_0 + \sum_{i=1}^{k-1} \frac{\beta_i}{\beta_k}\mathbb{A}^i\boldsymbol{r}_0. \tag{6.38}$$

Then we easily find that

$$\mathbb{A}^k\boldsymbol{r}_0 = \underbrace{\frac{1}{\beta_k}\boldsymbol{r}_0 - \sum_{i=1}^{k-1} \frac{\beta_i}{\beta_k}\mathbb{A}^i\boldsymbol{r}_0}_{\in\mathcal{K}_k(\mathbb{A},\boldsymbol{r}_0)}. \tag{6.39}$$

It means that $\dim\mathcal{K}_{k+1}(\mathbb{A}, \boldsymbol{r}_0) = \dim\mathcal{K}_k(\mathbb{A}, \boldsymbol{r}_0)$. Hence, we get a contradiction with the definition of ϑ.

ad (iii) and (iv)

From the identity $\dim\mathcal{K}_\vartheta(\mathbb{A}, \boldsymbol{r}_0) = \dim\mathcal{K}_{\vartheta+1}(\mathbb{A}, \boldsymbol{r}_0)$ we immediately obtain by an induction that $\dim\mathcal{K}_{\vartheta+1}(\mathbb{A}, \boldsymbol{r}_0) = \dim\mathcal{K}_{\vartheta+2}(\mathbb{A}, \boldsymbol{r}_0) = \ldots$. It implies that $\boldsymbol{x}_i = \boldsymbol{x}_\vartheta = \boldsymbol{x}^*$ for all $i > \vartheta$.

■

6.3.3 Projections

As mentioned in the beginning, the GMRES method is a projective method. In this section we shall discuss these projections.

First we present two theorems that ilustrate the construction of the vector with minimal residual in the Krylov subspace. It is the main idea of the GMRES method.

Theorem 6.2: *Let $m \geq k$, $\mathbb{B} \in I\!R^{m\times k}$, $\operatorname{rank}\mathbb{B} = k$. Then there exists only one couple of matrices $\mathbb{Q} \in I\!R^{m\times k}$ and $\mathbb{R} \in I\!R^{k\times k}$ with the following properties: $\mathbb{Q}^T\mathbb{Q}$ is a diagonal matrix with positive diagonal elements, $\mathbb{R}$ is an upper triangular matrix with 1 on the diagonal and*

$$\mathbb{B} = \mathbb{Q}\mathbb{R}. \tag{6.40}$$

Let us denote by $\boldsymbol{Q}_i$ the i-th column of the matrix $\mathbb{Q}$. Then $\mathbb{Q} = (\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_k)$ can be obtained as the matrix $\mathbb{B}^{(k)}$ in the sequence $\mathbb{B}^{(0)} = \mathbb{B}$, $\mathbb{B}^{(1)}, \ldots, \mathbb{B}^{(k)}$, where we can compute $\mathbb{B}^{(j)} = (\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots, \boldsymbol{Q}_j, \boldsymbol{B}_{j+1}^{(j)}, \ldots, \boldsymbol{B}_k^{(j)})$, $j = 1, \ldots, k$ from $\mathbb{B}^{(j-1)} = (\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots, \boldsymbol{Q}_{j-1}, \boldsymbol{B}_j^{(j-1)}, \ldots, \boldsymbol{B}_k^{(j-1)})$ in this way:

$$\begin{aligned}
&\boldsymbol{Q}_j = \boldsymbol{B}_j^{(j-1)},\ d_j = \boldsymbol{Q}_j^T\boldsymbol{Q}_j,\ r_{jj} = 1,\ r_{ji} = (\boldsymbol{Q}_j^T\boldsymbol{B}_i^{(j-1)})/d_j \\
&\boldsymbol{B}_i^{(j)} = \boldsymbol{B}_i^{(j-1)} - r_{ji}\boldsymbol{Q}_j, \quad i = j+1, \ldots, k.
\end{aligned}$$

Here we can suppose that $\mathbb{R} = (r_{\alpha\beta})_{\alpha,\beta=1}^k$.

Theorem 6.3: *Let us have a system*

$$\mathbb{B}\boldsymbol{x} = \boldsymbol{c} \tag{6.41}$$

with $\mathbb{B} \in I\!R^{m\times k}$ and $\operatorname{rank}\mathbb{B} = k$. If $\mathbb{B} = \mathbb{Q}\mathbb{R}$ is the factorization of the matrix $\mathbb{B}$ in accordance with Theorem 6.2, then the solution $\boldsymbol{x}$ of the system

$$\mathbb{R}\boldsymbol{x} = \mathbb{D}^{-1}\mathbb{Q}^T\boldsymbol{c}, \text{ where } \mathbb{D} = \operatorname{diag}\{d_1, \ldots, d_k\} \tag{6.42}$$

satisfies the identity

$$\mathbb{B}^T(\boldsymbol{c} - \mathbb{B}\hat{\boldsymbol{x}}) = 0. \tag{6.43}$$

This identity characterizes vectors $\hat{\boldsymbol{x}}$ with minimal Euclidean norm of residual $\boldsymbol{c} - \mathbb{B}\boldsymbol{x}$:

$$\|\boldsymbol{c} - \mathbb{B}\hat{\boldsymbol{x}}\| \leq \|\boldsymbol{c} - \mathbb{B}\boldsymbol{x}\| \tag{6.44}$$

for each vector $\boldsymbol{x}$.

The proofs of these two theorems can be found in [8].

Now, we shall discuss the condition of the minimal residual (6.32), i.e.

$$\boldsymbol{x}_i = \arg\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_i(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{b} - \mathbb{A}\boldsymbol{x}\|.$$

This relation can be rewritten in the form

$$\boldsymbol{x}_i = \boldsymbol{x}_0 + \arg\min_{\boldsymbol{u}\in\mathcal{K}_i(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{b} - \mathbb{A}(\boldsymbol{x}_0 + \boldsymbol{u})\| = \boldsymbol{x}_0 + \underbrace{\arg\min_{\boldsymbol{u}\in\mathcal{K}_i(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{r}_0 - \mathbb{A}\boldsymbol{u}\|}_{\boldsymbol{u}_i}. \tag{6.45}$$

We define the matrix $\mathbb{Q}_{\boldsymbol{r}_0,i} \in I\!R^{n\times i}$, $\mathbb{Q}_{\boldsymbol{r}_0,i} = (\mathbb{A}\boldsymbol{r}_0, \mathbb{A}^2\boldsymbol{r}_0, \ldots, \mathbb{A}^i\boldsymbol{r}_0)$ with the maximal rank. Because $\mathbb{A}\boldsymbol{u} \in \mathbb{A}\mathcal{K}_i(\mathbb{A}, \boldsymbol{r}_0)$, there exists $\boldsymbol{z} \in I\!R^i$ such that $\mathbb{A}\boldsymbol{u} = \mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}$. This implies the existence of $\mathbb{A}\boldsymbol{u}_i = \mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}_i$, where

$$\boldsymbol{z}_i = \arg\min_{\boldsymbol{z}\in I\!R^i} \|\boldsymbol{r}_0 - \mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}\|. \tag{6.46}$$

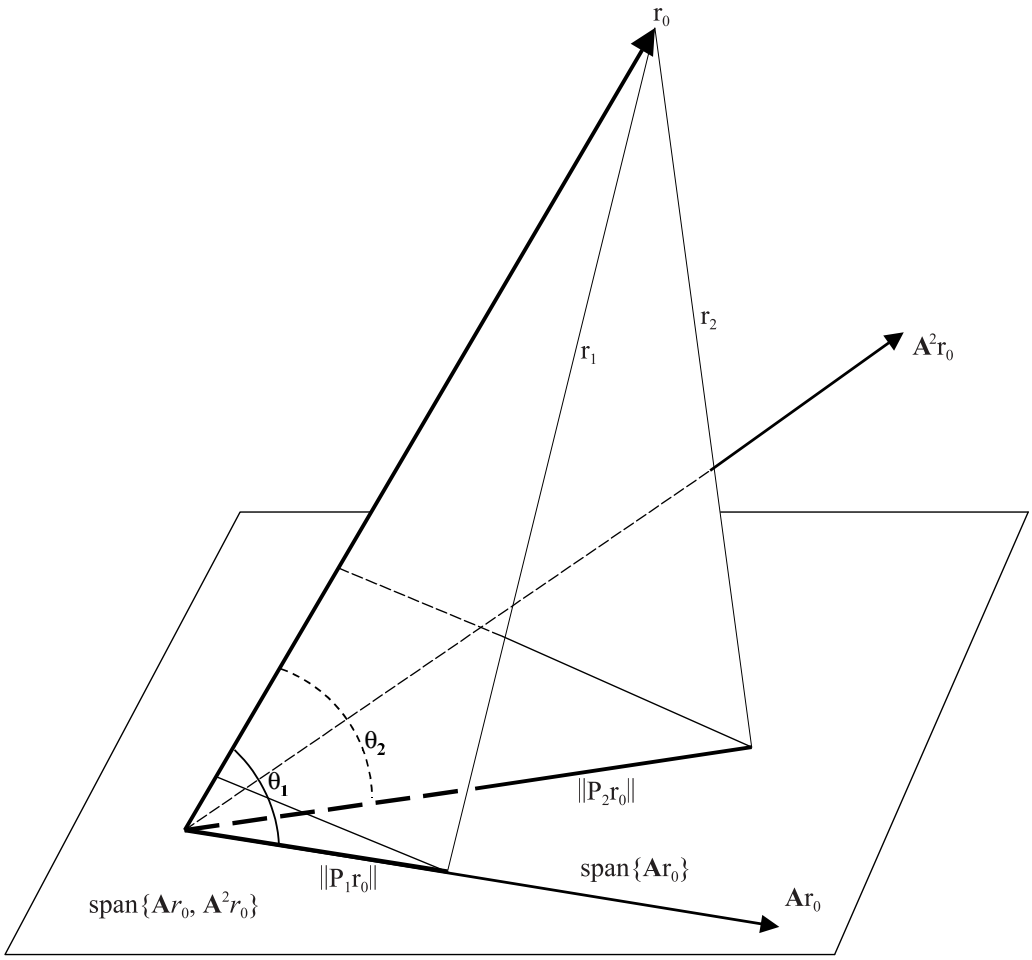


Figure 6.1: Residuals r_i in 3D

It allows us to say that $\mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}_i$ is the best approximation of $\boldsymbol{r}_0$ in the space $\mathbb{A}\mathcal{K}_i(\mathbb{A}, \boldsymbol{r}_0)$ in the sense of the *Least Square Method.* With the aid of Theorem 6.3 the following relation holds

$$\boldsymbol{z}_i = (\mathbb{Q}_{\boldsymbol{r}_0,i}^T \mathbb{Q}_{\boldsymbol{r}_0,i})^{-1} \mathbb{Q}_{\boldsymbol{r}_0,i}^T \boldsymbol{r}_0. \tag{6.47}$$

Then for the i-th residual,i.e. $\boldsymbol{r}_i = \boldsymbol{b} - \mathbb{A}\boldsymbol{x}_i$, we obtain $\boldsymbol{r}_i = \boldsymbol{r}_0 - \mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}_i$. If we replace $\boldsymbol{z}_i$ by the form (6.47), we get

$$\boldsymbol{r}_i = \boldsymbol{r}_0 - \mathbb{Q}_{\boldsymbol{r}_0,i}\boldsymbol{z}_i = \boldsymbol{r}_0 - \underbrace{\mathbb{Q}_{\boldsymbol{r}_0,i}(\mathbb{Q}_{\boldsymbol{r}_0,i}^T \mathbb{Q}_{\boldsymbol{r}_0,i})^{-1} \mathbb{Q}_{\boldsymbol{r}_0,i}^T \boldsymbol{r}_0}_{=\mathbb{P}_i}. \tag{6.48}$$

It can be shown that $\mathbb{P}_i = \mathbb{P}_i^2$ and $\mathbb{P}_i = \mathbb{P}_i^T$. It means that $\mathbb{P}_i\boldsymbol{r}_0$ is the orthogonal projection of vector $\boldsymbol{r}_0$ to the space $\mathcal{K}_i(\mathbb{A}, \mathbb{A}\boldsymbol{r}_0)$ and $\boldsymbol{r}_i = \boldsymbol{r}_0 - \mathbb{P}_i\boldsymbol{r}_0$ is orthogonal to the space $\mathcal{K}_i(\mathbb{A}, \mathbb{A}\boldsymbol{r}_0)$, especially $\boldsymbol{r}_i \perp \mathbb{P}_i\boldsymbol{r}_0$.

Figure 6.1 ilustrate the residuals $\boldsymbol{r}_i$ in three dimensions.

6.3.4 Construction of the orthonormal basis of the Krylov space

Numerical experience in work with large matrices show that the vectors of Krylov sequence become linearly dependent for larger number of iterations. It is caused by the cumulation of the round-off errors. In practice we avoid this problem by the construction of the orthonormal basis in the Krylov space $\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$. This is called *Arnoldi process* and it is a special case of the *Gramm-Schmidt othogonalization.*

We are searching vectors $\boldsymbol{v}_i,\ i = 1, \ldots, m$, that are mutually orthogonal and the relation

$$\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0) = \text{span}\,\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \tag{6.49}$$

holds. On the basis of Theorem 6.1 we shall assume that $\dim \mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0) = m$. For a better demonstration we present first two steps of the process in detail and then the whole algorithm will be given.

We choose

$$\boldsymbol{v}_1 = \boldsymbol{r}_0 / \left\|\boldsymbol{r}_0\right\| \tag{6.50}$$

and construct the vector $\boldsymbol{v}_2 \perp \boldsymbol{v}_1$ that satisfies

$$\text{span}\,\{\boldsymbol{v}_1, \boldsymbol{v}_2\} = \text{span}\,\{\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0\}\,. \tag{6.51}$$

Then we set

$$\hat{\boldsymbol{v}}_2 = \mathbb{A}\boldsymbol{v}_1 - h_{11}\boldsymbol{v}_1, \tag{6.52}$$

where h_{11} should be determined so that $\boldsymbol{v}_1^T \hat{\boldsymbol{v}}_2 = 0$. This gives us

$$\boldsymbol{v}_1^T \mathbb{A}\boldsymbol{v}_1 - h_{11}\boldsymbol{v}_1^T \boldsymbol{v}_1 = 0$$

and defines h_{11} as

$$h_{11} = \boldsymbol{v}_1^T \mathbb{A}\boldsymbol{v}_1.$$

Further we set $h_{21} = \|\hat{\boldsymbol{v}}_2\|$ and $\boldsymbol{v}_2 = \hat{\boldsymbol{v}}_2 / h_{21}$. There is no danger with division by zero, becauce the vectors $\boldsymbol{r}_0$ and $\mathbb{A}\boldsymbol{r}_0$ are linearly independent and hence, $h_{21} \neq 0$. If we use (6.50) and (6.52), we obtain

$$\boldsymbol{v}_2 = \frac{1}{h_{21}}\hat{\boldsymbol{v}}_2 = \frac{1}{h_{21}\left\|\boldsymbol{r}_0\right\|}\mathbb{A}\boldsymbol{r}_0 - \frac{h_{11}}{h_{21}\left\|\boldsymbol{r}_0\right\|}\boldsymbol{r}_0.$$

It is a linear combination of the elements from the space $\mathcal{K}_2(\mathbb{A}, \boldsymbol{r}_0)$. Hence $\boldsymbol{v}_2 \in \mathcal{K}_2(\mathbb{A}, \boldsymbol{r}_0)$ and $\text{span}\,\{\boldsymbol{v}_1, \boldsymbol{v}_2\} = \text{span}\,\{\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0\}$. This approach is further repeated:

In the j-th step of the process we already know the orthonormal vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j,\ j < \vartheta$, forming the basis of $\mathcal{K}_j(\mathbb{A}, \boldsymbol{r}_0)$. The vector $\boldsymbol{v}_k,\ 1 \leq k \leq j$, is a linear combination of the vectors $\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{k-1}\boldsymbol{r}_0$, where the coefficients at the terms $\mathbb{A}^{k-1}$ are $1/(\|\boldsymbol{r}_0\|\, h_{k,k-1})$. We figure out the vector $\hat{\boldsymbol{v}}_{j+1}$ as

$$\hat{\boldsymbol{v}}_{j+1} = \mathbb{A}\boldsymbol{v}_j - \sum_{i=1}^{j} h_{ij}\boldsymbol{v}_i, \tag{6.53}$$

where $h_{ij} = \boldsymbol{v}_i^T \mathbb{A}\boldsymbol{v}_j$. This implies that $\hat{\boldsymbol{v}}_{j+1} \perp \boldsymbol{v}_i,\ i = 1, \ldots, j$. We denote the normed vector $\hat{\boldsymbol{v}}_{j+1}$ as $\boldsymbol{v}_{j+1} = \hat{\boldsymbol{v}}_{j+1} / h_{j+1,j}$, where $h_{j+1,j} = \|\hat{\boldsymbol{v}}_{j+1}\|$.

We can summarize the above ideas in the following algorithm for the construction of the orthonormal basis in the space $\mathcal{K}_k(\mathbb{A}, \boldsymbol{r}_0)$.

Arnoldi process:

$$
\begin{aligned}
&\text{Input data}: \mathbb{A},\ \boldsymbol{b},\ \boldsymbol{x}_0,\ k \leq \vartheta \\
&\boldsymbol{r}_0 = \boldsymbol{b} - \mathbb{A}\boldsymbol{x}_0 \\
&\boldsymbol{v}_1 = \boldsymbol{r}_0 / \left\|\boldsymbol{r}_0\right\| \\
&\text{do } j = 1,\ k-1 \\
&\quad \hat{\boldsymbol{v}}_{j+1} = \mathbb{A}\boldsymbol{v}_j \\
&\quad \text{do } i = 1,\ j \\
&\qquad\qquad h_{i,j} = \boldsymbol{v}_i^T \mathbb{A}\boldsymbol{v}_j \\
&\qquad\qquad \hat{\boldsymbol{v}}_{j+1} = \hat{\boldsymbol{v}}_{j+1} - h_{i,j}\boldsymbol{v}_i \\
&\quad \text{end do} \\
&\quad h_{j+1,j} = \left\|\hat{\boldsymbol{v}}_{j+1}\right\| \qquad (6.54) \\
&\quad \boldsymbol{v}_{j+1} = \hat{\boldsymbol{v}}_{j+1} / h_{j+1,j} \qquad (6.55) \\
&\text{end do}
\end{aligned}
$$

Theorem 6.4: (Arnoldi process) *Using the Arnoldi process, we obtain for $k \leq \vartheta$ the orthonormal basis of the space $\mathcal{K}_k(\mathbb{A}, \boldsymbol{r}_0)$. It holds that $h_{j+1,j} \neq 0$ for $1 \leq j \leq k-1$. If $k = \vartheta + 1$, then $h_{\vartheta+1,\vartheta} = 0$ and the Arnoldi process collapses.*
Proof: If $\|\hat{\boldsymbol{v}}_{j+1}\| = 0$ for any $j+1 \leq \vartheta$, then $\hat{\boldsymbol{v}}_{j+1} = \boldsymbol{0}$ and

$$\mathbb{A}\boldsymbol{v}_j = \sum_{i=1}^{j} h_{ij}\boldsymbol{v}$$

according to (6.53). The vector $\boldsymbol{v}_j$ is a linear combination of vectors $\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{j-1}\boldsymbol{r}_0$ and the coefficient at the vector $\mathbb{A}^{j-1}\boldsymbol{r}_0$ is equal to $(\|\boldsymbol{r}_0\| \cdot \|\hat{\boldsymbol{v}}_2\| \cdot \ldots \cdot \|\hat{\boldsymbol{v}}_j\|)^{-1}$ as shown in (6.55). Hence, the vector $\boldsymbol{v}_j \neq 0$. The vector $\mathbb{A}\boldsymbol{v}_j$ is a linear combination of the vectors $\mathbb{A}\boldsymbol{r}_0, \mathbb{A}^2\boldsymbol{r}_0, \ldots, \mathbb{A}^j\boldsymbol{r}_0$ and the above lines imply that the coeffiecient at the vector $\mathbb{A}^j\boldsymbol{r}_0$ is nonzero. This means that $\mathbb{A}^j\boldsymbol{r}_0 \in \text{span}\,\{\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{j-1}\boldsymbol{r}_0\}$ and vectors $\boldsymbol{r}_0, \mathbb{A}\boldsymbol{r}_0, \ldots, \mathbb{A}^{j-1}\boldsymbol{r}_0, \mathbb{A}^j\boldsymbol{r}_0$ are linearly dependent, which is a contradiction to the definition of ϑ. Thus $h_{j+1,j} \neq 0$ for $j = 1, \ldots, \vartheta - 1$ and the Arnoldi process does not collapse for this j.

By induction it can be proved that for $j \leq \vartheta$ the following holds: $\boldsymbol{v}_j \in \mathcal{K}_j(\mathbb{A}, \boldsymbol{r}_0)$, $\boldsymbol{v}_i^T\boldsymbol{v}_j = 0$ $(i \neq j;\ i, j \leq \vartheta)$, $\|\boldsymbol{v}_j\| = 1$. Hence span $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j\} = \mathcal{K}_j(\mathbb{A}, \boldsymbol{r}_0)$.

Now, we shall prove $h_{\vartheta+1,\vartheta} = 0$. The coefficients in (6.53) are chosen in such way that the vector $\sum_{i=1}^{\vartheta} h_{i\vartheta}\boldsymbol{v}_\vartheta$ is a projection of the vector $\mathbb{A}\boldsymbol{v}_\vartheta$ to the space $\mathcal{K}_\vartheta(\mathbb{A}, \boldsymbol{r}_0)$. $\mathbb{A}\boldsymbol{v}_\vartheta$ is in the space $\mathcal{K}_\vartheta(\mathbb{A}, \boldsymbol{r}_0)$ and hence $\hat{\boldsymbol{v}}_{\vartheta+1} = \boldsymbol{0}$ and $h_{\vartheta+1,\vartheta} = 0$.

■

For each $k \leq \vartheta$ the Arnoldi process figures out the elements h_{ij} with $i \leq j + 1$.

From these elements we can construct the matrices $\mathbb{H}_k \in I\!R^{k\times k}$ and $\hat{\mathbb{H}}_k \in I\!R^{k\times k+1}$

$$\mathbb{H}_k := \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots 0 & h_{1,k} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,k} \\ 0 & h_{31} & h_{33} & \dots & h_{3,k} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_{k,k-1} & h_{k,k} \end{pmatrix}, \hat{\mathbb{H}}_k := \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots 0 & h_{1,k} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,k} \\ 0 & h_{31} & h_{33} & \dots & h_{3,k} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \ddots & h_{k,k} \\ 0 & \dots & 0 & 0 & h_{k+1,k} \end{pmatrix}.$$

Let define the matrix $\mathbb{V}_k \in I\!R^{n\times k}$ with columns $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$, i.e. $\mathbb{V}_k := (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ and the vector $\boldsymbol{v}_{\vartheta+1} := \boldsymbol{0}$. Then we can write

$$\mathbb{A}\mathbb{V}_k = \mathbb{V}_{k+1}\hat{\mathbb{H}}_k, \quad \text{for } k \leq \vartheta. \tag{6.56}$$

If we multiply this identity by the matrix $\mathbb{V}_k^T$ from the left-hand side and use the orthogonality of the vectors $\boldsymbol{v}_i$, we obtain

$$\mathbb{V}_k^T \mathbb{A}\mathbb{V}_k = \mathbb{H}_k.$$

6.3.5 $\mathbb{QR}$ factorization

Now we derive the self GMRES method. The orthonormal basis $\{\boldsymbol{v}_i\}_{i=1}^m$ of the Krylov space $\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$ is already constructed. Let come back to the condition of the minimal residual (6.32), i.e.

$$\boldsymbol{x}_m = \arg\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_m(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{b} - \mathbb{A}\boldsymbol{x}\|.$$

If we search $\boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$, then there exists exactly one $\boldsymbol{u} \in \mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$ such that $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{u}$. It allows us to rewrite (6.32) in the form

$$\boldsymbol{u}_m = \arg\min_{\boldsymbol{u}_m\in\mathcal{K}_m(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{b} - \mathbb{A}(\boldsymbol{x}_0 + \boldsymbol{u})\| = \arg\min_{\boldsymbol{u}_m\in\mathcal{K}_m(\mathbb{A},\boldsymbol{r}_0)} \|\boldsymbol{r}_0 - \mathbb{A}\boldsymbol{u}\|.$$

The vector $\boldsymbol{u}$ is an element of $\mathcal{K}_m(\mathbb{A}, \boldsymbol{r}_0)$ so that it can be rewritten as a linear combination of the vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$. We use the matrix representation $\boldsymbol{u} = \mathbb{V}_m\boldsymbol{z}$, where $\boldsymbol{z} \in I\!R^m$ and $\mathbb{V}_m = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_m)$. These considerations lead us to the identity

$$\boldsymbol{z}_m = \arg\min_{\boldsymbol{z}_m\in I\!R^m} \|\boldsymbol{r}_0 - \mathbb{A}\mathbb{V}_m\boldsymbol{z}\|.$$

Further we can write $\boldsymbol{r}_0$ as $\beta\boldsymbol{v}_1$, because from the Arnoldi process we know that $\boldsymbol{v}_1 = \boldsymbol{r}_0/\|\boldsymbol{r}_0\|$. In the same way we can write $\beta\boldsymbol{v}_1$ as $\beta\mathbb{V}_{m+1}\boldsymbol{e}_1^{m+1}$ and $\mathbb{A}\mathbb{V}_m$ as $\mathbb{V}_{m+1}\hat{\mathbb{H}}_m$, where $\hat{\mathbb{H}}_m \in I\!R^{(m+1)\times m}$ and its elements h_{ij} are obtained from the Arnoldi process. The described considerations imply that

$$\|\boldsymbol{r}_0 - \mathbb{A}\mathbb{V}_m\boldsymbol{z}\|^2 = \left\|\beta\mathbb{V}_{m+1}\boldsymbol{e}_1^{m+1} - \mathbb{V}_{m+1}\hat{\mathbb{H}}_m\boldsymbol{z}\right\|^2 = \left\|\beta\boldsymbol{e}_1^{m+1} - \hat{\mathbb{H}}_m\boldsymbol{z}\right\|^2.$$

Now we apply the $\mathbb{QR}$ factorization of the matrix $\hat{\mathbb{H}}_m$, $\hat{\mathbb{H}}_m = \mathbb{Q}_m\hat{\mathbb{R}}_m$, where $\mathbb{Q}_m \in I\!R^{(m+1)\times(m+1)}$ is a unitary matrix and $\hat{\mathbb{R}}_m \in I\!R^{(m+1)\times m}$ is an upper triangular matrix (for more details about the $\mathbb{QR}$ factorization see [9]):

$$\begin{aligned} \left\|\beta\boldsymbol{e}_1^{m+1} - \hat{\mathbb{H}}_m\boldsymbol{z}\right\|^2 &= \left\|\beta\boldsymbol{e}_1^{m+1} - \mathbb{Q}_m\hat{\mathbb{R}}_m\boldsymbol{z}\right\|^2 = \left\|\mathbb{Q}_m(\beta\mathbb{Q}_m^T\boldsymbol{e}_1^{m+1} - \hat{\mathbb{R}}_m\boldsymbol{z})\right\|^2 \\ &= \left\|\beta\mathbb{Q}_m^T\boldsymbol{e}_1^{m+1} - \hat{\mathbb{R}}_m\boldsymbol{z}\right\|^2. \end{aligned} \tag{6.57}$$

Let us denote $\hat{\boldsymbol{g}}_m = \beta \mathbb{Q}_m^T \boldsymbol{e}_1^{m+1}$ and write $\hat{\boldsymbol{g}}_m^T = (\boldsymbol{g}_m^T, \eta_{m+1})$ in the following. We obtain

$$\left\| \beta \mathbb{Q}_m^T \boldsymbol{e}_1^{m+1} - \hat{\mathbb{R}}_m \boldsymbol{z} \right\|^2 = \left\| \hat{\boldsymbol{g}}_m - \hat{\mathbb{R}}_m \boldsymbol{z} \right\|^2 = \| \boldsymbol{g}_m - \mathbb{R}_m \boldsymbol{z} \|^2 + |\eta_{m+1}|^2 ,$$

where $\mathbb{R}_m$ is the matrix $\hat{\mathbb{R}}_m$ without the last row. We transformed the relation (6.32) to the form

$$\boldsymbol{z}_m = \arg\min_{\boldsymbol{z}_m \in I\!R^m} \| \boldsymbol{g}_m - \mathbb{R}_m \boldsymbol{z} \| , \quad \boldsymbol{x}_m = \boldsymbol{x}_0 + \mathbb{V}_m \boldsymbol{z}_m . \tag{6.58}$$

In view of the relation (6.58), $\boldsymbol{z}_m$ is the solution of the system $\mathbb{R}_m \boldsymbol{z} = \boldsymbol{g}_m$, where $\mathbb{R}_m$ is the nonsingular upper triangular matrix. Our problem of searching the minimum in the relation (6.58) was reduced to the solution of a system with an upper triangular matrix.

At the end we present the algorithm of the GMRES method, where the $\mathbb{QR}$ factorization is carried out by the *Givens rotations*. The Givens matrices will be denoted by $\mathbb{G}_i^{(k)} \in I\!R^{(k+1)\times(k+1)}$, that anulate the term on the position $(i+1, i)$. For more details about the Givens matrices and $\mathbb{QR}$ factorization alghoritm see [9].

Algorithm of the GMRES method:

Input data : $\mathbb{A}$, $\boldsymbol{b}$, $\boldsymbol{x}_0$ initial guess, max maximum number of iterations, tol toleration
$\boldsymbol{r} = \boldsymbol{b} - \mathbb{A}\boldsymbol{x}_0$
$\beta = \|\boldsymbol{r}\|$
$\boldsymbol{v}_1 = \boldsymbol{r}/\beta$
do $k = 1,\ max$
  $\hat{\boldsymbol{v}}_{k+1} = \mathbb{A}\boldsymbol{v}_k$
  do $i = 1,\ k$
    $h_{i,k} = \boldsymbol{v}_i^T \mathbb{A} \boldsymbol{v}_k$
    $\hat{\boldsymbol{v}}_{k+1} = \hat{\boldsymbol{v}}_{k+1} - h_{i,k}\boldsymbol{v}_i$
  end do
  $h_{k+1,k} = \|\hat{\boldsymbol{v}}_{k+1}\|$
  $\boldsymbol{v}_{k+1} = \hat{\boldsymbol{v}}_{k+1}/h_{k+1,k}$
  do $i = 1,\ k-1$
    $(h_{1,k}, \ldots, h_{k+1,k})^T = \mathbb{G}_i^{(k)} (h_{1,k}, \ldots, h_{k+1,k})^T$
  end do
  Computation of $\mathbb{G}_k^{(k)}$
  $(h_{1,k}, \ldots, h_{k+1,k})^T = \mathbb{G}_k^{(k)} (h_{1,k}, \ldots, h_{k+1,k})^T$
  $\boldsymbol{g}^{k+1} = \mathbb{G}_k^{(k)} \begin{pmatrix} \boldsymbol{g}^k \\ 0 \end{pmatrix}$
  $\begin{pmatrix} \boldsymbol{g}^{(k)} \\ 0 \end{pmatrix} = \boldsymbol{g}^{(k+1)}$

if $\left|g_{k+1}^{(k+1)}\right| < tol$ then

Computation of $\boldsymbol{y}$ from equation $\mathbb{H}\boldsymbol{y} = \boldsymbol{g}^{(k)}$

$\boldsymbol{x} = \boldsymbol{x}_0 + y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + \ldots + y_k\boldsymbol{v}_k$

"End of computation, $\boldsymbol{x}$ is the searching solution"

else

do $i = 1, k$

$$\mathbb{G}_i^{(k+1)} = \begin{pmatrix} \mathbb{G}_i^{(k)} & 0 \\ 0^T & 1 \end{pmatrix}$$

end do

end if

end do

Chapter 7

Description of the program and input data

In this chapter we shall describe the most important part of the program used for our numerical expriments. In our simulation we use the C program created by Václav Kučera that was originally designed for the flow in the time-independent domains. In this program we modified some parts in order to allow computations in time-dependent domains using the ALE method. This means that now the program can be used for solving the flow problems in time-independent as well as time-dependent domains.

7.1 `main.c`

The main function `int main()` starts with a declaration of constants described in Table 7.1. The function `gettriang(vert,el,ed,tri1)` reads the information about a triangulation saved in `tri1`.

Further needed variables are initialized together with the chosen linear solver and variables used in the case of solving a time-dependent problem. For details see Section 7.2. The initialization of the initial conditions is carried out by the functions `setinitialconditions(el,0)` and `setinitialconditions(el,1)` that set the initial conditions as constant state read from the data set `constants.h`. We shall describe this important data set in the Section 7.2.

In the case that we want to start the computation from the state that we obtained from the previous runs of the program, we use the function `loadstate(el,act,statex,`

Type	Notation	Use
integer	`stepsave`	Each stepsave-th time level will be saved.
integer	`iterations`	Total number of time levels.
double	`tau`	Initial time step.
char	`tri1[]`	Name of the data set where a triangulation is saved.

Table 7.1: 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 7.2: Variables.

`&time,0)`.

Now, we shall be concerned with the main part of the main function `int main()`. It means that we describe the most important parts of the loop throught all time levels that directs the whole process done by the program. If the ALE method is used, it is necessary to set a new triangulation at each time step. The function `renewtriangulation(el,ed,vert,time)` set a new parameters of the triangulation according to the change of the domain. Then the function `iteration(el,ed,&mat,x,b,act,tau,20,1E-5,&error,&gmiters,btemp,0,time)` computes the solution on the next time level. The program is able to increase the time step by the multiplication of the time step from the previous time level by 1,3. The saving of the chosen variable is carried out by the function `savesolution(el, vert,i/stepsave,act, time,X)`, where `X` is the chosen variable. The notation of the variables see Table 7.2.

After this loop the function `savestate(el,ed,act,statex,time)` allows us to save the final state to the data set `statex`. A new computation can later start from these data.

7.2 Constants.h

In this data set the majority of variables is defined. The conditions of a computation are set here. Below we describe the most important constant for setting the computation.

In Tables 7.4 - 7.12 the possible setting of each constant can be found.

7.3 ALE.c

The code of `ALE.c` is designed for dealing with the time-dependent domain. In our simulation we use the ALE method in two different formulations. As mentioned in Chapter 5, we work with the motion of the domain that can be prescribed as a graph of a function. In the `ALE.c` we prescribe this function in the function `void ALEvalues (x,y,t,x2,y2)`, where `x,y` are the Cartesian reference coordinates, `x2,y2` are ALE coordinates in Ω_t and `t` is time. The ALE velocity, i.e. the time derivative of the ALE mapping, is applied by the function `void ALEvelocity (x,y,t,ax,ay)`, where `ax` and `ay` are the components of the ALE velocity.

Notation	Use
CLF	Constant for the CLF condition.
ISO	Setting for dealing with a curved boundary.
MAXN	Number of inner iterations in GMRES.
LINEAR_SOLVER	Setting of a linear solver.
ALE	Setting for dealing with a time-dependent domain.
PRECOND	Setting of the GMRES preconditioner.
STABILISE	Setting of a stabilization.
BCS	Handling of boundary condition on inlet and outlet.
WALLBCS	Handling of the wall boundary condition.

Table 7.3: Meaning of the constants.

ISO	Setting
0	Edges on the curved boundary are not treated isoparametrically.
1	Edges on the curved boundary are treated isoparametrically.

Table 7.4: Setting of the constant ISO.

LINEAR_SOLVER	Setting
0	GMRES.
1	UMFPACK.

Table 7.5: Setting of the constant LINEAR_SOLVER.

ALE	Setting	Type of problem
0	The ALE method is not used.	Time-independent domain.
1	The ALE formulation I is used and treated implicitly.	Time-dependent domain.
2	The ALE formulation I is used and treated explicitly.	Time-dependent domain.
11	The ALE formulation II is used and treated implicitly.	Time-dependent domain.

Table 7.6: Setting of the constant ALE.

PRECOND	Setting
0	No preconditioner for GMRES.
1	Diagonal preconditioner for GMRES.
2	Block diagonal preconditioner for GMRES.

Table 7.7: Setting of the constant PRECOND.

STABILISE	Setting
0	No stabilization.
1	Stabilization by a projection.
2	Stabilization by a shock capturing.

Table 7.8: Setting of the constant STABILISE.

BCS	Setting
1	Linearization.
2	Exact Riemann Solver.

Table 7.9: Setting of the constant BCS.

WALLBCS	Setting
1	Standart conditions described in Chapter 4.
2	Use of the velocity reflection.

Table 7.10: Setting of the constant WALLBCS.

Notation	Initial condition
RHO0	Initial density.
VX0	Initial x-th component of velocity.
VY0	Initial y-th component of velocity.
P0	Initial pressure.

Table 7.11: Notation of initial conditions.

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 7.12: Notation of boundary conditions.

Chapter 8

Examples

In this chapter we shall present and compare our numerical results obtained by two different ALE formulations of the governing equations as described in Chapter 3. All our computations were done in the rectangular channel $[-2, 2] \times [0, 1]$, where the lower wall was moving in the interval $[-1, 1]$. The motion of this part of the lower wall was prescribed by the function

$$\alpha \sin 0.4t \ (\cos(\pi X_1) + 1), \tag{8.1}$$

where the coefficient α represents the height of the closure of the channel. This movement was interpolated to the rest of the domain resulting in the ALE mapping. The ALE mapping was equal of identity in the sets $[-2, -1] \times [0, 1]$ and $[1, 2] \times [0, 1]$.

For the application of the developed method, we worked out a computer program. We went out from the program by Václav Kučera for the solution of compressible flow in time-independent domains. This program was modified and adapted to the solution of flow in time-dependent domains. For more information about setting of this program see Chapter 7.

8.1 Comparison of the ALE formulations

We shall compare the results obtained by both ALE formulations described in Chapter 3. In the computation was chosen the constant $\alpha = 0.34$ in formula (8.1). The computational domain was devided by the triangular mesh of 631 vertices, i.e 1160 elements. For the setting of the program and initial conditions see Tables 8.1 and 8.2. (For explanation of the notation see Chapter 7.) On the inlet part of the boundary was prescribed the same condition for the velocity and density as the initial conditions. The pressure on the outlet boundary has the same value as the initial pressure. Both schemes of the ALE formulation were treated implicitly. The chosen time step was 0.02.

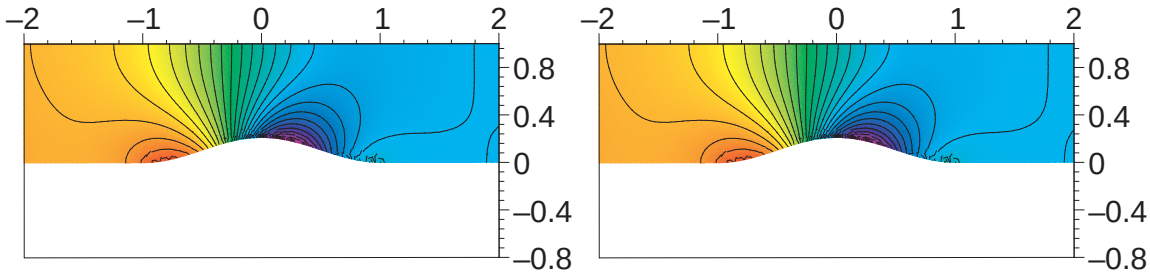
We can see the the results obtained by both formulations are very similar. It is important to mention that the same results were obtained with the second formulation without using the shock capturing technique. On the other hand, when we tried to solve the problem with the first formulation without the shock capturing technique, the computation collapsed. A further important fact is that the first formulation collapses for a higher coefficient α, even if the stabilization technique is apllied. Contrary to this, the second formulation works well also for a higher elevation.

Name of constant	Setting
CLF	0.75
ISO	0
MAXN	10
LINEAR SOLVER	0
PRECOND	2
STABILISE	2
BSC	2
WALLBSC	1

Table 8.1: Setting of the comparing computation.

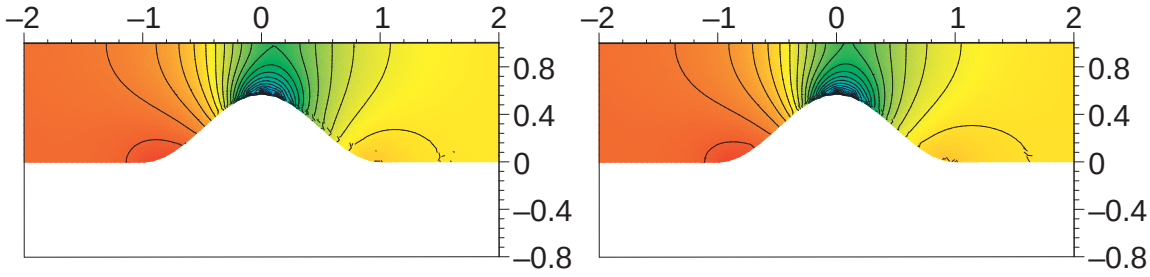
Name of constant	Setting
RHOO	1.0
VXO	1.0
VYO	0
PO	159.11912

Table 8.2: Setting of initial conditions.



(a) ALE formulation I, $t \doteq 0.5392$

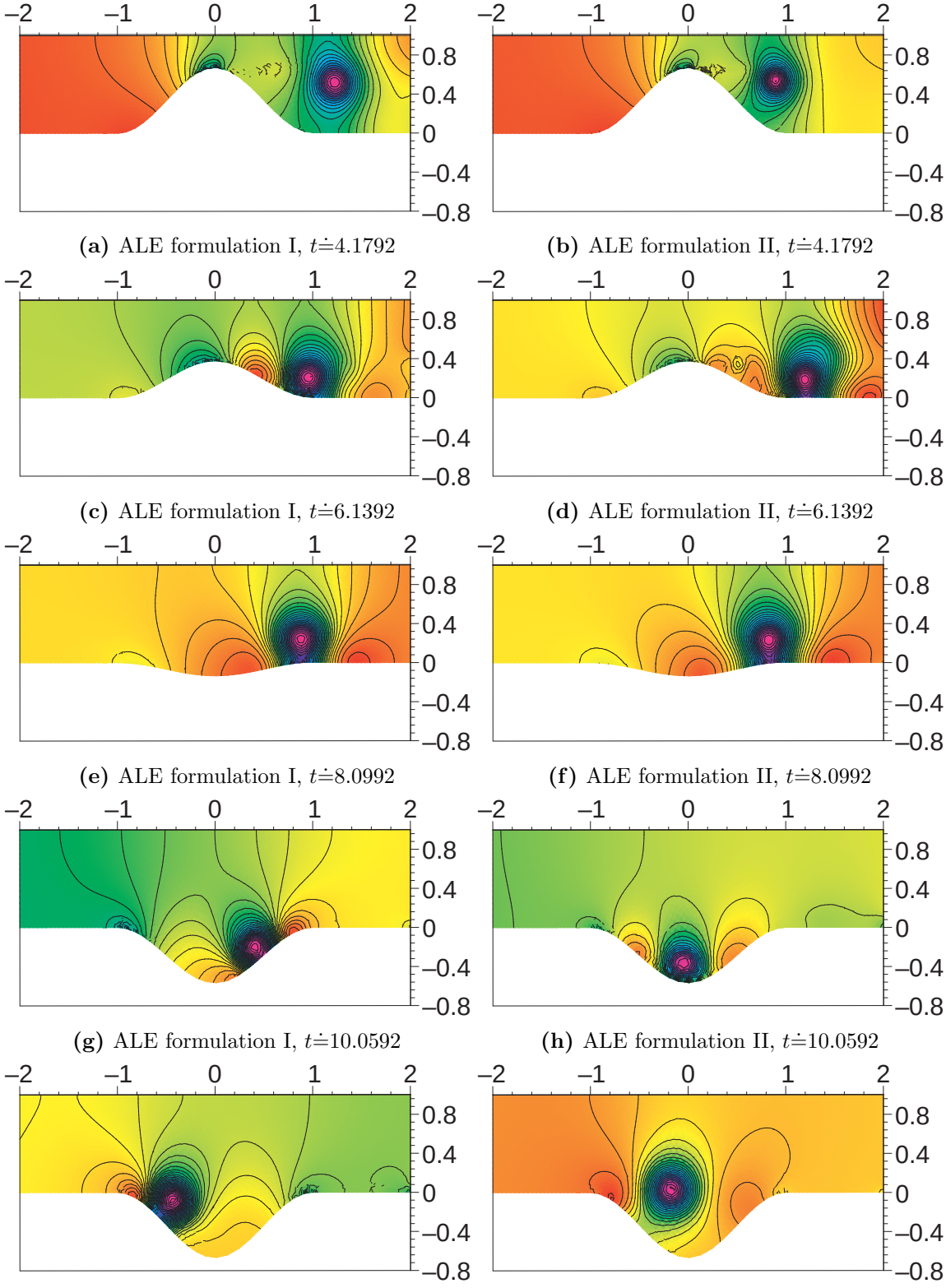
(b) ALE formulation II, $t \doteq 0.5392$



(c) ALE formulation I, $t \doteq 2.2192$

(d) ALE formulation II, $t \doteq 2.2192$

Figure 8.1: Comparison of the pressure isolines for ALE formulation I and II - the first part



(a) ALE formulation I, $t \doteq 4.1792$

(b) ALE formulation II, $t \doteq 4.1792$

(c) ALE formulation I, $t \doteq 6.1392$

(d) ALE formulation II, $t \doteq 6.1392$

(e) ALE formulation I, $t \doteq 8.0992$

(f) ALE formulation II, $t \doteq 8.0992$

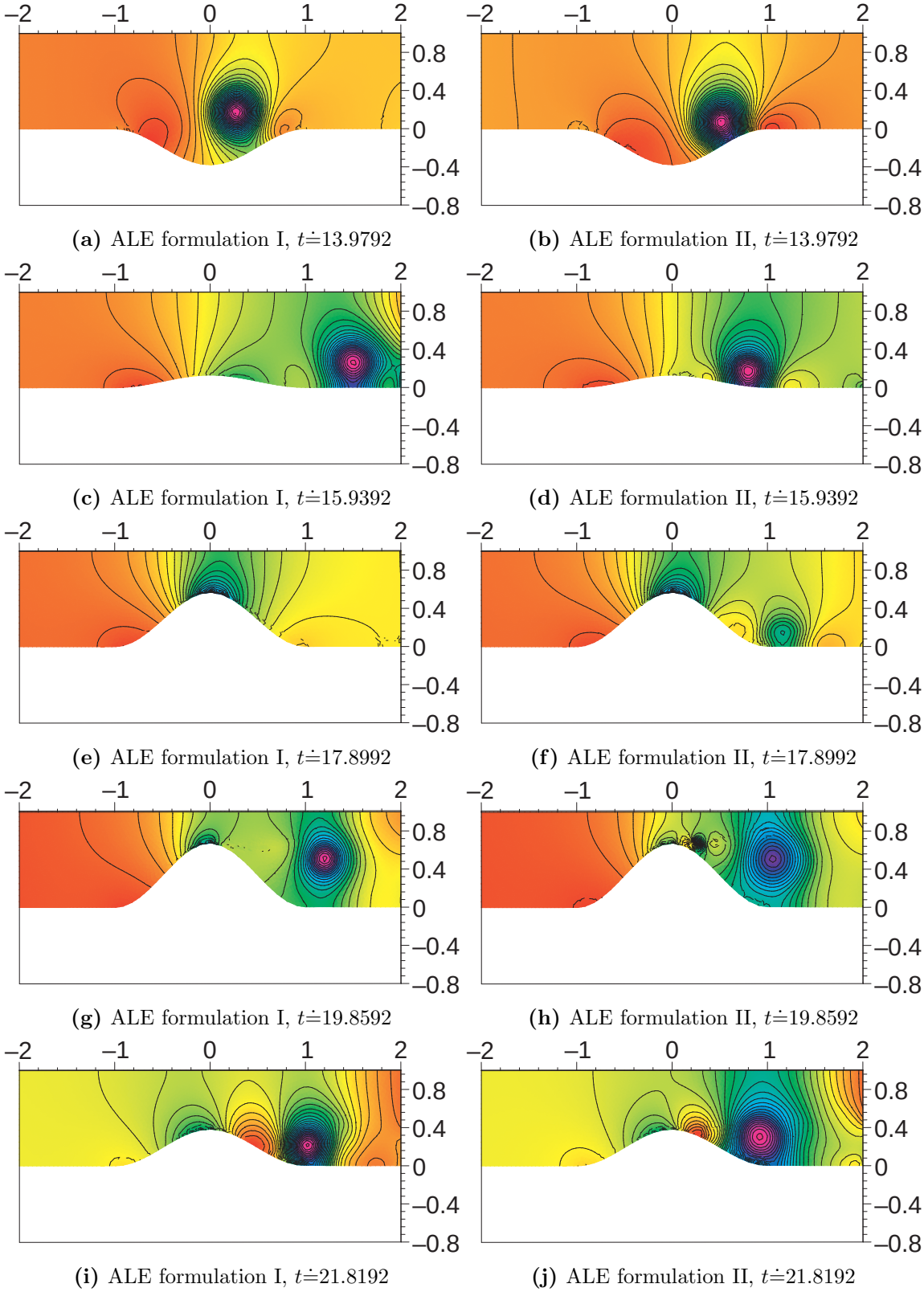
(g) ALE formulation I, $t \doteq 10.0592$

(h) ALE formulation II, $t \doteq 10.0592$

(i) ALE formulation I, $t \doteq 12.0192$

(j) ALE formulation II, $t \doteq 12.0192$

Figure 8.2: Comparison of the pressure isolines for ALE formulation I and II - the second part



(a) ALE formulation I, $t \doteq 13.9792$

(b) ALE formulation II, $t \doteq 13.9792$

(c) ALE formulation I, $t \doteq 15.9392$

(d) ALE formulation II, $t \doteq 15.9392$

(e) ALE formulation I, $t \doteq 17.8992$

(f) ALE formulation II, $t \doteq 17.8992$

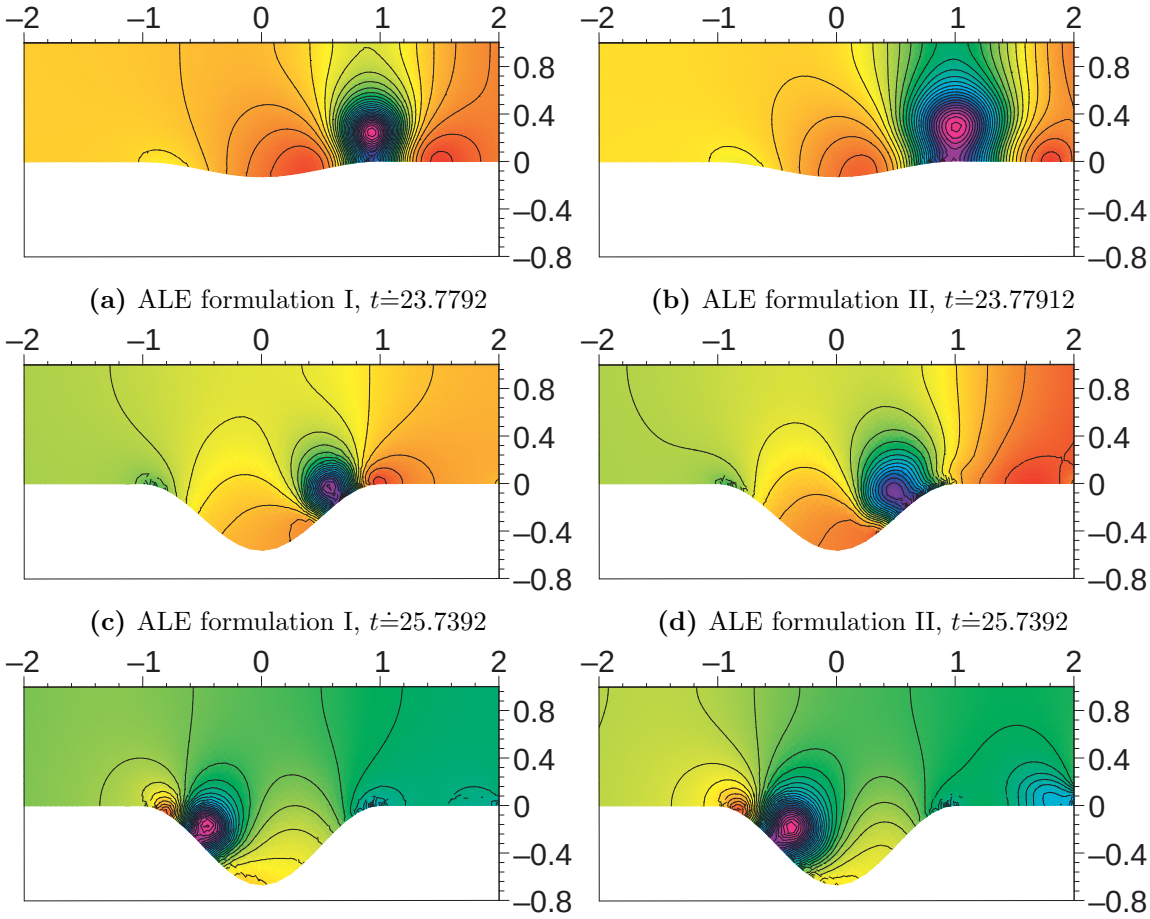
(g) ALE formulation I, $t \doteq 19.8592$

(h) ALE formulation II, $t \doteq 19.8592$

(i) ALE formulation I, $t \doteq 21.8192$

(j) ALE formulation II, $t \doteq 21.8192$

Figure 8.3: Comparison of the pressure isolines for ALE formulation I and II - the third part



(a) ALE formulation I, $t \doteq 23.7792$

(b) ALE formulation II, $t \doteq 23.77912$

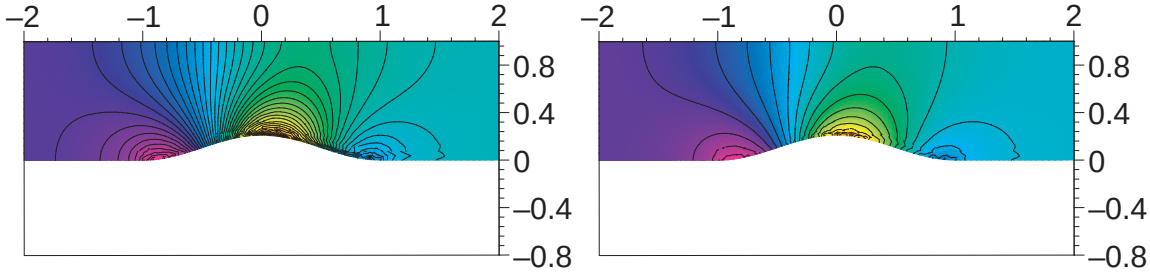
(c) ALE formulation I, $t \doteq 25.7392$

(d) ALE formulation II, $t \doteq 25.7392$

(e) ALE formulation I, $t \doteq 27.6992$

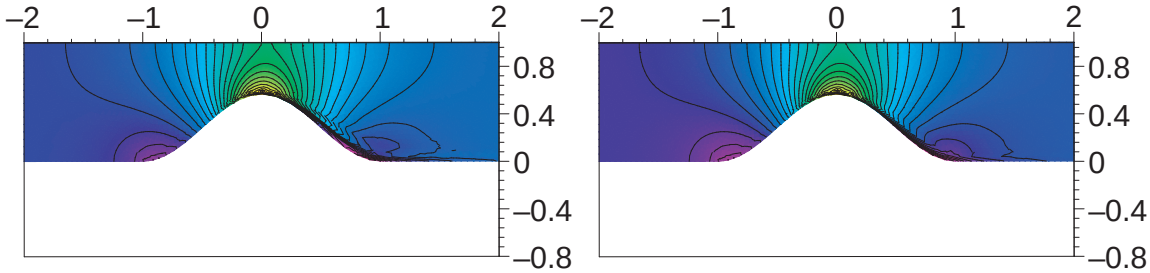
(f) ALE formulation II, $t \doteq 27.6992$

Figure 8.4: Comparison of the pressure isolines for ALE formulation I and II - the fourth part



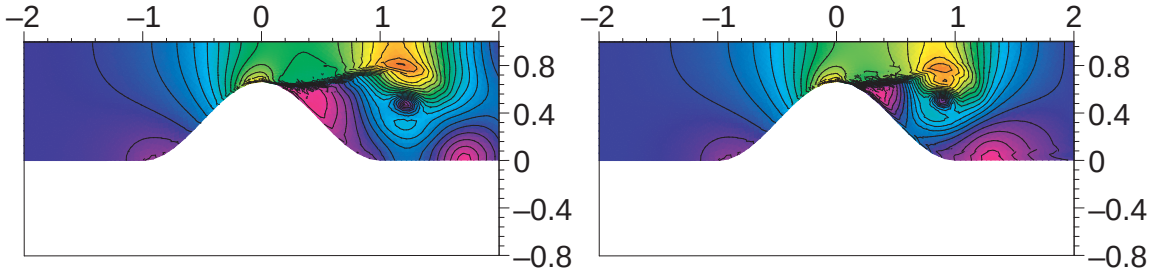
(a) ALE formulation I, $t \doteq 0.5392$

(b) ALE formulation II, $t \doteq 0.5392$



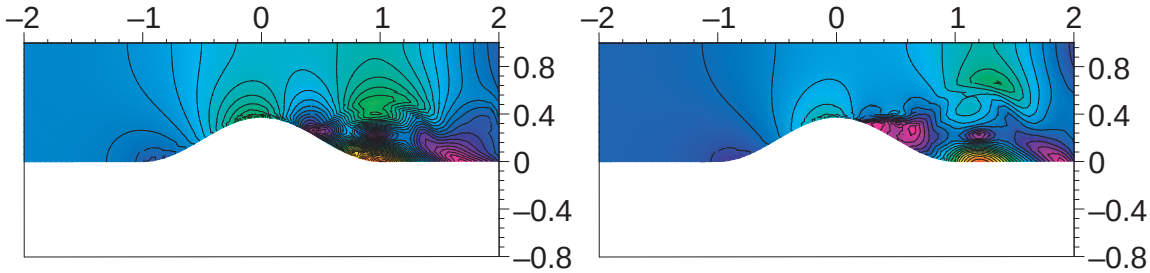
(c) ALE formulation I, $t \doteq 2.2192$

(d) ALE formulation II, $t \doteq 2.2192$



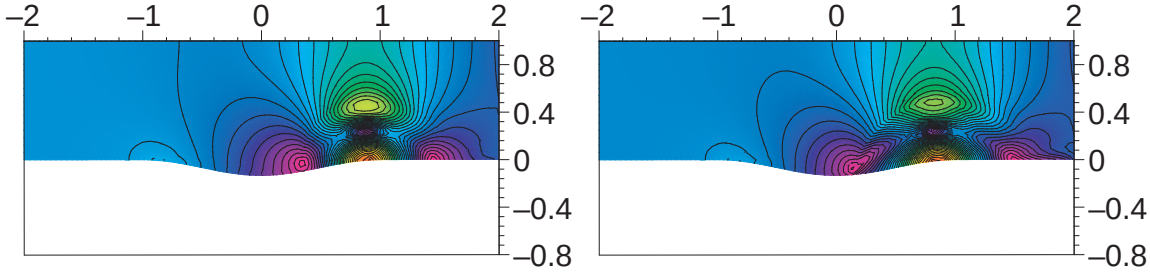
(e) ALE formulation I, $t \doteq 4.1792$

(f) ALE formulation II, $t \doteq 4.1792$



(g) ALE formulation I, $t \doteq 6.1392$

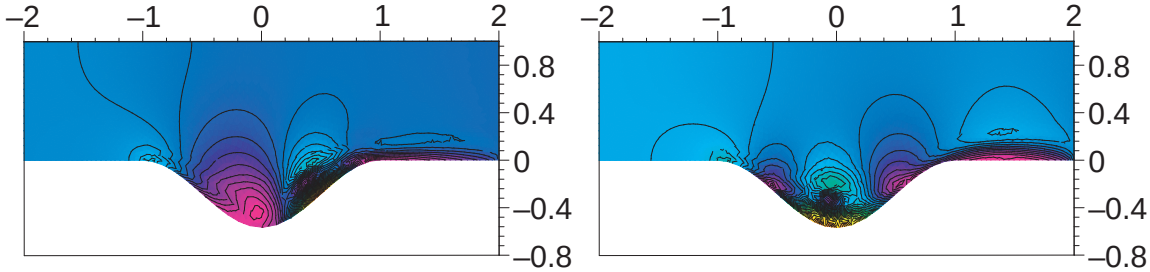
(h) ALE formulation II, $t \doteq 6.1392$



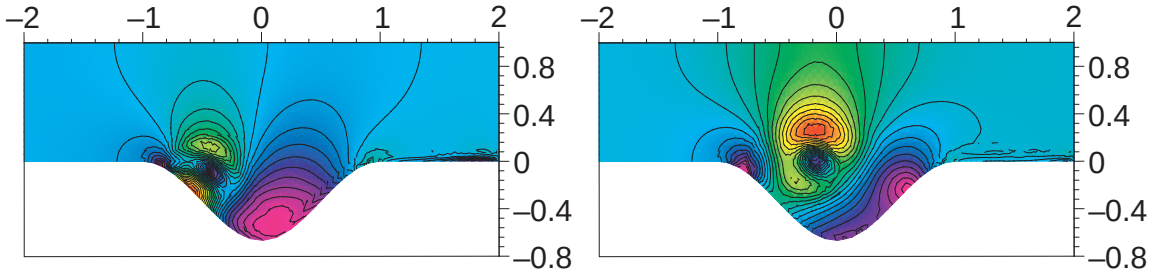
(i) ALE formulation I, $t \doteq 8.0992$

(j) ALE formulation II, $t \doteq 8.0992$

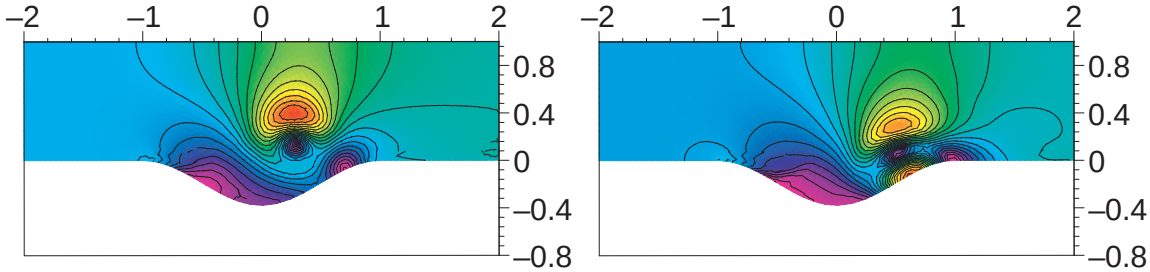
Figure 8.5: Comparison of the velocity isolines for ALE formulation I and II - the first part



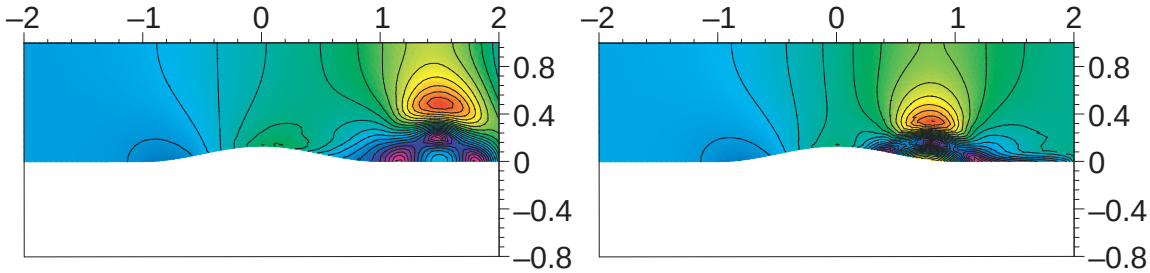
(a) ALE formulation I, $t \doteq 10.0592$ **(b)** ALE formulation II, $t \doteq 10.0592$



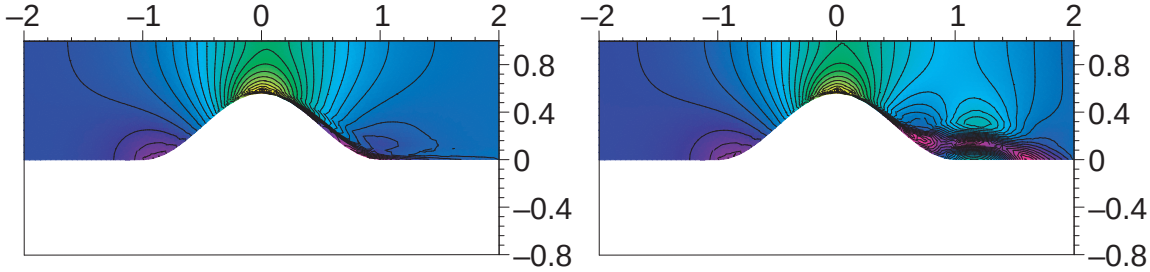
(c) ALE formulation I, $t \doteq 12.0192$ **(d)** ALE formulation II, $t \doteq 12.0192$



(e) ALE formulation I, $t \doteq 13.9792$ **(f)** ALE formulation II, $t \doteq 13.9792$

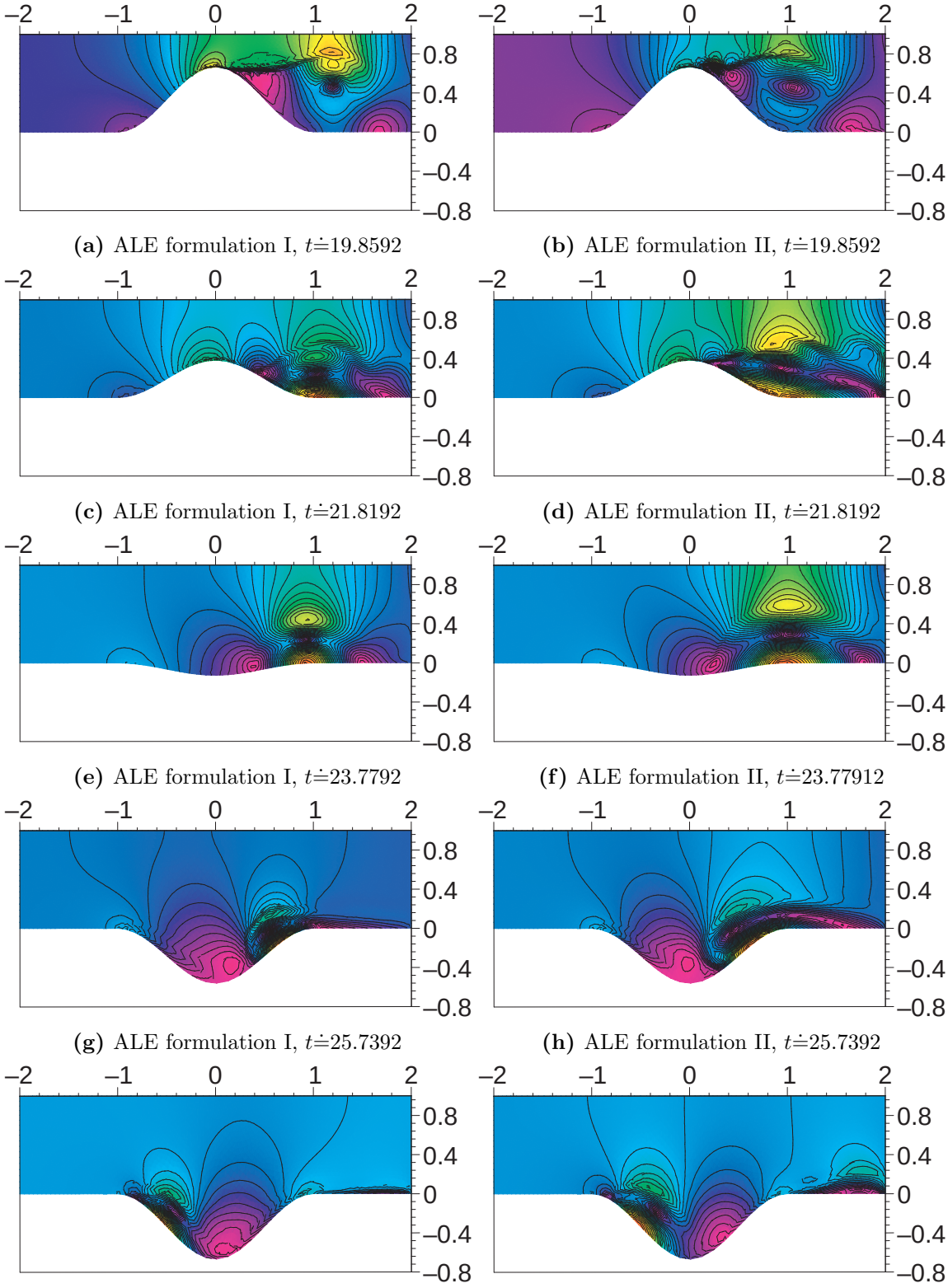


(g) ALE formulation I, $t \doteq 15.9392$ **(h)** ALE formulation II, $t \doteq 15.9392$



(i) ALE formulation I, $t \doteq 17.8992$ **(j)** ALE formulation II, $t \doteq 17.8992$

Figure 8.6: Comparison of the velocity isolines for ALE formulation I and II - the second part



(a) ALE formulation I, $t \doteq 19.8592$

(b) ALE formulation II, $t \doteq 19.8592$

(c) ALE formulation I, $t \doteq 21.8192$

(d) ALE formulation II, $t \doteq 21.8192$

(e) ALE formulation I, $t \doteq 23.7792$

(f) ALE formulation II, $t \doteq 23.77912$

(g) ALE formulation I, $t \doteq 25.7392$

(h) ALE formulation II, $t \doteq 25.7392$

(i) ALE formulation I, $t \doteq 27.6992$

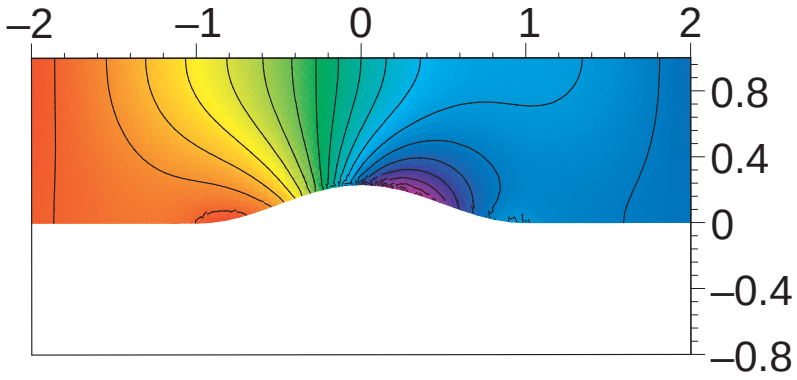
(j) ALE formulation II, $t \doteq 27.6992$

Figure 8.7: Comparison of the velocity isolines for ALE formulation I and II - the third part

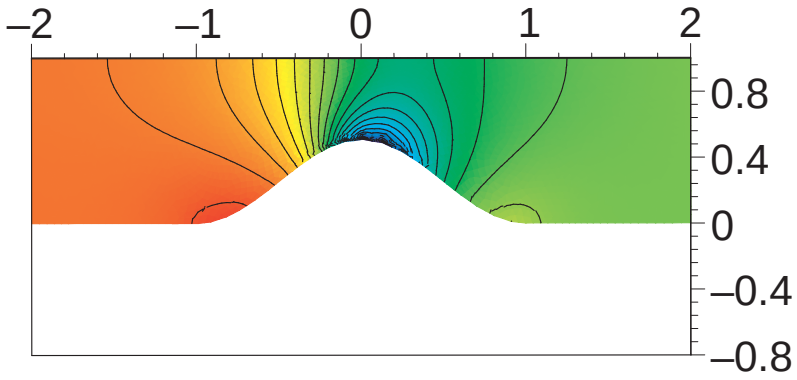
8.2 Results of the ALE formulation II

As mentioned in the previous section, the application of the ALE formulation I is limited by the elevation of the lower wall of the channel. The ALE formulation II does not suffer from this problem as will be seen in following figures. There is another problem that appears when the elavation of the lower wall of the channel is too large. In this case the elements of the computational mesh degenerate in the narrowest part of the channel. Of course, it causes the collaps of our computation.

Here we shall presented the maximal elevation of the bottom of the channel that we were able to compute with the use of the ALE formulation II. The computation was carried out with the same conditions and setting as in Section 8.1. The maximal possible value of the constant α was 0.45.

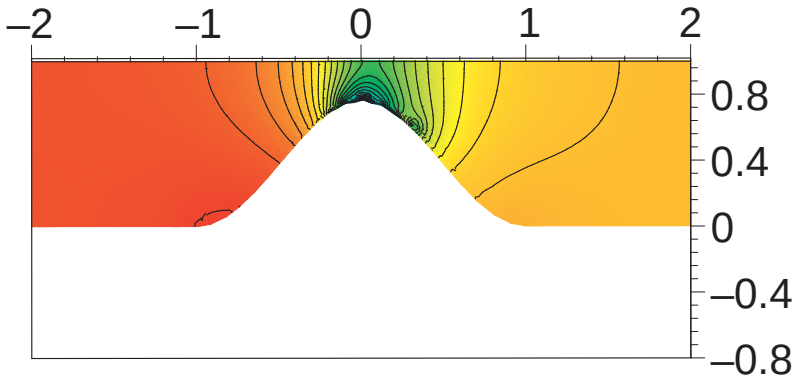


(a) $t \doteq 0.3992$

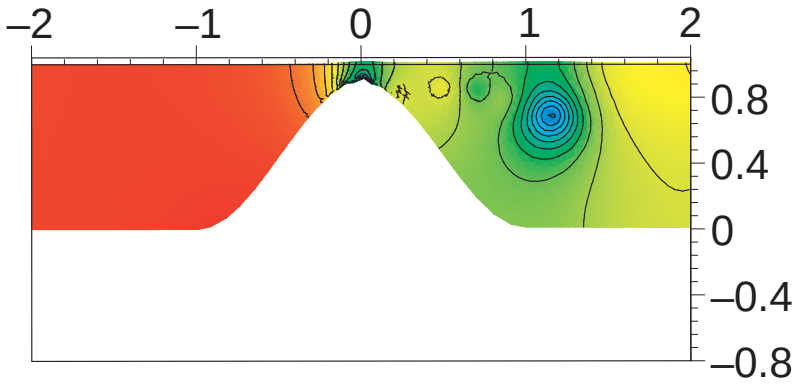


(b) $t \doteq 1.2392$

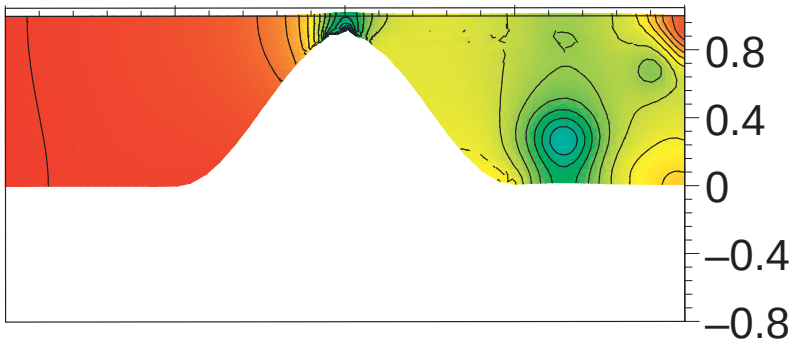
Figure 8.8: Pressure isolines for ALE formulation II - the first part



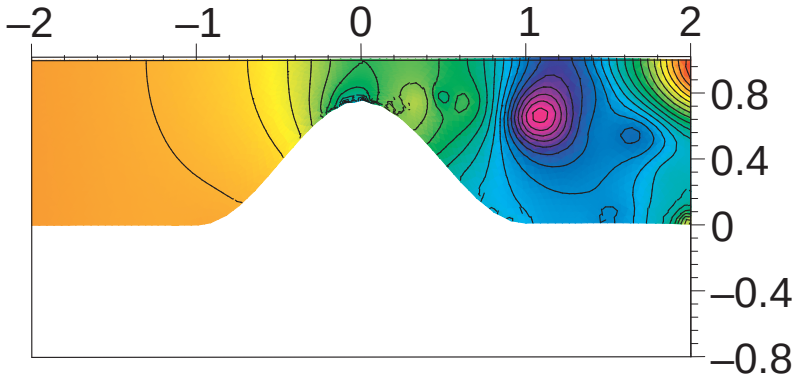
(a) $t \doteq 2.2192$



(b) $t \doteq 3.1992$

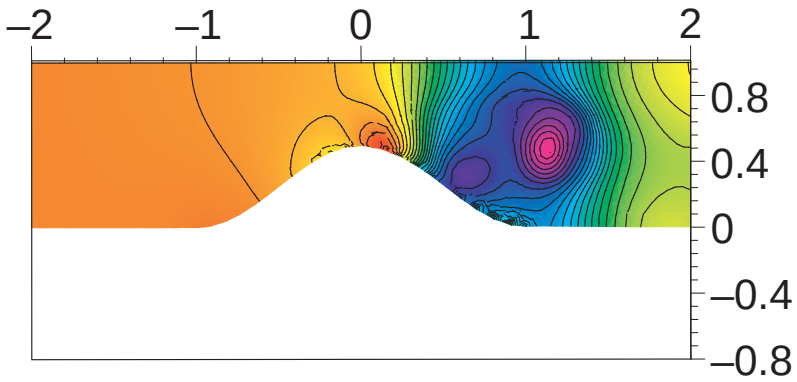


(c) $t \doteq 4.1792$

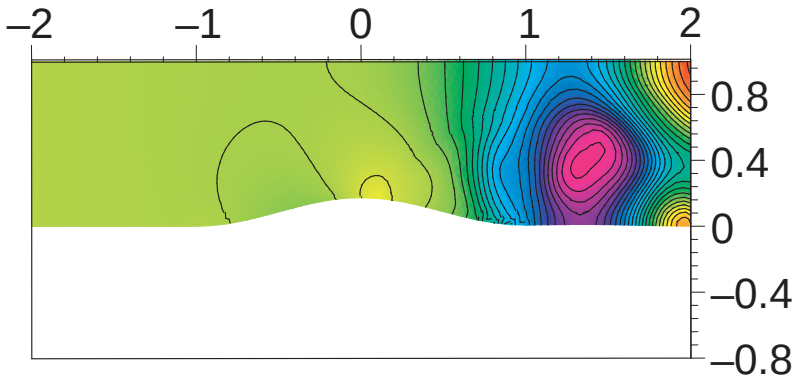


(d) $t \doteq 5.1592$

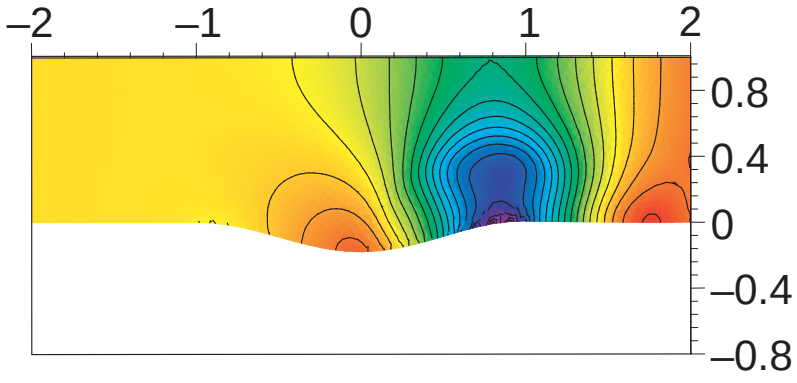
Figure 8.9: Pressure isolines for ALE formulation II - the second part



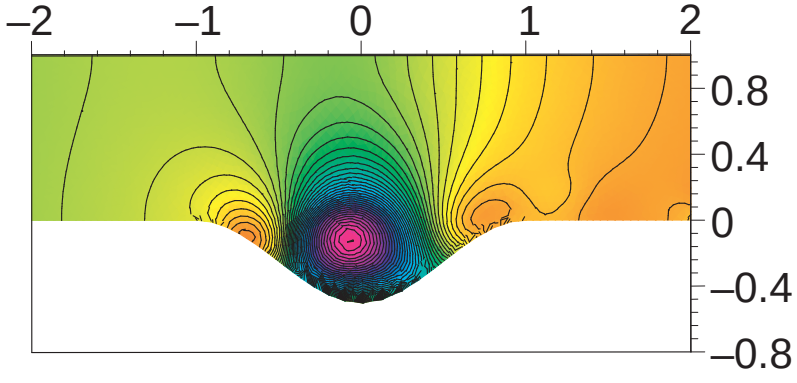
(a) $t \doteq 6.1392$



(b) $t \doteq 7.1192$

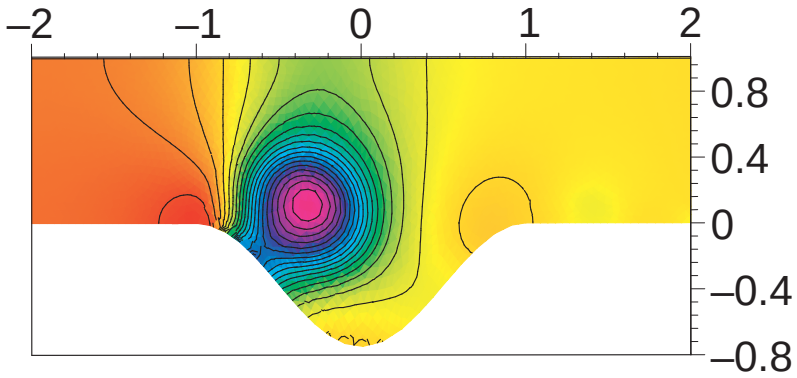


(c) $t \doteq 8.0992$

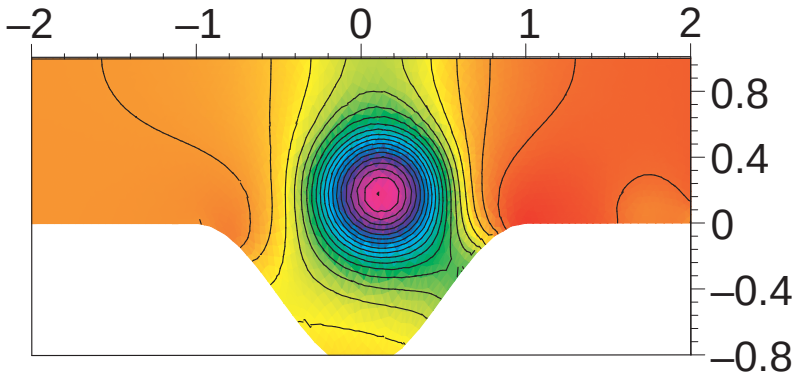


(d) $t \doteq 9.0792$

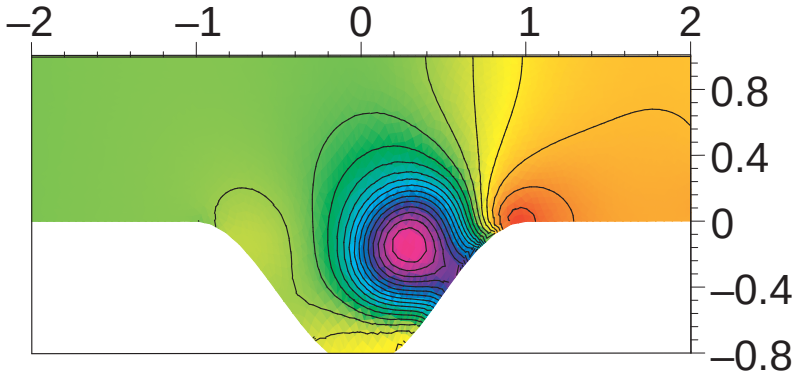
Figure 8.10: Pressure isolines for ALE formulation II - the third part



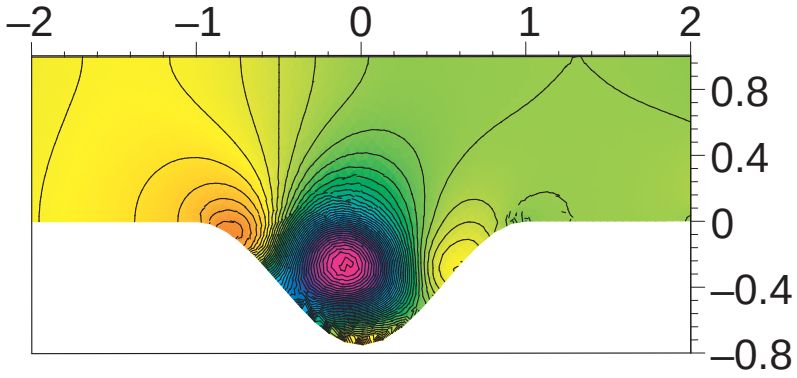
(a) $t \doteq 10.0592$



(b) $t \doteq 11.0392$

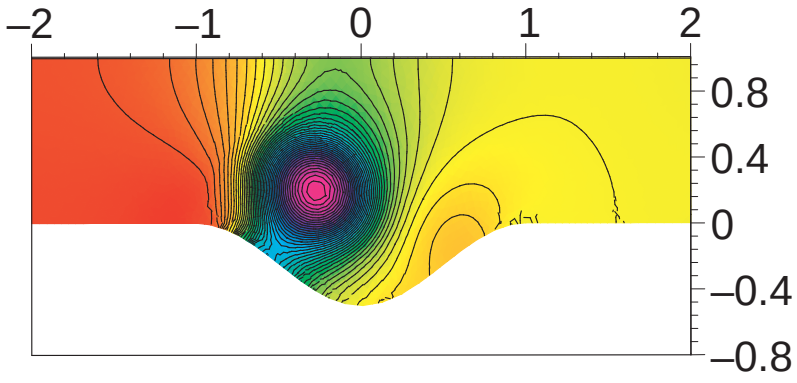


(c) $t \doteq 12.0192$

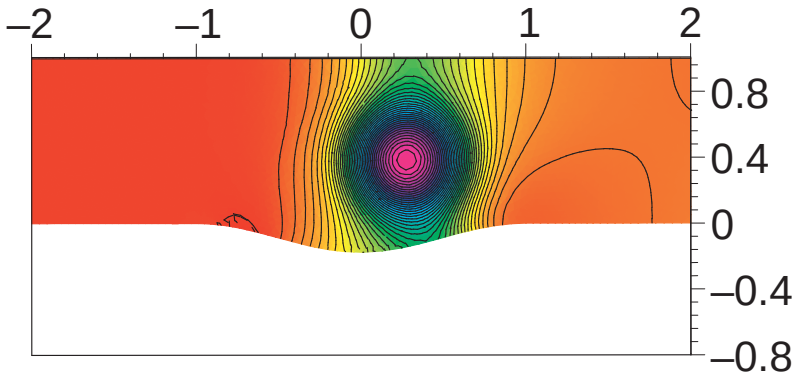


(d) $t \doteq 12.9992$

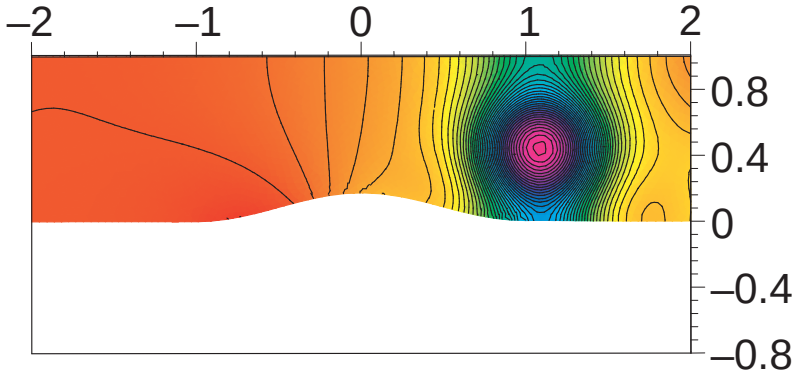
Figure 8.11: Pressure isolines for ALE formulation II - the fourth part



(a) $t \doteq 13.9792$

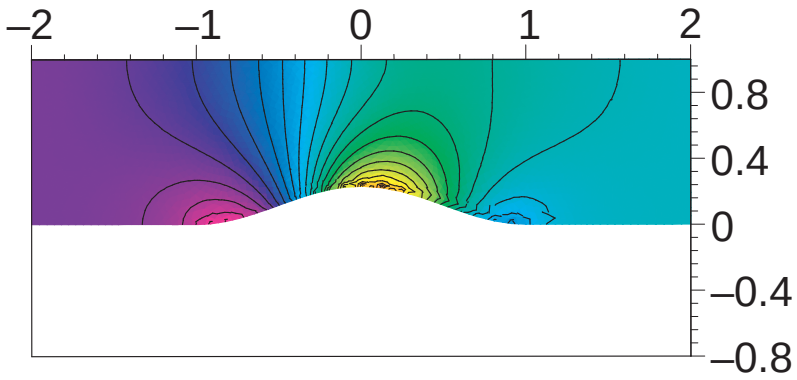


(b) $t \doteq 14.9592$

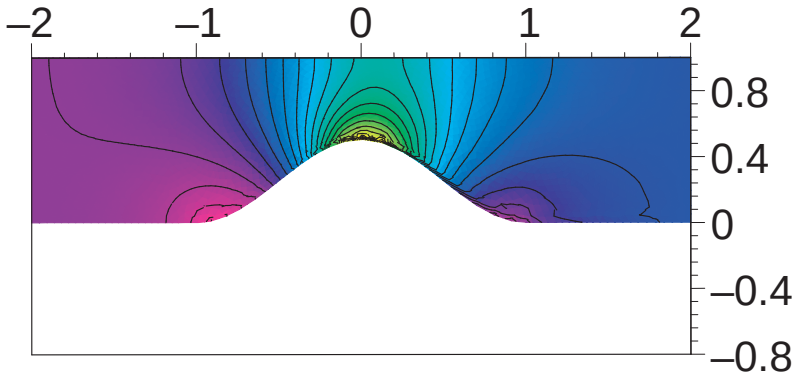


(c) $t \doteq 15.9392$

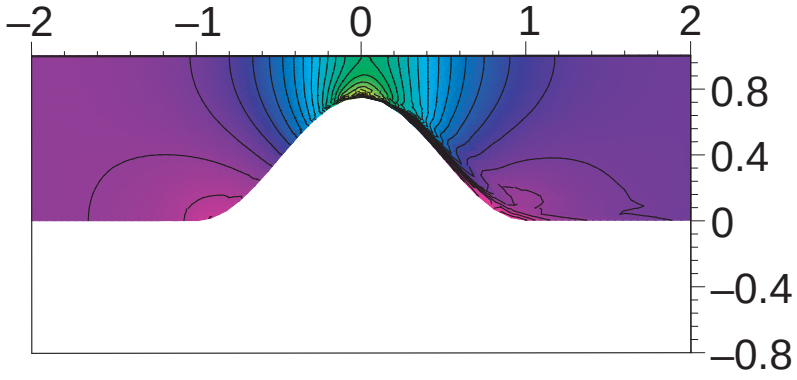
Figure 8.12: Pressure isolines for ALE formulation II - the fifth part



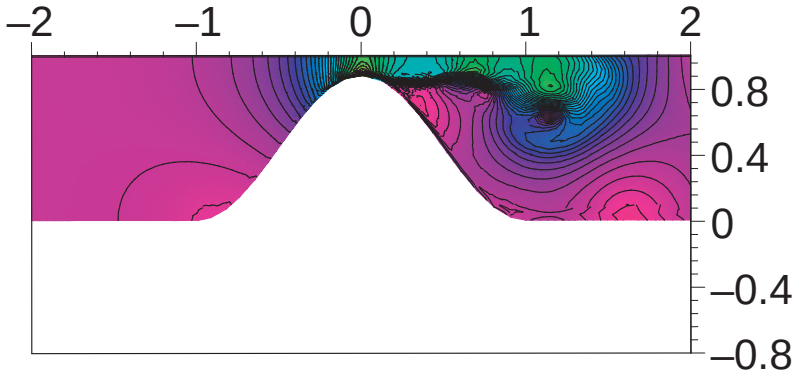
(a) $t \doteq 0.3992$



(b) $t \doteq 1.2392$

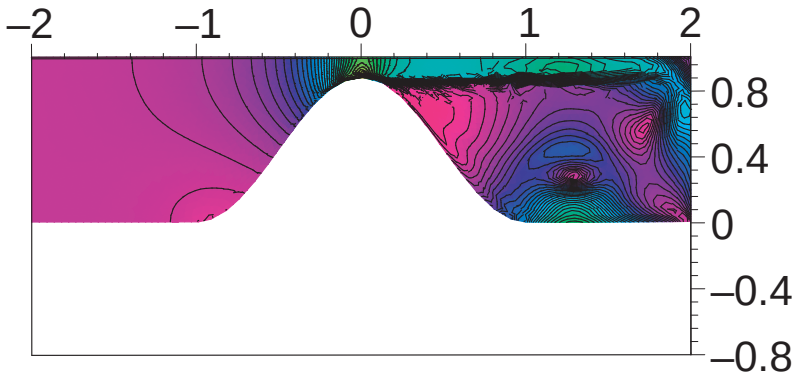


(c) $t \doteq 2.2192$

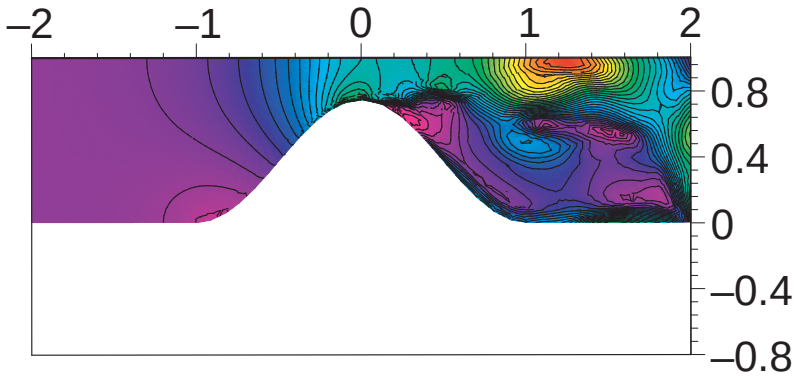


(d) $t \doteq 3.1992$

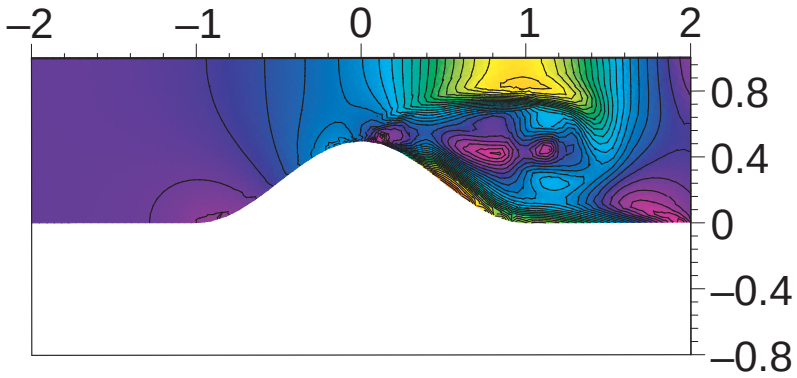
Figure 8.13: Velocity isolines for ALE formulation II - the first part



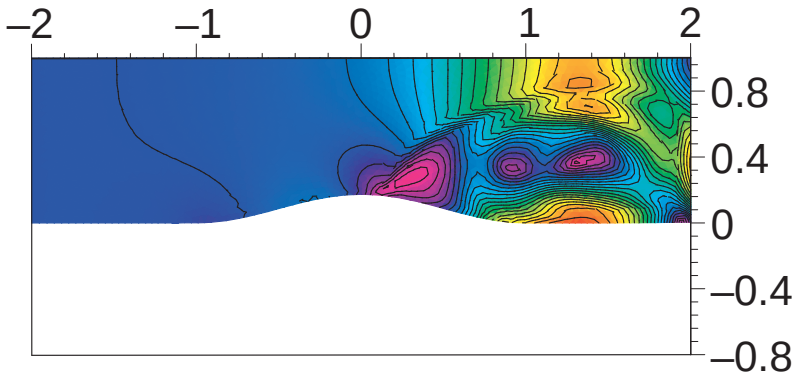
(a) $t \doteq 4.1792$



(b) $t \doteq 5.1592$

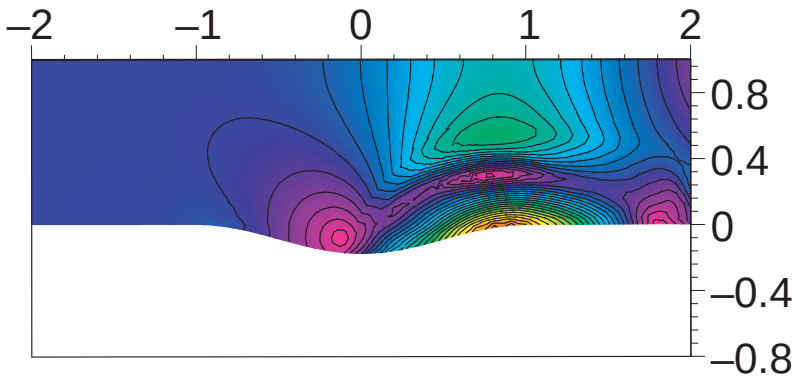


(c) $t \doteq 6.1392$

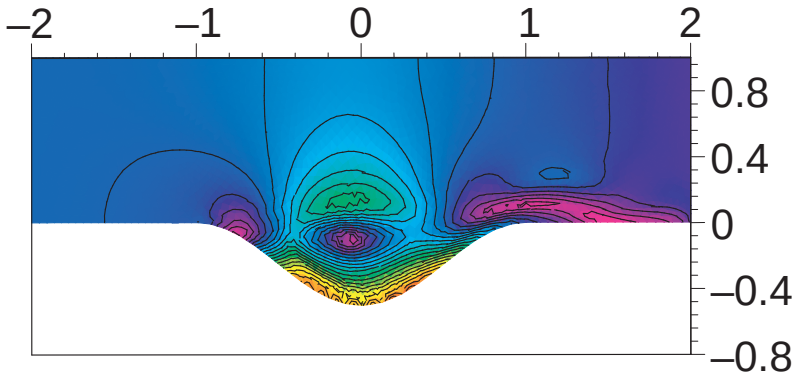


(d) $t \doteq 7.1192$

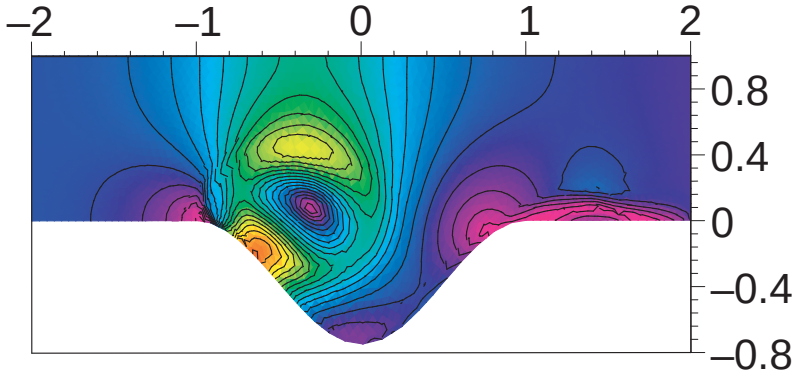
Figure 8.14: Velocity isolines for ALE formulation II - the second part



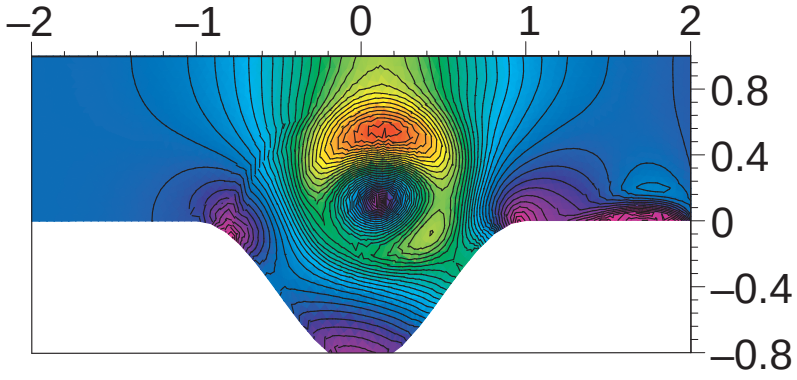
(a) $t \doteq 8.0992$



(b) $t \doteq 9.0792$

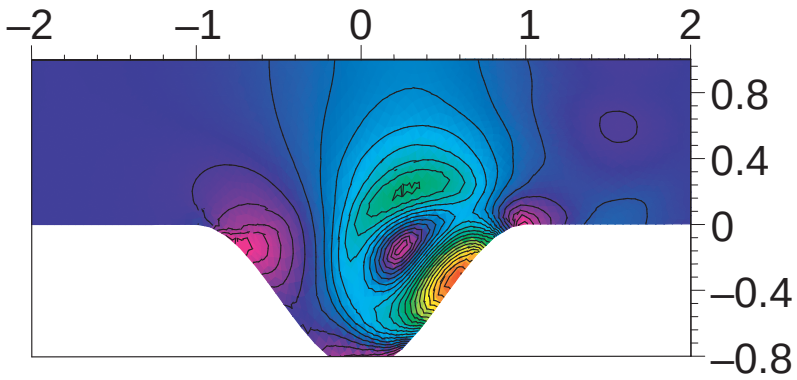


(c) $t \doteq 10.0592$

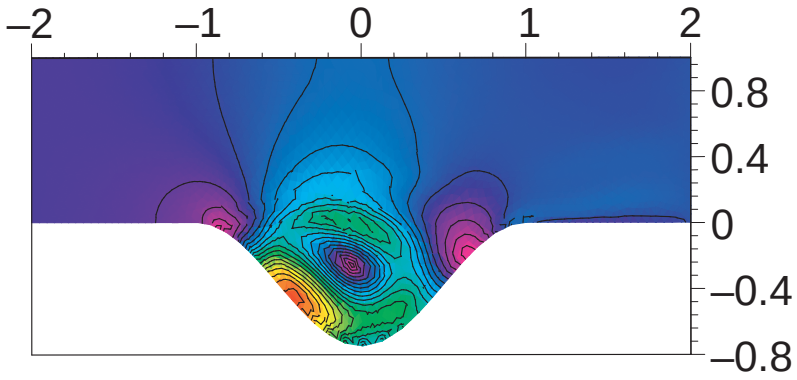


(d) $t \doteq 11.0392$

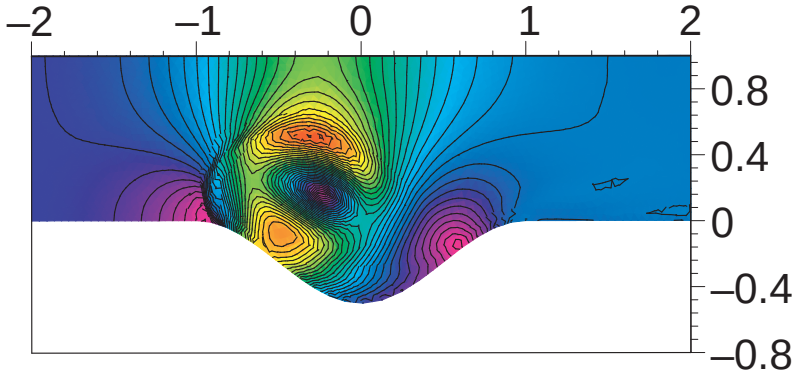
Figure 8.15: Velocity isolines for ALE formulation II - the third part



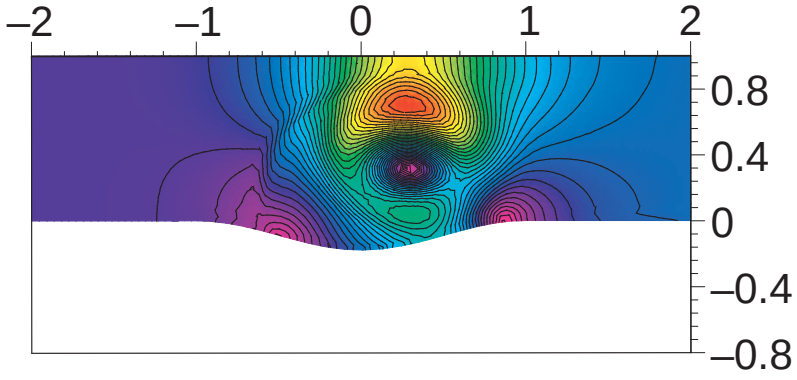
(a) $t \doteq 12.0192$



(b) $t \doteq 12.9992$

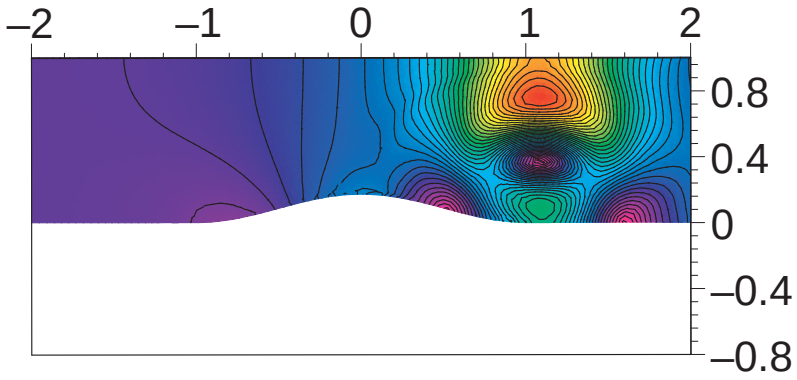


(c) $t \doteq 13.9792$



(d) $t \doteq 14.9592$

Figure 8.16: Velocity isolines for ALE formulation II - the fourth part



(a) $t \doteq 15.9392$

Figure 8.17: Velocity isolines for ALE formulation II - the fifth part

Conclusion

In the first chapter of this thesis, we have formulated and derived the governing equations describing the compressible flow. The chapter starts with the description of the flow and the transport theorem that is used for the derivation of the continuity equation and the equations of motion. Further, the Navier-Stokes equations, the equation of energy and some termodynamical relations are presented.

The second chapter is concerned with the Euler equations. The properties of this system are presented as well as the conservative form of these equations. Since the whole thesis is interested in the 2D compressible flow, the special care is paid to the properties of the Euler equations in conservative form in a two-dimensional domain that are used in the following chapter for a discretization of the system.

The third chapter discusses the problem of the Euler equations in a time-dependent domain, which is the main aim of this thesis. First, the Arbitrary Lagrangian-Eulerian (ALE) method allowing us to deal with the problem of a time-dependent domain is introduce. Further, two different formulations of the Euler equations are derived. The suitability of each of these formulations is discussed in Chapter 8, where also some results allowing a comparison of these formulations are presented.

In the following chapters the discretization of the compressible Euler equations by the discontinuous Galerkin finite element method (DGFEM) can be found. The space semidiscretization and time discretization using the linearization with the aid of the Vijayasundaram numerical flux are performed for both time-independent and time-dependent problem. These discretizations are separetly treated for both ALE formulations of the problem. In addition, the boundary conditions, a stabilization of the obtained schemes and an application of the isoparametric elements are discussed. It is important to mention that the shock capturing technique plays an important role, because the closing of the channel in our numerical simulations changes a character of the flow. The flow with a low inlet velocity is developed to the transonic flow and thus, the computation without suitable stabilization collapses for the motion with a higher amplitude.

The fifth chapter is devoted to the construction of the ALE mapping. Especially, the example used in our numerical experiments is described and its properties are investigated.

In the following we are concerned with the construction of a linear discrete system obtained with the aid of linear and quadratic finite elements. The description of the Generalized Minimal Residual (GMRES) method is given together with a basic algorithm proceeding the $\mathbb{QR}$ factorization with the aid of the Givens rotation. It is necessary to note that there are more possibilities how to effect the $\mathbb{QR}$ factorization. Some possibilities can be found in [9]. There also exist computational packages that can be implemented to the program and ensure the optimality of the algorithm.

In the seventh chapter we shortly described the used program with which all our results were obtained. It should allow the reader to run the program.

The last chapter is devoted to the presentation of our results. In the first section we compare the results of both ALE formulations. The presented closure of the channel was maximal possible in the case of the ALE formulation I even if the stabilisation was applied. On contrary, the ALE formulation II worked well also without the stabilisation. It shows a better suitability of the second formulation. In the second section of this chapter we have tried to completely close the channel. Based on the experience with the behaviour of the first ALE formulation, these tests were carried out only with the second ALE formulation. This formulation worked well. Only the used time step had to be reduced. Unfortunately, for the amplitudes larger than 0.9 the computation collapsed, because the elements in the narrowest part of the channel degenerated. Hence, we can assume that the limitation of the scheme is not caused by the ALE formulation II, but by the mesh. The construction of new mesh at each time level could lead to the better results. On contrary, the remeshing leads to larger round-off errors and the computation would be much longer.

Bibliography

[1] J. Kurzweil. *Ordinary Differential Equations*. Elsevier, Amsterdam, 1986.

[2] M. Feistauer. *Mathematical Methods in Fluid Dynamics*. Longman Scientific & Technical, Harlow, 1993.

[3] W. Rudin. *Real and Complex Analysis*. McGraw Hill, New York, 1974.

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

[5] J. Jaffre, C. Johnson, and A. Szepessy. Convergence of the discontinuous Galerkin finite elements for hyperbolic conservation laws. *Math. Models Methods Appl. Sci.*, 5:367–386, 1995.

[6] V. Dolejší, M. Feistauer, and C. Schwab. On some aspects of the discontinuous Galerkin finite elements mehod for conservation laws. *Math. Comput. Simul.*, 61:333–346, 2003.

[7] V. Dolejší and M. Feistauer. A Semi-Implicit Discontinuous Galerkin Finite Element Method for the Numerical Solution of Inviscid Compressible Flow. *Journal of Computational Physics*, 198:727–746, 2004.

[8] M. Fiedler. *Speciální matice a jejich použití v numerické matematice*. SNTL, Praha, 1981.

[9] G. H. Golub and Ch. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996.

[10] M. Feistauer and V. Kučera. On a robust discontinuous Galerkin technique for the solution of compressible flow. *Journal of Computational Physics*, 224:208–221, 2007.

[11] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J.Sci.Stat.Comput.*, 7(3):856–869, 1986.

[12] V. Kučera. *Higher order methods of the solution of compressible flows. Doctoral Thesis*. Faculty of mathematics and physics, Charles university in Prague, 2007.

[13] P. Sváček, M. Feistauer, and J. Horáček. Numerical simulation of flow induced airfoil vibrations with large amplitudes. *Journal of Fluid and Structures*, 7:391–411, 2007.

[14] V. Šidlof. *Fluid-structure interaction in human vocal folds. Doctoral Thesis.* Faculty of mathematics and physics, Charles university in Prague, 2007.

[15] P. Punčochářová, J. Horáček, K. Kozel, and J. Fürst. Numerical simulation of airflow through the oscillating glottis. In *5th International workshop: Model and Analysis of Vocal Emissions for Biomedical Applications*, 2007.