On the Method of Lines

Doctoral Thesis

August 1999 Pavel Šolín

Faculty of Mathematics and Physics
Charles University, Prague

On the Method of Lines

(Applications in Fluid Dynamics and A Posteriori Error Estimates)

Doctoral Thesis

Pavel Šolín

Supervisor: Doc. Dr. Jiří Felcman, CSc.
Consultant: Prof. Dr. Miloslav Feistauer, DrSc.
Branch: Mathematical and Computer Modelling

Contents

Acknowledgments

I would like to express my deep thanks to Prof. Dr. Miloslav Feistauer, DrSc. and Doc. Dr. Jiří Felcman, CSc. who have introduced me to Computational Fluid Dynamics and who have been supervising my work for more than five years. I deeply appreciate lots of hours that they spent with me discussing numerical schemes, revising my work and introducing me to new interesting open problems. Their enthusiasm, knowledge and readiness to answer my questions have been helping me a lot during the whole period of my Ph.D. study.

I am very grateful to Prof. Dr. Karel Kozel, DrSc. who not only discussed with me my work but who also provided me with contacts to other people dealing with a similar kind of problems. Their experience helped me then to treat succesfully several problems that I met during my work.

At the beginning of my study I met Doc. Dr. Josef Málek, CSc. who introduced me to the theory of partial differential equations. Thanks to him all the time I have been able to learn new analytical methods and to follow new trends and new results in the field of analytical investigation of flows, as an excellent complement to my numerical work.

Further I wish to thank Doc. Dr. Karel Segeth, DrSc. who advised me to use suitable numerical software for a part of my numerical problems. This software proved to be very efficient and reliable and it has considerably improved several of my numerical schemes.

At last, to keep the alphabetical order, I thank to Dr. Oldřich Ulrych, CSc. who has been maintaining the computer network. Thanks to him I have been able to perform series of extraordinary large numerical experiments that considerably influenced quality of numerical results introduced in this dissertation.

Prague, August 1999 Pavel Šolín

Preface

One of the basic principles of mathematics is to split complex problems into a number of less complicated ones, which have been already resolved or which are easier to treat.

Numerical solution of the compressible Euler and Navier-Stokes equations studied in the first three chapters of the thesis is such a complex problem. We apply the well-known Method of lines to both systems of partial differential equations to split their numerical solution into the following partial problems:

- To approximate the system of partial differential equations by a system of ordinary differential equations (ODE's),
- to analyze existence and uniqueness of solution of the arising system of ODE's,
- to solve efficiently and accurately the arising system of ODE's.

One can see that using the Method of lines, one works not only in the area of numerical solution of partial differential equations, but a considerable part of work is also done in the area of the analysis and numerical solution of ordinary differential equations. One can profit from many results which have been obtained in both fields.

The last chapter of the thesis, a result of a fruitful collaboration with Prof. Ivo Babuška and Prof. Miloslav Feistauer, introduces a new approach to a posteriori error estimates derived for multidimensional, generally nonlinear parabolic equations and linear hyperbolic problems.

Chapter 1

On the Finite Volume Semi-Discretization of Compressible Euler Equations

Abstract: Euler equations describing the motion of compressible inviscid fluid flow are semi–discretized in space by the Finite Volume Method. The semi–discrete problem is written as a system of non–linear ordinary differential equations of the form $\dot{\boldsymbol{y}}(t) = \boldsymbol{F}(\boldsymbol{y}(t), t)$, $\boldsymbol{y}(0) = \boldsymbol{y}^0$. Analysis of the Lipschitz continuity of the right–hand–side vector function $\boldsymbol{F}$ is performed and statements on unique solvability of the system are made. The system of ODE's is integrated in time using an implicit higher-order scheme suitable for stiff systems, based on the so-called *Backward Differentiation Formula* (BDF). One-, two- and three-dimensional numerical examples illustrating its advantages concerning precision and CPU efficiency are presented.

1.1 Introduction

Let the domain $\Omega \subset \mathbb{R}^d$ occupied by the fluid be bounded and have a piecewise smooth boundary $\partial\Omega$. The d-dimensional compressible Euler equations consist of the continuity equation, d Euler momentum equations and energy equation. We will consider them in the space-time cylinder $Q_T = \Omega \times (0,T)$ $(T > 0)$ and write them in the form

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{s=1}^{d} \frac{\partial \boldsymbol{f}_s(\boldsymbol{w})}{\partial x_s} = 0, \tag{1.1}$$

where

$$\boldsymbol{w} = \begin{pmatrix} \rho \\ \rho v_1 \\ \vdots \\ \rho v_d \\ e \end{pmatrix}, \quad \boldsymbol{f}_s(\boldsymbol{w}) = \begin{pmatrix} \rho v_s \\ \rho v_s v_1 + \delta_{s1} p \\ \vdots \\ \rho v_s v_d + \delta_{sd} p \\ (e+p)\, v_s \end{pmatrix}, \quad s = 1, \dots, d \tag{1.2}$$

and

$$e = \frac{p}{\kappa - 1} + \frac{1}{2}\rho|\boldsymbol{v}|^2. \tag{1.3}$$

We use the standard notation: t – time, $x_1, \dots, x_d$ – Cartesian coordinates in $\mathbb{R}^d$, $\boldsymbol{x} = (x_1, \dots, x_d)$, ρ – density, $\boldsymbol{v} = (v_1, \dots, v_d)$ – velocity vector with components $v_1, \dots, v_d$ in the directions $x_1, \dots, x_d$, p – pressure, e – total energy, δ_{sj} – Kronecker delta, $\kappa > 1$ – Poisson adiabatic constant. Due to physical reasons it is also suitable to require $\rho > 0, p > 0$. The functions $\boldsymbol{f}_s$, $s = 1, \dots, d$ are called inviscid (Euler) fluxes and are defined in the set

$$\mathcal{Q} = \{(w_1, \dots, w_{d+2}) \in \mathbb{R}^{d+2};\ w_1 > 0, w_{d+2} - \frac{w_2^2 + \dots + w_{d+1}^2}{2w_1} > 0\}. \tag{1.4}$$

System (1.1) (1.3) is equipped with the initial conditions

$$\boldsymbol{w}(\boldsymbol{x}, 0) = \boldsymbol{w}^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{1.5}$$

and a set of boundary conditions which will be specified later.

1.2 Semi–Discrete Problem

Integrating equation (1.1) over a finite volume Ω_i, $1 \leq i \leq N$ where N denotes number of finite volumes and using Green's theorem and the so-called *rotational invariance* of the Euler equations (for detail description see e.g. [28]), we obtain

$$
\begin{aligned}
\int_{\Omega_i} \frac{\partial \boldsymbol{w}}{\partial t}(\boldsymbol{x}, t)\ \mathrm{d}\boldsymbol{x} = & - \sum_{j \in S_{INT}(i)} \int_{\Gamma_{ij}} \mathbb{T}_d^{-1}(\nu_{ij}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{ij}) \boldsymbol{w}(\boldsymbol{x}, t))\ \mathrm{d}S \\
& - \sum_{\alpha \in S_{SOL}(i)} \int_{\Gamma_{i\alpha}} \mathbb{T}_d^{-1}(\nu_{i\alpha}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{i\alpha}) \boldsymbol{w}(\boldsymbol{x}, t))\ \mathrm{d}S \\
& - \sum_{i\alpha \in S_{I/O}(i)} \int_{\Gamma_{i\alpha}} \mathbb{T}_d^{-1}(\nu_{i\alpha}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{i\alpha}) \boldsymbol{w}(\boldsymbol{x}, t))\ \mathrm{d}S
\end{aligned}
\tag{1.6}
$$

for all $t \in (0, T)$ and $1 \leq i \leq N$. The index set $S_{INT}(i)$ contains indices of finite volumes adjacent to the finite volume Ω_i for all $1 \leq i \leq N$. Common edge (face) of adjacent finite volumes Ω_i and Ω_j is denoted Γ_{ij}. Analogously, we define the index sets $S_{SOL}(i)$ and $S_{I/O}(i)$: by $\Gamma_{i\alpha}$, $\alpha \in S_{SOL}(i)$ and $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}(i)$ are denoted edges (faces) of the finite volume Ω_i representing solid wall and inlet/outlet, respectively. Symbols $\nu_{ij}, \nu_{i\alpha} \in \mathbb{R}^d$ denote unit normal vectors to $\Gamma_{ij}, \Gamma_{i\alpha}$, respectively, pointing outward of Ω_i. The rotational matrices $\mathbb{T}_d(\nu)$, $\nu \in \mathbb{R}^d$, $|\nu| = 1$, $d = 1, 2, 3$ are defined as follows:

In 1D

$$
\mathbb{T}_1(\nu) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.7}
$$

In 2D, expressing ν in polar coordinates as $\nu = [\cos(\alpha), \sin(\alpha)]^{\mathrm{T}}, \alpha \in \langle -\pi, \pi \rangle$, we define

$$
\mathbb{T}_2(\nu) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) & 0 \\ 0 & -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1.8}
$$

In 3D, expressing ν in spherical coordinates as

$$
\nu = [\cos(\alpha)\cos(\beta), \sin(\alpha)\cos(\beta), \sin(\beta)]^{\mathrm{T}}, \alpha \in \langle -\pi, \pi \rangle, \beta \in \langle -\frac{\pi}{2}, \frac{\pi}{2} \rangle, \tag{1.9}
$$

the matrix $\mathbb{T}_3(\nu)$ has the form

$$
(1.10) \quad \mathbb{T}_3(\nu) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos(\alpha)\cos(\beta) & \sin(\alpha)\cos(\beta) & \sin(\beta) & 0 \\ 0 & -\sin(\alpha) & \cos(\alpha) & 0 & 0 \\ 0 & -\cos(\alpha)\sin(\beta) & -\sin(\alpha)\sin(\beta) & \cos(\beta) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.
$$

Replacing in (1.6) the function $\boldsymbol{w}(\boldsymbol{x}, t)$ with the element average function $\boldsymbol{w}_h(\boldsymbol{x}, t)$ which is piecewise constant in space on the finite volumes $\Omega_1, \Omega_2, \ldots, \Omega_N$ and approximating the right-hand-side integrals by integrals of the so-called *numerical fluxes* H_{INT}, H_{SOL} and $H_{I/O}$, we obtain

$$
\begin{aligned}
\frac{\partial \boldsymbol{w}_h}{\partial t}(\boldsymbol{x},\, t)|_{\Omega_i} = \; & - \frac{1}{|\Omega_i|} \sum_{j \in S_{INT}(i)} H_{INT}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_j}, \nu_{ij}(\boldsymbol{x}))|\Gamma_{ij}| \\
(1.11) \qquad & - \frac{1}{|\Omega_i|} \sum_{\alpha \in S_{SOL}(i)} H_{SOL}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \nu_{i\alpha}(\boldsymbol{x}))|\Gamma_{i\alpha}| \\
& - \frac{1}{|\Omega_i|} \sum_{\alpha \in S_{I/O}(i)} H_{I/O}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \boldsymbol{w}_h^B(\boldsymbol{x},\, t)|_{\Gamma_{i\alpha}}, \nu_{i\alpha}(\boldsymbol{x}))|\Gamma_{i\alpha}|,
\end{aligned}
$$

$t \in (0, T)$, $1 \le i \le N$, where $\nu_{i\alpha}$ denotes the unit normal vector to $\Gamma_{i\alpha}$ pointing outward of Ω. Symbols $|\Omega_i|$ and $|\Gamma_{i\alpha}|$ denote d-dimensional measure of $|\Omega_i|$ and $(d-1)$-dimensional measure of $\Gamma_{i\alpha}$, respectively. The function $\boldsymbol{w}_h^B$ represents a boundary state corresponding to the inlet/outlet boundary face $\Gamma_{i\alpha}$ which will be discussed later.

1.2.1 Numerical Flux

The numerical flux H_{INT} through interior faces Γ_{ij} has the form

$$
(1.12) \qquad H_{INT}(\boldsymbol{w}_i, \boldsymbol{w}_j, \nu_{ij}) = \mathbb{T}_d^{-1}(\nu_{ij}) \boldsymbol{f}_R(\mathbb{T}_d(\nu_{ij})\boldsymbol{w}_i, \mathbb{T}_d(\nu_{ij})\boldsymbol{w}_j),
$$

where the so-called *approximate Riemann solver* $\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)$ approximates the constant value $\boldsymbol{f}_1(\boldsymbol{q}_r(0, \tilde{t}))$, $\boldsymbol{q}_r(\tilde{x}_1, \tilde{t})$ being the exact solution to the Riemann problem

$$
(1.13) \qquad \frac{\partial \boldsymbol{q}}{\partial \tilde{t}}(\tilde{x}_1, \tilde{t}) + \frac{\partial \boldsymbol{f}_1(\boldsymbol{q})}{\partial \tilde{x}_1}(\tilde{x}_1, \tilde{t}) = 0 \qquad \text{in } \mathbb{R} \times (0, +\infty),
$$

$$q(\tilde{x}_1, 0) = \begin{cases} q_L, \ \tilde{x}_1 < 0, \\ \\ q_R, \ \tilde{x}_1 > 0, \end{cases} \tag{1.14}$$

as described in detail e.g. in [28].

The numerical flux H_{SOL} through the solid-wall faces $\Gamma_{i\alpha}$ has the form

$$H_{SOL}(\boldsymbol{w}_i, \nu_{i\alpha}) = [0, p(\boldsymbol{w}_i), 0, \dots, 0]^T \tag{1.15}$$

where $p(\boldsymbol{w}_i)$ denotes the pressure corresponding to the state $\boldsymbol{w}_i$. In this approximation, the impermeability condition $\boldsymbol{v} \cdot \nu_{i\alpha} = 0$ for solid walls was used, see e.g. [27] for details. Finally, the numerical flux $H_{I/O}$ through the inlet/outlet faces $\Gamma_{i\alpha}$ has the form

$$H_{I/O}(\boldsymbol{w}_i, \boldsymbol{w}_{i\alpha}, \nu_{i\alpha}) = \mathbb{T}_d^{-1}(\nu_{i\alpha}) \boldsymbol{f}_1(\boldsymbol{w}^B(\mathbb{T}_d(\nu_{i\alpha})\boldsymbol{w}_i, \mathbb{T}_d(\nu_{i\alpha})\boldsymbol{w}_{i\alpha})) \tag{1.16}$$

where $\boldsymbol{w}_{i\alpha}$ represents the boundary state corresponding to the face $\Gamma_{i\alpha}$. The function $\boldsymbol{w}^B$ represents a switch deciding how much information is diffused from the boundary into the domain, depending on the number of outgoing characteristics of the approximate solution $\boldsymbol{w}_h$. Switches of this type are very characteristic for hyperbolic systems. For our purpose, let us consider a function

$$\boldsymbol{w}^B(\boldsymbol{w}_L, \boldsymbol{w}_R) = \begin{cases} \boldsymbol{w}_L \text{ if } w_{L,2} - \sqrt{\kappa \frac{p_L}{\varrho_L}} \geq 0 \quad (\textit{supersonic outlet}), \\ \\ [w_{L,1}, \dots, w_{L,d+1}, w_{R,d+2}]^T \\ \text{if } w_{L,2} - \sqrt{\kappa \frac{p_L}{\varrho_L}} < 0 \text{ and } w_{L,2} \geq 0 \quad (\textit{subsonic outlet}), \\ \\ [w_{R,1}, \dots, w_{L,d+1}, w_{L,d+2}]^T \\ \text{if } w_{L,2} + \sqrt{\kappa \frac{p_L}{\varrho_L}} \geq 0 \text{ and } w_{L,2} < 0 \quad (\textit{subsonic inlet}), \\ \\ \boldsymbol{w}_R \text{ if } w_{L,2} + \sqrt{\kappa \frac{p_L}{\varrho_L}} < 0 \quad (\textit{supersonic inlet}), \end{cases} \tag{1.17}$$

extrapolating energy e at subsonic inlet and extrapolating density ϱ and all components of momentum $\varrho \boldsymbol{v}$ at subsonic outlet.

1.2.2 System of Non-Linear Ordinary Differential Equations

Denote

$$\boldsymbol{y}^0 = \begin{bmatrix} \boldsymbol{y}_1^0 \\ \boldsymbol{y}_2^0 \\ \vdots \\ \boldsymbol{y}_N^0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_1} \\ \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_2} \\ \vdots \\ \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_N} \end{bmatrix} \in \mathcal{Q}^N \tag{1.18}$$

and

$$\boldsymbol{F}_{INT}(\boldsymbol{y}) = \begin{bmatrix} -\frac{1}{|\Omega_1|}\sum_{j\in S_{INT}(1)} H_{INT}(\boldsymbol{y}_1, \boldsymbol{y}_j, \nu_{1j})|\Gamma_{1j}| \\ -\frac{1}{|\Omega_2|}\sum_{j\in S_{INT}(2)} H_{INT}(\boldsymbol{y}_2, \boldsymbol{y}_j, \nu_{2j})|\Gamma_{2j}| \\ \vdots \\ -\frac{1}{|\Omega_N|}\sum_{j\in S_{INT}(N)} H_{INT}(\boldsymbol{y}_N, \boldsymbol{y}_j, \nu_{Nj})|\Gamma_{Nj}| \end{bmatrix}, \tag{1.19}$$

$$\boldsymbol{F}_{SOL}(\boldsymbol{y}) = \begin{bmatrix} -\frac{1}{|\Omega_1|}\sum_{\alpha\in S_{SOL}(1)} H_{SOL}(\boldsymbol{y}_1, \nu_{1\alpha})|\Gamma_{1\alpha}| \\ -\frac{1}{|\Omega_2|}\sum_{\alpha\in S_{SOL}(2)} H_{SOL}(\boldsymbol{y}_2, \nu_{2\alpha})|\Gamma_{2\alpha}| \\ \vdots \\ -\frac{1}{|\Omega_N|}\sum_{\alpha\in S_{SOL}(N)} H_{SOL}(\boldsymbol{y}_N, \nu_{N\alpha})|\Gamma_{N\alpha}| \end{bmatrix}, \tag{1.20}$$

$$\boldsymbol{F}_{I/O}(\boldsymbol{y},t) = \begin{bmatrix} -\frac{1}{|\Omega_1|}\sum_{\alpha\in S_{I/O}(1)} H_{I/O}(\boldsymbol{y}_1, \boldsymbol{w}_{1\alpha}, \nu_{1\alpha})|\Gamma_{1\alpha}| \\ -\frac{1}{|\Omega_2|}\sum_{\alpha\in S_{I/O}(2)} H_{I/O}(\boldsymbol{y}_2, \boldsymbol{w}_{2\alpha}, \nu_{2\alpha})|\Gamma_{2\alpha}| \\ \vdots \\ -\frac{1}{|\Omega_N|}\sum_{\alpha\in S_{I/O}(N)} H_{I/O}(\boldsymbol{y}_N, \boldsymbol{w}_{N\alpha}, \nu_{N\alpha})|\Gamma_{N\alpha}| \end{bmatrix}. \tag{1.21}$$

where $\boldsymbol{w}_{i\alpha}(t) = \boldsymbol{w}_h^B(\boldsymbol{x},\, t)|_{\Gamma_{i\alpha}}$ represents boundary conditions at the faces $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}$, $1 \le i \le N$. Our aim is to find coefficients $\boldsymbol{y} \in C^1(0,T;\mathcal{Q}^N)$ to the approximate solution

$$\boldsymbol{w}_h(\boldsymbol{x},\, t) = \sum_{i=1}^{N} \boldsymbol{y}_i(t)\chi_i(\boldsymbol{x}) \tag{1.22}$$

satisfying

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{F}(\boldsymbol{y}(t),t) \text{ for all } t \in (0,T), \tag{1.23}$$

$$\boldsymbol{y}(0) = \boldsymbol{y}^0. \tag{1.24}$$

The symbol $\chi_i(\boldsymbol{x})$, $1 \le i \le N$ denotes Ω_i-characteristic functions. Note that the right-hand-side vector function

$$\boldsymbol{F}(\boldsymbol{y}(t),t) = \boldsymbol{F}_{INT}(\boldsymbol{y}(t)) + \boldsymbol{F}_{SOL}(\boldsymbol{y}(t)) + \boldsymbol{F}_{I/O}(\boldsymbol{y}(t),t) \tag{1.25}$$

has the form corresponding to (1.11).

1.3 Solvability of the Semi–Discrete Problem

The existence and uniqueness of solution to the semi–discrete problem is a consequence of a basic theorem in the theory of ODE's (see e.g. [20], [45], [46]):

Theorem:

Let $\omega \subset \mathbb{R}^m$ *be an open set,* $T > 0$, $(\boldsymbol{y}_0, t_0) \in \omega \times (0, T)$. *Let us consider the initial-value problem*

$$\begin{aligned} \dot{\boldsymbol{y}}(t) &= \Phi(\boldsymbol{y}(t), t),\ t \in (0, T), \\ \boldsymbol{y}(t_0) &= \boldsymbol{y}_0,\ \boldsymbol{y}_0 \in \omega. \end{aligned}$$

If $\Phi \in C(\omega \times (0, T), \mathbb{R}^m)$ *then there exists at least one maximal solution* $\boldsymbol{y}$, $\boldsymbol{y}(t_0) = \boldsymbol{y}_0$. *In addition, if* Φ *satisfies the Lipschitz condition*

$$\forall \boldsymbol{y}_1 \in \omega\ \exists\, \delta > 0, K > 0\ |\boldsymbol{y}_2 - \boldsymbol{y}_1| < \delta \Rightarrow\ |\Phi(\boldsymbol{y}_2, t) - \Phi(\boldsymbol{y}_1, t)| \le K |\boldsymbol{y}_2 - \boldsymbol{y}_1| \tag{1.26}$$

for all $t \in (0, T)$ *then the maximal solution is unique.*

Interior component $\boldsymbol{F}_{INT}(\boldsymbol{y})$

In this contribution, we will analyze the well known approximate Riemann solvers $\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)$ of Steger-Warming (see [68], [28]), Vijayasundaram (see [77], [28]), Van Leer (see [28]) and Osher-Solomon (see [56], [67], [28]). Modern, modified schemes such as e.g. HLLE (see [22], [23]), AUSM (see [49]) or recently appeared AUSMD, AUSMV, AUSMDV (see [78]) inherit the essential properties implying the unique solvability of the semi–discrete problem and can be analyzed in the same way.

Approximate Riemann solvers of Steger-Warming, Vijayasundaram and Van Leer are continuous in $\mathcal{Q}^2$. Their Jacobi matrices

$$\frac{\mathrm{D}\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)}{\mathrm{D}\boldsymbol{q}_L}, \quad \frac{\mathrm{D}\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)}{\mathrm{D}\boldsymbol{q}_R} \tag{1.27}$$

are continuous in $\mathcal{Q}_0 \equiv \mathcal{Q}^2 \setminus (\mathcal{Q}_{v_1-c=0} \cup \mathcal{Q}_{v_1=0} \cup \mathcal{Q}_{v_1+c=0})^2$ where

$$\mathcal{Q}_{v_1-c=0} = \left\{ \boldsymbol{q} \in \mathcal{Q};\ q_2 - \sqrt{\frac{\kappa(\kappa-1)\left(q_{d+2} - \frac{\sum_{i=2}^{d-1} q_i^2}{2q_1}\right)}{q_1}} = 0 \right\}, \tag{1.28}$$

$$\mathcal{Q}_{v_1=0} = \{\boldsymbol{q} \in \mathcal{Q};\ q_2 = 0\}, \tag{1.29}$$

$$\mathcal{Q}_{v_1+c=0} = \left\{ \boldsymbol{q} \in \mathcal{Q};\ q_2 + \sqrt{\frac{\kappa(\kappa-1)\left(q_{d+2} - \frac{\sum_{i=2}^{d-1} q_i^2}{2q_1}\right)}{q_1}} = 0 \right\} \tag{1.30}$$

and bounded in each compact subset of $\mathcal{Q}_0$. Approximate Riemann solver by Osher-Solomon and its Jacobi matrices

$$\frac{\mathrm{D}\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)}{\mathrm{D}\boldsymbol{q}_L}, \quad \frac{\mathrm{D}\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)}{\mathrm{D}\boldsymbol{q}_R} \tag{1.31}$$

are continuous in $\mathcal{Q}^2$ because it is constructed by integration of a continuous vector function along a continuous curve in $\mathcal{Q}$. Definitions of mentioned approximate Riemann solvers can be also found e.g. in [28].

It means that the vector function $\boldsymbol{F}_{INT}(\boldsymbol{y})$ defined by (1.19) is for mentioned approximate Riemann solvers continuous and satisfies the Lipschitz condition (1.26) in $\omega \equiv \mathcal{Q}^N$.

Solid-wall component $\boldsymbol{F}_{SOL}(\boldsymbol{y})$

As the function $p(\boldsymbol{w})$ expressing the pressure corresponding to the state $\boldsymbol{w}$ is smooth in $\mathcal{Q}$, the vector function $\boldsymbol{F}_{SOL}(\boldsymbol{y})$ defined by (1.20) is smooth, i.e. it satisfies the Lipschitz condition (1.26) in $\omega \equiv \mathcal{Q}^N$.

Inlet/outlet component $\boldsymbol{F}_{I/O}(\boldsymbol{y}, t)$

Treatment of boundary conditions plays an important role in the solvability analysis of the semi–discrete problem because the switch (1.17) is not continuous. For each boundary face $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}(i)$, $1 \leq i \leq N$, boundary conditions are represented by the corresponding boundary state $\boldsymbol{w}_{i\alpha}(t) = \boldsymbol{w}_h^B(\boldsymbol{x}, t)|_{\Gamma_{i\alpha}}$ which is considered continuous in time.

In general, if we know nothing else about the function $\boldsymbol{w}_h(\boldsymbol{x}, t)$, only the following statement can be made: Consider the switch (1.17) and denote all inlet/outlet boundary faces $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}(i)$, $1 \leq i \leq N$ according to the initial data $\boldsymbol{w}_h^0(\boldsymbol{x})$ as either supersonic inlet, subsonic inlet, subsonic outlet or supersonic outlet, respectively. The system (1.23) (1.24) is uniquely solvable until computed flow through one of inlet/outlet boundary faces changes its type. At this time-instant, due to discontinuity of the switch (1.17), discontinuity occurs in the vector function $\boldsymbol{F}_{I/O}$ and thus, solvability of the system can no longer be guaranteed.

It is always possible to treat boundary conditions in such a way that the corresponding system of ordinary differential equations is uniquely solvable in the time-interval $(0, +\infty)$. One possibility is to replace the boundary numerical flux $H_{I/O}(\boldsymbol{w}_i, \boldsymbol{w}_{i\alpha}, \nu_{i\alpha})$ with the internal flux $H_{INT}(\boldsymbol{w}_i, \boldsymbol{w}_{i\alpha}, \nu_{i\alpha})$ but this approach may lead to non-physical solutions. For example – all characteristics of the solution are leaving the domain at supersonic outlet, but as the reader can easily verify, using this approach, in this case values from the prescribed boundary state $\boldsymbol{w}_{i\alpha}$ influence the solution.

In our opinion, a better approach is to monitor type of flow through the inlet/outlet boundary faces and to adjust boundary states $\boldsymbol{w}_{i\alpha}$ at all inlet/outlet boundary faces $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}(i)$ according to the outgoing characteristics of the approximate solution. It means in our case to keep $\boldsymbol{w}_{i\alpha} \equiv \boldsymbol{w}_i$ if $\boldsymbol{w}_i$ represents supersonic outlet, $\boldsymbol{w}_{i\alpha,1} \equiv \boldsymbol{w}_{i,1}$, $\boldsymbol{w}_{i\alpha,2} \equiv \boldsymbol{w}_{i,2}$, $\ldots$, $\boldsymbol{w}_{i\alpha,d+1} \equiv \boldsymbol{w}_{i,d+1}$ if $\boldsymbol{w}_i$ represents subsonic outlet etc. Then, the switch (1.17) does not cause discontinuity in the right-hand-side function $\boldsymbol{F}_{I/O}$. Using this approach, components of the boundary states $\boldsymbol{w}_{i\alpha}$ corresponding to outgoing characteristics of the initial solution are not prescribed. This approach will be used in all presented computations.

1.4 Numerical Solution of the Semi-Discrete Problem

For compressible Euler equations, there are not many opportunities to compare results of numerical schemes with analytical solutions. One of them, which we will use in this paper, is the so-called *Sod's channel* (see [66]).

1.4.1 Sod's Channel

Let us consider the 1D Euler equations with the initial data

$$
\begin{aligned}
v(x,0) &= 0,\\
\rho(x,0) &= \begin{cases} \rho_L, & x < x_0 \\ \rho_R, & x > x_0 \end{cases},\\
p(x,0) &= \begin{cases} p_L, & x < x_0 \\ p_R, & x > x_0 \end{cases},
\end{aligned}
\tag{1.32}
$$

with $p_L > p_R$. The exact analytical solution to this 1D Riemann problem can be expressed in the following way:

$$
p(x,t) = \begin{cases}
p_L \quad , & x < x_0 - c_L t \\
p_L \left[1 - \frac{\kappa-1}{\kappa+1}\left(1 + \frac{x-x_0}{c_L t}\right)\right]^{\frac{2\kappa}{\kappa-1}}, & x_0 - c_L t < x < x_1 \\
p_2, & x_1 < x < x_o + v_c t \\
p_2, & x_0 + v_c t < x < x_0 + v_s t \\
p_R, & x > x_0 + v_s t
\end{cases},
\tag{1.33}
$$

$$
v(x,t) = \begin{cases}
0, & x < x_0 - c_L t \\
\frac{2}{\kappa+1}\left(\frac{x-x_0}{t} + c_L\right), & x_0 - c_L t < x < x_1 \\
v_c, & x_1 < x < x_o + v_c t \\
v_c, & x_0 + v_c t < x < x_0 + v_s t \\
0, & x > x_0 + v_s t
\end{cases},
\tag{1.34}
$$

$$
\rho(x,t) = \begin{cases}
\rho_L, & x < x_0 - c_L t \\
\rho_L \left[1 - \frac{\kappa-1}{\kappa+1}\left(1 + \frac{x-x_0}{c_L t}\right)\right]^{\frac{2}{\kappa-1}}, & x_0 - c_L t < x < x_1 \\
\rho_L \left(\frac{p_2}{p_L}\right)^{\frac{1}{\kappa}}, & x_1 < x < x_o + v_c t \\
\rho_R \frac{1 + \frac{\kappa+1}{\kappa-1}\frac{p_2}{p_R}}{\frac{\kappa+1}{\kappa-1} + \frac{p_2}{p_R}}, & x_0 + v_c t < x < x_0 + v_s t \\
\rho_R, & x > x_0 + v_s t
\end{cases}.
\tag{1.35}
$$

Here

$$
x_1 = x_0 + \left(\frac{\kappa+1}{2} v_c - c_L\right) t,
\tag{1.36}
$$

for the speed of the shock wave v_s we have

$$
v_s = c_R^2 \frac{\frac{p_2}{p_R} - 1}{\kappa v_c}
\tag{1.37}
$$

and for the speed of contact discontinuity it holds

$$v_c = c_R \sqrt{\frac{2}{\kappa(\kappa-1)}} \frac{\frac{p_2}{p_R} - 1}{\sqrt{1 + \frac{\kappa+1}{\kappa-1}\frac{p_2}{p_R}}}. \tag{1.38}$$

Moreover the pressure p_2 has to be determined. For p_2 one can find the following equation

$$\sqrt{\frac{2}{\kappa(\kappa-1)}} \frac{\frac{p_2}{p_R} - 1}{\sqrt{1 + \frac{\kappa+1}{\kappa-1}\frac{p_2}{p_R}}} = \frac{2}{\kappa-1}\frac{c_L}{c_R}\left[1 - \left(\frac{p_2}{p_L}\right)^{\frac{\kappa-1}{2\kappa}}\right], \tag{1.39}$$

which can be solved numerically. The speeds of sound c_L and c_R are obtained from the initial data.

Let $x_o = 0$, $x_L = -3$ and $x_R = 3$. The interval (x_L, x_R) is covered with an equidistant mesh consisting of $N = 300$ finite volumes. The initial data are chosen as follows:

$$\boldsymbol{w}_L = \begin{bmatrix} 1.0 \\ 0.0 \\ 2.0/(\kappa-1) \end{bmatrix}, \quad \boldsymbol{w}_R = \begin{bmatrix} 1.101463 \\ 0.0 \\ 1.0/(\kappa-1) \end{bmatrix}, \tag{1.40}$$

$$\boldsymbol{w}_h^0(x) = \begin{cases} \boldsymbol{w}_L & x_L < x < x_0, \\ \boldsymbol{w}_R & x_0 < x < x_R. \end{cases} \tag{1.41}$$

With the initial vector $\boldsymbol{y}^0 \in \mathcal{Q}^N$ corresponding to (1.41) both inlet/outlet edges are by (1.17) denoted as subsonic outlet. Thus, we prescribe

$$\boldsymbol{w}_{h,3}^B(x_L, t) = \boldsymbol{w}_{L,3}, \quad \boldsymbol{w}_{h,3}^B(x_R, t) = \boldsymbol{w}_{R,3}. \tag{1.42}$$

For the construction of the right-hand-side function $\boldsymbol{F}_{INT}$ we have chosen the 1D approximate Riemann solver by Osher-Solomon.

Figure 1.1 shows density distribution $\varrho(x, t_1)$ of solutions of the continuous and semi–discrete problem corresponding to the time $t_1 = 1$ (to be absolutely correct, the "exact solution to the semi-discrete problem" is a highly precise numerical solution obtained by an ODE-solver which will be introduced later, identified with the analytic one for the purpose of the figure). Difference between the exact solution of the continuous problem and the exact solution of the semi-discrete problem is caused by Finite Volume space approximations. Integrating the system (1.23) (1.24) in time, we will further approximate the exact solution of the semi-discrete problem.

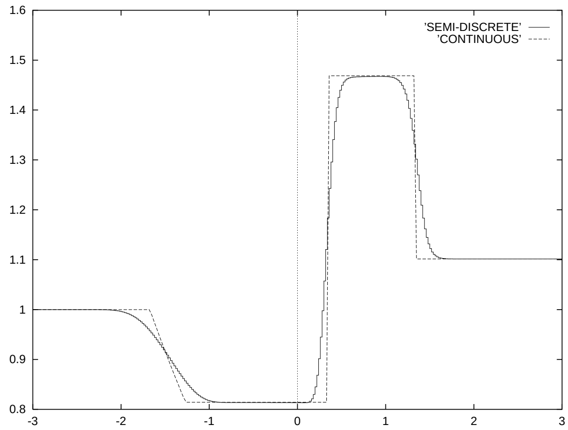

Figure 1.1: Solutions of the continuous and semi–discrete problem at $t = 1$s

1.4.2 Disadvantages of Explicit Time-Integration

In the last years a lot of work was done towards the construction of efficiently-evaluable and robust approximate Riemann solvers (e.g. [67], [21], [47], [68], [56], [28], [74], [22], [23], [49], [78]). Using these approximate Riemann solvers, the right-hand-side vector function $\boldsymbol{F}$ can be computed explicitly and this is the reason why explicit schemes for the numerical integration of (1.23) (1.24) are still popular in spite of their considerable disadvantages.

One of these disadvantages is instability unless the time-step is very small, controlled by the so-called *CFL condition.* The CFL condition contains the so-called *CFL constant*, a positive real number, which must be chosen in such a way that the computational process is stable.

In Figure 1.2 approximate solutions of the semi–discrete problem from the previous subsection at $t = 1$ are shown, computed by the explicit Euler method. The time-step was controlled using the CFL constant equal to 0.6 and 0.05, respectively. The reader notices the considerable difference between the above two approximate solutions.

These reasons motivated our effort to find a scheme which would be more suitable for the numerical solution of the system (1.23) (1.24).

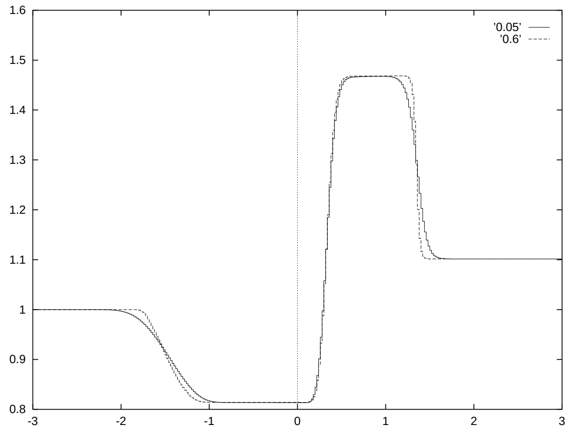

Figure 1.2: Difference of two approximate solutions computed with different CFL constants

1.4.3 Backward Differentiation Formula

Problem (1.23) (1.24), as numerical experiments indicate, is stiff. It means that it contains the so-called *strongly damped modes*, more exactly there are eigenvalues of the Jacobi matrix

$$\boldsymbol{J}(\boldsymbol{y}) = \frac{\mathrm{D}F}{\mathrm{D}\boldsymbol{y}} \tag{1.43}$$

whose real parts are negative and large in magnitude. It is known that explicit schemes are not suitable for the numerical solution of stiff problems. In the numerical analysis of ODE's a special treatment of stiff problems based on the so-called *Backward Differentiation Formula (BDF)* is widely used. There are two important properties of this scheme: First, it is absolutely stable. Second, the global error is essentially just the sum of local, truncation errors. For a detail description of BDF including the stability and error analysis see e.g. [64].

For the numerical solution of the semi–discrete problem we use an implementation of the BDF found in the Fortran programming package ODEPACK (see [39], [13] and [14]) based on explicit evaluation of the right-hand-side vector function $\boldsymbol{F}$. This implementation provides local absolute and relative error control. Efficiency of the computation can be considerably improved by supplying the structure of the sparse Jacobi matrix $J(\boldsymbol{y})$ which is as follows:

$$(1.44)\ J_{(d+2)(k-1)+k_1,(d+2)(l-1)+l_1} \begin{cases} \neq 0 & \text{in general for } k \text{ and } l \in S_{INT}(k) \\ \\ = 0 & \text{otherwise} \end{cases}$$

where $1 \leq k, l \leq N$ and $1 \leq k_1, l_1 \leq d+2$. It is also possible to supply a function which evaluates the Jacobi matrix explicitly – otherwise the Jacobi matrix is approximated numerically.

1.5 Numerical Examples

The right-hand-side vector function $F(\boldsymbol{y}, t)$ is evaluated explicitly by an original C++ code based on [28], $d = 1, 2$ or 3. Time-integration is performed using the Fortran library ODEPACK. Meshes introduced in this section were generated automatically by an original C++ grid generator. Meshes, 1D results and convergence histograms were visualized by Gnuplot, 2D results were visualized by the C programming tool GRAPE developed at the Friedrich-Wilhelm-University of Bonn. All computations were performed on a DEC Alpha workstation with a RISC 21164/550MHz processor in double precision.

1.5.1 Transonic Flow in the 2D GAMM Channel

Consider the 2D Euler equations. The polygonal domain Ω is placed between $x_L = -1$ and $x_R = 1$. It is 2 meters long, 1 meter high, with a 10% circular bump in the middle of its bottom as shown in Figure 1.3. The upper side and the lower one including the bump represent solid walls, sides corresponding to x_L and x_R represent inlet and outlet. The unstructured grid consists of $N = 11048$ triangles.
Initial data are chosen as a constant state

$$\boldsymbol{w}_h^0(\boldsymbol{x}) = \boldsymbol{w}^0 \tag{1.45}$$

with values $\varrho^0 = 1.5\text{kg/m}^3$, $\boldsymbol{v}^0 = [205.709227,\ 0]^T\text{m/s}$ and $p^0 = 101000\text{Pa}$.

According to the initial data (1.45), boundary edges corresponding to x_L and x_R are denoted as subsonic inlet and subsonic outlet, respectively. Inlet and outlet boundary conditions are prescribed as

$$\boldsymbol{w}_{h,1}^B(x_L, t) = \boldsymbol{w}_1^0,\ \ \boldsymbol{w}_{h,2}^B(x_L, t) = \boldsymbol{w}_2^0,\ \ \boldsymbol{w}_{h,3}^B(x_L, t) = \boldsymbol{w}_3^0 \text{ and } \boldsymbol{w}_{h,4}^B(x_R, t) = \boldsymbol{w}_4^0.$$

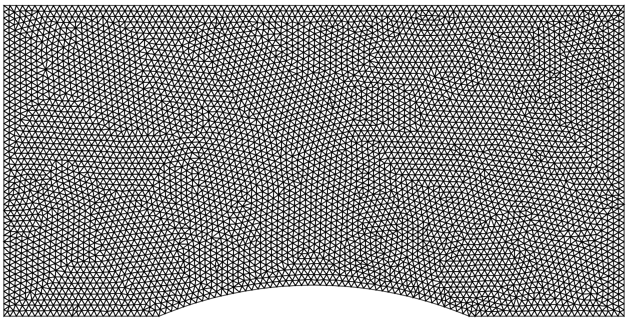

Figure 1.3: Unstructured triangular grid on the GAMM channel

For the construction of the right-hand-side function $\boldsymbol{F}_{INT}$ the approximate Riemann solver by Osher-Solomon is used. For the sake of comparison, two computations are performed – by the BDF and by the explicit first order Runge-Kutta (RK1) method with $CFL = 0.8$. Convergence to the steady state is measured by the discrete values

$$\log \|\frac{\partial}{\partial t}\boldsymbol{w}_h^\tau\|_{L^\infty(\Omega)} \quad \text{and} \quad \log \|\frac{\partial}{\partial t}\boldsymbol{w}_h^\tau\|_{L^1(\Omega)}, \tag{1.46}$$

where $\boldsymbol{w}_h^\tau$ denotes the computed piecewise linear approximation of the solution of the system of ordinary differential equations.

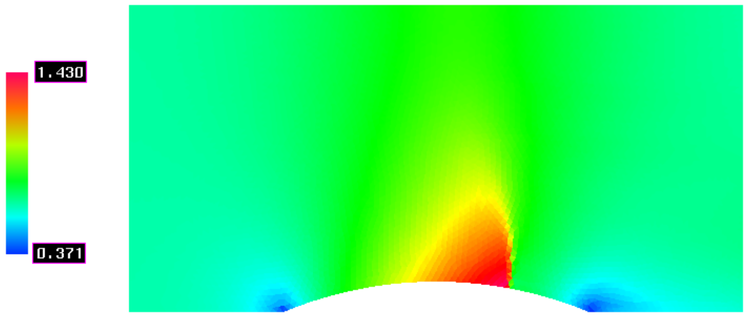

Figure 1.4: Steady state Mach number distribution

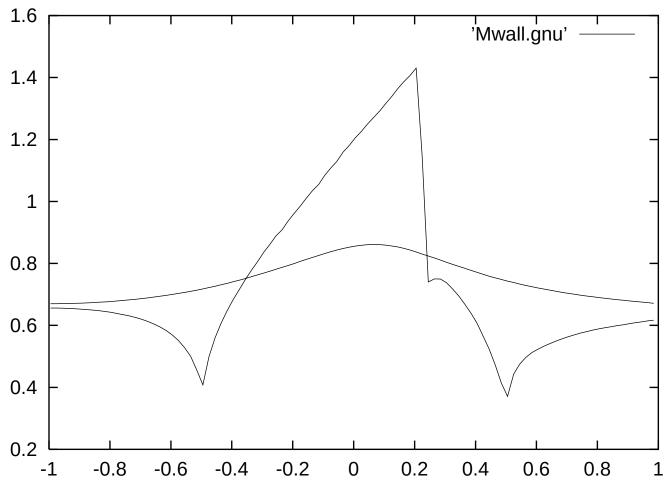

Figure 1.5: Mach number on the solid walls

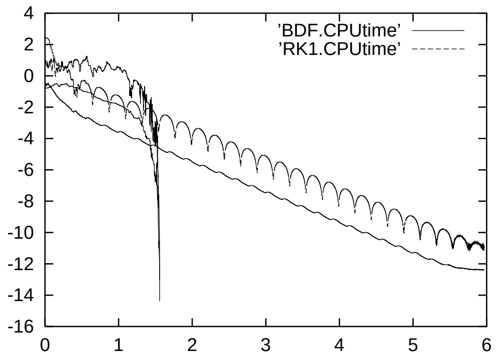

Figure 1.6: Convergence to the steady state, x-axis: CPU hours,

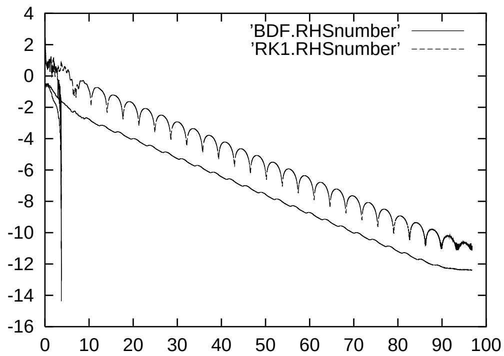

Figure 1.7: Convergence to the steady state, x-axis: (number of calls to $F(\boldsymbol{y}, t))/1000$

There are considerable differences between the computation performed by the explicit Euler scheme and by the BDF. One can see that BDF needs several times less calls to the right-hand-side function $\boldsymbol{F}$ and that it needs several times less CPU time to reach the steady state. Moreover, due to the adaptivity and higher order of accuracy of the mentioned implementation of the BDF, results obtained using this time integrator are more reliable.

1.5.2 Subsonic Flow in a 3D Curved Channel

In this subsection, we will deal with the numerical solution of subsonic flow in a 3D domain Ω described by

$$(1.47)\ \Omega = \left\{(x_1, x_2, x_3) = \left(\alpha_1, \frac{1}{4}\sin\alpha_1\pi + \alpha_2, \frac{1}{4}\cos\alpha_1\pi + \alpha_3\right) \in \mathbb{R}^3\right\}$$

where $0 < \alpha_1 < 2,\ 0 < \alpha_2, \alpha_3 < \frac{1}{2}$, shown in Figure 1.8. Boundary faces corresponding to $x_L = 0$ and $x_R = 2$ represent inlet and outlet and remaining boundary faces represent solid wall.

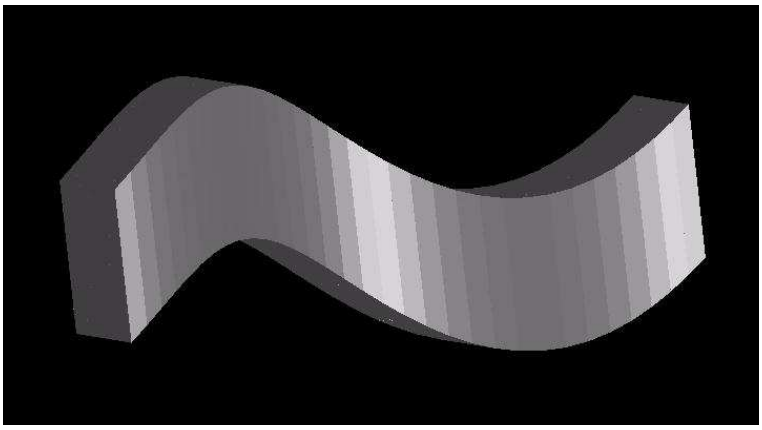

Figure 1.8: 3D curved channel

Initial data are chosen for simplicity constant

$$\boldsymbol{w}_h^0(\boldsymbol{x}) = \boldsymbol{w}^0 \tag{1.48}$$

with values $\varrho^0 = 1.5\mathrm{kg/m^3}$, $\boldsymbol{v}^0 = [4,\ 3.14,\ 0]^T\mathrm{m/s}$ and $p^0 = 101000\mathrm{Pa}$. According to initial data, boundary faces corresponding to $x_L = 0$ and $x_R = 2$ are denoted as subsonic inlet and subsonic outlet, respectively. Inlet and outlet boundary conditions are prescribed as

$$\boldsymbol{w}_{h,1}^B(x_L, t) = \boldsymbol{w}_1^0,\ \boldsymbol{w}_{h,2}^B(x_L, t) = \boldsymbol{w}_2^0,\ \boldsymbol{w}_{h,3}^B(x_L, t) = \boldsymbol{w}_3^0 \text{ and } \boldsymbol{w}_{h,4}^B(x_R, t) = \boldsymbol{w}_4^0.$$

The domain is covered with a structured tetrahedral grid consisting of $N = 12288$ finite volumes.

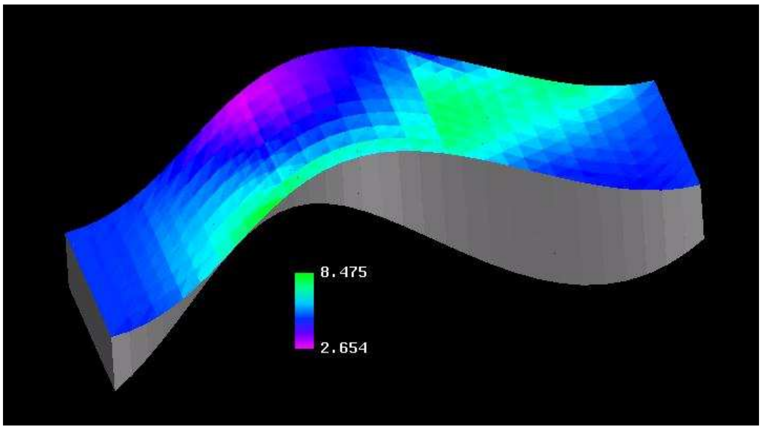

Figure 1.9: Velocity distribution, xy–cut, $z = 0.25$

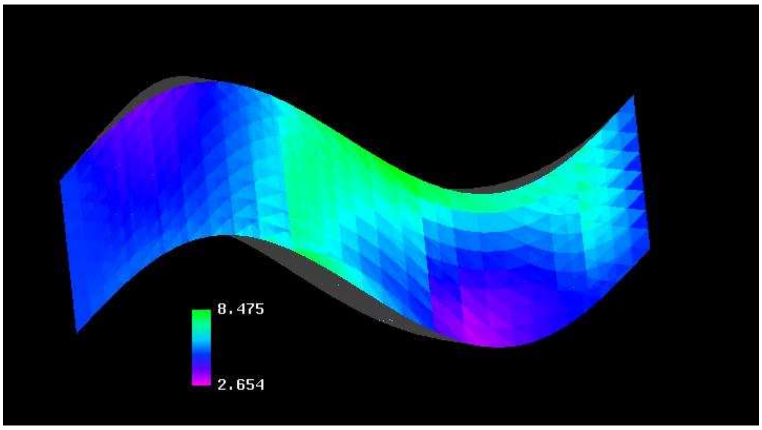

Figure 1.10: Velocity distribution, xz–cut, $y = 0$

For the construction of the right-hand-side function $\boldsymbol{F}_{INT}$ the approximate Riemann solver by Osher-Solomon is used. In Figures 1.9 and 1.10 we show the approximate velocity distribution corresponding to the steady state obtained with a high prescribed precision by the BDF. In the last two figures we once more document effectivity of BDF in comparison with the explicit Euler method.

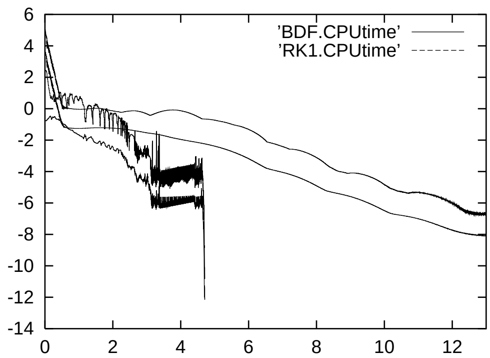

Figure 1.11: Convergence to the steady state, x-axis: CPU hours,

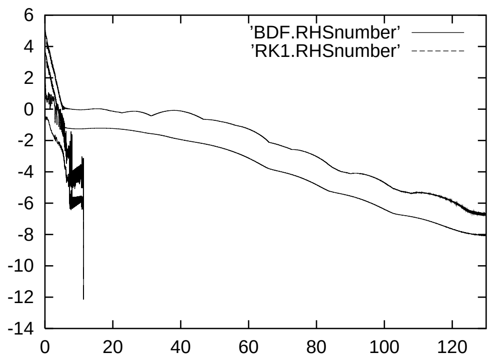

Figure 1.12: Convergence to the steady state, x-axis: (number of calls to $F(\boldsymbol{y}, t))/1000$

Chapter 2

On a Model of Polluted Gas Flow

Abstract: This part of the thesis is devoted to the numerical simulation of motion of a compressible inviscid fluid polluted by small particles of mass. We assume that the particles are so small that they do not influence the fluid flow. The compressible Euler equations are semi–discretized in space by the Finite Volume Method and transformed into a system of first-order ordinary differential equations. This system is coupled with a system of second-order ordinary differential equations describing the motion of particles within the fluid. Resulting system of ordinary differential equations is solved by suitable implicit and explicit solvers for systems of ordinary differential equations. Two and three-dimensional numerical examples are presented.

2.1 Position of the Problem

We consider flow in a space-time cylinder $Q_T = \Omega \times (0,T)$ where $\Omega \subset \mathbb{R}^d$, $d = 2,3$ is a bounded polygonal domain representing the region occupied by the fluid and $T > 0$.

Let us consider a finite number P of spherical mass particles $X_1, X_2, \ldots, X_P$ with initial positions $\boldsymbol{x}_1^0, \boldsymbol{x}_2^0, \ldots, \boldsymbol{x}_P^0 \in \Omega$ and initial velocities $\boldsymbol{u}_1^0, \boldsymbol{u}_2^0, \ldots, \boldsymbol{u}_P^0 \in \mathbb{R}^d$. Let us by r_k, m_k denote radius and mass of particle X_k, respectively. We assume that the particles are so small that we can neglect their influence on the fluid flow. Motion of the fluid is governed by compressible Euler equations introduced in section **2.2**. Frictional forces between the fluid and the particles, which are the source of their motion, are described in section **2.3**.

2.2 Equations Describing Motion of Fluid

The compressible Euler equations consist of the continuity equation, d momentum equations and of the energy equation. We will consider this system in the form

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{s=1}^{d} \frac{\partial \boldsymbol{f}_s(\boldsymbol{w})}{\partial x_s} = 0, \tag{2.1}$$

where

$$\boldsymbol{w} = \begin{pmatrix} \rho \\ \rho v_1 \\ \vdots \\ \rho v_d \\ e \end{pmatrix}, \quad \boldsymbol{f}_s(\boldsymbol{w}) = \begin{pmatrix} \rho v_s \\ \rho v_s v_1 + \delta_{s1} p \\ \vdots \\ \rho v_s v_d + \delta_{sd} p \\ (e+p)\, v_s \end{pmatrix}, \quad s = 1, \ldots, d \tag{2.2}$$

and

$$e = \frac{p}{\kappa - 1} + \frac{1}{2}\rho |\boldsymbol{v}|^2. \tag{2.3}$$

We use the standard notation: t – time, $x_1, \ldots, x_d$ – Cartesian coordinates in $\mathbb{R}^d$, $\boldsymbol{x} = (x_1, \ldots, x_d)^T$, ρ – density, $\boldsymbol{v} = (v_1, \ldots, v_d)^T$ – velocity vector with components $v_1, \ldots, v_d$ in the directions $x_1, \ldots, x_d$, p – pressure, e – total energy, δ_{sj} – Kronecker delta, $\kappa > 1$ – Poisson adiabatic constant. The gravity acceleration vector will be denoted $\boldsymbol{g}$, $\boldsymbol{g} \in \mathbb{R}^d$. Due to physical reasons it is also suitable to require $\rho > 0, p > 0$. The functions $\boldsymbol{f}_s, s = 1, \ldots, d$ are called the inviscid (Euler) fluxes and are defined in the set

$$(2.4)\quad \mathcal{Q} = \{(w_1, \ldots, w_{d+2}) \in \mathbb{R}^{d+2};\ w_1 > 0, w_{d+2} - \frac{w_2^2 + \ldots + w_{d+1}^2}{2w_1} > 0\}.$$

System (2.1), (2.3) is equipped with the initial conditions

$$(2.5)\qquad \boldsymbol{w}(\boldsymbol{x}, 0) = \boldsymbol{w}^0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega$$

and a set of boundary conditions which will be specified later.

2.3 Equations Describing Motion of Pollution

In general, it is difficult to find a reliable relation describing force by which the fluid acts on the particles. We introduce here two approximations found in [12] depending on the type of flow.

The so-called *Stokes approximation*

$$(2.6)\qquad \boldsymbol{f}_S(r, \varrho, \boldsymbol{v}_{rel}) = 6\pi\nu r\varrho\boldsymbol{v}_{rel}$$

where ϱ is density of the fluid, ν kinematic viscosity of the fluid, r radius of the particle and $\boldsymbol{v}_{rel}$ velocity of the particle relative to the velocity of the fluid, is suitable for laminar viscous flows polluted with very small particles.

The so-called *Newton approximation*

$$(2.7)\qquad \boldsymbol{f}_N(S, \varrho, \boldsymbol{v}_{rel}) = \frac{C}{2}\varrho S\boldsymbol{v}_{rel}^2$$

where C is a constant depending on the shape of particles (see [12] for examples) and S is the active profile size of the particle (e.g. πr^2 for spherical particles), works better for higher Reynolds number flows or when the particles are larger.

In this part of the thesis, motion of mass particles X_k, $k = 1, 2, \ldots, P$ is modeled using the system of second-order ordinary differential equations

$$(2.8)\qquad m_k\ddot{\boldsymbol{x}}_k(t) = \boldsymbol{f}_S\left(r_k, \varrho(\boldsymbol{x}_k(t), t), \dot{\boldsymbol{x}}_k(t) - \boldsymbol{v}(\boldsymbol{x}_k(t), t)\right) + \boldsymbol{g}$$

$k = 1, 2, \ldots, P$ with initial conditions

$$\dot{\boldsymbol{x}}_k(0) = \boldsymbol{u}_k^0, \tag{2.9}$$

$$\boldsymbol{x}_k(0) = \boldsymbol{x}_k^0, \tag{2.10}$$

$k = 1, 2, \ldots, P$, where $\boldsymbol{x}_k(t)$ denotes position of the particle X_k and $\varrho(\boldsymbol{x}_k(t), t)$, $\boldsymbol{v}(\boldsymbol{x}_k(t), t)$ denote density and velocity of the fluid interacting with the particle, respectively.

Applying substitution

$$\boldsymbol{u}_k(t) = \dot{\boldsymbol{x}}_k(t), \tag{2.11}$$

where $\boldsymbol{u}_k(t)$ denotes velocity of the mass particle X_k, we transform the second-order system (2.8) into a system of first-order ordinary differential equations

$$m_k \dot{\boldsymbol{u}}_k(t) = \boldsymbol{f}_S(r_k, \varrho(\boldsymbol{x}_k(t), t), \boldsymbol{u}_k(t) - \boldsymbol{v}(\boldsymbol{x}_k(t), t)) + \boldsymbol{g}, \tag{2.12}$$

$$\dot{\boldsymbol{x}}_k(t) = \boldsymbol{u}_k(t), \tag{2.13}$$

$k = 1, 2, \ldots, P$, with initial conditions

$$\boldsymbol{u}_k(0) = \boldsymbol{u}_k^0, \tag{2.14}$$

$$\boldsymbol{x}_k(0) = \boldsymbol{x}_k^0, \tag{2.15}$$

$k = 1, 2, \ldots, P$.

2.4 Approximation of Motion of Fluid

2.4.1 Finite Volume Method

Let us consider a finite volume grid $\mathcal{T}_h$ consisting of grid points

$$\mathcal{P} = \{P_1, P_2, \ldots, P_M\} \subset \overline{\Omega} \tag{2.16}$$

and finite volumes

$$\mathcal{T}_h = \{\Omega_1, \Omega_2, \ldots, \Omega_N\} \subset \overline{\Omega}, \tag{2.17}$$

having standard properties

(i) $\Omega = \bigcup_{i=1}^{N} T_i$,

(ii) if $T_i, T_j \in \mathcal{T}_h$, $i \neq j$ then exactly one of the following options is valid:
 (a) $T_i \cup T_j = \emptyset$,
 (b) T_i and T_j have exactly one common vertex,
 (c) T_i and T_j have exactly one common edge,
 (c) T_i and T_j have exactly one common face (only in 3D).

We use triangular finite volumes in $2D$ and tetrahedral finite volumes in $3D$. Integrating equation (2.1) over a finite volume Ω_i and using Green's theorem and the so-called *rotational invariance* of the Euler equations (for detail description see e.g. [28]), we obtain

$$\begin{aligned} \int_{\Omega_i} \frac{\partial \boldsymbol{w}}{\partial t}(\boldsymbol{x},t)\ \mathrm{d}\boldsymbol{x} = & - \sum_{j \in S_{INT}(i)} \int_{\Gamma_{ij}} \mathbb{T}_d^{-1}(\nu_{ij}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{ij}) \boldsymbol{w}(\boldsymbol{x},t))\ \mathrm{d}S \\ & - \sum_{\alpha \in S_{SOL}(i)} \int_{\Gamma_{i\alpha}} \mathbb{T}_d^{-1}(\nu_{i\alpha}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{i\alpha}) \boldsymbol{w}(\boldsymbol{x},t))\ \mathrm{d}S \\ & - \sum_{\alpha \in S_{I/O}(i)} \int_{\Gamma_{i\alpha}} \mathbb{T}_d^{-1}(\nu_{i\alpha}) \boldsymbol{f}_1(\mathbb{T}_d(\nu_{i\alpha}) \boldsymbol{w}(\boldsymbol{x},t))\ \mathrm{d}S \end{aligned} \tag{2.18}$$

for all $t \in (0,T)$ and $1 \leq i \leq N$. The index set $S_{INT}(i)$ contains indices of finite volumes adjacent to the finite volume Ω_i for all $1 \leq i \leq N$. Common edge (face) of adjacent finite volumes Ω_i and Ω_j is denoted Γ_{ij}. Analogously, we define the index sets $S_{SOL}(i)$ and $S_{I/O}(i)$: by $\Gamma_{i\alpha}$, $\alpha \in S_{SOL}(i)$ and $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}(i)$ are denoted edges (faces) of the finite volume Ω_i representing solid wall and inlet/outlet, respectively. Symbols $\nu_{ij}, \nu_{i\alpha} \in \mathbb{R}^d$ denote unit normal vectors to $\Gamma_{ij}, \Gamma_{i\alpha}$, respectively, pointing outward of Ω_i. The rotational matrices $\mathbb{T}_d(\nu)$, $\nu \in \mathbb{R}^d$, $|\nu| = 1$, $d = 1, 2, 3$ are defined as follows:

In 1D

$$\mathbb{T}_1(\nu) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.19}$$

In 2D, expressing ν in polar coordinates as $\nu = [\cos(\alpha), \sin(\alpha)]^{\mathrm{T}}$, $\alpha \in \langle -\pi, \pi)$, we define

$$\mathbb{T}_2(\nu) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) & 0 \\ 0 & -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.20}$$

In 3D, expressing ν in spherical coordinates as

$$\nu = [\cos(\alpha)\cos(\beta), \sin(\alpha)\cos(\beta), \sin(\beta)]^{\mathrm{T}}, \alpha \in \langle -\pi, \pi), \beta \in \langle -\frac{\pi}{2}, \frac{\pi}{2}), \tag{2.21}$$

the matrix $\mathbb{T}_3(\nu)$ has the form

$$\mathbb{T}_3(\nu) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos(\alpha)\cos(\beta) & \sin(\alpha)\cos(\beta) & \sin(\beta) & 0 \\ 0 & -\sin(\alpha) & \cos(\alpha) & 0 & 0 \\ 0 & -\cos(\alpha)\sin(\beta) & -\sin(\alpha)\sin(\beta) & \cos(\beta) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.22}$$

Replacing in (2.18) the function $\boldsymbol{w}(\boldsymbol{x}, t)$ with the element average function $\boldsymbol{w}_h(\boldsymbol{x}, t)$ which is piecewise constant in space on the finite volumes $\Omega_1, \Omega_2, \ldots, \Omega_N$ and approximating the right-hand-side integrals by integrals of the so-called *numerical fluxes* H_{INT}, H_{SOL} and $H_{I/O}$, we obtain

$$\begin{aligned} |\Omega_i| \frac{\partial \boldsymbol{w}_h}{\partial t}(\boldsymbol{x},\, t)|_{\Omega_i} = & - \sum_{j \in S_{INT}(i)} H_{INT}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_j}, \nu_{ij}(\boldsymbol{x}))|\Gamma_{ij}| \\ & - \sum_{\alpha \in S_{SOL}(i)} H_{SOL}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \nu_{i\alpha}(\boldsymbol{x}))|\Gamma_{i\alpha}| \\ & - \sum_{\alpha \in S_{I/O}(i)} H_{I/O}(\boldsymbol{w}_h(\boldsymbol{x},\, t)|_{\Omega_i}, \boldsymbol{w}_h^B(\boldsymbol{x},\, t)|_{\Gamma_{i\alpha}}, \nu_{i\alpha}(\boldsymbol{x}))|\Gamma_{i\alpha}|, \end{aligned} \tag{2.23}$$

$t \in (0, T)$, $1 \le i \le N$, where $\nu_{i\alpha}$ denotes the unit normal vector to $\Gamma_{i\alpha}$ pointing outward of Ω. Symbols $|\Omega_i|$ and $|\Gamma_{i\alpha}|$ denote d-dimensional measure of $|\Omega_i|$ and $(d-1)$-dimensional measure of $\Gamma_{i\alpha}$, respectively. The function $\boldsymbol{w}_h^B$ represents a boundary state corresponding to the inlet/outlet boundary face $\Gamma_{i\alpha}$ which will be discussed later.

2.4.2 Numerical Flux

The numerical flux H_{INT} through interior faces Γ_{ij} has the form

$$H_{INT}(\boldsymbol{w}_i, \boldsymbol{w}_j, \nu_{ij}) = \mathbb{T}_d^{-1}(\nu_{ij}) \boldsymbol{f}_R(\mathbb{T}_d(\nu_{ij})\boldsymbol{w}_i, \mathbb{T}_d(\nu_{ij})\boldsymbol{w}_j), \tag{2.24}$$

where the so-called *approximate Riemann solver* $\boldsymbol{f}_R(\boldsymbol{q}_L, \boldsymbol{q}_R)$ approximates the constant value $\boldsymbol{f}_1(\boldsymbol{q}_r(0,\tilde{t}))$, $\boldsymbol{q}_r(\tilde{x}_1,\tilde{t})$ being the exact solution to the Riemann problem

$$\frac{\partial \boldsymbol{q}}{\partial \tilde{t}}(\tilde{x}_1,\tilde{t}) + \frac{\partial \boldsymbol{f}_1(\boldsymbol{q})}{\partial \tilde{x}_1}(\tilde{x}_1,\tilde{t}) = 0 \qquad \text{in } \mathbb{R}\times(0,+\infty), \tag{2.25}$$

$$\boldsymbol{q}(\tilde{x}_1, 0) = \begin{cases} \boldsymbol{q}_L, & \tilde{x}_1 < 0, \\ \boldsymbol{q}_R, & \tilde{x}_1 > 0, \end{cases} \tag{2.26}$$

as described in detail e.g. in [28].

The numerical flux H_{SOL} through the solid-wall faces $\Gamma_{i\alpha}$ has the form

$$H_{SOL}(\boldsymbol{w}_i, \nu_{i\alpha}) = [0, p(\boldsymbol{w}_i), 0, \dots, 0]^T \tag{2.27}$$

where $p(\boldsymbol{w}_i)$ denotes the pressure corresponding to the state $\boldsymbol{w}_i$. In this approximation, the impermeability condition $\boldsymbol{v}\cdot\nu_{i\alpha} = 0$ for solid walls was used, see e.g. [27] for details. Finally, the numerical flux $H_{I/O}$ through the inlet/outlet faces $\Gamma_{i\alpha}$ has the form

$$H_{I/O}(\boldsymbol{w}_i, \boldsymbol{w}_{i\alpha}, \nu_{i\alpha}) = \mathbb{T}_d^{-1}(\nu_{i\alpha})\boldsymbol{f}_1(\boldsymbol{w}^B(\mathbb{T}_d(\nu_{i\alpha})\boldsymbol{w}_i, \mathbb{T}_d(\nu_{i\alpha})\boldsymbol{w}_{i\alpha})) \tag{2.28}$$

where $\boldsymbol{w}_{i\alpha}$ represents the boundary state corresponding to the face $\Gamma_{i\alpha}$. The function $\boldsymbol{w}^B$ represent a switch deciding how much information is transported from the boundary into the domain, depending on the number of outgoing characteristics of the approximate solution $\boldsymbol{w}_h$. Switches of this type are very characteristic for hyperbolic systems. For our purpose, let us consider a function

$$\boldsymbol{w}^B(\boldsymbol{w}_L, \boldsymbol{w}_R) = \begin{cases} \boldsymbol{w}_L \text{ if } w_{L,2} - \sqrt{\kappa\frac{p_L}{\varrho_L}} \ge 0 \quad (\textit{supersonic outlet}), \\[2ex] [w_{L,1}, \dots, w_{L,d+1}, w_{R,d+2}]^T \\ \text{if } w_{L,2} - \sqrt{\kappa\frac{p_L}{\varrho_L}} < 0 \text{ and } w_{L,2} \ge 0 \quad (\textit{subsonic outlet}), \\[2ex] [w_{R,1}, \dots, w_{L,d+1}, w_{L,d+2}]^T \\ \text{if } w_{L,2} + \sqrt{\kappa\frac{p_L}{\varrho_L}} \ge 0 \text{ and } w_{L,2} < 0 \quad (\textit{subsonic inlet}), \\[2ex] \boldsymbol{w}_R \text{ if } w_{L,2} + \sqrt{\kappa\frac{p_L}{\varrho_L}} < 0 \quad (\textit{supersonic inlet}), \end{cases} \tag{2.29}$$

extrapolating energy e at subsonic inlet and extrapolating density ϱ and all components of momentum $\varrho\boldsymbol{v}$ at subsonic outlet. For more details see [28].

2.4.3 System of Non-Linear Ordinary Differential Equations

Denote

$$\boldsymbol{W}^0 = \begin{bmatrix} \boldsymbol{w}_1^0 \\ \boldsymbol{w}_2^0 \\ \vdots \\ \boldsymbol{w}_N^0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_1} \\ \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_2} \\ \vdots \\ \boldsymbol{w}_h(\boldsymbol{x},0)|_{\Omega_N} \end{bmatrix} \in \mathcal{Q}^N \tag{2.30}$$

and

$$\boldsymbol{F}_{INT}(\boldsymbol{W}) = \begin{bmatrix} -\sum_{j\in S_{INT}(1)} H_{INT}(\boldsymbol{w}_1, \boldsymbol{w}_j, \nu_{1j})|\Gamma_{1j}| \\ -\sum_{j\in S_{INT}(2)} H_{INT}(\boldsymbol{w}_2, \boldsymbol{w}_j, \nu_{2j})|\Gamma_{2j}| \\ \vdots \\ -\sum_{j\in S_{INT}(N)} H_{INT}(\boldsymbol{w}_N, \boldsymbol{w}_j, \nu_{Nj})|\Gamma_{Nj}| \end{bmatrix}, \tag{2.31}$$

$$\boldsymbol{F}_{SOL}(\boldsymbol{W}) = \begin{bmatrix} -\sum_{\alpha\in S_{SOL}(1)} H_{SOL}(\boldsymbol{w}_1, \nu_{1\alpha})|\Gamma_{1\alpha}| \\ -\sum_{\alpha\in S_{SOL}(2)} H_{SOL}(\boldsymbol{w}_2, \nu_{2\alpha})|\Gamma_{2\alpha}| \\ \vdots \\ -\sum_{\alpha\in S_{SOL}(N)} H_{SOL}(\boldsymbol{w}_N, \nu_{N\alpha})|\Gamma_{N\alpha}| \end{bmatrix}, \tag{2.32}$$

$$\boldsymbol{F}_{I/O}(\boldsymbol{W},t) = \begin{bmatrix} -\sum_{\alpha\in S_{I/O}(1)} H_{I/O}(\boldsymbol{w}_1, \boldsymbol{w}_{1\alpha}, \nu_{1\alpha})|\Gamma_{1\alpha}| \\ -\sum_{\alpha\in S_{I/O}(2)} H_{I/O}(\boldsymbol{w}_2, \boldsymbol{w}_{2\alpha}, \nu_{2\alpha})|\Gamma_{2\alpha}| \\ \vdots \\ -\sum_{\alpha\in S_{I/O}(N)} H_{I/O}(\boldsymbol{w}_N, \boldsymbol{w}_{N\alpha}, \nu_{N\alpha})|\Gamma_{N\alpha}| \end{bmatrix}. \tag{2.33}$$

where $\boldsymbol{w}_{i\alpha}(t) = \boldsymbol{w}_h^B(\boldsymbol{x},\, t)|_{\Gamma_{i\alpha}}$ represents boundary conditions at the faces $\Gamma_{i\alpha}$, $\alpha \in S_{I/O}$, $1 \le i \le N$. Our aim is to find a function $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_N)^T \in C^1(0,T;\mathcal{Q}^N)$, $\boldsymbol{w}_i \in C^1(0,T;\mathcal{Q})$ representing coefficients to the approximate solution

$$\boldsymbol{w}_h(\boldsymbol{x},\, t) = \sum_{i=1}^{N} \boldsymbol{w}_i(t)\chi_i(\boldsymbol{x}) \tag{2.34}$$

satisfying

$$\boldsymbol{A}\dot{\boldsymbol{W}}(t) = \boldsymbol{F}(\boldsymbol{W}(t),t), \tag{2.35}$$

$$\boldsymbol{W}(0) = \boldsymbol{W}^0. \tag{2.36}$$

where $\boldsymbol{A} = \{a_{i,j}\}_{i,j=1}^{(d+2)N}$ is a diagonal matrix with

$$a_{(d+2)i,(d+2)i} = a_{(d+2)i+1,(d+2)i+1} = \ldots = a_{(d+2)i+d+1,(d+2)i+d+1} = |\Omega_i|, \tag{2.37}$$

$i = 1, 2, \ldots, N$. The symbol $\chi_i(\boldsymbol{x})$, $1 \leq i \leq N$ denotes Ω_i-characteristic functions. Note that the right-hand-side vector function

$$\boldsymbol{F}(\boldsymbol{W}(t), t) = \boldsymbol{F}_{INT}(\boldsymbol{W}(t)) + \boldsymbol{F}_{SOL}(\boldsymbol{W}(t)) + \boldsymbol{F}_{I/O}(\boldsymbol{W}(t), t) \tag{2.38}$$

has the form corresponding to (2.23).

2.5 Coupling of the Finite Volume Scheme with Equations Describing Motion of Pollution

Let us denote $\varrho_h(\boldsymbol{x}_k(t), t)$, $\boldsymbol{v}_h(\boldsymbol{x}_k(t), t)$ piecewise constant approximations of fluid density and fluid velocity given by the finite volume scheme, corresponding to the point $\boldsymbol{x}_k(t)$, respectively. In this section we re-construct quantities $\varrho_h(\boldsymbol{x}_k(t), t)$, $\boldsymbol{v}_h(\boldsymbol{x}_k(t), t)$ to piecewise linear continuous functions $\hat{\varrho}_h(\boldsymbol{x}_k(t), \boldsymbol{W}(t))$, $\hat{\boldsymbol{v}}_h(\boldsymbol{x}_k(t), \boldsymbol{W}(t))$, respectively, using solution $\boldsymbol{W}(t)$ corresponding to the finite volume scheme (2.35)(2.36), and substitute them into the ordinary differential equations (2.12)(2.13) describing motion of particles.

For all $i = 1, 2, \ldots, N$, let us construct index sets $K_i \subset \{1, 2, \ldots, M\}$, $\mathrm{card}(K_i) = k_i$ of all vertices of finite volume Ω_i, . Let us further construct a set $N_j \subset \{1, 2, \ldots, N\}$, $\mathrm{card}(N_j) = n_j$ of all finite volumes having the grid point P_j in common, $j = 1, 2, \ldots, M$.

In order to get a solvable system of ordinary differential equations with continuous right-hand-side, we reconstruct the piecewise constant finite volume approximation to an approximation which is continuous in Ω. In our case we use piecewise linear reconstruction. Piecewise constant finite volume approximations of flow density and velocity corresponding to centroids of finite volumes $\Omega_1, \Omega_2, \ldots, \Omega_N$ are transmitted into grid points $P_1, P_2, \ldots, P_M$ by means of weighted averaging as follows:

$$\hat{\boldsymbol{v}}_j(\boldsymbol{W}(t)) = \frac{\sum_{i \in N_j} \boldsymbol{v}_i(\boldsymbol{W}(t)) |\Omega_i|}{\sum_{i \in N_j} |\Omega_i|}. \tag{2.39}$$

Here $\hat{\boldsymbol{v}}_j(\boldsymbol{W}(t))$ is the reconstructed value of the piecewise linear velocity at the grid point P_j, $j = 1, 2, \ldots, M$ and $\boldsymbol{v}_i(\boldsymbol{W}(t))$ is the piecewise constant

finite volume velocity approximation corresponding to centroids of finite volumes Ω_i, $i = 1, 2, \ldots, N$.
Analogously, we reconstruct also the density:

$$\hat{\varrho}_j(\boldsymbol{W}(t)) = \frac{\sum_{i \in N_j} \varrho_i(\boldsymbol{W}(t))|\Omega_i|}{\sum_{i \in N_j} |\Omega_i|}. \tag{2.40}$$

Let us remark that with the piecewise linear reconstruction we could only proceed with triangular finite volumes in $2D$ and with tetrahedral finite volumes in $3D$.

By $\lambda_{i,j}(\boldsymbol{x})$, $i = 1, 2, \ldots, N$, $j \in K_i$ we denote barycentric co-ordinates of a point $\boldsymbol{x} \in \Omega$ corresponding to the triangle (or tetrahedron) Ω_i and its vertex P_j. If $\boldsymbol{x} \in \Omega_i$, it is

$$\sum_{j \in K_i} \lambda_{i,j}(\boldsymbol{x}) = 1. \tag{2.41}$$

If $\boldsymbol{x} \notin \Omega_i$, we define $\lambda_{i,j}(\boldsymbol{x}) = 0$ for all $j \in K_i$.

With the above introduced barycentric co-ordinates, we can finally define piecewise linear functions

$$\hat{\boldsymbol{v}}_h(\boldsymbol{x}, \boldsymbol{W}(t)) = \sum_{i=1}^{N} \sum_{j \in K_i} \lambda_{i,j}(\boldsymbol{x}) \hat{\boldsymbol{v}}_j(\boldsymbol{W}(t)) \tag{2.42}$$

and

$$\hat{\varrho}_h(\boldsymbol{x}, \boldsymbol{W}(t)) = \sum_{i=1}^{N} \sum_{j \in K_i} \lambda_{i,j}(\boldsymbol{x}) \hat{\varrho}_j(\boldsymbol{W}(t)), \tag{2.43}$$

suitable for substitution into the system (2.12)(2.13). Substituting then (2.42) (2.43) into (2.12)(2.13) with $\boldsymbol{x} := \boldsymbol{x}_k(t)$ and using system of ordinary differential equations (2.35)(2.36) constructed in subsection **2.4.3**, we obtain a system of first order ordinary differential equations of the form

$$\boldsymbol{A}\dot{\boldsymbol{W}}(t) = \boldsymbol{F}(\boldsymbol{W}(t), t), \tag{2.44}$$

$$\boldsymbol{M}\dot{\boldsymbol{U}}(t) = \boldsymbol{F}_S(\boldsymbol{X}(t), \boldsymbol{U}(t), \boldsymbol{W}(t)), \tag{2.45}$$

$$\dot{\boldsymbol{X}}(t) = \boldsymbol{U}(t), \tag{2.46}$$

$t \in (0, T)$, with the initial condition

$$\boldsymbol{W}(0) = \boldsymbol{W}^0, \tag{2.47}$$
$$\boldsymbol{U}(0) = \boldsymbol{U}^0, \tag{2.48}$$
$$\boldsymbol{X}(0) = \boldsymbol{X}^0. \tag{2.49}$$

Initial data have according to (2.14)(2.15) the form

$$\boldsymbol{U}^0 = (\boldsymbol{u}_1^0, \boldsymbol{u}_2^0, \dots, \boldsymbol{u}_P^0)^T, \tag{2.50}$$

$$\boldsymbol{X}^0 = (\boldsymbol{x}_1^0, \boldsymbol{x}_2^0, \dots, \boldsymbol{x}_P^0)^T. \tag{2.51}$$

Matrix $\boldsymbol{M} = \{m_{k,l}\}_{k,l=1}^{dP}$ is a diagonal matrix containing d-times weight of the particle X_1, d-times weight of the particle X_2, …, d-times weight of the particle X_P on its diagonal.

The right-hand-side function $\boldsymbol{F}_S$ has according to (2.12)(2.13) the form

$$\boldsymbol{F}_{S,k}(\boldsymbol{X}(t), \boldsymbol{U}(t), \boldsymbol{W}(t)) = \boldsymbol{f}_S(r_k, \hat{\varrho}(\boldsymbol{x}_k(t), \boldsymbol{W}(t)), \boldsymbol{u}_k(t) - \hat{\boldsymbol{v}}(\boldsymbol{x}_k(t), \boldsymbol{W}(t))) + \boldsymbol{g}, \tag{2.52}$$

$k = 1, 2, \dots, P$, and vector functions $\boldsymbol{U}(t)$, $\boldsymbol{X}(t)$ have according to (2.12)(2.13) the form

$$\boldsymbol{U}(t) = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_P)^T(t), \tag{2.53}$$

$$\boldsymbol{X}(t) = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_P)^T(t). \tag{2.54}$$

2.6 Numerical Examples

Right-hand-side vector functions $\boldsymbol{F}$ and $\boldsymbol{F}_S$ have been evaluated explicitly by original C++ codes EULER[1] based on [28], working for $d = 1, 2, 3$ and SEPAR[2], working for $d = 2, 3$. Time-integration is performed using the free Fortran library ODEPACK downloaded from Netlib. For generation of $2D$ meshes we use original C++ grid generator XGEN[3]. Flow has been visualized by the C programming tool GRAPE developed at the Friedrich-Wilhelm-University of Bonn. All computations were performed on a DEC Alpha workstation with a RISC 21164/550MHz processor in double precision.

[1] Free to download at http://www.karlin.mff.cuni.cz/katedry/knm/euler
[2] Free to download at http://www.karlin.mff.cuni.cz/katedry/knm/separ
[3] Free to download at http://www.karlin.mff.cuni.cz/katedry/knm/xgen

2.6.1 Motion of Air Polluted by Water Droplets in a 2D Channel

In this subsection, we will compute stationary solution of motion of air polluted by water droplets in a 2D channel. The domain Ω is described by

$$\Omega = \left\{ (x_1, x_2) = \left(\alpha_1, \frac{1}{2} \cos \alpha_1 \pi - \alpha_2 \right) \in \mathbb{R}^2 \right\} \tag{2.55}$$

where $0 < \alpha_1 < 2$, $0 < \alpha_2 < \frac{1}{2}$, shown in Figure 2.1. Boundary edges corresponding to $x_L = 0$ and $x_R = 2$ represent the inlet and outlet and remaining boundary edges represent solid wall. The air containing small water droplets is entering the domain with velocity $\boldsymbol{v}_{inlet} = [10,\, 0]^T$m/s and density $\varrho_{inlet} = 1.5$kg/m^3. Outlet pressure is specified as $p_{outlet} = 101000$Pa. We consider further the kinematic viscosity of air $\nu = 3.52$e-6 and density of the water droplets $\varrho_i = 986.8$kg/m^3, $i = 1, 2, \ldots, P$.

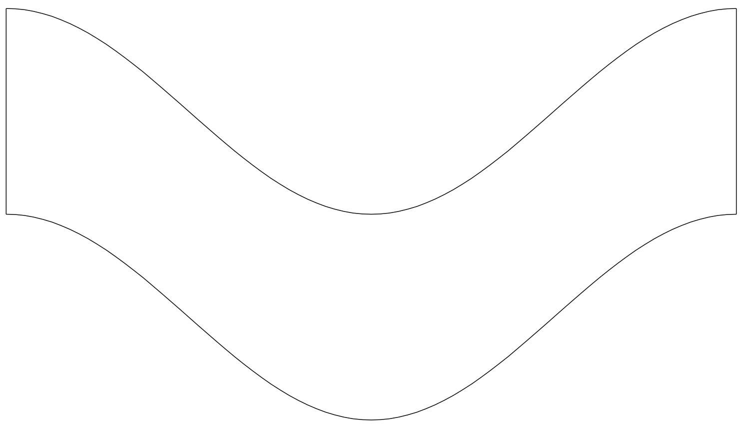

Figure 2.1: 2D curved channel

Finite volume grid containing 3200 triangles is shown in Figure 2.2.

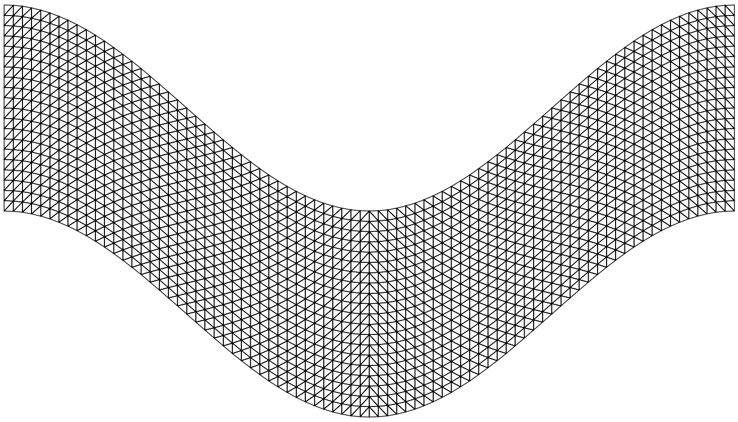

Figure 2.2: Finite volume grid

We are interested in motion of the pollution in stationary flow, therefore we use stationary solution of system (2.35)(2.36) as initial data $\boldsymbol{W}^0$ for the system (2.44)(2.45)(2.46). Solution $\boldsymbol{W}$ of the system (2.35)(2.36) is called stationary if there is a $t_0 > 0$ such that $\boldsymbol{F}(\boldsymbol{W}, t) = 0$ for all $t > t_0$. Stationary solution corresponding to our example is shown in Figures 2.3, 2.4.

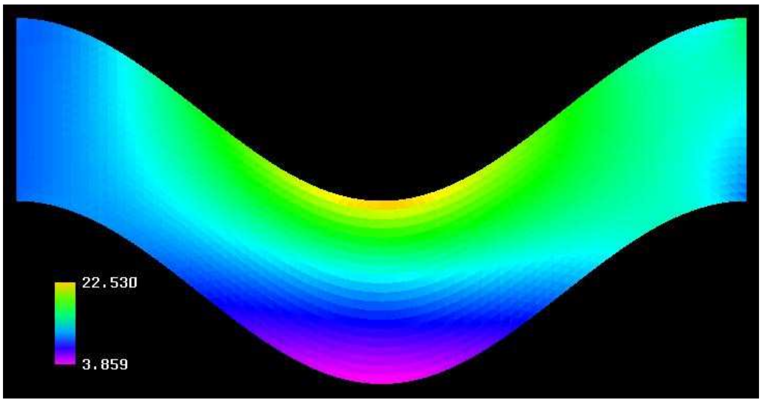

Figure 2.3: Velocity distribution corresponding to the steady state

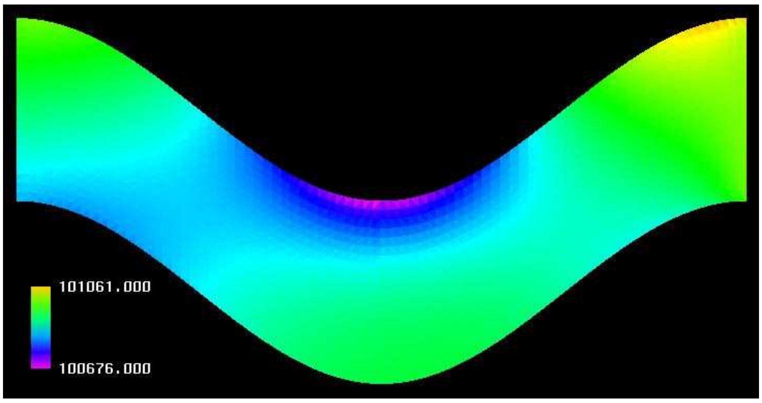

Figure 2.4: Pressure distribution corresponding to the steady state

Water droplets are simulated by means of $P = 49$ particles X_k placed equidistantly on the inlet edge with initial velocities $\boldsymbol{v}_k^0 = \boldsymbol{v}_{inlet}$, $k = 1, 2, \ldots, P$. Traces of particles for different values of their radius r_k are shown in Figures 2.5 – 2.7.

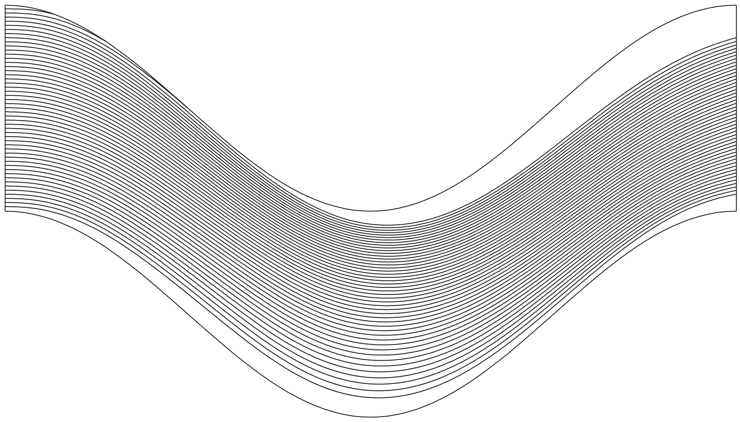

Figure 2.5: Traces of water droplets, radius r = 1e-5m.

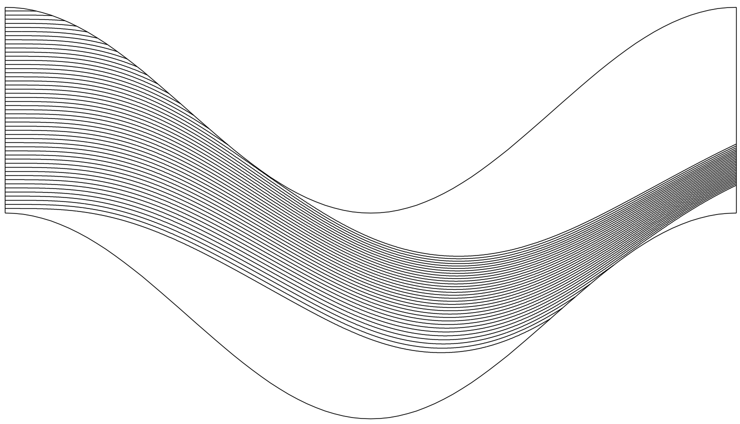

Figure 2.6: Traces of water droplets, radius r = 5e-5m.

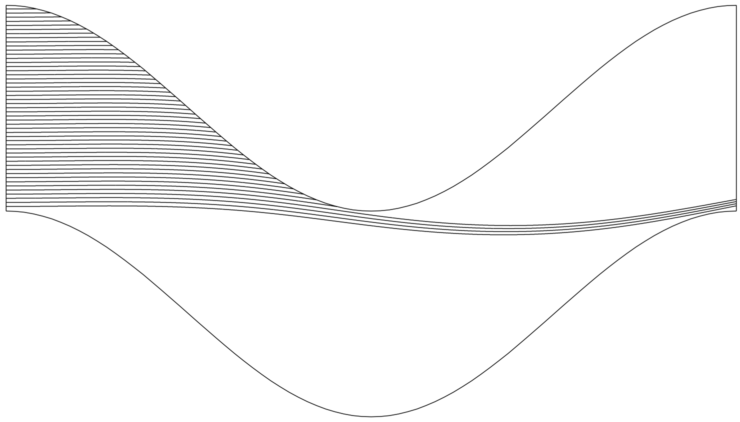

Figure 2.7: Traces of water droplets, radius r = 3e-4m.

2.6.2 Motion of Air Polluted with Water Droplets in a 3D Channel

We perform a $3D$ analogy of the previous $2D$ computation. The domain Ω is described by

$$(2.56)\ \Omega = \left\{ (x_1, x_2, x_3) = \left(\alpha_1, \frac{1}{4} \sin \alpha_1 \pi + \alpha_2, \frac{1}{4} \cos \alpha_1 \pi + \alpha_3 \right) \in \mathbb{R}^3 \right\}$$

where $0 < \alpha_1 < 2,\ 0 < \alpha_2, \alpha_3 < \frac{1}{2}$, shown in Figure 2.8. Boundary faces corresponding to $x_L = 0$ and $x_R = 2$ represent inlet and outlet and remaining boundary faces represent solid wall. The air containing small water droplets is entering the domain with velocity $\boldsymbol{v}_{inlet} = [4,\ 3.14,\ 0]^T$m/s and density $\varrho_{inlet} = 1.5$kg/m^3. Outlet pressure is specified as $p_{outlet} = 101000$Pa.

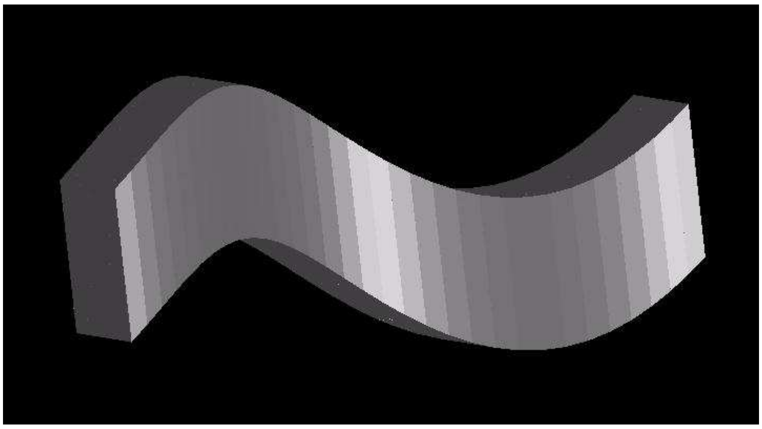

Figure 2.8: 3D curved channel

The domain is covered with a structured tetrahedral grid consisting of $N = 12288$ finite volumes. We consider again the kinematic viscosity of air $\nu = 3.52$e-6 and density of the water droplets $\varrho_i = 986.8$kg/m^3, $i = 1, 2, \ldots, P$.

Again, we intend to simulate motion of the pollution in stationary flow, therefore we use stationary solution of system (2.35)(2.36) as initial data $\boldsymbol{W}^0$ for the system (2.44)(2.45)(2.46). Stationary solution corresponding to our example is shown in Figures 2.9, 2.10

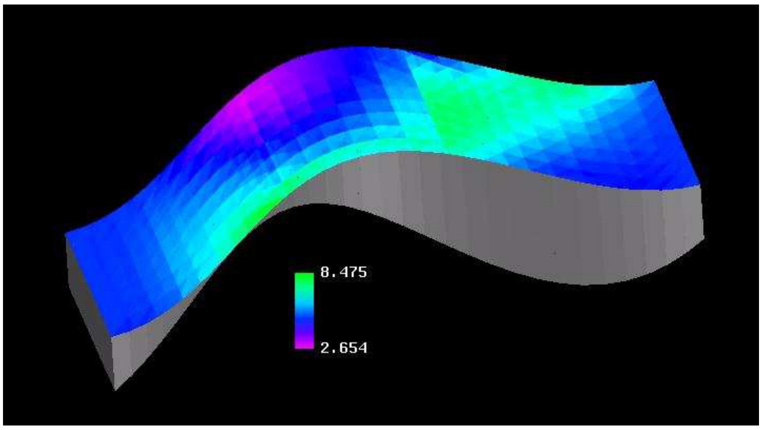

Figure 2.9: Velocity distribution, xy–cut, $z = 0.25$

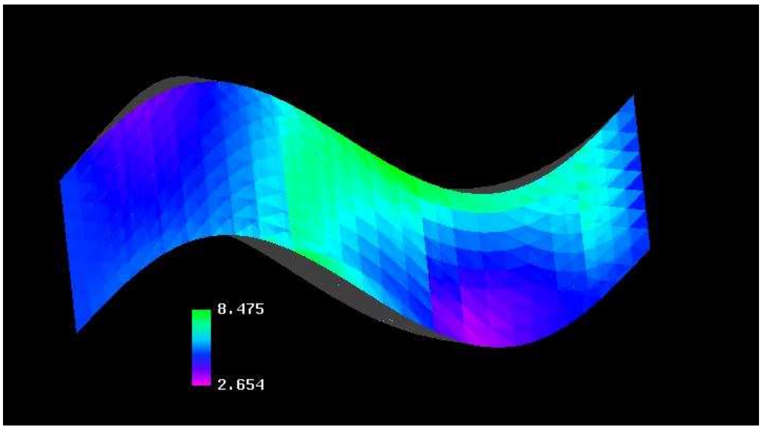

Figure 2.10: Velocity distribution, xz–cut, $y = 0$

Water droplets are simulated by means of $P = 100$ particles X_k placed equidistantly on the inlet face with initial velocities $\boldsymbol{v}_k^0 = \boldsymbol{v}_{inlet}$, $k = 1, 2, \ldots, P$. Traces of particles with the radius r_k = 2e-5m are shown in Figures 2.11 – 2.14.

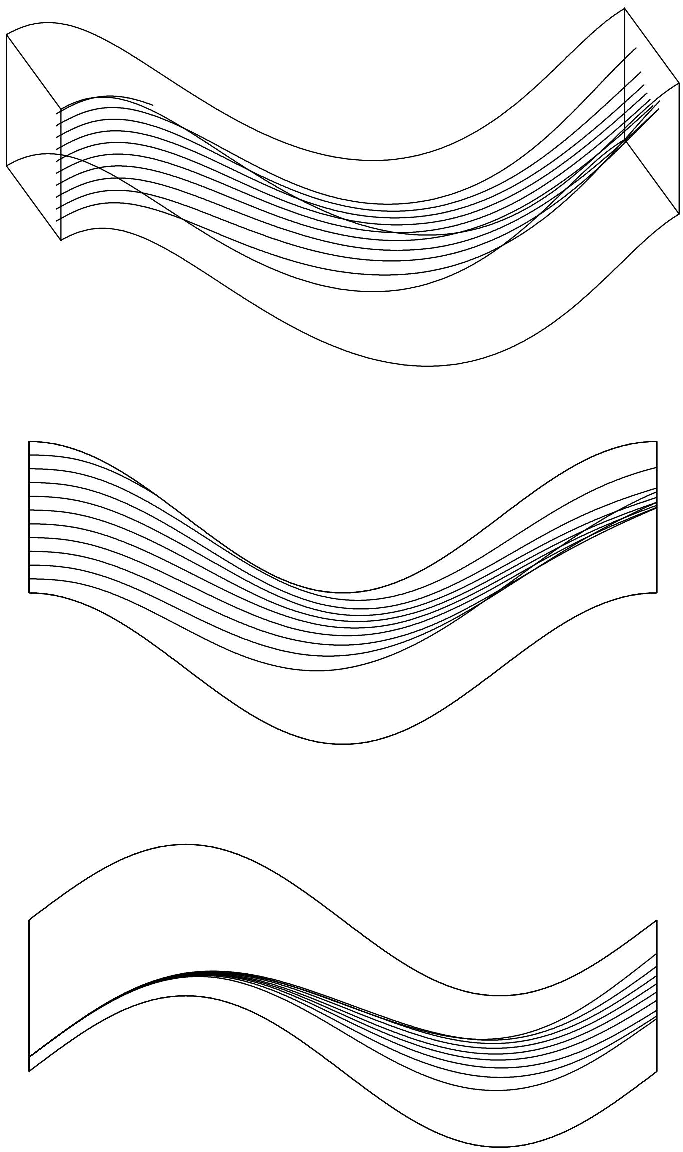

Figure 2.11: Traces of water droplets; perspective, x_2 - axis and x_3 - axis view.

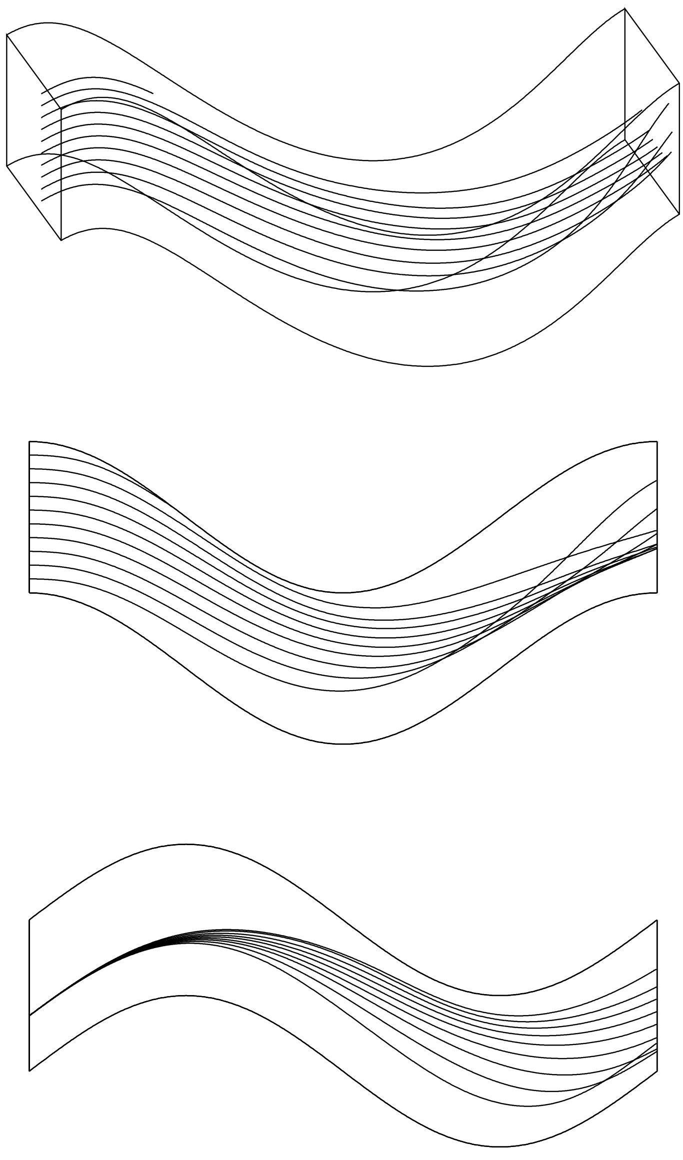

Figure 2.12: Traces of water droplets, perspective; x_2 - axis and x_3 - axis view.

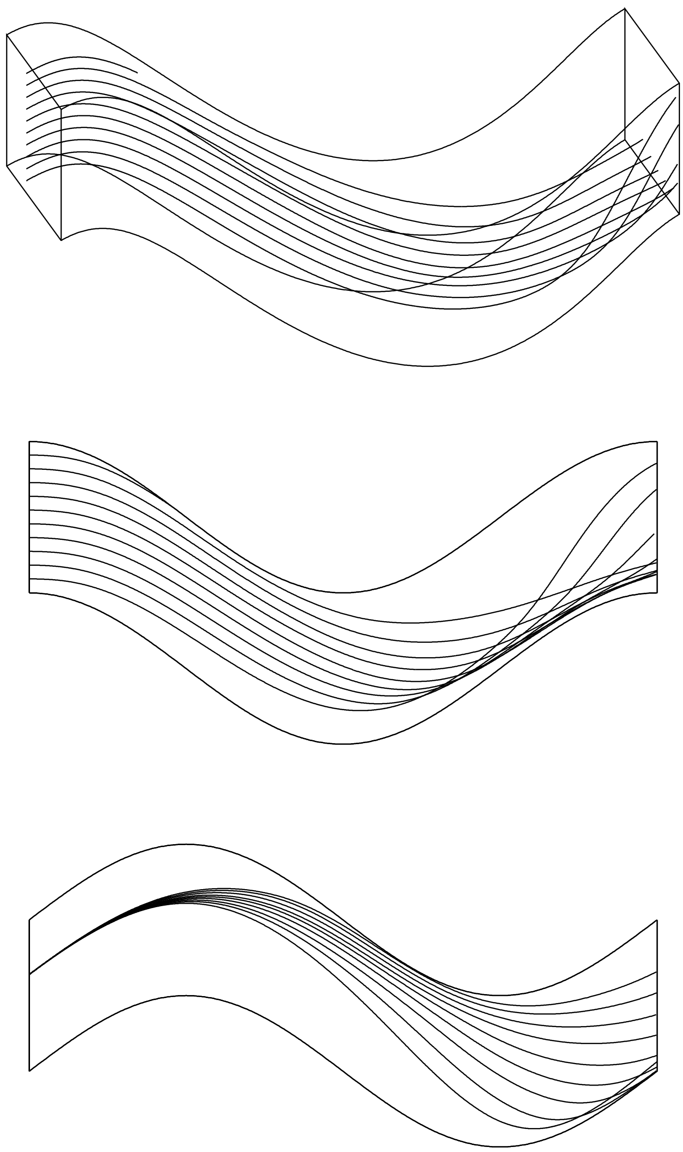

Figure 2.13: Traces of water droplets, perspective; x_2 - axis and x_3 - axis view.

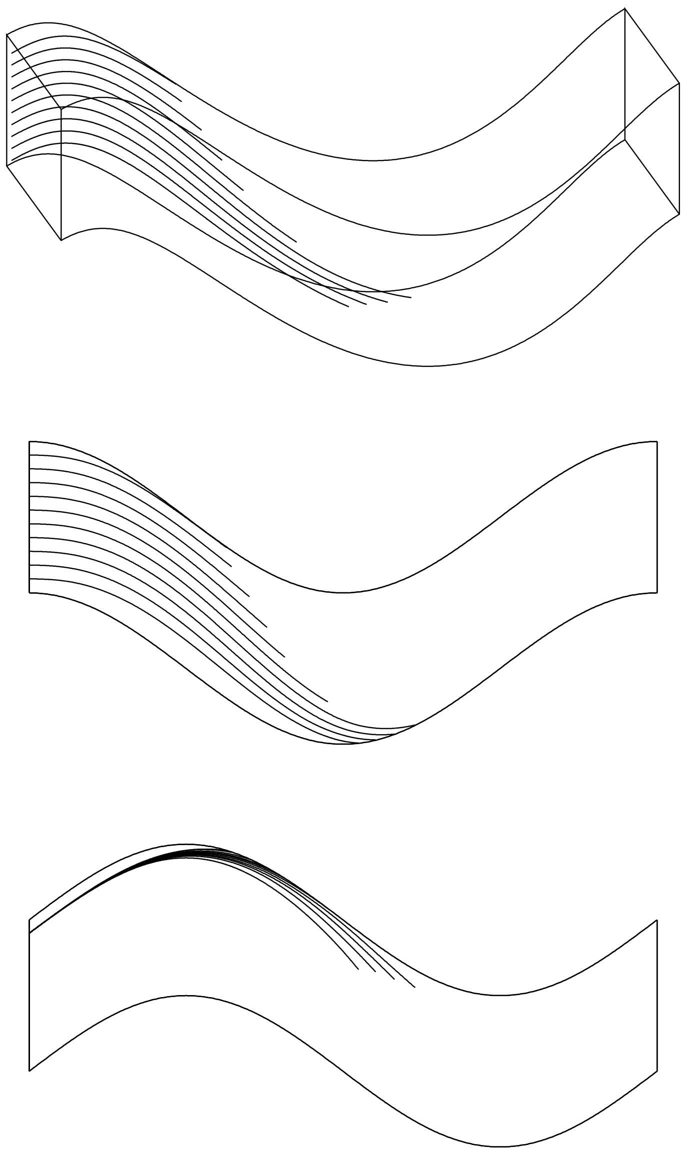

Figure 2.14: Traces of water droplets, perspective; x_2 - axis and x_3 - axis view.

2.7 Conclusion

We have developed a mathematical model of motion of small mass particles within flowing fluids. The model has been implemented in the programming language C++ and several 2D and 3D numerical experiments presenting its possible application have been presented. The model can be applied for many problems in ecology (simulation of distribution of pollutions in the air), turbomachinery (simulation of cleaning of steam), automotive industry (simulation of car catalyzers), air-conditioning (simulation of motion of dust in buildings) and others.

Chapter 3

On the $h - p$ FEM Semi-Discretization of Compressible Navier-Stokes Equations

Abstract: This part of dissertation is devoted to the higher-order finite element solution of compressible Navier-Stokes equations. In space, the partial differential equations are semi-discretized according to the so-called $h-p$ strategy while in time the known Method of lines is applied. We formulate the problem in the weak sense and construct the so-called *hierarchical basis* of the approximative polynomial spaces of functions. We formulate the semi-discrete problem as a non-linear system of ordinary differential equations and use a suitable numerical software for its solution. We construct a non-trivial stationary problem for compressible Navier-Stokes equation in a square channel having analytical solution. Numerical results verifying correctness of implementation of the relatively complicated numerical algorithm on this analytically solvable example are presented.

3.1 Introduction

This part of dissertation is devoted to the numerical solution of viscous compressible gas flow in domains with piecewise polynomial boundaries.

System of compressible Navier-Stokes equations in conservation form is approximated by the method of lines using higher-order finite element semi-discretization in space. Resulting system of ordinary differential equations has the form

$$\tilde{\boldsymbol{A}}\dot{\boldsymbol{y}}(t) = \tilde{\Phi}(\boldsymbol{y}(t), t) \tag{3.1}$$

$$\boldsymbol{y}(0) = \boldsymbol{y}^0 \tag{3.2}$$

where the so-called *mass matrix* $\tilde{\boldsymbol{A}}$ is a symmetric positive definite real matrix and $\tilde{\Phi}$ is a non-linear vector function.

It is well known that standard Galerkin finite element methods break down if the solution develops discontinuities, which is caused by the fact that global conservation of entropy is achieved in the Galerkin variational formulation (see e.g. [35]). For the purpose of numerical solution, the finite element method must be stabilized by artificial dissipation terms. In this dissertation the so-called *Streamline-diffusion method* (see e.g. [35], [36], [65], [17]) is used, imposing entropy inequality by means of additional least squares artificial viscosity terms. These terms change both the right-hand-side and the left-hand-side of the system (3.1) and make the mass matrix $\tilde{\boldsymbol{A}}$ non-symmetric and time-dependent. So, system (3.1)(3.2) comes over to

$$\boldsymbol{A}(t)\dot{\boldsymbol{y}}(t) = \Phi(\boldsymbol{y}(t), t) \tag{3.3}$$

$$\boldsymbol{y}(0) = \boldsymbol{y}^0 \tag{3.4}$$

Analysis of continuity and Lipschitz continuity of the right hand side vector function Φ is performed and existence and uniqueness of solution of the system (3.3)(3.4) is discussed.

Implementation of the above mentioned method requires several numerical techniques. Higher-order one- and two-dimensional numerical quadratures of the Gauss type are used for the construction of the mass matrix $\boldsymbol{A}$ and for the evaluation of the non-linear right-hand-side vector function Φ. The vector $\dot{\boldsymbol{y}}$ is resolved by efficient bi-conjugate iterative gradient solvers with incomplete LU preconditioning. Time-integration of the system (3.3)(3.4) can be carried out by explicit multistep higher-order schemes as well as by an implicit multistep higher-order scheme based on the so-called *Backward Difference Formula* (see e.g. [64]) both implemented in the Fortran programming package ODEPACK (see [39], [13], [14]).

Notation

We use standard characters for scalar values $(\varrho, p, \theta, \ldots)$ and boldface characters for vectors and vector valued functions $(\boldsymbol{v}, \boldsymbol{\varphi}, \ldots)$. Components of vectors are denoted by $\boldsymbol{U} = (u_1, u_2, \ldots, u_N)^T$ where N is the space dimension. Throughout this dissertation, we will use the following notation:

t, $\boldsymbol{x} = (x_1, x_2)^T$ ... time and space variable,
$\varrho(\boldsymbol{x}, t)$, $\boldsymbol{v}(\boldsymbol{x}, t) = (v_1, v_2)^T(\boldsymbol{x}, t)$, $p(\boldsymbol{x}, t)$ ... density, velocity and pressure,
$\theta(\boldsymbol{x}, t)$, $e(\boldsymbol{x}, t) = \varrho(c_v\theta + \frac{1}{2}|\boldsymbol{v}|^2)$ ... temperature and density of energy,
$\boldsymbol{g}(\boldsymbol{x}, t) = (g_1, g_2)^T(\boldsymbol{x}, t)$ ... density of volume forces relative to unit of mass,
$\tau(\boldsymbol{x}, t)$ ... stress tensor,
c_p, c_v ... specific heats at constant pressure and volume,
$q(\boldsymbol{x}, t)$ ... density of heat sources,
$\underline{\boldsymbol{q}}(\boldsymbol{x}, t)$ ... heat flux,
$R = c_p - c_v$ ... the gas constant,
$\kappa = \frac{c_p}{c_v}$ ... the Poisson constant,
μ, λ ... viscosity coefficients,
k ... Fourier heat conductivity coefficient,
$P_r = \mu c_p / k$ ... Prandtl number.

Operators

Let a be a scalar, $\boldsymbol{b}$ be a vector and $\boldsymbol{C}$ be a tensor. By $\boldsymbol{C}_i$, $i = 1, 2$ we denote the $i-$th column of $\boldsymbol{C}$. Nabla is applied to scalars and vectors with the following conventions:

$\cdot$ ... (dot) denotes scalar product,
$\times$ ... denotes vector product,
$\otimes$ ... denotes tensor product,
$\nabla = (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2})^T$ ... nabla. $\nabla a = (\frac{\partial a}{\partial x_1}, \frac{\partial a}{\partial x_2})^T$ ... gradient of a scalar value
$\nabla \cdot \boldsymbol{b} = \frac{\partial b_1}{\partial x_1} + \frac{\partial b_2}{\partial x_2}$ ... divergence of a vector,
$\nabla \boldsymbol{b} = \nabla \otimes \boldsymbol{b} = \{\frac{\partial b_j}{\partial x_i}\}_{i,j=1,2}$ ... gradient of a vector,

$\triangle a = \frac{\partial^2 a_1}{\partial x_1^2} + \frac{\partial^2 a_2}{\partial x_2^2}$ ... Laplacian operator,
$\nabla \cdot \boldsymbol{C} = (\nabla \cdot \boldsymbol{C}_1, \nabla \cdot \boldsymbol{C}_2)^T$ ... divergence of a tensor.

3.2 Governing Equations

3.2.1 Conservation Laws

The system of compressible Navier-Stokes equations consists of the law of conservation of mass

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot (\varrho \boldsymbol{v}) = 0, \tag{3.5}$$

law of conservation of momentum

$$\frac{\partial \varrho v_i}{\partial t} + \nabla \cdot (\varrho v_i \boldsymbol{v}) = \varrho g_i + \nabla \cdot \tau_i, \quad i = 1, 2 \tag{3.6}$$

and of law of conservation of energy

$$\frac{\partial e}{\partial t} + \nabla \cdot (e \boldsymbol{v}) = \varrho \boldsymbol{g}.\boldsymbol{v} + \nabla \cdot (\tau \boldsymbol{v}) + \varrho q - \nabla \cdot \underline{\boldsymbol{q}}. \tag{3.7}$$

Heat flux $\underline{\boldsymbol{q}}$ is governed by the Fourier law

$$\underline{\boldsymbol{q}} = -k \nabla \theta \tag{3.8}$$

where k is a real constant. Moreover, perfect gas state equation

$$p = \varrho R \theta \tag{3.9}$$

and equation

$$e = \varrho (c_v \theta + \frac{1}{2} |\boldsymbol{v}|^2) \tag{3.10}$$

defining total energy density are considered in order to balance number of unknowns (which are ϱ, $\boldsymbol{v}$, p, e, θ, $\underline{\boldsymbol{q}}$) with number of equations. The stress tensor τ has the form

$$\tau = (-p + \lambda \nabla \cdot \boldsymbol{v}) \boldsymbol{I} + 2\mu \boldsymbol{D} \tag{3.11}$$

where

$$3\lambda + 2\mu = 0 \tag{3.12}$$

and

$$\boldsymbol{D} = \frac{\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T}{2} \tag{3.13}$$

denotes the deformation velocity tensor. In this dissertation, we neglect volume forces $\boldsymbol{g}$ and density of heat sources q. Equations (3.5) – (3.10) are considered in a domain Ω which is described in detail in the next paragraph.

3.2.2 Assumptions on Domain Occupied by the Fluid

Let us consider a bounded domain Ω with Lipschitz continuous, positively oriented boundary.

We assume existence of a finite number of points $X_1, X_2, \dots X_{N_x} \in \partial\Omega$, $N_x > 2$, $X_i \neq X_j$ for $1 \leq i < j \leq N_x$ with the following properties:

We say that two points X_i and X_j, $i \neq j$ are *neighbours*, if it is possible to connect them with a continuous curve $\phi \subset \partial\Omega$ which does not contain any other point X_k, $k \neq i \neq j$. For each pair of neighbouring points X_i, X_j, $1 \leq i, j \leq N_x$ let us construct a local co-ordinate system $\mathcal{S}_{i,j}$ obtained from the original one by rotation and translation so that both points X_i, X_j lie in the x-axis of $\mathcal{S}_{i,j}$ and one of the points X_i, X_j is the origin $\mathcal{O}$ of $\mathcal{S}_{i,j}$ so that the boundary $\partial\Omega$ has at $\mathcal{O}$ the same orientation as the x-axis of $\mathcal{S}_{i,j}$.

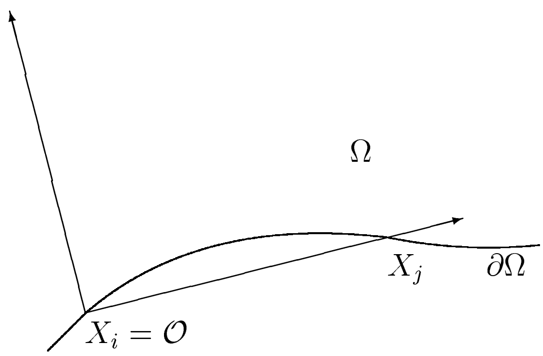

Figure 3.1: Local co-ordinate system $\mathcal{S}_{i,j}$

We assume that in each of the above constructed local co-ordinate systems $\mathcal{S}_{i,j}$ it is possible to express the boundary $\partial\Omega$ as a function $f_{i,j}(x'_{i,j})$, $x'_{i,j} \in$

$\langle 0, |X_j - X_i| \rangle$, $f_{i,j}(0) = f_{i,j}(|X_j - X_i|) = 0$ and the function $f_{i,j}$ is a polynomial of degree at most p_B.

In numerical schemes for this type of problems, one usually approximates the domain Ω with another domain which is more suitable for the application of the numerical scheme. With the above introduced assumptions, we can omit this step.

3.2.3 Transformation to Conservation Variables

It is convenient to turn the system of compressible Navier-Stokes equations into a more suitable form

$$\frac{\partial}{\partial t}\boldsymbol{w}(\boldsymbol{x},t) + \sum_{i=1}^{2}\frac{\partial}{\partial x_i}\boldsymbol{f}_i(\boldsymbol{w}(\boldsymbol{x},t)) = \sum_{i=1}^{2}\frac{\partial}{\partial x_i}\boldsymbol{R}_i(\boldsymbol{w}(\boldsymbol{x},t))prd \tag{3.14}$$

using the so-called *conservation variables*

$$\boldsymbol{w}(\boldsymbol{x},t) = (w_1, w_2, w_3, w_4)^T(\boldsymbol{x},t) = (\varrho, \varrho v_1, \varrho v_2, e)^T(\boldsymbol{x},t). \tag{3.15}$$

The vector $\boldsymbol{w}(\boldsymbol{x},t)$ representing physical quantities is defined in the so-called *set of admissible states*

$$\mathcal{Q} = \left\{\boldsymbol{w} \in \mathbb{R}^4;\ w_1 > 0,\ w_4 > \frac{w_2^2 + w_3^2}{2w_1}\right\}. \tag{3.16}$$

The inviscid fluxes $\boldsymbol{f}_i(\boldsymbol{w}) : \mathcal{Q} \to \mathbb{R}^4$ and the viscous fluxes $\boldsymbol{R}_i(\boldsymbol{w}) : \mathcal{Q}\times\mathbb{R}^4 \to \mathbb{R}^4$ have the following form:

$$\boldsymbol{f}_1(\boldsymbol{w}) = \begin{pmatrix} w_2 \\ \frac{{w_2}^2}{w_1} + p \\ \frac{w_2 w_3}{w_1} \\ \frac{w_2}{w_1}e \end{pmatrix}, \tag{3.17}$$

$$\boldsymbol{f}_2(\boldsymbol{w}) = \begin{pmatrix} w_3 \\ \frac{w_2 w_3}{w_1} \\ \frac{{w_3}^2}{w_1} + p \\ \frac{w_3}{w_1}e \end{pmatrix}, \tag{3.18}$$

$$\boldsymbol{R}_1(\boldsymbol{w}) = \begin{pmatrix} 0 \\ \tau_{11} \\ \tau_{12} \\ \tau_{11}\frac{w_2}{w_1} + \tau_{12}\frac{w_3}{w_1} + k\frac{\partial\theta}{\partial x_1} \end{pmatrix}, \tag{3.19}$$

$$\boldsymbol{R}_2(\boldsymbol{w}) = \begin{pmatrix} 0 \\ \tau_{21} \\ \tau_{22} \\ \tau_{21}\frac{w_2}{w_1} + \tau_{22}\frac{w_3}{w_1} + k\frac{\partial\theta}{\partial x_2} \end{pmatrix} \tag{3.20}$$

where

$$\boldsymbol{v} = \left(\frac{w_2}{w_1}, \frac{w_3}{w_1}\right)^T, \tag{3.21}$$

$$p = (\kappa - 1)\left(w_4 - \frac{{w_2}^2 + {w_3}^2}{2w_1}\right), \tag{3.22}$$

$$e = \kappa w_4 - (\kappa - 1)\frac{{w_2}^2 + {w_3}^2}{2w_1}, \tag{3.23}$$

$$\boldsymbol{D} = \frac{\nabla\boldsymbol{v} + (\nabla\boldsymbol{v})^T}{2}, \tag{3.24}$$

$$\tau = \lambda\left(\frac{\partial}{\partial x_1}\frac{w_2}{w_1} + \frac{\partial}{\partial x_2}\frac{w_3}{w_1}\right)\boldsymbol{I} + 2\mu\boldsymbol{D}, \tag{3.25}$$

$$\theta = \frac{1}{c_v}\left(\frac{w_4}{w_1} - \frac{w_2^2 + w_3^2}{2w_1^2}\right). \tag{3.26}$$

See Appendix **3.9.1** for 3D form of compressible Navier-Stokes equations in conservation variables. Let us remark that if we set $\mu = \lambda = 0$ and neglect density of volume sources $\boldsymbol{g}$, heat sources q and heat flux $\underline{\boldsymbol{q}}$, the system is simplified to the so-called *Euler equations*

$$\frac{\partial}{\partial t}\boldsymbol{w}(\boldsymbol{x}, t) + \sum_{i=1}^{2}\frac{\partial}{\partial x_i}\boldsymbol{f}_i(\boldsymbol{w}(\boldsymbol{x}, t)) = 0. \tag{3.27}$$

For the implementation of the streamline diffusion strategy in section **3.5** we will need also the explicit form of the Jacobi matrices $\boldsymbol{A}_1(\boldsymbol{w}) = \boldsymbol{f}_1^{'}(\boldsymbol{w})$, $\boldsymbol{A}_2(\boldsymbol{w}) = \boldsymbol{f}_2^{'}(\boldsymbol{w})$ of the inviscid fluxes $\boldsymbol{f}_1(\boldsymbol{w})$, $\boldsymbol{f}_2(\boldsymbol{w})$. They have the form

(3.28) $$\boldsymbol{A}_1(\boldsymbol{w}) =$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{\kappa-1}{2}\frac{w_2^2+w_3^2}{w_1^2}-\frac{w_2^2}{w_1^2} & (3-\kappa)\frac{w_2}{w_1} & (1-\kappa)\frac{w_3}{w_1} & \kappa-1 \\ -\frac{w_2w_3}{w_1^2} & \frac{w_3}{w_1} & \frac{w_2}{w_1} & 0 \\ \frac{w_2}{w_1}\left(\frac{(\kappa-1)(w_2^2+w_3^2)}{w_1^2}-\kappa\frac{w_4}{w_1}\right) & \kappa\frac{w_4}{w_1}-\frac{(\kappa-1)w_2^2}{w_1^2}-\frac{\kappa-1}{2}\frac{w_2^2+w_3^2}{w_1^2} & \frac{(1-\kappa)w_2w_3}{w_1^2} & \kappa\frac{w_2}{w_1} \end{pmatrix}$$

and

(3.29) $$\boldsymbol{A}_2(\boldsymbol{w}) =$$

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ -\frac{w_2w_3}{w_1^2} & \frac{w_3}{w_1} & \frac{w_2}{w_1} & 0 \\ \frac{\kappa-1}{2}\frac{w_2^2+w_3^2}{w_1^2}-\frac{w_3^2}{w_1^2} & (1-\kappa)\frac{w_2}{w_1} & (3-\kappa)\frac{w_3}{w_1} & \kappa-1 \\ \frac{w_3}{w_1}\left(\frac{(\kappa-1)(w_2^2+w_3^2)}{w_1^2}-\kappa\frac{w_4}{w_1}\right) & \frac{(1-\kappa)w_2w_3}{w_1^2} & \kappa\frac{w_4}{w_1}-\frac{(\kappa-1)w_3^2}{w_1^2}-\frac{\kappa-1}{2}\frac{w_2^2+w_3^2}{w_1^2} & \kappa\frac{w_3}{w_1} \end{pmatrix}.$$

See Appendix **3.9.2** for Jacobi matrices corresponding to 3D compressible Euler und Navier-Stokes equations.

3.3 Boundary Conditions

3.3.1 Partition of $\partial\Omega$

In the following subsection we specify boundary conditions which will be used in the numerical analysis of compressible Navier-Stokes equations performed in this dissertation. First, we split boundary of the computational domain into several parts. For this purpose, we introduce the following physical quantities:

Speed of sound

(3.30) $$c = \sqrt{\kappa\frac{p}{\varrho}}$$

is the speed of acoustic waves.

Mach number

$$M = \frac{|\boldsymbol{v}|}{c} \tag{3.31}$$

is ratio of speed of fluid to the speed of sound which has in mechanics of compressible flows special importance.

This dissertation is devoted mainly to flows through channels, past airfoils and to driven cavity problems. Therefore, we assume that boundary of the computational domain consists only of impermeable walls and parts where the fluid is coming in (inlet) or leaving the domain (outlet). We use the following terminology: *supersonic inlet* is a part of $\partial\Omega$ where the fluid is coming in at Mach number greater than one, *subsonic inlet* is a part of $\partial\Omega$ where the fluid is coming in at Mach number lower than one, *subsonic outlet* is a part of $\partial\Omega$ where the fluid is leaving the domain at Mach number lower than one and finally, *supersonic outlet* is a part of $\partial\Omega$ where the fluid is leaving the domain at Mach number higher than one. We use the no-slip boundary conditions $\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{v}_{wall}(\boldsymbol{x}, t)$ at impermeable walls moving with the velocity $\boldsymbol{v}_{wall}$.

Two-dimensional Navier-Stokes equations consist of four partial differential equations (3.5) – (3.7). Equation of continuity is a first order hyperbolic equation. Remaining three equations – law of conservation of momentum and law of conservation of energy – are second order parabolic equations. Therefore compressible Navier-Stokes equations have the form of an *incompletely parabolic system* in the sense of Strikwerda (see [70]). Results of Strickwerda's analysis of initial boundary value problems for general incompletely parabolic systems have been applied also to the compressible Navier-Stokes equations. As the mentioned analysis is relatively extensive, we confine ourselves only to cite results which will be used in this dissertation.

Except for supersonic inlet, Navier-Stokes equations require more boundary conditions than Euler equations (see Table 3.1).

2D case	Euler equations	Navier-Stokes equations
Subsonic inlet	3	4
Supersonic inlet	4	4
Subsonic outlet	1	3
Supersonic outlet	0	3
Impermeable wall	1	3

Table 3.1: Number of required boundary conditions for 2D Euler and Navier-Stokes equations

See Appendix **3.9.3** for number of required boundary conditions for 3D compressible Euler and Navier-Stokes equations.

Important results concerning the construction of suitable boundary conditions were obtained by Gustafsson and Sundström [33], Oliger and Sundström [55], Nordström [52] and Poinsot and Lele [58].

Let the boundary $\partial\Omega$ be divided into several disjoint parts $\partial\Omega^{Isup}$, $\partial\Omega^{Isub}$, $\partial\Omega^{Osup}$, $\partial\Omega^{Osub}$, $\partial\Omega^{Wadi}$, $\partial\Omega^{Wiso}$, representing supersonic inlet, subsonic inlet, supersonic outlet, subsonic outlet, impermeable adiabatic wall and impermeable isothermal wall, respectively. so that

$$\partial\Omega = \overline{\partial\Omega^{Isup}} \cup \overline{\partial\Omega^{Isub}} \cup \overline{\partial\Omega^{Osup}} \cup \overline{\partial\Omega^{Osub}} \cup \overline{\partial\Omega^{Wadi}} \cup \overline{\partial\Omega^{Wiso}}. \tag{3.32}$$

Let us remark that this partition is in practical application for non-stationary problems determined by the initial state of the flow. We assume that different boundary parts can touch only at points $X_1, X_2, \dots X_{N_x}$ introduced in paragraph **3.2.2**.

In the following paragraph we give a few examples of admissible boundary conditions. For simplicity, we restrict ourselves to inlets and outlets perpendicular to the x-axis. Impermeable walls can have arbitrary direction. The boundary conditions can be chosen as follows:

3.3.2 Supersonic Inlet

One prescribes four boundary conditions, all “inviscid”. It is possible to apply Dirichlet boundary conditions

$$\varrho = g_1, \tag{3.33}$$

$$v_1 = g_2, \tag{3.34}$$

$$v_2 = g_3 \tag{3.35}$$

and

$$p = g_4 \tag{3.36}$$

specifying also the inlet temperature θ. Another possibility is (see [58]) to prescribe

$$\theta = g_4. \tag{3.37}$$

instead of (3.36). Then, obviously, the inlet pressure p is determined.

3.3.3 Subsonic Inlet

One prescribes four boundary conditions, three "inviscid" and one "viscous". Gustaffson and Sundström [33] proposed the following combination: Dirichlet boundary conditions

$$\varrho = g_1, \tag{3.38}$$

$$v_1 = g_2, \tag{3.39}$$

$$v_2 = g_3 \tag{3.40}$$

and Neumann boundary condition

$$\frac{\partial \theta}{\partial x_1} = g_4 \tag{3.41}$$

for inlets perpendicular to the x-axis.

A non-standard approach prescribing three Dirichlet conditions

$$v_1 = g_1, \tag{3.42}$$

$$v_2 = g_2, \tag{3.43}$$

$$\theta = g_3 \tag{3.44}$$

was tested by Poinsot and Lele (see [58]) with satisfactory results but we don't use this case in this dissertation.

3.3.4 Supersonic Outlet

One prescribes three “viscous” boundary conditions. Gustafsson and Sundström [33] and Oliger and Sundström [55] proposed to prescribe Neumann boundary conditions

$$\frac{\partial v_1}{\partial x_1} = g_1, \tag{3.45}$$

$$\frac{\partial v_2}{\partial x_1} = g_2 \tag{3.46}$$

and

$$\frac{\partial \theta}{\partial x_1} = g_3. \tag{3.47}$$

3.3.5 Subsonic Outlet

One prescribes three boundary conditions, one ”inviscid” and two “viscous”. It is possible (see [58]) to prescribe

$$p = g_1 \tag{3.48}$$

and Neumann boundary conditions

$$\frac{\partial v_2}{\partial x_1} = g_2 \tag{3.49}$$

and

$$\frac{\partial \theta}{\partial x_1} = g_3. \tag{3.50}$$

We have also good experience with the Dirichlet boundary condition

$$e = g_1. \tag{3.51}$$

instead of (3.48). Pair of Neumann boundary conditions

$$\frac{\partial v_1}{\partial x_1} = g_2 \tag{3.52}$$

and

$$\frac{\partial v_2}{\partial x_1} = g_3. \tag{3.53}$$

can be prescribed instead of (3.49)(3.50).

Impermeable adiabatic wall:

We consider a wall moving with the velocity $\boldsymbol{v}_{wall} = (g_1, g_2)^T$. One prescribes three boundary conditions. It is possible to apply Dirichlet (no-slip) boundary conditions

$$v_1 = g_1, \tag{3.54}$$

$$v_2 = g_2 \tag{3.55}$$

and (adiabatic) Neumann boundary condition

$$\frac{\partial \theta}{\partial \nu} = g_3. \tag{3.56}$$

Impermeable isothermal wall:

In this case, the boundary conditions (3.54)(3.55) are used and the Neumann condition (3.56) is replaced with a Dirichlet condition

$$\theta = g_3 \tag{3.57}$$

characterizing an isothermal character of the impermeable wall.

3.4 Weak Formulation

3.4.1 Function Spaces

Let us introduce the following spaces of functions:

$$\boldsymbol{C}^\infty(\Omega) = \left[C^\infty(\overline{\Omega})\right]^4, \tag{3.58}$$

$$\boldsymbol{L}^\infty(\Omega) = \left[L^\infty(\overline{\Omega})\right]^4, \tag{3.59}$$

$$\boldsymbol{U} = \overline{\boldsymbol{C}^\infty(\Omega)}^{W^{2,2}(\Omega)} \cap \boldsymbol{L}^\infty(\Omega), \tag{3.60}$$

Test functions will be taken in a space $\boldsymbol{V}$, the form of which depends on combination of prescribed Dirichlet (essential) boundary conditions. Let us define the space $\boldsymbol{V}_0$ as

$$(3.61)\quad \boldsymbol{V}_0 = \{\boldsymbol{\varphi} \in \boldsymbol{C}^\infty(\Omega); \varphi_{1,2,3,4} = 0 \text{ on } \partial\Omega^{Isup}, \varphi_{1,2,3} = 0 \text{ on } \partial\Omega^{Isub},$$

$$\varphi_4 = 0 \text{ on } \partial\Omega^{Osub},\ \varphi_{2,3} = 0 \text{ on } \partial\Omega^{Wadi},\ \varphi_{2,3,4} = 0 \text{ on } \partial\Omega^{Wiso}\}.$$

Then, space $\boldsymbol{V}$ is defined as

$$\boldsymbol{V} = \overline{\boldsymbol{V}_0}^{W^{1,2}(\Omega)} \cap \boldsymbol{L}^\infty(\Omega). \tag{3.62}$$

Neumann (natural) boundary conditions do not influence form of the space $\boldsymbol{V}$. They will be incorporated into boundary integrals in the weak formulation of the compressible Navier-Stokes equations.

3.4.2 Integral Forms

Multiplying (3.14) by a test function $\boldsymbol{\varphi} \in \boldsymbol{V}$, integrating over Ω and using Green's theorem we obtain the scalar identity

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{w}(t), \boldsymbol{\varphi}) + f(\boldsymbol{w}(t), \boldsymbol{\varphi}) = -R(\boldsymbol{w}(t), \nabla\boldsymbol{\varphi}) + R^B(\boldsymbol{w}(t), \boldsymbol{\varphi}) \tag{3.63}$$

where $\boldsymbol{w}(t) : \Omega \to \boldsymbol{R}^4$ for all $t \in (0, T)$, $\boldsymbol{w}(t)(\boldsymbol{x}) = \boldsymbol{w}(\boldsymbol{x}, t)$ for all $\boldsymbol{x} \in \Omega$ and $t \in (0, T)$. We define scalar product

$$(\boldsymbol{w}, \boldsymbol{\varphi}) = \int_\Omega \boldsymbol{w}(\boldsymbol{x}) \cdot \boldsymbol{\varphi}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{3.64}$$

in $[L^2(\Omega)]^4 \times [L^2(\Omega)]^4$ and integral forms

$$f(\boldsymbol{w}, \boldsymbol{\varphi}) = \int_\Omega \sum_{k=1}^{2} \frac{\partial}{\partial x_k} \boldsymbol{f}_k(\boldsymbol{w}(\boldsymbol{x})) \cdot \boldsymbol{\varphi}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{3.65}$$

$$R(\boldsymbol{w}, \nabla\boldsymbol{\varphi}) = \int_\Omega \sum_{k=1}^{2} \boldsymbol{R}_k(\boldsymbol{w}(\boldsymbol{x})) \cdot \frac{\partial}{\partial x_k} \boldsymbol{\varphi}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{3.66}$$

in $L^2(\Omega \to \mathcal{Q}) \times [L^2(\Omega)]^4$. Let us remark that forms f, R and R^B are linear only in the second variable.

In the following paragraph we define the form R^B which depends on prescribed Neumann boundary conditions. For simplicity, **zero Neumann boundary conditions** will be considered.

3.4.3 Interpretation of Neumann Boundary Conditions

The form R^B, which is based on integration along the boundary $\partial\Omega$, will be defined as a sum

$$R^B = R^{Isup} + R^{Isub} + R^{Osup} + R^{Osub} + R^{Wadi} + R^{Wiso} \tag{3.67}$$

corresponding to the partition of $\partial\Omega$ constructed in **3.3**.

Form R^{Isup}:

It is

$$R^{Isup} = 0 \tag{3.68}$$

because all components of the test function $\boldsymbol{\varphi}$ are zero on the part $\partial\Omega^{Isup}$ of $\partial\Omega$ (see definition (3.61)).

Form R^{Isub}:

With the definition (3.61) of the space $\boldsymbol{V}_0$, the first three components of the test function $\boldsymbol{\varphi}$ are zero on the part $\partial\Omega^{Isub}$ of $\partial\Omega$. With subsonic inlet boundary conditions (3.38) – (3.40), (3.41), one obtains

$$R^{Isub}(\boldsymbol{w},\boldsymbol{\varphi}) = \int_{\partial\Omega^{Isub}} (\boldsymbol{R}_1(\boldsymbol{w})\cdot\boldsymbol{\varphi})\,\nu_1 \mathrm{dS} = \tag{3.69}$$

$$= \int_{\partial\Omega} \left\{ \lambda \left[\frac{\partial}{\partial x_1}\left(\frac{w_2}{w_1}\right) + \frac{\partial}{\partial x_2}\left(\frac{w_3}{w_1}\right)\right] + 2\mu \frac{\partial}{\partial x_1}\left(\frac{w_2}{w_1}\right)\right\} \frac{w_2}{w_1}\nu_1\varphi_4 +$$

$$+\mu\left[\frac{\partial}{\partial x_2}\left(\frac{w_2}{w_1}\right) + \frac{\partial}{\partial x_1}\left(\frac{w_3}{w_1}\right)\right]\frac{w_3}{w_1}\nu_1\varphi_4 \mathrm{dS}$$

for inlets perpendicular to the x-axis.

Form R^{Osup}:

Since no Dirichlet boundary condition is imposed at supersonic outlet, viscous parts of equations of momentum and of equation for energy will contribute to R^{Osup}. With supersonic outlet boundary conditions (3.45) – (3.47) one obtains

$$R^{Osup}(\boldsymbol{w},\boldsymbol{\varphi}) = \int_{\partial\Omega^{Osup}} (\boldsymbol{R}_1(\boldsymbol{w})\cdot\boldsymbol{\varphi})\,\nu_1 \mathrm{dS} = \tag{3.70}$$

$$= \int_{\partial\Omega} \lambda \frac{\partial}{\partial x_2}\left(\frac{w_3}{w_1}\right)\nu_1\varphi_2 + \mu\frac{\partial}{\partial x_2}\left(\frac{w_2}{w_1}\right)\nu_1\varphi_3 +$$

$$+\left[\lambda\frac{\partial}{\partial x_2}\left(\frac{w_3}{w_1}\right)\frac{w_2}{w_1}+\mu\frac{\partial}{\partial x_2}\left(\frac{w_2}{w_1}\right)\frac{w_3}{w_1}\right]\nu_1\varphi_4\mathrm{dS}$$

for outlets perpendicular to the x-axis.

Form R^{Osub}:

In this case only viscous parts of momentum equations play a role. With subsonic outlet boundary conditions (3.51) – (3.53), one obtains

$$R^{Osub}(\boldsymbol{w},\boldsymbol{\varphi})=\int_{\partial\Omega^{Osub}}\left(\boldsymbol{R}_1(\boldsymbol{w})\cdot\boldsymbol{\varphi}\right)\nu_1\mathrm{dS}= \tag{3.71}$$

$$=\int_{\partial\Omega}\lambda\frac{\partial}{\partial x_2}\left(\frac{w_3}{w_1}\right)\nu_1\varphi_2+\mu\frac{\partial}{\partial x_2}\left(\frac{w_2}{w_1}\right)\nu_1\varphi_3\mathrm{dS}$$

for outlets perpendicular to the x-axis.

Form R^{Wadi}:

For simplicity, let us consider an impermeable adiabatic wall which is not moving. In case of boundary conditions (3.54),(3.55),(3.56) we easily compute that

$$R^{Wadi}=0 \tag{3.72}$$

independently of the direction of the impermeable adiabatic wall.

Form R^{Wiso}:

Again, for simplicity let us consider an impermeable isothermal wall which is not moving. In case of boundary conditions (3.54),(3.55),(3.57) we again easily compute that

$$R^{Wiso}=0 \tag{3.73}$$

independently of the direction of the impermeable isothermal wall.

3.4.4 Formulation of the Weak Problem

Let Ω be a bounded domain with Lipschitz continuous, piecewise polynomial boundary introduced in paragraph **3.2.2**.

Let the boundary $\partial\Omega$ be divided into several disjoint parts $\partial\Omega^{Isup}$, $\partial\Omega^{Isub}$, $\partial\Omega^{Osup}$, $\partial\Omega^{Osub}$, $\partial\Omega^{Wadi}$ and $\partial\Omega^{Wiso}$ representing supersonic inlet, subsonic inlet, supersonic outlet, subsonic outlet, impermeable adiabatic wall and impermeable isothermal wall, respectively, as described in paragraph **3.3.1**. In order to allow several different inlet and outlet boundary conditions, these open boundary parts are further splitted into a finite number of open arcs

$$
\begin{aligned}
\overline{\partial\Omega^{Isup}} &= \bigcup_{i=1}^{N_{Isup}} \overline{\Gamma_i^{Isup}}, \\
\overline{\partial\Omega^{Isub}} &= \bigcup_{i=1}^{N_{Isub}} \overline{\Gamma_i^{Isub}}, \\
\overline{\partial\Omega^{Osup}} &= \bigcup_{i=1}^{N_{Osup}} \overline{\Gamma_i^{Osup}}, \\
\overline{\partial\Omega^{Osub}} &= \bigcup_{i=1}^{N_{Osub}} \overline{\Gamma_i^{Osub}}, \\
\overline{\partial\Omega^{Wadi}} &= \bigcup_{i=1}^{N_{Wadi}} \overline{\Gamma_i^{Wadi}}. \\
\overline{\partial\Omega^{Wiso}} &= \bigcup_{i=1}^{N_{Wiso}} \overline{\Gamma_i^{Wiso}}.
\end{aligned} \tag{3.74}
$$

We prescribe the following boundary conditions:

- (3.33)(3.34)(3.35)(3.36) or (3.33)(3.34)(3.35)(3.37) for all Γ_i^{Isup}, $i = 1, 2, \ldots N_{Isup}$,
- (3.38)(3.39)(3.40)(3.41) for all Γ_i^{Isub}, $i = 1, 2, \ldots N_{Isub}$,
- (3.45)(3.46)(3.47) for all Γ_i^{Osup}, $i = 1, 2, \ldots N_{Osup}$,
- (3.48)(3.49)(3.50), (3.51)(3.49)(3.50), (3.48)(3.52)(3.53) or (3.51)(3.52)(3.53) for all Γ_i^{Osub}, $i = 1, 2, \ldots N_{Osub}$,
- (3.54)(3.55)(3.56) for all Γ_i^{Wadi}, $i = 1, 2, \ldots N_{Wadi}$,
- (3.54)(3.55)(3.57) for all Γ_i^{Wadi}, $i = 1, 2, \ldots N_{Wadi}$.

We remark that it is necessary to ensure compatibility of prescribed Dirichlet boundary conditions (i.e. continuity of corresponding functions) at points where different boundary arc touch. We assume that different boundary arcs touch only at points $X_1, X_2, \ldots, X_{N_x}$ introduced in paragraph **3.2.2**.

Let us consider spaces of functions which have been defined in paragraph **3.4.1**.

Consider an initial data $\boldsymbol{w}^0 \in \boldsymbol{U}$ satisfying the above chosen combinations of Dirichlet and Neumann boundary conditions which are used for supersonic inlet, subsonic inlet, supersonic outlet, subsonic outlet, impermeable adiabatic wall and impermeable isothermal wall, respectively. Moreover, type of flow represented by $\boldsymbol{w}^0$ must match the prescribed partition of inlet and outlet boundary parts to supersonic inlet, subsonic inlet, supersonic outlet and subsonic outlet, respectively (e.g. along $\partial\Omega^{Isup}$ $\boldsymbol{w}^0$ must represent flow coming into the domain with the Mach number $M > 1$ etc.).

Further, let $\boldsymbol{w}^* \in \boldsymbol{U}$ be a function satisfying the corresponding Dirichlet boundary conditions (often we set $\boldsymbol{w}^* \equiv \boldsymbol{w}^0$).

It is our aim to find a function $\boldsymbol{w}(\boldsymbol{x}, t) \in \boldsymbol{U}$ such that

1. $\boldsymbol{w} - \boldsymbol{w}^* \in \boldsymbol{V}$,
2. $\boldsymbol{w}(\boldsymbol{x}, 0) = \boldsymbol{w}^0(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$,
3. $\boldsymbol{w}$ satisfies the integral identity

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{w}, \boldsymbol{\varphi}) = -f(\boldsymbol{w}, \boldsymbol{\varphi}) - R(\boldsymbol{w}, \nabla\boldsymbol{\varphi}) + R^B(\boldsymbol{w}, \boldsymbol{\varphi}) \tag{3.75}$$

for all $\boldsymbol{\varphi} \in \boldsymbol{V}$ and all $t \in (0, T)$. The forms f, R and R^B have been defined in paragraphs **3.4.2**, **3.4.3**.

3.5 Semi-Discrete Problem

3.5.1 Finite Element Grid

By $\mathcal{T}_h$ we denote a triangulation of the domain Ω with the following properties:

(i) Triangulation $\mathcal{T}_h$ consists of closed linear or curvilinear triangles T_i, $i = 1, 2, \ldots, M$,

(ii) $\Omega = \bigcup_{i=1}^{M} T_i$,

(iii) if $T_i, T_j \in \mathcal{T}_h$, $i \neq j$ then $T_i \cup T_j = \emptyset$ or T_i and T_j have exactly one common vertex or T_i and T_j have exactly one common side.

Without loss of generality, we assume that the set of grid points lying on the boundary $\partial\Omega$ is identical with the set consisting of points $X_1, X_2, \ldots, X_{N_x}$ introduced in paragraph **3.2.2**. Further we assume that internal grid sides are linear, i.e. only grid triangles having a side on $\partial\Omega$ can be non-linear.

The index h is defined as

$$h = \max_{i=1,\ldots,M} \mathrm{diam}(T_i). \tag{3.76}$$

3.5.2 Polynomial Element Maps

Let us define a reference triangle $\hat{T}$ as shown in Figure 3.2.

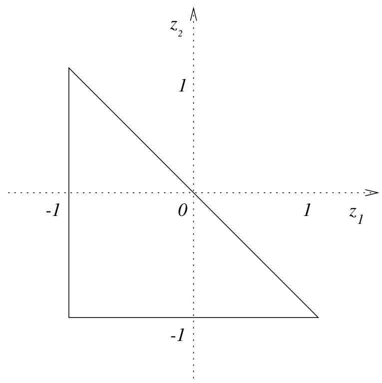

Figure 3.2: Reference triangle $\hat{T}$

In this paragraph we construct for each grid triangle $T_i \in \mathcal{T}_h$ a polynomial vector function $\boldsymbol{F}_{T_i} : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\boldsymbol{F}_{T_i}(\hat{T}) = T_i. \tag{3.77}$$

Let T_i have polynomial sides of the maximal degree p_i. The function $\boldsymbol{F}_{T_i}$ mapping the reference triangle $\hat{T}$ onto T_i is sought in the form

$$\boldsymbol{F}_{T_i}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} a_{N_{p_i}} z_1^{p_i} + a_{N_{p_i}-1} z_1^{p_i-1} z_2 + a_{N_{p_i}-2} z_1^{p_i-2} z_2^2 + \ldots + a_1 \\ b_{N_{p_i}} z_1^{p_i} + b_{N_{p_i}-1} z_1^{p_i-1} z_2 + b_{N_{p_i}-2} z_1^{p_i-2} z_2^2 + \ldots + b_1 \end{pmatrix} \tag{3.78}$$

where $N_{p_i} = (p_i+1)(p_i+2)/2$. The reference triangle $\hat{T}$ is covered with N_{p_i} characteristic points

$$P_{p_i,j,k} = \left[-1 + (k-1)\frac{2}{p_i},\ -1 + (j-1)\frac{2}{p_i}\right], \tag{3.79}$$

$j = 1, 2, \ldots, p_i + 1$; $k = 1, 2, \ldots, p_i + 2 - j$ as shown in Figures 3.3, 3.4 and 3.5.

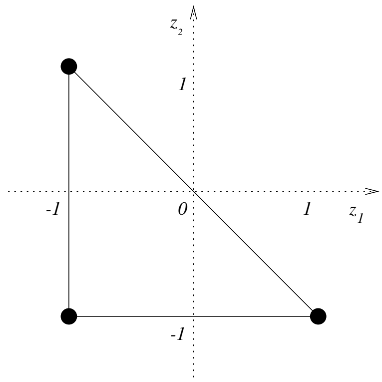

Figure 3.3: Characteristic points on $\hat{T}$, $p_i = 1$

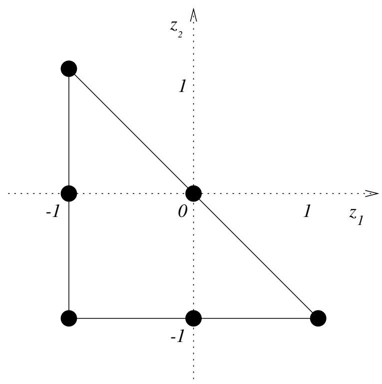

Figure 3.4: Characteristic points on $\hat{T}$, $p_i = 2$

By $Q_{p_i,j,k} \in T_i$ we denote suitably chosen points onto which the reference triangle points $P_{p_i,j,k}$ will be mapped, as shown in Figure 3.6.

Let us now focus on the construction of the points $Q_{p_i,j,k}$. We consider a point $P \in \hat{T}$ and a curvilinear triangle $T \in \mathcal{T}_h$. By A, B, C we denote vertices of T as shown in Figure 3.7. Let us denote by T_{lin} the linear triangle

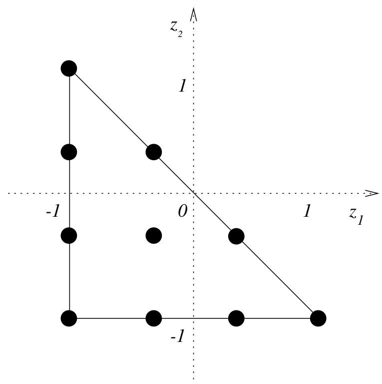

Figure 3.5: Characteristic points on $\hat{T}$, $p_i = 3$

Figure 3.6: Mapping of characteristic points, $p = 2$

specified by points A, B, C. We can easily construct a linear mapping $F_{T_{lin}}$ which maps the reference triangle $\hat{T}$ onto T_{lin}. We denote $Q_{lin} = F_{T_{lin}}(P)$. To the pair of vertices $[A, B]$ we construct points Q_{lin}^{AB} and Q^{AB} (see Figure 3.7) in the following way:

Denote Q_{lin}^{AB} the intersection of the line CQ_{lin} and the side AB of the linear triangle T_{lin}. Denote Q^{AB} the intersection of the line CQ_{lin} and the side AB of the curvilinear triangle T. If the side AB is linear, it is $Q^{AB} = Q_{lin}^{AB}$ (this is true e.g. for all internal grid sides, see paragraph **3.5.1**). If the side AB is non-linear, the intersection can be computed numerically using polynomial functions describing the side introduced in paragraph **3.2.2**. Analogously we construct points Q_{lin}^{BC}, Q^{BC}, Q_{lin}^{CA} and Q^{CA} corresponding to sides BC and CA, respectively. Finally we define

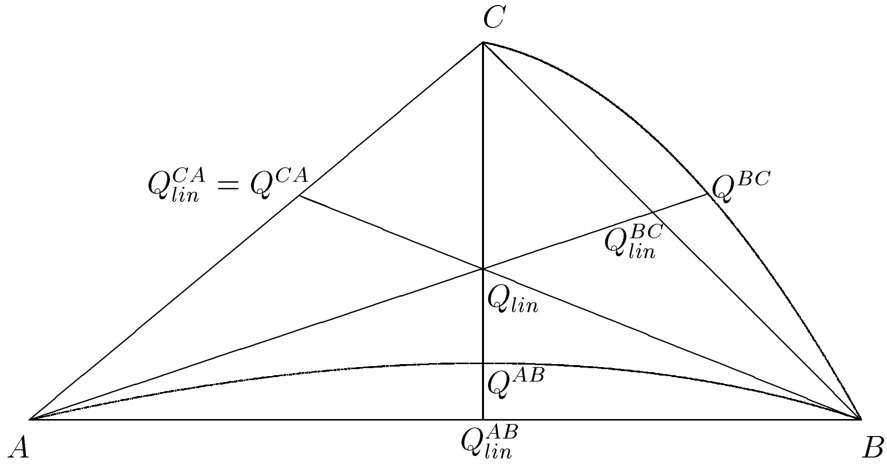

Figure 3.7: Construction of points $Q_{p_i,j,k}$ inside of a grid triangle T

$$Q = Q_{lin} + (C - Q_{lin})\frac{|Q^{AB} - Q_{lin}^{AB}|}{|C - Q_{lin}^{AB}|} + \tag{3.80}$$
$$+(A - Q_{lin})\frac{|Q^{BC} - Q_{lin}^{BC}|}{|A - Q_{lin}^{BC}|} + (B - Q_{lin})\frac{|Q^{CA} - Q_{lin}^{CA}|}{|B - Q_{lin}^{CA}|}.$$

Substituting now $T = T_i$ and $P = P_{p_i,j,k}$, for $i = 1, 2, \ldots M$, $j = 1, 2, \ldots, p_i + 1$; $k = 1, 2, \ldots, p_i + 2 - j$, we obtain desired points $Q_{p_i,j,k}$, $j = 1, 2, \ldots, p_i + 1$; $k = 1, 2, \ldots, p_i + 2 - j$.

In the end, we solve the linear system

$$Q_{p_i,j,k} = \boldsymbol{F}_{T_i}(P_{p_i,j,k}) \qquad j = 1, 2, \ldots, p_i + 1;\ k = 1, 2, \ldots, p_i + 2 - j \tag{3.81}$$

for the unknown coefficients a_l, b_l, $l = 1, 2, \ldots, N_{p_i}$. This approach to the construction of the polynomial map $\boldsymbol{F}_{T_i}$ is advantageous from the point of view of computer implementation – the map itself and its Jacobian are easily storable and evaluable. In Figures 3.8, 3.9 and 3.10 we show an example of approximation of a circular obstacle by polynomials of various degrees.

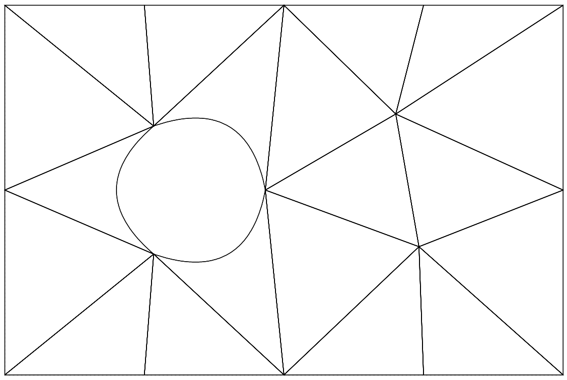

Figure 3.8: Approximation of a circular obstacle, $p = 2$

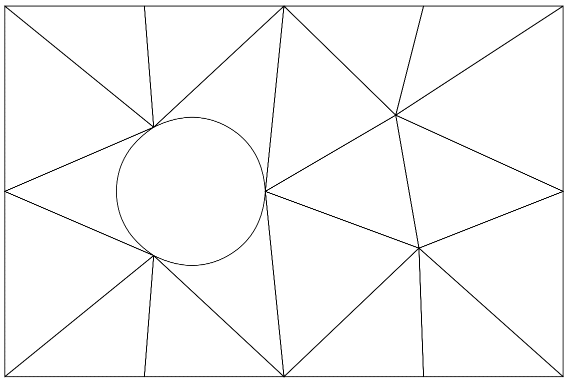

Figure 3.9: Approximation of a circular obstacle, $p = 4$

From Figures 3.8 – 3.10 one can see the improvement of the quality of the approximation of the circle with increasing p.

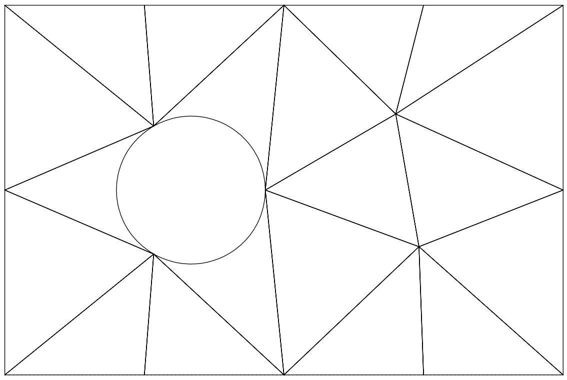

Figure 3.10: Approximation of a circular obstacle, $p = 6$

3.5.3 Approximation of Function Spaces

In this paragraph we will approximate the function spaces $\boldsymbol{U}$, $\boldsymbol{V}$ defined in subsection **3.4.1** by spaces of piecewise polynomial functions.

For this purpose, we use polynomial mappings $\boldsymbol{F}_{T_i}$, $i = 1, 2, \ldots, M$ constructed in the previous paragraph. For each finite element $T_i \in \mathcal{T}_h$, we denote by $p_i \geq 1$ its polynomial degree. Further we define a multi-index $\boldsymbol{p} = \{p_1, p_2, \ldots, p_M\}$ characterizing distribution of polynomial degree on grid triangles $T_i \in \mathcal{T}_h$. Let us denote by $P_k(\hat{T})$, $k \geq 1$ space of polynomial functions of the maximal degree k on the reference triangle $\hat{T}$.

Spaces $\boldsymbol{U}$ and $\boldsymbol{V}$ are approximated by spaces of piecewise polynomial functions

(3.82) $\boldsymbol{U}_{h\boldsymbol{p}} = \{\boldsymbol{\varphi} \in \boldsymbol{U};\ \varphi_i \circ F_{T_j}^{-1}|T_j \in P_{p_j}(T_j),\ 1 \leq i \leq 4,\ 1 \leq j \leq M\}$,

(3.83) $\boldsymbol{V}_{h\boldsymbol{p}} = \{\boldsymbol{\varphi} \in \boldsymbol{V};\ \varphi_i \circ F_{T_j}^{-1}|T_j \in P_{p_j}(T_j),\ 1 \leq i \leq 4,\ 1 \leq j \leq M\}$,

respectively.

3.5.4 Streamline Diffusion Stabilization Term

Stabilization of finite element methods for hyperbolic problems is particularly important. Therefore we consider suitable to introduce a detailed construction of the streamline diffusion term which will be used in this dissertation. For the 2D case we refer to [35], [36] and for the 3D case given in Appendix **3.9.4** we refer to [69].

Consider a function $\boldsymbol{w}_{h\boldsymbol{p}} \in \boldsymbol{U}_{h\boldsymbol{p}}$ and the elementwise residual

$$\mathcal{R}(\boldsymbol{w}_{h\boldsymbol{p}}) = \frac{\partial}{\partial t}\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t) + \sum_{k=1}^{2} \frac{\partial}{\partial x_k}\boldsymbol{f}_k(\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t)) - \sum_{k=1}^{2} \frac{\partial}{\partial x_k}\boldsymbol{R}_k(\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t)) \tag{3.84}$$

on each $T_i \in \mathcal{T}_h$, obtained after substitution of the function $\boldsymbol{w}_{h\boldsymbol{p}}$ to the original compressible Navier-Stokes equations (3.14). It is suitable to split this residual into two parts

$$\mathcal{R}(\boldsymbol{w}_{h\boldsymbol{p}}) = \mathcal{R}_t(\boldsymbol{w}_{h\boldsymbol{p}}) + \mathcal{R}_{\boldsymbol{x}}(\boldsymbol{w}_{h\boldsymbol{p}}) \tag{3.85}$$

where

$$\mathcal{R}_t(\boldsymbol{w}_{h\boldsymbol{p}}) = \frac{\partial}{\partial t}\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t), \tag{3.86}$$

$$\mathcal{R}_{\boldsymbol{x}}(\boldsymbol{w}_{h\boldsymbol{p}}) = \sum_{k=1}^{2} \frac{\partial}{\partial x_k}\boldsymbol{f}_k(\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t)) - \sum_{k=1}^{2} \frac{\partial}{\partial x_k}\boldsymbol{R}_k(\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x},t)). \tag{3.87}$$

The streamline diffusion term $\mathcal{S}$ (see e.g. [35], [36], [69]) has the form

$$\mathcal{S}(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h) = \sum_{i=1}^{M} \int_\Omega \left(\sum_{k=1}^{2} \boldsymbol{A}_k^T(\boldsymbol{w}_{h\boldsymbol{p}}) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k} \right) \cdot \Theta_i(\boldsymbol{w}_{h\boldsymbol{p}})\mathcal{R}(\boldsymbol{w}_{h\boldsymbol{p}}). \tag{3.88}$$

where $\boldsymbol{\varphi}_h \in \boldsymbol{V}_{h\boldsymbol{p}}$. For our purposes, we also split it into two parts

$$\mathcal{S}_t(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h) = \sum_{i=1}^{M} \int_\Omega \left(\sum_{k=1}^{2} \boldsymbol{A}_k^T(\boldsymbol{w}_{h\boldsymbol{p}}) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k} \right) \cdot \Theta_i(\boldsymbol{w}_{h\boldsymbol{p}})\mathcal{R}_t(\boldsymbol{w}_{h\boldsymbol{p}}), \tag{3.89}$$

$$\mathcal{S}_{\boldsymbol{x}}(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h) = \sum_{i=1}^{M} \int_\Omega \left(\sum_{k=1}^{2} \boldsymbol{A}_k^T(\boldsymbol{w}_{h\boldsymbol{p}}) \frac{\partial \boldsymbol{\varphi}_h}{\partial x_k} \right) \cdot \Theta_i(\boldsymbol{w}_{h\boldsymbol{p}})\mathcal{R}_{\boldsymbol{x}}(\boldsymbol{w}_{h\boldsymbol{p}}). \tag{3.90}$$

Matrices $\boldsymbol{A}_k$ have been defined in (3.28), (3.29) and the matrix

$$\Theta_i = \frac{1}{2}\,\mathrm{diam}(T_i)\xi \left(\sum_{k=1}^{2} \boldsymbol{A}_k^2 \right)^{-1/2} \tag{3.91}$$

for $i = 1, 2, \ldots, M$. The parameter ξ is a real number from the interval $\langle 0, 1\rangle$ influencing measure of the stabilization. In 2D, the inverse square root in (3.91) can be computed using the spectral decomposition

$$\left(\sum_{k=1}^{2} \boldsymbol{A}_k^2\right)^{-1/2} = \boldsymbol{R}\Lambda\boldsymbol{L} \tag{3.92}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues

$$\begin{aligned}
\lambda_1 &= 1/\sqrt{|\boldsymbol{v}|^2}, \\
\lambda_2 &= 1/\sqrt{|\boldsymbol{v}|^2 + c^2}, \\
\lambda_3 &= 1/\sqrt{|\boldsymbol{v}|^2 + 3c^2/2 - c\tilde{c}/2}, \\
\lambda_4 &= 1/\sqrt{|\boldsymbol{v}|^2 + 3c^2/2 + c\tilde{c}/2}
\end{aligned}$$

(i.e. inverse square roots of eigenvalues of the matrix $\sum_{k=1}^{2} \boldsymbol{A}_k^2$) on its diagonal, $\mathbf{L}$ is the regular matrix containing the left eigenvectors

$$\begin{aligned}
\boldsymbol{l}_1 &= \left(1 - \frac{\tilde{\kappa}|\boldsymbol{v}|^2}{2c^2}, \frac{\tilde{\kappa}v_1}{c^2}, \frac{\tilde{\kappa}v_2}{c^2}, -\frac{\tilde{\kappa}}{c^2}\right)^T, \\
\boldsymbol{l}_2 &= \left(0, -\frac{v_1v_2}{\varrho|\boldsymbol{v}|^2}, \frac{v_1^2}{\varrho|\boldsymbol{v}|^2}, 0\right)^T, \\
\boldsymbol{l}_3 &= \left(\frac{2c|\boldsymbol{v}|^2}{\tilde{c}} + \frac{4\tilde{\kappa}|\boldsymbol{v}|^4}{\tilde{c}^2 + c\tilde{c}}, -\frac{2cv_1}{\tilde{c}} - \frac{8\tilde{\kappa}v_1|\boldsymbol{v}|^2}{\tilde{c}^2 + c\tilde{c}}, -\frac{2cv_2}{\tilde{c}} - \frac{8\tilde{\kappa}v_2|\boldsymbol{v}|^2}{\tilde{c}^2 + c\tilde{c}}, \frac{8\tilde{\kappa}|\boldsymbol{v}|^2}{\tilde{c}^2 + c\tilde{c}}\right)^T, \\
\boldsymbol{l}_4 &= \left(-\frac{2c|\boldsymbol{v}|^2}{\tilde{c}} + \frac{4\tilde{\kappa}|\boldsymbol{v}|^4}{\tilde{c}^2 - c\tilde{c}}, \frac{2cv_1}{\tilde{c}} - \frac{8\tilde{\kappa}v_1|\boldsymbol{v}|^2}{\tilde{c}^2 - c\tilde{c}}, \frac{2cv_2}{\tilde{c}} - \frac{8\tilde{\kappa}v_2|\boldsymbol{v}|^2}{\tilde{c}^2 - c\tilde{c}}, \frac{8\tilde{\kappa}|\boldsymbol{v}|^2}{\tilde{c}^2 - c\tilde{c}}\right)^T
\end{aligned}$$

of the matrix $\sum_{k=1}^{2} \boldsymbol{A}_k^2$ in its rows and $\boldsymbol{R}$ is the regular matrix containing the right eigenvectors

$$\begin{aligned}
\boldsymbol{r}_1 &= \left(1, v_1, v_2, \frac{|\boldsymbol{v}|^2}{2}\right)^T, \\
\boldsymbol{r}_2 &= \left(0, -\frac{\varrho v_2}{v_1}, \varrho, 0\right)^T,
\end{aligned}$$

$$\boldsymbol{r}_3 = \left(\frac{1}{c^2}, \frac{v_1}{c^2} + \frac{4v_1}{c^2 - c\tilde{c}}, \frac{v_2}{c^2} + \frac{4v_2}{c^2 - c\tilde{c}}, \frac{1}{\tilde{\kappa}} + \frac{|\boldsymbol{v}|^2}{2c^2} + \frac{4|\boldsymbol{v}|^2}{c^2 - c\tilde{c}}\right)^T,$$
$$\boldsymbol{r}_4 = \left(\frac{1}{c^2}, \frac{v_1}{c^2} + \frac{4v_1}{c^2 + c\tilde{c}}, \frac{v_2}{c^2} + \frac{4v_2}{c^2 + c\tilde{c}}, \frac{1}{\tilde{\kappa}} + \frac{|\boldsymbol{v}|^2}{2c^2} + \frac{4|\boldsymbol{v}|^2}{c^2 + c\tilde{c}}\right)^T$$

of the matrix $\sum_{k=1}^{2} \boldsymbol{A}_k^2$ in its columns. Obviously $\mathbf{LR} = \boldsymbol{I}$. The symbol $\tilde{\kappa}$ denotes $\kappa - 1$ and

$$\tilde{c} = \sqrt{c^2 + 16|\boldsymbol{v}|^2}. \tag{3.93}$$

All physical quantities $\boldsymbol{v}, \varrho, p, c, \tilde{c}$ used in this paragraph correspond to the function $\boldsymbol{w}_{h\boldsymbol{p}} \in \boldsymbol{U}_{h\boldsymbol{p}}$.

3.5.5 Formulation of the Approximate Weak Problem

In agreement with **3.4.4** let us consider boundary conditions to all open arcs Γ_i^{Isup}, Γ_i^{Isub}, Γ_i^{Osup}, Γ_i^{Osub}, Γ_i^{wall}, respectively.

Let us approximate the initial data $\boldsymbol{w}^0 \in \boldsymbol{U}$ and the function $\boldsymbol{w}^* \in \boldsymbol{U}$ introduced in **3.4.4** by functions $\boldsymbol{w}_{h\boldsymbol{p}}^0 \in \boldsymbol{U}_{h\boldsymbol{p}}$, $\boldsymbol{w}_{h\boldsymbol{p}}^* \in \boldsymbol{U}_{h\boldsymbol{p}}$, respectively.

It is our aim to find a function $\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x}, t) \in \boldsymbol{U}_{h\boldsymbol{p}}$ such that

1. $\boldsymbol{w}_{h\boldsymbol{p}} - \boldsymbol{w}_{h\boldsymbol{p}}^* \in \boldsymbol{V}_{h\boldsymbol{p}}$,
2. $\boldsymbol{w}_{h\boldsymbol{p}}(\boldsymbol{x}, 0) = \boldsymbol{w}_{h\boldsymbol{p}}^0(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$,
3. $\boldsymbol{w}_{h\boldsymbol{p}}$ satisfies the integral identity

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{w}_{h\boldsymbol{p}}, \boldsymbol{\varphi}_h) + \mathcal{S}_t(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h) = \tag{3.94}$$

$$= -f(\boldsymbol{w}_{h\boldsymbol{p}}, \boldsymbol{\varphi}_h) - R(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h) + R^B(\boldsymbol{w}_{h\boldsymbol{p}}, \boldsymbol{\varphi}_h) - \mathcal{S}_{\boldsymbol{x}}(\boldsymbol{w}_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_h)$$

for all $\boldsymbol{\varphi}_h \in \boldsymbol{V}_{h\boldsymbol{p}}$ and all $t \in (0, T)$.

3.6 Higher-Order Space Discretization

In this section we are going to turn the approximate weak problem defined in **3.5.5** into a semi-discrete one with a finite number of time-dependent spatial unknowns, using higher-order polynomial finite elements.

3.6.1 Shape Functions

There is a number of ways how to choose a basis in $\boldsymbol{V}_{h\boldsymbol{p}}$. We will follow the so-called *hierarchical approach* using three types of continuous basis-functions: the so-called *grid nodal functions* related to grid points, *grid side-functions* related to grid edges and *grid bubble-functions* related to grid elements. Usually, one constructs these functions on the refernce triangle $\hat{T}$.

Barycentric coordinates of a point $[z_1, z_2]^T \in \mathbb{R}^2$ with respect to the reference triangle $\hat{T}$ will be denoted $[\lambda_1, \lambda_2, \lambda_3]$ where

$$\begin{aligned} \lambda_1(z_1, z_2) &= -\frac{z_1 + z_2}{2}, \\ \lambda_2(z_1, z_2) &= \frac{z_1 + 1}{2}, \\ \lambda_3(z_1, z_2) &= \frac{z_2 + 1}{2}. \end{aligned} \tag{3.95}$$

Further, we introduce Legendre polynomials

$$P_k(\xi) = \frac{1}{2^k k!} \frac{\mathrm{d}^k}{\mathrm{d}\xi^k} \left(\xi^2 - 1\right)^k, \tag{3.96}$$

$\xi \in [-1, 1]$, $k = 1, 2, \ldots$, which form an orthogonal system in $L^2([-1, 1])$. Especially, it is

$$\int_{-1}^{1} P_0(\xi) P_i(\xi) \, \mathrm{d}\xi = 0 \tag{3.97}$$

for all $i \geq 1$ which allows us to define function

$$\tilde{\phi}_i(x) = \int_{-1}^{x} P_{i-1}(\xi) \, \mathrm{d}\xi \tag{3.98}$$

for all $i \geq 2$ such that

$$\tilde{\phi}_i(-1) = \tilde{\phi}_i(1) = 0 \tag{3.99}$$

for all $i \geq 2$. Thus, we can construct another polynomial function

$$\phi_i(x) = \frac{\tilde{\phi}_i(x)}{(x^2 - 1)} \tag{3.100}$$

$x \in [-1, 1]$, $i \geq 2$ by eliminating roots 1, -1 from the polynomial $\tilde{\phi}_i(x)$. Functions $\phi_i(x)$ will play an important role in construction of reference triangle shape functions.

For a polynomial degree p, the hierarchical basis

$$B_{\hat{T},p} = \{\nu_k\}_{k=1}^{(p+1)(p+2)/2} \tag{3.101}$$

of the space of polynomials of degree at most p on the reference triangle $\hat{T}$ is constructed in the following way:

For $p = 1$ we have a basis with three reference triangle nodal functions

$$B_{\hat{T},1} = B^n_{\hat{T},1} = \{\lambda_1, \lambda_2, \lambda_3\}. \tag{3.102}$$

For $p = 2$ we use all basis functions corresponding to the degree $p = 1$ and add three quadratic functions. We obtain

$$B_{\hat{T},2} = B^n_{\hat{T},1} \cup B^s_{\hat{T},2} \tag{3.103}$$

where

$$B^s_{\hat{T},2} = \{\lambda_1\lambda_2\phi_2(\lambda_1 - \lambda_2),\ \lambda_2\lambda_3\phi_2(\lambda_2 - \lambda_3),\ \lambda_3\lambda_1\phi_2(\lambda_3 - \lambda_1)\}. \tag{3.104}$$

Further, for all polynomial degrees $p \geq 3$ we enrich the basis $B_{\hat{T},p-1}$ with three $p-$th order side-functions $B^s_{\hat{T},p}$ and with $p - 2$ $p-$th order bubble functions $B^b_{\hat{T},p}$. Thus,

$$B_{\hat{T},p} = B_{\hat{T},p-1} \cup B^s_{\hat{T},p} \cup B^b_{\hat{T},p} \tag{3.105}$$

where

$$B^s_{\hat{T},p} = \{\lambda_1\lambda_2\phi_p(\lambda_1 - \lambda_2),\ \lambda_2\lambda_3\phi_p(\lambda_2 - \lambda_3),\ \lambda_3\lambda_1\phi_p(\lambda_3 - \lambda_1)\} \tag{3.106}$$

and

$$B^b_{\hat{T},p} = \{\lambda_1^k\lambda_2^l\lambda_3\}_{k+l=p-1,\ k,l\geq 1} \tag{3.107}$$

3.6.2 Basis Functions

In this section we will construct a hierarchical basis $B_{\boldsymbol{p}}$ of the space $\boldsymbol{V}_{h\boldsymbol{p}}$. First, for a grid edge l we denote by p_l the polynomial degree of the adjacent grid triangle (if $l \subset \partial\Omega$), or the maximal polynomial degree of the two adjacent grid triangles (if $l \not\subset \partial\Omega$). The basis $B_{\boldsymbol{p}}$ consists of three types of polynomials:

- *Grid nodal functions* $\omega_P^{(k)}$, P being a grid point and $1 \leq k \leq 4$ the vector component.
- *Grid side-functions* $\omega_{l,q}^{(k)}$, l being a grid edge, $2 \leq q \leq p_l$ the polynomial degree and $1 \leq k \leq 4$ the vector component. Set of grid side-functions is not empty only if there is a $T_i \in \mathcal{T}_h$ with $p_i \geq 2$.
- *Grid bubble-functions* $\omega_{T_i,q,m}^{(k)}$, $T_i \in \mathcal{T}_h$ being a grid triangle, $3 \leq q \leq p_i$ the polynomial degree, $1 \leq m \leq q-2$ and $1 \leq k \leq 4$ the vector component. Set of grid bubble-functions is not empty only if there is a $T_i \in \mathcal{T}_h$ with $p_i \geq 3$.

Grid nodal functions

First, we will deal with the construction of the grid nodal functions. Consider a grid point P and an index set

$$I_P = \{1 \leq i \leq M,\ P \in T_i\}. \tag{3.108}$$

of all triangles $T_i \in \mathcal{T}_h$ having P in common. For each one of triangles or curvilinear triangles T_i, $i \in I_P$, there is exactly one reference triangle nodal function

$$\psi_{P,i} \in B^n_{\hat{T},1},\ \psi_{P,i}(F_{T_i}^{-1}(P)) = 1. \tag{3.109}$$

For each $i \in I_P$ we define a function

$$\omega_{P,i}(\boldsymbol{x}) = \begin{cases} \psi_{P,i}(\boldsymbol{F}_{T_i}^{-1}(\boldsymbol{x})) & \boldsymbol{x} \in T_i \\ 0 \quad \boldsymbol{x} \notin T_i. \end{cases} \tag{3.110}$$

Further, we define the function

$$\omega_P(\boldsymbol{x}) = \sum_{i \in I_P} \omega_{P,i}(\boldsymbol{x}). \tag{3.111}$$

If P does not lie on the boundary of Ω, we consider four grid nodal functions corresponding to P:

$$\omega_P^{(1)}(\boldsymbol{x}) = (\omega_P(\boldsymbol{x}),\, 0,\, 0,\, 0)^T\,, \tag{3.112}$$

$$\omega_P^{(2)}(\boldsymbol{x}) = (0,\, \omega_P(\boldsymbol{x}),\, 0,\, 0)^T\,, \tag{3.113}$$

$$\omega_P^{(3)}(\boldsymbol{x}) = (0,\, 0,\, \omega_P(\boldsymbol{x}),\, 0)^T\,, \tag{3.114}$$

$$\omega_P^{(4)}(\boldsymbol{x}) = (0,\, 0,\, 0,\, \omega_P(\boldsymbol{x}))^T\,, \tag{3.115}$$

Otherwise, we consider only grid nodal functions $\omega_P^{(k)}$, $1 \le k \le 4$ such that the kth component of the solution $\boldsymbol{w_{hp}}$ is not prescribed at P by a Dirichlet boundary condition.

Grid side-functions

Let us demonstrate construction of a grid side-functions of a degree $p \ge 2$. First, we consider an internal grid edge $l \not\subset \partial\Omega$. There are exactly two triangles $T_i, T_j \in \mathcal{T}_h$, $i < j$, such that $l = T_i \cap T_j$.

It follows from (3.106) that there is exactly one reference element side-function $\psi_{l,i,p} \in B^s_{\hat{T},p}$ such that $\psi_{l,i,p}(\boldsymbol{F}_{T_i}^{-1}(\boldsymbol{x})) = 0$ for all $\boldsymbol{x} \in \partial T_i \setminus l$. Analogously, there is exactly one reference element side-function $\psi_{l,j,p} \in B^s_{\hat{T},p}$ such that $\psi_{l,j,p}(\boldsymbol{F}_{T_j}^{-1}(\boldsymbol{x})) = 0$ for all $\boldsymbol{x} \in \partial T_j \setminus l$.

We define functions

$$\omega_{l,i,p}(\boldsymbol{x}) = \begin{cases} \psi_{l,i,p}(\boldsymbol{F}_{T_i}^{-1}(\boldsymbol{x})) & \boldsymbol{x} \in T_i \\ 0 \quad \boldsymbol{x} \notin T_i \end{cases} \tag{3.116}$$

and

$$\omega_{l,j,p}(\boldsymbol{x}) = \begin{cases} \psi_{l,j,p}(\boldsymbol{F}_{T_j}^{-1}(\boldsymbol{x})) & \boldsymbol{x} \in T_j \\ 0 \quad \boldsymbol{x} \notin T_j. \end{cases} \tag{3.117}$$

Further, we define the function

$$\omega_{l,p}(\boldsymbol{x}) = \omega_{l,i,p}(\boldsymbol{x}) + \omega_{l,j,p}(\boldsymbol{x}) \tag{3.118}$$

We consider four grid side-functions corresponding to the internal edge l:

$$\omega_{l,p}^{(1)}(\boldsymbol{x}) = (\omega_{l,p}(\boldsymbol{x}),\, 0,\, 0,\, 0)^T\,, \tag{3.119}$$

$$\omega_{l,p}^{(2)}(\boldsymbol{x}) = (0,\, \omega_{l,p}(\boldsymbol{x}),\, 0,\, 0)^T\,, \tag{3.120}$$

$$\omega_{l,p}^{(3)}(\boldsymbol{x}) = (0,\, 0,\, \omega_{l,p}(\boldsymbol{x}),\, 0)^T\,, \tag{3.121}$$

$$\omega_{l,p}^{(4)}(\boldsymbol{x}) = (0,\, 0,\, 0,\, \omega_{l,p}(\boldsymbol{x}))^T\,. \tag{3.122}$$

If the edge l is a part of $\partial\Omega$, there is exactly one triangle $T_i \in \mathcal{T}_h$ such that $l = \partial T_i \cap \partial\Omega$. There is exactly one reference element side-function $\psi_{l,i,p} \in B^s_{\hat{T},p}$ such that $\psi_{l,i,p}(\boldsymbol{F}_{T_i}^{-1}(\boldsymbol{x})) = 0$ for all $\boldsymbol{x} \in \partial T_i \setminus l$.

We define the function

$$\omega_{l,i,p}(\boldsymbol{x}) = \begin{cases} \psi_{l,i,p}(\boldsymbol{F}_{T_i}^{-1}(\boldsymbol{x})) & \boldsymbol{x} \in T_i \\ 0 \quad \boldsymbol{x} \notin T_i. \end{cases} \tag{3.123}$$

The function $\omega_{l,p}$ is defined as

$$\omega_{l,p}(\boldsymbol{x}) = \omega_{l,i,p}(\boldsymbol{x}). \tag{3.124}$$

We consider only grid side-functions $\omega_{l,p}^{(k)}$, $1 \le k \le 4$ such that the kth component of the solution $\boldsymbol{w}_{h\boldsymbol{p}}$ is on l not prescribed by a Dirichlet boundary condition.

Grid bubble-functions

Grid bubble-functions of a degree p corresponding to a grid triangle $T_i \in \mathcal{T}_h$ are constructed in the following way: The set $B^b_{\hat{T},p}$ (see (3.107)) has $p - 2$ elements which can be assigned $\psi_{\hat{T},p,1}, \psi_{\hat{T},p,2}, \ldots, \psi_{\hat{T},p,p-2}$. We define the functions

$$\omega_{T_i,p,j}(\boldsymbol{x}) = \begin{cases} \psi_{\hat{T},p,j}(\boldsymbol{F}_T^{-1}(\boldsymbol{x})),\, 1 \le j \le p-2 & \boldsymbol{x} \in T_i, \\ 0 \quad \boldsymbol{x} \notin T_i. \end{cases} \tag{3.125}$$

We consider $4(p-2)$ grid bubble-functions of the degree p corresponding to the triangle T_i:

$$\omega^{(1)}_{T_i,p,j}(\boldsymbol{x}) = (\omega_{T_i,p,j}(\boldsymbol{x}),\, 0,\, 0,\, 0)^T\,, \tag{3.126}$$

$$\omega^{(2)}_{T_i,p,j}(\boldsymbol{x}) = (0,\, \omega_{T_i,p,j}(\boldsymbol{x}),\, 0,\, 0)^T\,, \tag{3.127}$$

$$\omega^{(3)}_{T_i,p,j}(\boldsymbol{x}) = (0,\, 0,\, \omega_{T_i,p,j}(\boldsymbol{x}),\, 0)^T\,, \tag{3.128}$$

$$\omega^{(4)}_{T_i,p,j}(\boldsymbol{x}) = (0,\, 0,\, 0,\, \omega_{T_i,p,j}(\boldsymbol{x}))^T \tag{3.129}$$

where $1 \le j \le p-2$.

3.6.3 Higher-Order Numerical Quadrature

Numerical quadrature over grid triangles $T_i \in \mathcal{T}_h$, $i = 1, 2, \ldots, M$ will be carried out on the reference triangle $\hat{T}$ by means of the Substitution Theorem. First let us introduce numerical quadrature on the reference triangle $\hat{T}$.

Quadrature on the reference triangle $\hat{T}$

There is a wide class of schemes proposed for the higher-order numerical quadrature. For our purpose the so-called *Gauss quadrature* using weighted function values at certain integration points inside of $\hat{T}$ is suitable.

Let us consider an order of accuracy p. We consider basis of the space of polynomials of degree $\le p$ on $\hat{T}$. It can be shown that this basis has $N_p = (p+1)(p+2)/2$ elements $\nu_1, \nu_2, \ldots, \nu_{N_p}$. It is natural to use the reference triangle shape functions $B_{\hat{T},p}$ defined in (3.101). Second, we consider a set of N_p suitably chosen integration points inside of the reference triangle $\hat{T}$. We can use the reference triangle characteristic points $P_{p,i,j}$ introduced in the subsection **3.5.2**. The corresponding weights $W_{p,i,j}$ are resolved from the linear system

$$\sum_{i=1}^{p+1} \sum_{j=1}^{p+2-i} \nu_k(P_{p,i,j}) W_{p,i,j} = \int_{\hat{T}} \nu_k \mathrm{d}\boldsymbol{z} \qquad k = 1, 2, \ldots, N_p \tag{3.130}$$

where ν_k, $k = 1, 2, \ldots, N_p$ are elements of $B_{\hat{T},p}$. With the known weights $W_{p,i,j}$ an arbitrary function ϕ which is continuous on $\overline{\hat{T}}$ can be integrated using the formula

$$I_{\hat{T},p}(\phi) = \sum_{i=1}^{p+1} \sum_{j=1}^{p+2-i} \phi(P_{p,i,j}) W_{p,i,j}. \tag{3.131}$$

Obviously, if the function ϕ is a polynomial of degree p, it holds exactly

$$I_{\hat{T},p}(\phi) = \int_{\hat{T}} \phi \mathrm{d}\boldsymbol{z}. \tag{3.132}$$

Quadrature on grid triangles $T_i \in \mathcal{T}_h$

Let f be a scalar function continuous in $\overline{T_i}$, $T_i \in \mathcal{T}_h$. Since the map $\boldsymbol{F}_{T_i}$ defined in (3.78) is polynomial, we can write

$$\int_{T_i} f \mathrm{d}\boldsymbol{x} = \int_{\hat{T}} f(\boldsymbol{F}_{T_i}) |J_{\boldsymbol{F}_{T_i}}| \mathrm{d}\boldsymbol{z} \tag{3.133}$$

where

$$J_{\boldsymbol{F}_{T_i}}(\boldsymbol{z}) = \frac{\mathrm{D}\boldsymbol{F}_{T_i}}{\mathrm{D}\boldsymbol{z}}. \tag{3.134}$$

Let p_i be polynomial degree chosen for the finite element T_i in subsection **3.5.3**. Using (3.133) we can define approximate integral of the function f over the triangle T_i with the order of accuracy p_i:

$$I_{T_i,p_i}(f) = I_{\hat{T},p_i}(f(\boldsymbol{F}_{T_i}) |J_{\boldsymbol{F}_{T_i}}|) \tag{3.135}$$

One-dimensional Gauss quadrature on the reference edge $\langle -1, 1 \rangle$ which is necessary for evaluation of the form R^B is constructed analogously. Let us give examples of 1D and 2D integration points and corresponding weights for some integration orders:

Order	**Integration points**	**Integration weights**
1	-1	1
	1	1
2	-1	1/3
	0	4/3
	1	1/3
3	-1	1/4
	-1/3	3/4
	1/3	3/4
	1	1/4

Table 3.2: examples of 1D integration points and weights

Order	**Integration points**	**Integration weights**
1	[-1, -1]	2/3
	[1, -1]	2/3
	[-1, 1]	2/3
2	[-1, -1]	0
	[0, -1]	2/3
	[1, -1]	0
	[-1, 0]	2/3
	[0, 0]	2/3
	[-1, 1]	0
3	[-1, -1]	1/15
	[-1/3, -1]	3/20
	[1/3, -1]	3/20
	[1, -1]	1/15
	[-1, -1/3]	3/20
	[-1/3, -1/3]	9/10
	[1/3, -1/3]	3/20
	[-1, 1/3]	3/20
	[-1/3, 1/3]	3/20
	[-1, 1]	1/15

Table 3.3: examples of 2D integration points and weights

3.6.4 Construction of the Non-Linear System

The basis $B_{\boldsymbol{p}}$ of $\boldsymbol{V}_{h\boldsymbol{p}}$ has a finite number N of elements. Let us denote them $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$. We return to the equation (3.94). Let us denote

$$\tilde{\boldsymbol{w}}_{h\boldsymbol{p}} = \boldsymbol{w}_{h\boldsymbol{p}} - \boldsymbol{w}^*_{h\boldsymbol{p}}. \tag{3.136}$$

The function $\tilde{\boldsymbol{w}}_{h\boldsymbol{p}} \in \boldsymbol{V}_{h\boldsymbol{p}}$ can be written as

$$\tilde{\boldsymbol{w}}_{h\boldsymbol{p}}(\boldsymbol{x}, t) = \sum_{j=1}^{N} y_j(t) \boldsymbol{\varphi}_j(\boldsymbol{x}), \tag{3.137}$$

y_i being real time-dependent coefficients. Substituting $\boldsymbol{\varphi}_h := \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$ and (3.137) into (3.94), we obtain a system of N non-linear ordinary differential equations of the form

$$(3.138)\qquad \sum_{j=1}^{N} \dot{y}_j(t)(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) + \sum_{j=1}^{N} \dot{y}_j(t) \left(\sum_{i=1}^{M} \Theta_i^T \sum_{k=1}^{2} \boldsymbol{A}_k^T \frac{\partial \boldsymbol{\varphi}_i}{\partial \boldsymbol{x}_k}, \boldsymbol{\varphi}_j \right) =$$

$$= -f\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^*, \boldsymbol{\varphi}_i \right) - R\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^*, \nabla\boldsymbol{\varphi}_i \right) +$$

$$+R^B\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^*, \boldsymbol{\varphi}_i \right) - \mathcal{S}_{\boldsymbol{x}}\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^*, \nabla\boldsymbol{\varphi}_i \right)$$

where

$$(3.139)\qquad \Theta_i^T = \Theta_i^T\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^* \right),$$

$$(3.140)\qquad \boldsymbol{A}_k^T = \boldsymbol{A}_k^T\left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}_{h\boldsymbol{p}}^* \right)$$

for $i = 1, 2, \ldots, N$. Applying higher order numerical quadrature introduced in the previous paragraph to integral forms defined in **3.4.2**, we obtain the approximate scalar product

$$(3.141)\qquad (\boldsymbol{w}, \boldsymbol{\varphi})_h = \sum_{i=1}^{M} I_{T_i,p_i}(\boldsymbol{w}(\boldsymbol{x}) \cdot \boldsymbol{\varphi}(\boldsymbol{x}))$$

in $[L^2(\Omega)]^4 \times [L^2(\Omega)]^4$ and integral forms

$$(3.142)\qquad f_h(\boldsymbol{w}, \boldsymbol{\varphi}) = \sum_{i=1}^{M} I_{T_i,p_i}(\sum_{k=1}^{2} \frac{\partial}{\partial x_k}\boldsymbol{f}_k(\boldsymbol{w}) \cdot \boldsymbol{\varphi}),$$

$$(3.143)\qquad R_h(\boldsymbol{w}, \nabla\boldsymbol{\varphi}) = \sum_{i=1}^{M} I_{T_i,p_i}(\sum_{k=1}^{2} \boldsymbol{R}_k(\boldsymbol{w}) \cdot \frac{\partial}{\partial x_k}\boldsymbol{\varphi})$$

in $L^2(\Omega \to \mathcal{Q}) \times [L^2(\Omega)]^4$. The remaining forms $\mathcal{S}_{\boldsymbol{x}}$ and R^B are approximated analogously.

The final system containing numerical quadrature has the form

$$(3.144)\qquad \sum_{j=1}^{N} \dot{y}_j(t)(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j)_h + \sum_{j=1}^{N} \dot{y}_j(t) \left(\sum_{i=1}^{M} \Theta_i^T \sum_{k=1}^{2} \boldsymbol{A}_k^T \frac{\partial \boldsymbol{\varphi}_i}{\partial \boldsymbol{x}_k}, \boldsymbol{\varphi}_j \right)_h =$$

$$= -f_h \left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}^*_{h\boldsymbol{p}}, \boldsymbol{\varphi}_i \right) - R_h \left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}^*_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_i \right) +$$

$$+ R_h^B \left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}^*_{h\boldsymbol{p}}, \boldsymbol{\varphi}_i \right) - \mathcal{S}_{\boldsymbol{x},h} \left(\sum_{j=1}^{N} y_j(t)\boldsymbol{\varphi}_j + \boldsymbol{w}^*_{h\boldsymbol{p}}, \nabla\boldsymbol{\varphi}_i \right)$$

and can be written simply as

$$\boldsymbol{A}(t)\dot{\boldsymbol{y}}(t) = \Phi(\boldsymbol{y}(t), t) \tag{3.145}$$

corresponding to (3.94). The matrix $\boldsymbol{A}(t)$ has the form

$$\boldsymbol{A}(t) = \{a_{ij}\}_{i,j=1}^{N}, \tag{3.146}$$

$$a_{ij} = (\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j)_h + \left(\sum_{i=1}^{M} \Theta_i^T \sum_{k=1}^{2} \boldsymbol{A}_k^T \frac{\partial \boldsymbol{\varphi}_i}{\partial \boldsymbol{x}_k}, \boldsymbol{\varphi}_j \right)_h \tag{3.147}$$

and

$$\boldsymbol{y}(t) = (y_1, y_2, \ldots, y_N)^T(t). \tag{3.148}$$

For the initial data $\tilde{\boldsymbol{w}}_{h\boldsymbol{p}}(\boldsymbol{x}, 0) = \tilde{\boldsymbol{w}}^0_{h\boldsymbol{p}}(\boldsymbol{x}) = \boldsymbol{w}^0_{h\boldsymbol{p}}(\boldsymbol{x}) - \boldsymbol{w}^*_{h\boldsymbol{p}}(\boldsymbol{x}) \in \boldsymbol{V}_{h\boldsymbol{p}}$ to the semi-discrete problem there is exactly one set of coefficients y_i^0, $i = 1, 2, \ldots, N$ such that

$$\tilde{\boldsymbol{w}}^0_{h\boldsymbol{p}}(\boldsymbol{x}) = \sum_{i=1}^{N} y_i^0 \boldsymbol{\varphi}_i(\boldsymbol{x}). \tag{3.149}$$

which yields the initial condition

$$\boldsymbol{y}(0) = \boldsymbol{y}^0 \tag{3.150}$$

where

$$\boldsymbol{y}^0 = (y_1^0, y_2^0, \ldots, y_N^0)^T \tag{3.151}$$

for the system (3.145).

3.7 The Non-Linear System

In this section we will discuss existence and uniqueness of solution of the system (3.145), (3.150) and deal with its numerical solution. It will be solved by a method with a higher order of accuracy. For the construction of the matrix $\boldsymbol{A}(t)$ and for the evaluation of the vector function Φ, higher-order numerical quadrature of the Gaussian type will be used. The vector $\dot{\boldsymbol{y}}$ will be resolved by efficient iterative solvers for large sparse linear systems. Finally, time-integration will be carried out by a suitable explicit and implicit higher-order method.

3.7.1 Solvability of the Non-Linear System

Semi-discretizing in space, we have moved from the area of partial differential equations into the area of ordinary differential equations. We will profit from that using the following basic existence and uniqueness theorem for systems of ordinary differential equations (see e.g. [20], [45], [46]):

Theorem 3.7.1:
Let $\omega \subset \mathbb{R}^m$ be an open set, $T > 0$, $(\boldsymbol{y}_0, t_0) \in \omega \times (0,T)$. Let us consider the initial-value problem

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{F}(\boldsymbol{y}(t), t), \ t \in (0,T), \tag{3.152}$$

$$\boldsymbol{y}(t_0) = \boldsymbol{y}_0. \tag{3.153}$$

If $\boldsymbol{F} \in C(\omega \times (0,T), \mathbb{R}^m)$ then there exists at least one maximal solution $\boldsymbol{y}$ (i.e. solution defined on the maximal possible time interval), $\boldsymbol{y}(t_0) = \boldsymbol{y}_0$. In addition, if $\boldsymbol{F}$ satisfies the Lipschitz condition

$$\forall \boldsymbol{y}_1 \in \omega \ \exists \, \delta > 0, K > 0 \ \forall \boldsymbol{y}_2 \in B(\boldsymbol{y}_1, \delta) \ |\boldsymbol{F}(\boldsymbol{y}_2, t) - \boldsymbol{F}(\boldsymbol{y}_1, t)| \le K|\boldsymbol{y}_2 - \boldsymbol{y}_1| \tag{3.154}$$

for all $t \in (0,T)$ then the maximal solution is unique.

The right-hand-side vector function Φ of the system (3.145), (3.150) is defined in an open set $\mathcal{D} \times (0,T)$ where

$$\mathcal{D} = \{\boldsymbol{y} \in \mathbb{R}^N; \ \boldsymbol{w}^*(\boldsymbol{x}) + \sum_{i=1}^{N} y_i \boldsymbol{\varphi}_i(\boldsymbol{x}) \in \mathcal{Q} \text{ for all } \boldsymbol{x} \in \Omega\}, \tag{3.155}$$

$\boldsymbol{\varphi}_i \in B_{\boldsymbol{p}}$, $1 \le i \le N$, and $\mathcal{Q}$ being the set of admissible states defined by (3.16). It is our aim to find the largest possible subset of $\mathcal{D}$ where the function Φ is Lipschitz continuous in the first variable to fulfill assumptions of the above Theorem 3.7.1.

Case without streamline-diffusion stabilization

First let us neglect the streamline-diffusion stabilization terms both on the right- and left-hand side of the system (3.145)(3.150) setting $\xi = 0$ in (3.91).

It is easy to see that the Jacobi matrices $\boldsymbol{A}_1(\boldsymbol{w})$, $\boldsymbol{A}_2(\boldsymbol{w})$ and the viscous fluxes $R_1(\boldsymbol{w})$, $R_2(\boldsymbol{w})$ have continuous all partial derivatives with respect to $w_1, w_2, \ldots, w_4$ in $\mathcal{Q}$, respectively. Therefore also the function Φ is smooth in $\mathcal{D} \times (0, T)$.

The fact that functions $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$ used for the construction of the system (3.145)(3.150) form a basis in $\boldsymbol{V}_{h\boldsymbol{p}}$ implies that the matrix $\boldsymbol{A}(t)$ (which is constant in this case) is regular.

As a consequence, the system (3.145), (3.150) is uniquely solvable while type of the flow through the inlet and outlet parts of the boundary $\partial\Omega$ matches the division of $\partial\Omega$ into $\partial\Omega^{Isup}$, $\partial\Omega^{Isub}$, $\partial\Omega^{Osup}$ and $\partial\Omega^{Osub}$. Otherwise (e.g if at a time t_{crit} flow through an edge belonging to Ω^{Osub} changes its direction and starts to stream into the domain Ω), division of the boundary $\partial\Omega$ must be reset, the problem must be reformulated and from that point on, a completely different system of ODE's is solved.

Contribution of stabilization terms

In the sense of relation (3.147), the streamline-diffusion stabilization term works as a small perturbation of the basis functions $\boldsymbol{\varphi}_i \in \boldsymbol{V}_{h\boldsymbol{p}}$. The matrix $\boldsymbol{A}(t)$ is regular while perturbed functions $\boldsymbol{\varphi}_i$ still form a basis of $\boldsymbol{V}_{h\boldsymbol{p}}$. Measure of the perturbation is given (see (3.91)) by the grid diameter h, the parameter ξ in (3.91) and by the velocity $|\boldsymbol{v}|$ of the flow at integration points. Streamline-diffusion stabilization terms are not defined for integration points where $|\boldsymbol{v}| = 0$ (see definition of the eigenvalue λ_1 used for the construction of the inverse square root matrix in (3.91)). On the other hand, flow at these integration points need not be stabilized (typically, one does not stabilize at integration points where zero velocity is prescribed as a Dirichlet condition – this can occur e.g. at impermeable walls). Formally, this would mean to introduce into the streamline-diffusion terms one more parameter depending on integration point and on the velocity at this integration point, with values e.g. 0 and 1 which would allow us to switch off the stabilization at certain integration points or when velocity approaches zero closer than a prescribed parameter. We decided not to do it in this dissertation because it would influence the form of the stabilization terms considerably and could lead to confusions. Nevertheless, this is the way how streamline-diffusion stabilization is implemented in our software.

Reasoning as above, we come to the conclusion that for a sufficiently fine mesh and a sufficiently small parameter ξ in (3.91) the matrix $\boldsymbol{A}(t)$ remains regular. The reader can verify that under these assumptions also assumptions of the theorem in paragraph **3.7.1** are fulfilled. Unfortunately, we are not yet able to make preciser statements which would e.g. specify required mesh diameter depending on a prescribed parameter ξ, initial and boundary conditions of the flow, geometry of the domain Ω etc.

3.7.2 Evaluation of $\dot{\boldsymbol{y}}$

Evaluation of the derivative $\dot{\boldsymbol{y}}$ from the system $\boldsymbol{A}(t)\dot{\boldsymbol{y}} = \Phi(\boldsymbol{y}, t)$ consists of three tasks:

- to compute the right-hand-side function $\Phi(\boldsymbol{y}, t)$ using a higher-order numerical quadrature,
- to construct the matrix $\boldsymbol{A}(t)$ using a higher-order numerical quadrature,
- to resolve the vector $\dot{\boldsymbol{y}}$ from the linear system $\boldsymbol{A}(t)\dot{\boldsymbol{y}} = \Phi(\boldsymbol{y}, t)$.

The kth row of the matrix $\boldsymbol{A}(t)$ and the kth component of the vector function $\Phi(\boldsymbol{y}, t)$ correspond to the left and right-hand-side of the kth equation of the system (3.145), (3.150), respectively. Therefore it is suitable to use numerical quadrature of the same order for their evaluation. Further it is suitable to choose an appropriate order of integration such that the "non-stabilized" part of the matrix $\boldsymbol{A}(t)$ is computed exactly. Otherwise the matrix due to approximate quadrature can become singular. It means that the minimal order of numerical quadrature for the kth equation in the system (3.145)(3.150) is equal to the maximal degree of the polynomial $\boldsymbol{\varphi}_k \cdot \boldsymbol{\varphi}_j$ such that $(\boldsymbol{\varphi}_k, \boldsymbol{\varphi}_j) \neq 0$,

$j = 1, 2, \ldots, N$.

If we order degrees of freedom according to physical quantities as e.g. first density, second the first component of momentum, third the second component of momentum and last energy, the matrix $\boldsymbol{A}(t)$ becomes block diagonal. Size of the square sparse blocks corresponds to numbers of degrees of freedom corresponding to physical quantities.

3.7.3 Solution of the Linear System

For numerical solution of the linear system, we prefer iterative solvers over the Gauss elimination. The reason is that truncation errors can become large in case of a larger number of equations and that the elimination lasts in general longer than the iterative solution. Further, the matrix $\boldsymbol{A}(t)$ is sparse and using the simple Gauss elimination, it must be stored in the full form which costs a lot of computer memory. We decided to use Fortran implementation of iterative sparse matrix solvers BiCG, BiCGMR, CGS, SCGS, DCGS, QMR, STAB, STAB2, CGM and MR, all of them with or without incomplete LU preconditioning, which have been investigated in the framework of the master thesis [73]. Description of these solvers can also be found in [34]. Among the above mentioned solvers, we have the best experience with preconditioned BiCG method.

3.7.4 Higher-Order Adaptive Time Integration

We have a lot of possibilities to integrate the system (3.145)(3.150) in time. According to our numerical experience we usually use the so-called *Backward difference formula* (BDF), which has been implemented as an implicit multistep higher-order adaptive scheme in the program package ODEPACK (see [39], [13], [14]). There are two important properties of BDF: First, it is unconditionally stable. Second, the global error is essentially just the sum of local, truncation errors. For a detail description of the BDF including the stability and error analysis see e.g. [64].

3.8 Numerical Examples

3.8.1 An Analytical Stationary Example

Very often it is advantageous to test numerical schemes on simple analytically solvable problems.

Motivation

It is our aim to find a set of non-trivial time-independent functions describing physical quantities $\varrho, \boldsymbol{v}, p, \theta$ which would satisfy the compressible Navier-Stokes equations (3.5) – (3.7). One possibility is to choose some of them, substitute into the system (3.5) – (3.7) and to compute the remaining ones so that equations (3.5) – (3.7) hold. For simplicity, we consider $\boldsymbol{g} \equiv \boldsymbol{0}$, $\underline{\boldsymbol{q}} \equiv \boldsymbol{0}$ and the coefficients c_v, μ, ν, k constant. Following this idea, we consider the velocity $\boldsymbol{v}$ in a linear form

$$v_1(\boldsymbol{x}, t) = C_1 x_2, \;\; v_2(\boldsymbol{x}, t) = 0, \;\; C_1 > 0. \tag{3.156}$$

Substituting (3.156) into (3.5) we obtain

$$\varrho(\boldsymbol{x}, t) = \varrho(x_2, t). \tag{3.157}$$

Equation (3.6) implies that

$$\frac{\partial}{\partial x_1} p(\boldsymbol{x}, t) = 0, \frac{\partial}{\partial x_2} p(\boldsymbol{x}, t) = 0, \tag{3.158}$$

which means that

$$p(\boldsymbol{x}, t) = p_0 = \text{ const.} \tag{3.159}$$

Equation (3.9) implies that

$$\theta(\boldsymbol{x}, t) = \theta(x_2, t). \tag{3.160}$$

The energy equation yields

$$\frac{\partial}{\partial x_1}\left[(\varrho c_v \theta + \frac{1}{2}\varrho v_1^2 + p) v_1\right] = \frac{\partial}{\partial x_1}(2\mu \frac{\partial v_1}{\partial x_1} v_1) + k\frac{\partial^2 \theta}{\partial x_1^2} + \frac{\partial}{\partial x_2}(\mu \frac{\partial v_1}{\partial x_2} v_1) + k\frac{\partial^2 \theta}{\partial x_2^2} \tag{3.161}$$

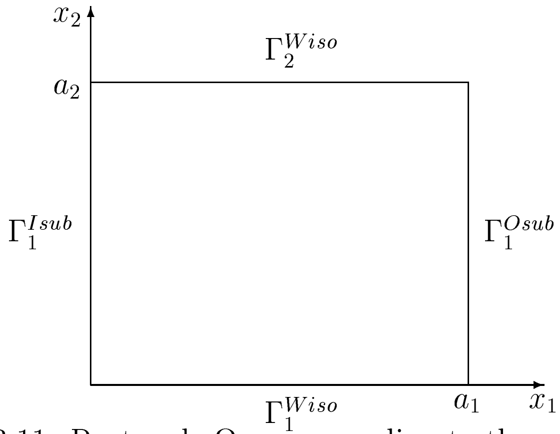

Figure 3.11: Rectangle Ω corresponding to the analytical example

which together with (3.156), (3.157), (3.160) gives

$$\mu C_1^2 + k\frac{\partial^2\theta}{\partial x_2^2} = 0. \tag{3.162}$$

It is not difficult to find a function

$$\theta(\boldsymbol{x}, t) = -\frac{\mu C_1^2}{2k}x_2^2 + C_2 x_2 + C_3 \tag{3.163}$$

satisfying (3.162). Constants C_2 and C_3 must be chosen in such a way that θ is positive in $\Omega \times (0, T)$. One possible choice is

$$C_2 > 0,\ C_3 > \frac{\mu C_1^2 a_2^2}{2k}. \tag{3.164}$$

Density

$$\varrho(\boldsymbol{x}, t) = \frac{p_0}{\left(-\frac{\mu C_1^2}{2k}x_2^2 + C_2 x_2 + C_3\right) R} \tag{3.165}$$

is computed from (3.9) using (3.159) and (3.163). The above considerations lead us to the formulation of the following problem:

Definition of the problem

Consider a rectangle Ω shown in Figure 3.11.

Boundary of the domain Ω is divided into four open arcs Γ_1^{Isub}, Γ_1^{Osub}, Γ_1^{Wiso} and Γ_2^{Wiso} where $\partial\Omega^{Isub} = \Gamma_1^{Isub}$ represents subsonic inlet, $\partial\Omega^{Osub} = \Gamma_1^{Osub}$

represents subsonic outlet and $\partial\Omega^{Wiso} = \Gamma_1^{Wiso} \cup \Gamma_2^{Wiso}$ represents impermeable isothermal wall.

Let the values of λ, μ and c_v correspond to air and the value p_0 to the atmospheric pressure. We prescribe subsonic inlet boundary conditions (3.38)(3.39)(3.40)(3.41) with values

$$
\begin{aligned}
g_1 &= \frac{p_0}{\left(-\frac{\mu C_1^2}{2k} x_2^2 + C_2 x_2 + C_3\right) R}, \\
g_2 &= C_1 x_2, \\
g_3 &= 0, \\
g_4 &= 0
\end{aligned}
\tag{3.166}
$$

at Γ_1^{Isub}. Subsonic outlet boundary conditions (3.48)(3.49)(3.50) with values

$$
\begin{aligned}
g_1 &= p_0, \\
g_2 &= 0, \\
g_3 &= 0
\end{aligned}
\tag{3.167}
$$

or subsonic outlet boundary conditions (3.48)(3.52)(3.53) with values

$$
\begin{aligned}
g_1 &= p_0, \\
g_2 &= 0, \\
g_3 &= 0
\end{aligned}
\tag{3.168}
$$

can be prescribed at Γ_1^{Osub}. At Γ_1^{Wiso} we prescribe isothermal wall boundary conditions (3.54)(3.55)(3.57) with values

$$
\begin{aligned}
g_1 &= 0, \\
g_2 &= 0, \\
g_3 &= C_3
\end{aligned}
\tag{3.169}
$$

and at Γ_2^{Wiso} we prescribe also the isothermal wall boundary conditions (3.54)(3.55)(3.57) with values

$$g_1 = C_1 a_2, \quad g_2 = 0, \quad g_3 = -\frac{\mu C_1^2}{2k} a_2^2 + C_2 a_2 + C_3. \tag{3.170}$$

Initial condition is defined as

$$\varrho(\boldsymbol{x}, 0) = \frac{p_0}{\left(-\frac{\mu C_1^2}{2k} x_2^2 + C_2 x_2 + C_3\right) R} \tag{3.171}$$

$$v_1(\boldsymbol{x}, 0) = C_1 x_2, \ \ v_2(\boldsymbol{x}, 0) = 0, \tag{3.172}$$

$$p(\boldsymbol{x}, 0) = p_0, \tag{3.173}$$

$$\theta(\boldsymbol{x}, 0) = -\frac{\mu C_1^2}{2k} x_2^2 + C_2 x_2 + C_3 \tag{3.174}$$

for all $\boldsymbol{x} \in \Omega$. According to the previous paragraph, this initial condition is at the same time a stationary solution to compressible Navier Stokes equations in the square Ω.

Results of computation

We have chosen a unit square domain Ω and covered it with a triangular grid G_1 shown in Figure 3.12. This grid has been several times sub-divided as shown in Figures 3.13 – 3.17.

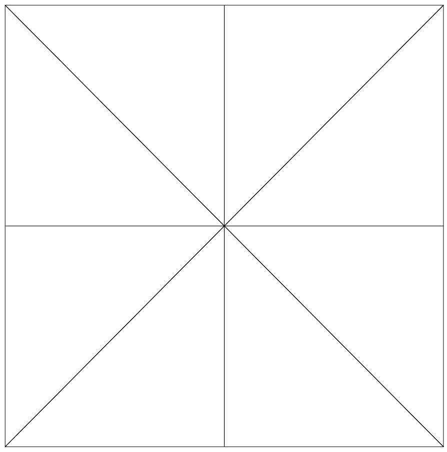

Figure 3.12: Basic grid G_1 with 9 grid points and 8 elements

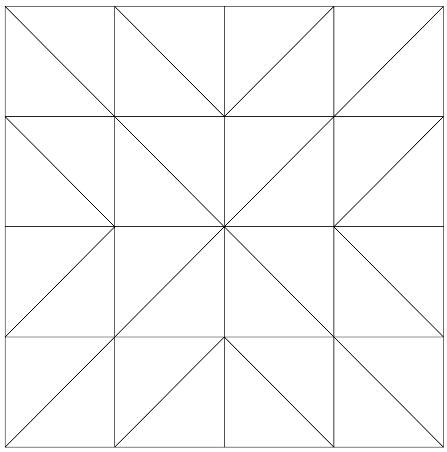

Figure 3.13: Refined grid G_2 with 26 grid points and 32 elements

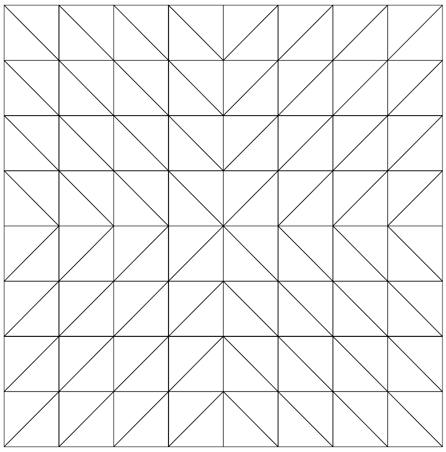

Figure 3.14: Refined grid G_3 with 81 grid points and 128 elements

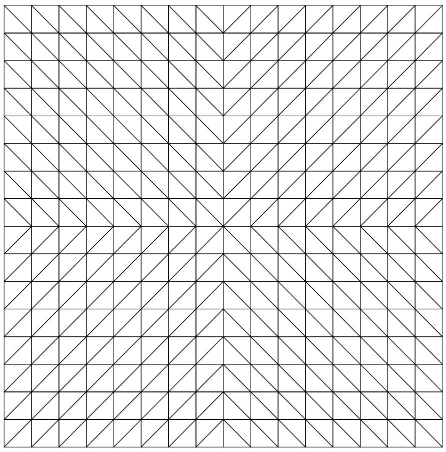

Figure 3.15: Refined grid G_4 with 289 grid points and 512 elements

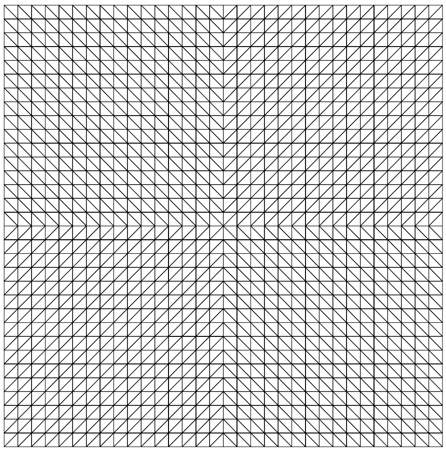

Figure 3.16: Refined grid G_5 with 1089 grid points and 2048 elements

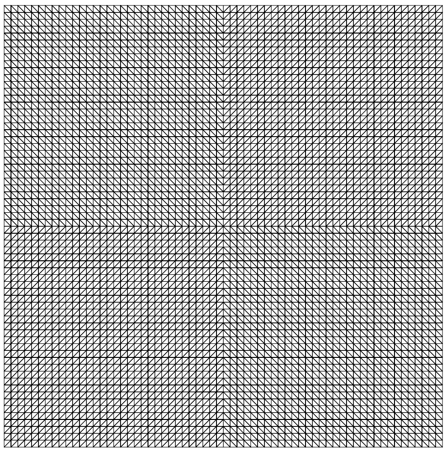

Figure 3.17: Refined grid G_6 with 4225 grid points and 8192 elements

We perform two numerical experiments – in the first one, we use linear finite elements on grids $G_1, G_2, \dots G_6$ and in the latter one, we use the grid G_1 with finite elements of degree $1, 2, \dots, 6$. It is our aim to show that

$$||\dot{\boldsymbol{y}}(0)||_{L^\infty\left(\boldsymbol{R}^{N_{dof}}\right)} \to 0 \tag{3.175}$$

as the grid diameter h goes to zero in the first case and as the polynomial degree of the finite elements increases to infinity in the latter case. The number N_{dof} denotes number of degrees of freedom.

Results of the first computation with linear finite elements are shown in the following Table 3.4.

Grid	**Diameter** h	N_{dof}	$\lVert\dot{\boldsymbol{y}}(0)\rVert_{L^\infty\left(\boldsymbol{R}^{N_{dof}}\right)}$
G_1	$\sqrt{2}/2$	12	15.95411
G_2	$\sqrt{2}/4$	56	8.91001
G_3	$\sqrt{2}/8$	240	5.10656
G_4	$\sqrt{2}/16$	992	1.47997
G_5	$\sqrt{2}/32$	4032	0.4289211
G_6	$\sqrt{2}/64$	16256	0.1243088

Table 3.4: results for linear finite elements, fixed polynomial degree $p = 1$

Results of the second computation with finite elements of higher polynomial degree on the grid G_1 are shown in the Table 3.5.

Grid	**Polynomial degree** p	N_{dof}	$\lVert\dot{\boldsymbol{y}}(0)\rVert_{L^\infty\left(\boldsymbol{R}^{N_{dof}}\right)}$
G_1	1	12	15.95411
G_1	2	56	1.3289046
G_1	3	132	0.1081782
G_1	4	240	0.0355803
G_1	5	380	0.0117025
G_1	6	552	0.003849

Table 3.5: results for polynomial finite elements, fixed diameter $h = \sqrt{2}/2$

3.9 Appendix

3.9.1 3D Form of Compressible Navier-Stokes Equations in Conservation Variables

In three dimensions, equation (3.14) has the form

$$\frac{\partial}{\partial t}\boldsymbol{w}(\boldsymbol{x},t)+\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\boldsymbol{f}_i(\boldsymbol{w}(\boldsymbol{x},t))=\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\boldsymbol{R}_i(\boldsymbol{w}(\boldsymbol{x},t)) \tag{3.176}$$

where

$$\boldsymbol{w}=(w_1,w_2,w_3,w_4,w_5)^T=(\varrho,\varrho v_1,\varrho v_2,\varrho v_3,e)^T. \tag{3.177}$$

Analogously as in 2D, we define the set of admissible states

$$\mathcal{Q}=\left\{\boldsymbol{w}\in\mathbb{R}^5;\ w_1>0,\ w_5>\frac{w_2^2+w_3^2+w_4^2}{2w_1}\right\}. \tag{3.178}$$

The inviscid fluxes $\boldsymbol{f}_i(\boldsymbol{w}):\mathcal{Q}\to\mathbb{R}^5$ and the viscous fluxes $\boldsymbol{R}_i(\boldsymbol{w}):\mathcal{Q}\times\mathbb{R}^5\to\mathbb{R}^5$ have the following form:

$$\boldsymbol{f}_1(\boldsymbol{w})=\begin{pmatrix} w_2 \\ \frac{w_2{}^2}{w_1}+p \\ \frac{w_2w_3}{w_1} \\ \frac{w_2w_4}{w_1} \\ \frac{w_2}{w_1}e \end{pmatrix}, \tag{3.179}$$

$$\boldsymbol{f}_2(\boldsymbol{w})=\begin{pmatrix} w_3 \\ \frac{w_3w_2}{w_1} \\ \frac{w_3{}^2}{w_1}+p \\ \frac{w_3w_4}{w_1} \\ \frac{w_3}{w_1}e \end{pmatrix}, \tag{3.180}$$

$$\boldsymbol{f}_3(\boldsymbol{w})=\begin{pmatrix} w_4 \\ \frac{w_4w_2}{w_1} \\ \frac{w_4w_3}{w_1} \\ \frac{w_4{}^2}{w_1}+p \\ \frac{w_4}{w_1}e \end{pmatrix}, \tag{3.181}$$

$$\boldsymbol{R}_1(\boldsymbol{w}) = \begin{pmatrix} 0 \\ \tau_{11} \\ \tau_{12} \\ \tau_{13} \\ \tau_{11}\frac{w_2}{w_1} + \tau_{12}\frac{w_3}{w_1} + \tau_{13}\frac{w_4}{w_1} + k\frac{\partial\theta}{\partial x_1} \end{pmatrix}, \tag{3.182}$$

$$\boldsymbol{R}_2(\boldsymbol{w}) = \begin{pmatrix} 0 \\ \tau_{21} \\ \tau_{22} \\ \tau_{23} \\ \tau_{21}\frac{w_2}{w_1} + \tau_{22}\frac{w_3}{w_1} + \tau_{32}\frac{w_4}{w_1} + k\frac{\partial\theta}{\partial x_2} \end{pmatrix}, \tag{3.183}$$

$$\boldsymbol{R}_3(\boldsymbol{w}) = \begin{pmatrix} 0 \\ \tau_{31} \\ \tau_{32} \\ \tau_{33} \\ \tau_{31}\frac{w_2}{w_1} + \tau_{32}\frac{w_3}{w_1} + \tau_{33}\frac{w_4}{w_1} + k\frac{\partial\theta}{\partial x_3} \end{pmatrix} \tag{3.184}$$

where

$$\boldsymbol{v} = \left(\frac{w_2}{w_1}, \frac{w_3}{w_1}, \frac{w_4}{w_1}\right)^T, \tag{3.185}$$

$$p = (\kappa - 1)\left(w_5 - \frac{{w_2}^2 + {w_3}^2 + {w_4}^2}{2w_1}\right), \tag{3.186}$$

$$e = \kappa w_5 - (\kappa - 1)\frac{{w_2}^2 + {w_3}^2 + {w_4}^2}{2w_1}, \tag{3.187}$$

$$\boldsymbol{D} = \frac{\nabla\boldsymbol{v} + (\nabla\boldsymbol{v})^T}{2}, \tag{3.188}$$

$$\tau = \lambda\left(\frac{\partial}{\partial x_1}\frac{w_2}{w_1} + \frac{\partial}{\partial x_2}\frac{w_3}{w_1} + \frac{\partial}{\partial x_3}\frac{w_4}{w_1}\right)\boldsymbol{I} + 2\mu\boldsymbol{D}, \tag{3.189}$$

$$\tilde{\theta} = \frac{1}{c_v}\left(\frac{w_5}{w_1} - \frac{w_2^2 + w_3^2 + w_4^2}{2w_1^2}\right). \tag{3.190}$$

3.9.2 Jacobi Matrices of Inviscid Fluxes for 3D Euler and Navier-Stokes Equations

The explicit form of the Jacobi matrices $\boldsymbol{A}_1(\boldsymbol{w}) = \boldsymbol{f}_1^{'}(\boldsymbol{w})$, $\boldsymbol{A}_2(\boldsymbol{w}) = \boldsymbol{f}_2^{'}(\boldsymbol{w})$, $\boldsymbol{A}_3(\boldsymbol{w}) = \boldsymbol{f}_3^{'}(\boldsymbol{w})$ of the inviscid fluxes $\boldsymbol{f}_1(\boldsymbol{w})$, $\boldsymbol{f}_2(\boldsymbol{w})$, $\boldsymbol{f}_3(\boldsymbol{w})$ are

$$\boldsymbol{A}_1(\boldsymbol{w}) = \tag{3.191}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} - \frac{w_2^2}{w_1^2} & (3-\kappa)\frac{w_2}{w_1} & (1-\kappa)\frac{w_3}{w_1} & (1-\kappa)\frac{w_4}{w_1} & \kappa-1 \\ -\frac{w_2w_3}{w_1^2} & \frac{w_3}{w_1} & \frac{w_2}{w_1} & 0 & 0 \\ -\frac{w_2w_4}{w_1^2} & \frac{w_4}{w_1} & 0 & \frac{w_2}{w_1} & 0 \\ \frac{w_2}{w_1}\left(\frac{(\kappa-1)(w_2^2+w_3^2+w_4^2)}{w_1^2} - \kappa\frac{w_5}{w_1}\right) & \kappa\frac{w_5}{w_1} - \frac{(\kappa-1)w_2^2}{w_1^2} - \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} & \frac{(1-\kappa)w_2w_3}{w_1^2} & \frac{(1-\kappa)w_2w_4}{w_1^2} & \kappa\frac{w_2}{w_1} \end{pmatrix},$$

$$\boldsymbol{A}_2(\boldsymbol{w}) = \tag{3.192}$$

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ -\frac{w_2w_3}{w_1^2} & \frac{w_3}{w_1} & \frac{w_2}{w_1} & 0 & 0 \\ \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} - \frac{w_3^2}{w_1^2} & (1-\kappa)\frac{w_2}{w_1} & (3-\kappa)\frac{w_3}{w_1} & (1-\kappa)\frac{w_4}{w_1} & \kappa-1 \\ -\frac{w_3w_4}{w_1^2} & 0 & \frac{w_4}{w_1} & \frac{w_3}{w_1} & 0 \\ \frac{w_3}{w_1}\left(\frac{(\kappa-1)(w_2^2+w_3^2+w_4^2)}{w_1^2} - \kappa\frac{w_5}{w_1}\right) & \frac{(1-\kappa)w_2w_3}{w_1^2} & \kappa\frac{w_5}{w_1} - \frac{(\kappa-1)w_3^2}{w_1^2} - \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} & \frac{(1-\kappa)w_3w_4}{w_1^2} & \kappa\frac{w_3}{w_1} \end{pmatrix},$$

$$\boldsymbol{A}_3(\boldsymbol{w}) = \tag{3.193}$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ -\frac{w_2w_4}{w_1^2} & \frac{w_4}{w_1} & 0 & \frac{w_2}{w_1} & 0 \\ -\frac{w_3w_4}{w_1^2} & 0 & \frac{w_4}{w_1} & \frac{w_3}{w_1} & 0 \\ \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} - \frac{w_4^2}{w_1^2} & (1-\kappa)\frac{w_2}{w_1} & (1-\kappa)\frac{w_3}{w_1} & (3-\kappa)\frac{w_4}{w_1} & \kappa-1 \\ \frac{w_4}{w_1}\left(\frac{(\kappa-1)(w_2^2+w_3^2+w_4^2)}{w_1^2} - \kappa\frac{w_5}{w_1}\right) & \frac{(1-\kappa)w_2w_4}{w_1^2} & \frac{(1-\kappa)w_3w_4}{w_1^2} & \kappa\frac{w_5}{w_1} - \frac{(\kappa-1)w_4^2}{w_1^2} - \frac{\kappa-1}{2}\frac{w_2^2+w_3^2+w_4^2}{w_1^2} & \kappa\frac{w_4}{w_1} \end{pmatrix}.$$

3.9.3 Number of Inlet and Outlet Boundary Conditions for 3D Euler and Navier-Stokes Equations

In the following table we introduce number of required boundary conditions for 3D Euler and Navier-Stokes equation.

3D case	Euler equations	Navier-Stokes equations
Subsonic inlet	4	5
Supersonic inlet	5	5
Subsonic outlet	1	4
Supersonic outlet	0	4
Impermeable wall	1	4

Table 3.6: Number of required boundary conditions for 3D Euler and Navier-Stokes equations

3.9.4 Streamline Diffusion Stabilization Term in 3D

In 3D (see [69]), the inverse square root in (3.91) can be computed using the spectral decomposition

$$\left(\sum_{k=1}^{2} \boldsymbol{A}_k^2\right)^{-1/2} = \boldsymbol{R}\Lambda\boldsymbol{L} \tag{3.194}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues

$$\begin{aligned}
\lambda_1 &= 1/\sqrt{|\boldsymbol{v}|^2}, \\
\lambda_2 &= 1/\sqrt{|\boldsymbol{v}|^2 + c^2}, \\
\lambda_3 &= 1/\sqrt{|\boldsymbol{v}|^2 + c^2}, \\
\lambda_4 &= 1/\sqrt{|\boldsymbol{v}|^2 + c^2(2 - \imath)}, \\
\lambda_5 &= 1/\sqrt{|\boldsymbol{v}|^2 + c^2(2 + \imath)}
\end{aligned}$$

on its diagonal, $\mathbf{L}$ is a regular matrix containing the left eigenvectors

$$\begin{aligned}
\boldsymbol{l}_1 &= \left(H - |\boldsymbol{v}|^2, v_1, v_2, v_3, -1\right)^T, \\
\boldsymbol{l}_2 &= (0, \alpha_1, \alpha_2, \alpha_3, 0)^T, \\
\boldsymbol{l}_3 &= (0, \beta_1, \beta_2, \beta_3, 0)^T, \\
\boldsymbol{l}_4 &= \left(\zeta\left(H - |\boldsymbol{v}|^2\right) - \frac{3+\imath}{2}, \left(\frac{1+\imath}{2|\boldsymbol{v}|^2} + \zeta\right)\boldsymbol{v}, -\zeta\right)^T, \\
\boldsymbol{l}_5 &= \left(\zeta\left(H - |\boldsymbol{v}|^2\right) - \frac{3-\imath}{2}, \left(\frac{1-\imath}{2|\boldsymbol{v}|^2} + \zeta\right)\boldsymbol{v}, -\zeta\right)^T
\end{aligned}$$

of the matrix $\sum_{k=1}^{2} \boldsymbol{A}_k^2$ in its rows and $\boldsymbol{R}$ is a regular matrix containing the right eigenvectors

$$
\begin{aligned}
\boldsymbol{r}_1 &= (\zeta, \zeta v_1, \zeta v_2, \zeta v_3, \zeta H - 1)^T\,, \\
\boldsymbol{r}_2 &= (0, \alpha_1, \alpha_2, \alpha_3, 0)^T\,, \\
\boldsymbol{r}_3 &= (0, \beta_1, \beta_2, \beta_3, 0)^T\,, \\
\boldsymbol{r}_4 &= \frac{1}{2\imath}\left(1 - \imath, (3 - \imath)\boldsymbol{v}, (1 - \imath)H + 2|\boldsymbol{v}|^2\right)^T\,, \\
\boldsymbol{r}_5 &= -\frac{1}{2\imath}\left(1 + \imath, (3 + \imath)\boldsymbol{v}, (1 + \imath)H + 2|\boldsymbol{v}|^2\right)^T
\end{aligned}
$$

of the matrix $\sum_{k=1}^{3} \boldsymbol{A}_k^2$ in its columns. Obviously $\mathbf{LR} = \boldsymbol{I}$. The symbol c denotes the speed of sound, $H = e + p/\varrho$ is the total enthalpy and α, β are normalized vectors such that $\alpha \cdot \beta = \alpha \cdot \boldsymbol{v} = \beta \cdot \boldsymbol{v} = 0$. Further,

$$\imath = \sqrt{1 + 4\left(\frac{|\boldsymbol{v}|}{c}\right)^2} \tag{3.195}$$

and

$$\zeta = \frac{\left(\frac{\partial p}{\partial S}\right)_\varrho}{\varrho \theta c^2}, \tag{3.196}$$

$$S = c_v \ln \frac{p/p_0}{(\varrho/\varrho_0)^\kappa} + \mathrm{const} \tag{3.197}$$

denoting the entropy. For perfect gases it is

$$\left(\frac{\partial p}{\partial S}\right)_\varrho = p/c_v \tag{3.198}$$

and therefore

$$\zeta = \frac{1}{\kappa c_v \theta}. \tag{3.199}$$

Chapter 4

On One Approach to A Posteriori Error Estimates for Evolution Problems Solved by the Method of Lines

Abstract: We develop a new technique for a posteriori error estimates suitable to parabolic and hyperbolic equations solved by the method of lines. This approach is applied to a general nonlinear parabolic problem with a strongly monotone elliptic operator, to a linear nonstationary convection-diffusion problem and a linear second order hyperbolic problem. The error is measured with the aid of the L^2-norm in the space-time cylinder combined with a special time-weighted energy norm. Theory as well as computational results are presented.

4.1 Introduction

Recently a number of various techniques for a posteriori error estimates has been developed. A posteriori error estimation serves for accepting the results obtained by the finite element method. The subject has been treated by many authors in various ways. Let us mention, e.g., the works [4, 6, 7, 9, 10, 24, 75], where elliptic problems are analyzed. Parabolic problems are investigated, e. g., in [1, 2, 11, 25, 26, 31, 50, 61, 62, 63, 76]. In [25, 26, 38] conforming finite elements are used for the space discretization and the time discretization is carried out with the aid of the discontinuous Galerkin method. A posteriori error estimates for fully discretized nonlinear parabolic equations are also derived in [76]. In [51] a posteriori error estimates are investigated for parabolic problems discretized with respect to time only. (Then, of course, the space discretization necessarily follows.)

An opposite approach applies the semidiscretization in space only, whereas time remains continuous. This technique, called the method of lines, converts a parabolic or hyperbolic problem to a system of ordinary differential equations, which can be solved by very efficient solvers for large systems of ordinary differential equations. These solvers, as, e. g., LSODE, LSODI [40] or DDASSL [57] possess high accuracy and use sophisticated adaptive strategies.

The a posteriori error estimates for parabolic problems with one space variable discretized by the method of lines are the subject matter of papers [2, 50, 61, 62, 63].

In the present study we are concerned with one approach to a posteriori error estimates derived for multidimensional, generally nonlinear parabolic problems. The error is estimated in L^2-norm over the space-time cylinder combined with a special time weighted energy norm. Moreover, we pay the attention to linear hyperbolic problems.

The aim of the study is to show that a posteriori error estimation for elliptic equations can be utilized for the parabolic or hyperbolic problems and the quality of the estimates for these time dependent problems is the same as for the elliptic ones. In [5] the ideas of this study are further developed with emphasis on practical computations for linear parabolic problems of heat transfer and viscoelasticity.

The outline of this part of thesis is as follows. In Section 4.2 we formulate the continuous problem in a weak sense. From the weak formulation we derive an integral identity which is the basis for our a posteriori error estimates.

Section 4.3 is concerned with the discretization of the problem by the method of lines. In Section 4.4 a posteriori upper bound for the error in terms of local residuals and interelement jumps is derived. Section 4.5 is devoted to the analysis of a linear convection-diffusion problem. Section 4.6 is concerned with the application of our approach to a hyperbolic problem. Finally, in Section 4.7 numerical experiments are presented.

4.2 Continuous Problem

Let us consider the problem to find $u : Q_T \to \mathbb{R}$, $Q_T = \Omega \times (0,T)$, such that

$$\frac{\partial u}{\partial t} - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} a_i(x,t,u,\nabla u) + a_0(x,t,u,\nabla u) = 0 \quad \text{in } Q_T, \tag{4.1}$$

$$u(x,0) = u^0(x), \quad x \in \Omega, \tag{4.2}$$

$$u(x,t) = 0, \qquad x \in \partial\Omega,\ t \in (0,T). \tag{4.3}$$

Here $d \geq 1$ is an integer, $\Omega \subset \mathbb{R}^d$ is a bounded domain with a Lipschitz-continuous boundary, $u = u(x,t), x \in \Omega, t \in (0,T), 0 < T < \infty$. Let the following assumptions be satisfied:

(4.4)

a) $a_i = a_i(x,t,\xi_0,\ldots,\xi_d),\ i = 0,\ldots,d,$ are continuous in $\overline{Q}_T \times \mathbb{R}^{d+1}$ and Lipschitz-continuous with respect to $\xi_0,\ldots,\xi_d$;

b) there exists a constant $c > 0$ such that

$$|a_i(x,t,\xi_0,\xi_1,\ldots,\xi_d)| \leq c\left(1 + \sum_{i=0}^{d} |\xi_i|\right),$$
$$(x,t) \in Q_T,\ \xi = (\xi_0,\ldots,\xi_d) \in \mathbb{R}^{d+1};$$

c) there exists a constant $\alpha > 0$ such that

$$\sum_{i=0}^{d}(a_i(x,t,\xi) - a_i(x,t,\sigma))(\xi_i - \sigma_i) \geq \alpha \sum_{i=1}^{d}(\xi_i - \sigma_i)^2,$$
$$(x,t) \in Q_T,\ \xi = (\xi_0,\ldots,\xi_d),\ \sigma = (\sigma_0,\ldots,\sigma_d) \in \mathbb{R}^{d+1},$$

d) $a_j,\ j = 1,\ldots,d,$ are continuously differentiable with respect to $x_1,\ldots,x_d$ and $\xi_0,\ldots,\xi_d$ in $\overline{Q}_T \times \mathbb{R}^{d+1}$.

Under conditions (4.4, a–b), problem (4.1) – (4.3) can be reformulated in a weak sense. To this end, we recall the standard notation $L^2(Q_T),\ L^2(\Omega),\ H^1(\Omega) =$

$W^{1,2}(\Omega)$ and $H_0^1(\Omega) = W_0^{1,2}(\Omega)$ for the Lebesgue spaces of square integrable functions, the Sobolev space and its subspace formed by functions with zero traces on $\partial\Omega$. For simplicity we set

$$V = H_0^1(\Omega). \tag{4.5}$$

In $H_0^1(\Omega)$ beside the standard norm $\|v\|_{H^1(\Omega)} = \left(\int_\Omega (v^2 + |\nabla v|^2)\,\mathrm{d}x\right)^{1/2}$ we can use the seminorm $|v|_{H^1(\Omega)} = \left(\int_\Omega |\nabla v|^2\,\mathrm{d}x\right)^{1/2}$, $v \in H^1(\Omega)$, which is an equivalent norm on $H_0^1(\Omega)$.

Furthermore, by $L^2(0,T;V)$ we will denote the Bochner space of square integrable mappings of the interval $(0,T)$ into V. If $u = u(x,t)$, then $u(t)$ will denote the function defined on Ω such that $u(t)\,(x) = u(x,t)$ for $x \in \Omega$ and $t \in (0,T)$. (For the definitions and properties of function spaces, see [44] or [32].)

Let

$$u^0 \in L^2(\Omega). \tag{4.6}$$

The *weak solution* of problem (4.1) – (4.3) is defined as a function $u : Q_T \to \mathbb{R}$ such that

(4.7) a) $\quad u \in L^2(0,T;V),$

b) $$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u(t)\,v\,\mathrm{d}x + \int_\Omega \sum_{i=1}^d a_i(\cdot,t,u(t),\nabla u(t))\,\frac{\partial v}{\partial x_i}\,dx + \\ + \int_\Omega a_0(\cdot,t,u(t),\nabla u(t))\,v\,\mathrm{d}x = 0 \\ \forall\, v \in V, \text{ in the sense of distributions over } (0,T),$$

c) $\quad u(0) = u^0.$

It is possible to verify that under assumptions (4.4, a–b) the above integrals have sense for $u \in L^2(0,T;V)$ and $v \in V$. Moreover, it can be shown that $u \in C([0,T];V^*)$, where V^* is the dual of V (cf. e.g. [32, 48]), i.e., u is a continuous mapping of the closed interval $[0,T]$ with values in V^*. Taking into account the inclusions $V \subset L^2(\Omega) \subset V^*$, we see that condition (4.7, c) has sense.

With the aid of techniques from [32] and [48] it is possible to establish the existence and uniqueness of the solution u of problem (4.7, a–c). It can be shown that $\partial u/\partial t \in L^2(0,T;V^*)$. In what follows we shall assume a *stronger regularity of* u:

$$\frac{\partial u}{\partial t} \in L^2(Q_T). \tag{4.8}$$

Then (4.7, b) can be rewritten as

$$\int_\Omega \frac{\partial u(t)}{\partial t}\, v \,\mathrm{d}x + \int_\Omega \sum_{i=1}^d a_i(\cdot,t,u(t),\nabla u(t))\, \frac{\partial v}{\partial x_i}\,\mathrm{d}x + \tag{4.9}$$

$$+\int_\Omega a_0(\cdot,t,u(t),\nabla u(t))\, v\,\mathrm{d}x = 0 \quad \forall\, v \in V, \quad \text{a.e. } t \in (0,T).$$

Now, we take an arbitrary $v \in L^2(0,T;V)$ and in (4.9) for a. e. $t \in (0,T)$ we use the test function $v := v(\cdot,t)$, integrate the obtained relation from 0 to $\tau \in (0,T)$ with respect to t and then carry out the integration over $(0,T)$ with respect to τ. We get the identity

$$\int_0^T \int_0^\tau \int_\Omega \frac{\partial u(t)}{\partial t}\, v(t)\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \tag{4.10}$$

$$+\int_0^T\int_0^\tau\int_\Omega \left\{ \sum_{i=1}^d a_i(\cdot,t,u(t),\nabla u(t)) \frac{\partial v(t)}{\partial x_i} + a_0(\cdot,t,u(t),\nabla u(t)) v(t) \right\} \mathrm{d}x\mathrm{d}t\mathrm{d}\tau = 0,$$

$$v \in L^2(0,T;V). \tag{4.11}$$

Let us notice that under conditions (4.4), (4.7, a) and (4.8), the integrals in (4.10) have sense for $v \in L^2(0,T;V)$.

With the aid of Fubini's theorem we obtain the relation

$$\int_0^T \int_0^\tau \varphi(t)\,\mathrm{d}t\mathrm{d}\tau = \int_0^T (T-t)\,\varphi(t)\,\mathrm{d}t, \quad \varphi \in L^1(0,T). \tag{4.12}$$

Then, if we define the spaces

$$\begin{aligned} \mathcal{U} &= \left\{ \varphi \in L^2(0,T;V);\ \frac{\partial \varphi}{\partial t} \in L^2(Q_T) \right\}, \\ \mathcal{V} &= L^2(0,T;V) \end{aligned} \tag{4.13}$$

(obviously, $\mathcal{U} \subset \mathcal{V}$) and the form $\mathcal{B} : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$,

$$\mathcal{B}(u,v) = \int_0^T \int_0^\tau \int_\Omega \frac{\partial u(t)}{\partial t}\, v(t)\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \tag{4.14}$$

$$+\int_0^T (T-t) \int_\Omega \left\{ \sum_{i=1}^d a_i(\cdot,t,u(t),\nabla u(t)) \,\frac{\partial v(t)}{\partial x_i} + a_0(\cdot,t,u(t),\nabla u(t))\, v(t) \right\} \mathrm{d}x\mathrm{d}t,$$

we see from (4.12) that (4.10) can be rewritten in the form

$$\mathcal{B}(u,v) = 0, \quad v \in \mathcal{V}. \tag{4.15}$$

The form $\mathcal{B}$, which is linear in v and (in general) nonlinear in u, has the following properties:

Theorem 4.1. *Under assumption* (4.4, c) *the form* $\mathcal{B}(u,v)$ *is strongly monotone in the following sense*:

$$\mathcal{B}(u,u-v) - \mathcal{B}(v,u-v) \geq \frac{1}{2}\|u-v\|^2_{L^2(Q_T)} + \tag{4.16}$$

$$+\alpha \int_0^T (T-t)\, |u(t)-v(t)|^2_{H^1(\Omega)} \mathrm{d}t - \frac{1}{2} T \|u(0)-v(0)\|^2_{L^2(\Omega)} \quad \forall\, u,\, v \in \mathcal{U}.$$

Proof. By (4.14), for $u,\, v \in \mathcal{U}$,

$$\mathcal{B}(u,u-v)-\mathcal{B}(v,u-v) = \int_0^T \int_0^\tau \int_\Omega \frac{\partial (u(t)-v(t))}{\partial t}\,(u(t)-v(t))\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \tag{4.17}$$

$$+\int_0^T (T-t) \int_\Omega \left\{ \sum_{i=1}^d [a_i(\cdot,t,u(t),\nabla u(t)) - a_i(\cdot,t,v(t),\nabla v(t))] \,\frac{\partial (u(t)-v(t))}{\partial x_i} + \right.$$

$$\left. + \left[a_0(\cdot,t,u(t),\nabla u(t)) - a_0(\cdot,t,v(t),\nabla v(t))\right](u(t)-v(t)) \right\} \mathrm{d}x\mathrm{d}t.$$

By Fubini's theorem,

$$\int_0^T \int_0^\tau \int_\Omega \frac{\partial (u(t)-v(t))}{\partial t}\,(u(t)-v(t))\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau = \tag{4.18}$$

$$= \frac{1}{2}\int_\Omega \int_0^T \int_0^\tau \frac{\partial}{\partial t}\,(u(t)-v(t))^2\,\mathrm{d}t\mathrm{d}\tau\mathrm{d}x =$$

$$= \frac{1}{2}\int_\Omega \int_0^T (u(\tau)-v(\tau))^2\,\mathrm{d}\tau\mathrm{d}x - \frac{1}{2}T\|u(0)-v(0)\|^2_{L^2(\Omega)} =$$

$$= \frac{1}{2}\|u-v\|^2_{L^2(Q_T)} - \frac{1}{2}T\|u(0)-v(0)\|^2_{L^2(\Omega)}.$$

Furthermore, in virtue of (4.4, c), the second integral in (4.17) is bounded from below by

$$\alpha \int_0^T (T-t) \int_\Omega \left\{ \sum_{i=1}^d \left[\frac{\partial (u(t) - v(t))}{\partial x_i} \right]^2 \right\} \mathrm{d}x\mathrm{d}t = \alpha \int_0^T (T-t)\, |u(t)-v(t)|^2_{H^1(\Omega)} \,\mathrm{d}t.$$

This and (4.18) immediately yield (4.16). *Q.E.D.*

If we introduce in the space $\mathcal{V}$ beside the standard norm $\|v\|_{L^2(0,T;V)} = \left(\int_0^T \|v(t)\|^2_{H^1(\Omega)} \,\mathrm{d}t\right)^{1/2}$ also the *weighted norm*

$$\begin{aligned} (4.19) \quad \|v\|_{\mathcal{V}} &= \left(\frac{1}{2} \int_{Q_T} v^2 \,\mathrm{d}x\mathrm{d}t + \alpha \int_0^T (T-t) \int_\Omega |\nabla v|^2 \,\mathrm{d}x\mathrm{d}t \right)^{1/2} = \\ &= \left(\frac{1}{2} \|v\|_{L^2(Q_T)} + \alpha \int_0^T (T-t)\, |v(t)|^2_{H^1(\Omega)} \,\mathrm{d}t \right)^{1/2}, \end{aligned}$$

then (4.16) can be written in the simple form

$$\mathcal{B}(u, u-v) - \mathcal{B}(v, u-v) \geq \|u-v\|^2_{\mathcal{V}} - \frac{1}{2} T \|u(0)-v(0)\|^2_{L^2(\Omega)}, \quad u,\, v \in \mathcal{U}. \tag{4.20}$$

4.3 Discrete Problem

We solve problem (4.1) – (4.3) or, in the weak formulation, (4.7) by the finite element method of lines. This means that we use the semidiscretization in space, whereas we consider the time to be continuous.

To this end, we assume that the closure $\overline{\Omega}$ of the domain Ω is a union of a finite number of simplexes $K \in \mathcal{T}_h$, where $\mathcal{T}_h$ is a triangulation of Ω. If $d = 1$ or 2 or 3, then Ω is an interval in $\mathbb{R}^1$ or a polygonal domain in $\mathbb{R}^2$ or a polyhedral domain in $\mathbb{R}^3$, respectively. We can allow here also domains with cracks or polyhedral domains in $\mathbb{R}^3$ which are not Lipschitzian. We assume that the triangulation $\mathcal{T}_h$ has standard properties used in the finite element method (see, e. g. [18]).

For each $t \in [0, T]$ the solution $u(t)$ will be approximated by a function $u_h(t) \in V_h$, where

$$V_h = \left\{ v_h \in C(\overline{\Omega}); v_h|_{\partial\Omega} = 0,\, v_h|_K \in P_p(K)\ \forall\, K \in \mathcal{T}_h \right\}. \tag{4.21}$$

Here $P_p(K)$ is the space of all polynomials on K of degree $\leq p$. Starting from indentity (4.9), we define the approximate solution as a function $u_h = u_h(x,t)$ satisfying the following conditions:

(4.22) a) $\quad u_h \in C^1([0,T]; V_h)$,

$$\text{b)} \quad \int_\Omega \frac{\partial u_h(t)}{\partial t} v_h \, \mathrm{d}x + \int_\Omega \sum_{i=1}^d a_i(\cdot, t, u_h(t), \nabla u_h(t)) \frac{\partial v_h}{\partial x_i} \, \mathrm{d}x + \\ + \int_\Omega a_0(\cdot, t, u_h(t), \nabla u_h(t)) \, v_h \, \mathrm{d}x \quad \forall \, v_h \in V_h, \ \forall \, t \in (0,T),$$

c) $\quad u_h(0) = u_h^0$,

where u_h^0 is a V_h-approximation of u^0.

Under assumptions (4.4), problem (4.22) has a unique solution. It is easy to see that this problem is equivalent to an initial value problem for a system of $n = n_h$ ordinary differential equations, where $n = \dim V_h$.

The semidiscretization in space is applied very often to evolutionary problems. (See, e. g., [72, 2, 50, 61, 62, 63].) The resulting system of ODE's is then solved with the aid of efficient ODE-solvers ([40, 57])

Similarly as in Section 4.2 we can reformulate the discrete problem in the form

$$\int_0^T \int_0^\tau \int_\Omega \frac{\partial u_h(t)}{\partial t} v_h(t) \, \mathrm{d}x \mathrm{d}t \mathrm{d}\tau + \\ + \int_0^T (T-t) \int_\Omega \Bigg\{ \sum_{i=1}^d a_i(\cdot, t, u_h(t), \nabla u_h(t)) \frac{\partial v_h(t)}{\partial x_i} + \\ + a_0(\cdot, t, u_h(t), \nabla u_h(t)) \, v_h(t) \Bigg\} \mathrm{d}x \mathrm{d}t \mathrm{d}\tau = 0, \quad v_h \in \mathcal{V}_h := L^2(0,T; V_h) \ (\subset \mathcal{V}) \tag{4.23}$$

or, simply

$$\mathcal{B}(u_h, v_h) = 0, \quad v_h \in \mathcal{V}_h. \tag{4.24}$$

This together with (4.15) is the basis for our further considerations.

4.4 A Posteriori Error Estimates

By e_h we denote the error of the method of lines (4.22):

$$e_h = u - u_h \in \mathcal{U}. \tag{4.25}$$

Our goal is to derive an a posteriori estimate for the error e_h. As usual we consider triangulations $\mathcal{T}_h$ of Ω, $h = \max_{K\in\mathcal{T}_h} h_K$ for $h \in (0, h_0), h_0 > 0$, and introduce standard assumptions.

Assumption 4.1. There exists a constant $\hat{c} > 0$ independent of h such that $h_K/\rho_K \le \hat{c}$ for all $K \in \mathcal{T}_h$, where $h_K = \mathrm{diam}(K) =$ diameter of K and ρ_K is the radius of the largest circle inscribed in K.

Let $v \in \mathcal{V}$. Then, in view of (4.15),

$$\mathcal{B}(u,v) - \mathcal{B}(u_h,v) = -\mathcal{B}(u_h,v) =: \mathcal{R}_h(v) = \text{residual}. \tag{4.26}$$

By (4.24), the residual $\mathcal{R}_h$ has the *orthogonality property*

$$\mathcal{R}_h(v_h) = 0, \quad v_h \in \mathcal{V}_h. \tag{4.27}$$

Let us express $\mathcal{R}_h(v)$ with the aid of local residuals and jumps. To this end, we denote by $\mathcal{E}_h$ the set of all sides S of all $K \in \mathcal{T}_h$ such that $S \not\subset \partial\Omega$. We set $h_S = \mathrm{diam}(S)$. By $\boldsymbol{n}_S = (n_{S1}, \ldots, n_{Sd})$ we denote a unit normal to $S \in \mathcal{E}_h$. For any $S \in \mathcal{E}_h$ let K_S^1 and K_S^2 be elements from $\mathcal{T}_h$ with a common side S. If φ is continuous in the interior $\mathring{K}_S^i$ of each K_S^i and continuously extendable from $\mathring{K}_S^i$ on S, then we denote by $[\varphi]_S$ the jump of φ across S:

$$[\varphi]_S(x) = \lim_{\varepsilon\to 0+} (\varphi(x+\varepsilon\boldsymbol{n}_S) - \varphi(x-\varepsilon\boldsymbol{n}_S)), \quad x \in S. \tag{4.28}$$

This jump depends, of course, on the orientation of $\boldsymbol{n}_S$ but the expressions $[\varphi]_S n_{Si}, i = 1, \ldots, d$, are independent of the orientation of $\boldsymbol{n}_S$. Now we set

$$\begin{aligned} r_K(x,t) &= -\frac{\partial u_h(x,t)}{\partial t} + \sum_{i=1}^d \frac{\partial}{\partial x_i} a_i(x,t,u_h(x,t),\nabla u_h(x,t)) - \\ &\quad - a_0(x,t,u_h(x,t),\nabla u_h(x,t)), \\ &\qquad x \in K,\ t \in (0,T),\ K \in \mathcal{T}_h \ \text{(local residuals)}, \end{aligned} \tag{4.29}$$

$$\begin{aligned} J_S(x,t) &= -\sum_{i=1}^d [a_i(x,t,u_h(x,t),\nabla u_h(x,t))]_S\, n_{Si} \\ &\qquad x \in S,\ t \in (0,T),\ S \in \mathcal{E}_h \ \text{(local jumps)}. \end{aligned}$$

Using (4.14), the relation $\int_\Omega = \sum_{K\in\mathcal{T}_h}\int_K$ and applying Green's theorem elementwise, we find that

$$\mathcal{R}_h(v) = \int_0^T (T-t)\left\{\sum_{K\in\mathcal{T}_h}\int_K r_K(t)v(t)\,\mathrm{d}x + \sum_{S\in\mathcal{E}_h}\int_S J_S(t)\,v(t)\,\mathrm{d}S\right\}\mathrm{d}t,\ \ v\in\mathcal{V}. \tag{4.30}$$

If we now set $v := e_h$ in (4.30) and use the properties of the form $\mathcal{B}$, we obtain an *abstract a posteriori error estimate*:

Theorem 4.2. *For the error e_h we have the following upper bound:*

$$\|e_h\|_{\mathcal{V}}^2 \le \frac{1}{2}T\sum_{K\in\mathcal{T}_h}\|e_h(0)\|_{L^2(K)}^2 + \tag{4.31}$$

$$+\int_0^T (T-t)\sum_{K\in\mathcal{T}_h}\Big(h_K\|r_K(t)\|_{L^2(K)}\Big)\Big(h_K^{-1}\|e_h(t)-v_h(t)\|_{L^2(K)}\Big)\,\mathrm{d}t +$$

$$+\int_0^T (T-t)\sum_{S\in\mathcal{E}_h}\Big(h_S^{1/2}\|J_S(t)\|_{L^2(S)}\Big)\Big(h_S^{-1/2}\|e_h(t)-v_h(t)\|_{L^2(S)}\Big)\,\mathrm{d}t,$$

$$v_h\in\mathcal{V}_h.$$

Proof. We set $v := u_h$ in (4.20) and use (4.26) – (4.27). Then we have

$$\begin{aligned}\|e_h\|_{\mathcal{V}}^2 &\le \mathcal{R}_h(e_h) + \frac{1}{2}T\|e_h(0)\|_{L^2(\Omega)}^2 = \\ &= \mathcal{R}_h(e_h - v_h) + \frac{1}{2}T\|e_h(0)\|_{L^2(\Omega)}^2 \text{ for any } v_h\in\mathcal{V}_h.\end{aligned} \tag{4.32}$$

Now, (4.30), (4.32), the Cauchy inequality and the relation

$$\|e_h(0)\|_{L^2(\Omega)}^2 = \sum_{K\in\mathcal{T}_h}\|e_h(0)\|_{L^2(K)}^2$$

yield (4.31). *Q.E.D.*

The application of the abstract a posteriori error estimate will be split into two cases: I) $d = 1$ (one space dimension) and II) $d \ge 1$ (more space dimensions).

I. The case $d = 1$. In this case the domain Ω reduces to an open bounded interval $(a, b) \subset \mathbb{R}$ and the triangulation $\mathcal{T}_h$ is its partition formed by intervals $K_i = [x_{i-1}, x_i],\ i = 1, \dots, N$, where $a = x_0 < x_1 < \dots < x_N = b$. We can set $\mathcal{E}_h = \{S_1, \dots, S_{N-1}\}$ with $S_i = \{x_i\}$. Of course, $h_{K_i} = x_i - x_{i-1}$.

By I_h we denote the operator of the piecewise linear Lagrange interpolation: $I_h : V \to V_h$,

$$I_h v \in V_h, \quad (I_h v)(x_i) = v(x_i), \quad i = 0, \dots, N,\ v \in V. \tag{4.33}$$

It is possible to show that

$$\|v - I_h v\|_{L^2(K_i)} \le C_1 h_{K_i} |v|_{H^1(K_i)}, \quad v \in V, \text{with } C_1 = 1/\pi, \tag{4.34}$$

where $|\cdot|_{H^1(K_i)}$ is the $H^1(K_i)$-seminorm. The choice of $v_h := I_h e_h$ in the bound (4.31) yields the following result.

Theorem 4.3. *If $d = 1$, then we have the a posteriori error estimate of the form*

$$\begin{aligned} &\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T - t)\, |e_h(t)|^2_{H^1(\Omega)}\, \mathrm{d}t \le \\ &\le T \sum_{i=1}^N \|e_h(0)\|^2_{L^2(K_i)} + \frac{1}{\alpha \pi^2} \int_0^T (T - t) \sum_{i=1}^N \|h_{K_i} r_{K_i}(t)\|^2_{L^2(K_i)}\, \mathrm{d}t, \quad h \in (0, h_0). \end{aligned} \tag{4.35}$$

Proof. We use (4.31) with $v_h := I_h e_h$ and apply the Cauchy inequality. Then, because $(e_h - I_h e_h)(x_i, t) = 0$, the terms with jumps $J_{S_i}(t)$ in (4.31) vanish and (4.34) yields the estimate

$$\begin{aligned} &\frac{1}{2}\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T - t)\, |e_h(t)|^2_{H^1(\Omega)}\, \mathrm{d}t \le \frac{1}{2} T \sum_{i=1}^N \|e_h(0)\|^2_{L^2(K_i)} + \\ &+ \frac{1}{\pi} \left(\int_0^T (T - t) \sum_{i=1}^N \|h_{K_i} r_{K_i}(t)\|^2_{L^2(K_i)}\, \mathrm{d}t \right)^{1/2} \left(\int_0^T (T - t) \sum_{i=1}^N |e_h(t)|^2_{H^1(K_i)}\, \mathrm{d}t \right)^{1/2}. \end{aligned} \tag{4.36}$$

Now, (4.36) and Young's inequality $ab \le \varepsilon a^2/2 + b^2/(2\varepsilon)$ for $a,\ b \ge 0,\ \varepsilon > 0$, imply that

$$\frac{1}{2}\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T - t)\, |e_h(t)|^2_{H^1(\Omega)}\, \mathrm{d}t \le$$

$$\leq \frac{1}{2}T\sum_{i=1}^{N}\|e_h(0)\|^2_{L^2(K_i)} + \frac{\alpha}{2}\int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t + $$
$$+\frac{1}{2\alpha\pi^2}\int_0^T (T-t)\sum_{i=1}^{N}\|h_{K_i} r_{K_i}(t)\|^2_{L^2(K_i)}\,\mathrm{d}t.$$

This already yields (4.35). *Q.E.D.*

II. The case $d \geq 2$. Now the situation is more complex, because $H^1(\Omega) \not\subset C(\overline{\Omega})$ and, hence, the Lagrange interpolation cannot be used. Moreover, the jump terms $J_S(t)$ may not disappear in (4.31).

One possibility is to use the so-called *Clement's interpolation* ([19] or [18], Exercise 3.2.3) $I_h : V \to V_h$. In [19], the following properties of I_h have been established:

Lemma 4.1. *Let*

$$\tilde{\omega}_K = \bigcup_{\substack{K' \in \mathcal{T}_h \\ K \cap K' \neq \emptyset}} K' \quad \text{for any } K \in \mathcal{T}_h. \tag{4.37}$$

Then there exist constants $\tilde{c}_1,\ \tilde{c}_2 > 0$ such that for all $K \in \mathcal{T}_h,\ S \in \mathcal{E}_h,\ S \subset \partial K,\ v \in H^1(\Omega)$ and $h \in (0, h_0)$ the following estimates hold:

$$\begin{aligned} &\text{a)} \qquad \|v - I_h v\|_{L^2(K)} \leq \tilde{c}_1 h_K |v|_{H^1(\tilde{\omega}_K)}, \\ &\text{b)} \qquad \|v - I_h v\|_{L^2(S)} \leq \tilde{c}_2 h_K^{1/2} |v|_{H^1(\tilde{\omega}_K)}. \end{aligned} \tag{4.38}$$

Because of estimates (4.38), we have to introduce the following geometric assumption on the meshes $\mathcal{T}_h,\ h \in (0, h_0)$:

Assumption 4.2. There exists a constant $\tilde{C} > 0$ such that

$$\operatorname{card}\{K' \in \mathcal{T}_h; K' \subset \tilde{\omega}_K\} \leq \tilde{C}\ \forall\, K \in \mathcal{T}_h\ \forall\, h \in (0, h_0).$$

If $d = 2$, then this property is an obvious consequence of Assumption 4.1. Now we can establish an a posteriori error estimate.

Theorem 4.4. *Under Assumptions 4.1 and 4.2 for $d \geq 2$ we have the following a posteriori error estimate:*

$$
\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T-t)\, |e_h(t)|^2_{H^1(\Omega)} \,\mathrm{d}t \leq \tag{4.39}
$$

$$
\leq T \sum_{K\in\mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \frac{c^*}{\alpha} \int_0^T (T-t) \sum_{K\in\mathcal{T}_h} \|h_K r_K(t)\|^2_{L^2(K)} \,\mathrm{d}t +
$$

$$
+ \frac{c^{**}}{\alpha} \int_0^T (T-t) \sum_{S\in\mathcal{E}_h} \|h_S^{1/2}\, J_S(t)\|^2_{L^2(S)} \,\mathrm{d}t, \quad h \in (0, h_0),
$$

where c^* and c^{**} are constants that depend on constants $\tilde{c}_1$, $\tilde{c}_2$, $\hat{c}$ and $\tilde{C}$, but are independent of h.

Proof. For each $S \in \mathcal{E}_h$ there exist two triangles K_S, $K'_S \in \mathcal{T}_h$ such that $S = \partial K_S \cap \partial K'_S$. Further, we have $h_S \geq \rho_{K_S}$. This and Assumption 4.1 imply that $h_{K_S} h_S^{-1} \leq \hat{c}$. Moreover, $\sum_{S\in\mathcal{E}_h} |e_h(t)|^2_{H^1(\tilde{\omega}_{K_S})} \leq (\tilde{C}/2) \sum_{K\in\mathcal{T}_h} |e_h(t)|^2_{H^1(K)}$. Then, starting from (4.31) with $v_h := I_h e_h$, using (4.38), and Young's inequality, we find that

$$
\frac{1}{2}\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T-t)\, |e_h(t)|^2_{H^1(\Omega)} \,\mathrm{d}t \leq
$$

$$
\leq \frac{1}{2} T \sum_{K\in\mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \int_0^T (T-t) \sum_{K\in\mathcal{T}_h} \Big(h_K \|r_k(t)\|_{L^2(K)} \Big) \Big(\tilde{c}_1 |e_h(t)|_{H^1(\tilde{\omega}_K)} \Big) \,\mathrm{d}t +
$$

$$
+ \int_0^T (T-t) \sum_{S\in\mathcal{E}_h} \Big(h_S^{1/2} \|J_S(t)\|_{L^2(S)} \Big) \Big(\tilde{c}_2\, \hat{c}^{1/2} |e_h(t)|_{H^1(\omega_{K_S})} \Big) \,\mathrm{d}t \leq
$$

$$
\leq \frac{1}{2} T \sum_{K\in\mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \left\{ \tilde{C}^{1/2} \tilde{c}_1 \left(\int_0^T (T-t) \sum_{K\in\mathcal{T}_h} \|h_K r_K(t)\|^2_{L^2(K)} \mathrm{d}t \right)^{1/2} + \right.
$$

$$
\left. + \ (\hat{c}\,\tilde{C}/2)^{1/2} \tilde{c}_2 \left(\int_0^T (T-t) \sum_{S\in\mathcal{E}_h} \|h_S^{1/2} J_S(t)\|^2_{L^2(K)} \mathrm{d}t \right)^{1/2} \right\} \times
$$

$$
\times \left(\int_0^T (T-t)\, |e_h(t)|^2_{H^1(\Omega)} \,\mathrm{d}t \right)^{1/2} \leq
$$

$$\leq \frac{1}{2} T \sum_{K \in \mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \frac{\alpha}{2} \int_0^T (T-t) \sum_{K \in \mathcal{T}_h} |e_h(t)|^2_{H^1(\Omega)} \, \mathrm{d}t +$$

$$+ \frac{\tilde{C}\, \tilde{c}_1^2}{\alpha} \int_0^T (T-t) \sum_{K \in \mathcal{T}_h} \|h_K r_K(t)\|^2_{L^2(K)} \, \mathrm{d}t +$$

$$+ \frac{\hat{c}\, \tilde{C}\, \tilde{c}_2^2}{2\alpha} \int_0^T (T-t) \sum_{S \in \mathcal{E}_h} \|h_S^{1/2} J_S(t)\|^2_{L^2(S)} \, \mathrm{d}t.$$

This already yields (4.39). *Q.E.D.*

Remark 4.1. As was said in the introduction, any elliptic estimator – the residual explicit, equilibrated one or based on the recovery – can be used. Here we utilize the simplest explicit estimator which gives good results in one space dimension, but has usually bad quality in higher dimensions. That is why we have introduced Theorem 4.4 only because of completness. It is an analogy to a posteriori error estimates obtained e.g. in [76] for elliptic problems. As follows from [16], the use of the explicit estimator combined with Clement's interpolation gives a very high overestimation of the error (the effectivity index index $\approx$ 30 for uniform meshes and $\approx 50 - 70$ for adaptive meshes, which is completely useless in practice). This implies the necessity to apply more sophisticated estimators. For more details, see [3], where constants in estimates are specified, and [5] using the equilibrated residual estimator combined with the technique treated here.

4.5 Application to a Convection–Diffusion Problem

Now we pay our attention to the problem

(4.40)
$$\text{a)} \quad \frac{\partial u}{\partial t} + \mathrm{div}(\boldsymbol{b}u) - \varepsilon \Delta u = f \quad \text{in } Q_T = \Omega \times (0,T),$$
$$\text{b)} \quad u|_{\partial\Omega \times (0,T)} = 0,$$
$$\text{c)} \quad u(x,0) = u^0(x), \quad x \in \Omega,$$

where Ω has the same properties as in Sections 1 and 2, $\boldsymbol{b} = (b_1, \ldots, b_d) : Q_T \to \mathbb{R}^d$, $\mathrm{div}(\boldsymbol{b}u) = \sum_{i=1}^d \frac{\partial(b_i u)}{\partial x_i}$, $\Delta u = \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2}$, $\varepsilon > 0$, $f : Q_T \to \mathbb{R}$, $u^0 : \Omega \to \mathbb{R}$. Here $\boldsymbol{b}$, f, u^0, ε are given data. In applications very often $\varepsilon \ll 1$.

Equation (4.1, a) describes, e. g., the heat conduction in a fluid moving with velocity $\boldsymbol{b}$ or the propagation of some material with concentration u in a moving fluid.

In works dealing with theoretical analysis of error estimates for the numerical solution of problem (4.40), usually the additional (unphysical) term $c\,u$ with $c \geq c_0 = \text{const} > 0$ is required to stand at the left-hand side of (4.1, a). Otherwise, there are difficulties with the choice of a reasonable norm (cf., e. g. [60]). (The naturally chosen energy norm $\|u\|_\varepsilon := \left(\varepsilon \int_0^T \int_\Omega |\nabla u|^2 \,\mathrm{d}x\mathrm{d}t\right)^{1/2}$ degenerates for $\varepsilon \to 0$.) Our approach treated in previous sections allows us to obtain the error estimate in $L^2(Q_T)-$ norm, which is suitable for problem (4.40).

We first reformulate problem (4.40) similarly as in (4.7) and under assumption (4.8) in the form analogous to (4.9):

$$
\begin{aligned}
(4.41)\qquad & \int_\Omega \frac{\partial u(t)}{\partial t} v \,\mathrm{d}x + \int_\Omega \operatorname{div}(\boldsymbol{b}(t)u(t))\, v \,\mathrm{d}x + \varepsilon \int_\Omega \nabla u(t) \cdot \nabla v \,\mathrm{d}x = \\
&= \int_\Omega f(t)\, v \,\mathrm{d}x, \quad v \in V, \text{ a. e. } t \in (0,T).
\end{aligned}
$$

We assume that $f \in L^2(Q_T)$, $\boldsymbol{b} \in L^2(0,T;W^{1,\infty}(\Omega))^d$, $u^0 \in L^2(\Omega)$. The analogy to (4.10) reads now

$$
\begin{aligned}
(4.42)\qquad & \int_0^T \int_0^\tau \int_\Omega \frac{\partial u(t)}{\partial t} v(t) \,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \\
&+ \int_0^T (T-t) \int_\Omega \{\operatorname{div}(\boldsymbol{b}(t)u(t))\, v(t) + \varepsilon \nabla u(t) \cdot \nabla v(t)\} \,\mathrm{d}x\mathrm{d}t = \\
&= \int_0^T (T-t) \int_\Omega f(t)\, v(t) \,\mathrm{d}x\mathrm{d}t, \quad v \in \mathcal{V} = L^2(0,T;V).
\end{aligned}
$$

The semidiscrete approximate solution $u_h \in \mathcal{U}_h$ satisfies now the identity

$$
\begin{aligned}
(4.43)\qquad & \int_0^T \int_0^\tau \int_\Omega \frac{\partial u_h(t)}{\partial t} v_h(t) \,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \\
&+ \int_0^T (T-t) \int_\Omega \{\operatorname{div}(\boldsymbol{b}(t)u_h(t))\, v_h(t) + \varepsilon \nabla u_h(t) \cdot \nabla v_h(t)\} \,\mathrm{d}x\mathrm{d}t = \\
&= \int_0^T (T-t) \int_\Omega f(t)\, v_h(t) \,\mathrm{d}x\mathrm{d}t, \quad v_h \in \mathcal{V}_h = L^2(0,T;V_h).
\end{aligned}
$$

Hence, defining the forms

$$
(4.44)\qquad \mathcal{B}(u,v) = \int_0^T \int_0^\tau \int_\Omega \frac{\partial u(t)}{\partial t} v(t) \,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \varepsilon \int_0^T (T-t) \int_\Omega \nabla u(t) \cdot \nabla v(t) \,\mathrm{d}x\mathrm{d}t +
$$

$$+\int_0^T (T-t)\int_\Omega \operatorname{div}(\boldsymbol{b}(t)\,u(t))\,v(t)\,\mathrm{d}x\mathrm{d}t,$$

$$\mathcal{L}(v) = \int_0^T (T-t)\int_\Omega f(t)\,v(t)\,\mathrm{d}x\mathrm{d}t, \quad u\in\mathcal{U},\ v\in\mathcal{V},$$

the continuous and discrete problems can be written as

$$\mathcal{B}(u,v) = \mathcal{L}(v), \quad v\in\mathcal{V}, \tag{4.45}$$

and

$$\mathcal{B}(u_h,v_h) = \mathcal{L}(v_h), \quad v_h\in\mathcal{V}_h, \tag{4.46}$$

respectively.

Let us establish the coercivity of the form $\mathcal{B}$.

Theorem 4.5. *Let*

$$(T-t)\operatorname{div}\boldsymbol{b}(t) \le c_b < 1 \quad \textit{in } \Omega \textit{ for a. e. } t\in(0,T). \tag{4.47}$$

Then

$$\mathcal{B}(u,u) \ge \frac{1}{2}(1-c_b)\|u\|^2_{L^2(Q_T)}+\varepsilon\int_0^T (T-t)\,|u(t)|^2_{H^1(\Omega)}\,\mathrm{d}t-\frac{1}{2}T\|u(0)\|^2_{L^2(\Omega)},\ u\in\mathcal{U}. \tag{4.48}$$

Proof. Substituting $v := u$ into $\mathcal{B}$, we obtain

$$\begin{aligned}\mathcal{B}(u,u) = \frac{1}{2}\int_0^T\int_\Omega u^2\,\mathrm{d}x\mathrm{d}t-\frac{1}{2}T\|u(0)\|^2_{L^2(\Omega)}+\varepsilon\int_0^T (T-t)\int_\Omega |\nabla u(t)|^2\,\mathrm{d}x\mathrm{d}t+\\ +\int_0^T (T-t)\int_\Omega \operatorname{div}(\boldsymbol{b}(t)\,u(t))\,u(t)\,\mathrm{d}x\mathrm{d}t.\end{aligned} \tag{4.49}$$

Using Green's theorem, we find that

$$\int_\Omega \operatorname{div}(\boldsymbol{b}(t)\,u(t))\,u(t)\,\mathrm{d}x = -\frac{1}{2}\int_\Omega u^2(t)\operatorname{div}\boldsymbol{b}(t)\,\mathrm{d}x.$$

This and assumption (4.47) imply that

$$\int_0^T (T-t)\int_\Omega \operatorname{div}(\boldsymbol{b}(t)\,u(t))\,u(t)\,\mathrm{d}x\mathrm{d}t \ge -\frac{c_b}{2}\int_{Q_T} u^2\,\mathrm{d}x\mathrm{d}t$$

and, hence, from (4.49) we immediately get (4.48). *Q.E.D.*

By (4.45) and (4.46), the error $e_h = u - u_h$ satisfies the relation

$$\mathcal{B}(e_h, v) = \mathcal{R}_h(v) = \mathcal{L}(v) - B(u_h, v), \quad v \in \mathcal{V}, \tag{4.50}$$

and, the orthogonality property holds:

$$\mathcal{R}_h(v_h) = 0, \quad v_h \in \mathcal{V}_h. \tag{4.51}$$

In what follows, we will again derive an a posteriori error bound in terms of local residuals

$$r_K(t) = -\frac{\partial u_h(t)}{\partial t} + \varepsilon \Delta u_h(t) - \operatorname{div}(\boldsymbol{b}(t)\, u_h(t)) + f(t) \text{ on } K,\ t \in (0,T), \tag{4.52}$$

and jumps of the normal derivatives on sides $S \in \mathcal{E}_h$:

$$J_S(t) = -\left[\frac{\partial u_h}{\partial n}(t)\right]_S = -[\nabla u_h(t)]_S \cdot \boldsymbol{n}_S. \tag{4.53}$$

Then we get

$$\begin{aligned} \mathcal{R}_h(v) \ &= \ \int_0^T (T-\tau) \left\{ \sum_{K \in \mathcal{T}_h} \int_K r_K(t)\, v(t)\, \mathrm{d}x + \right. \\ &\left. + \ \sum_{S \in \mathcal{E}_h} \varepsilon \int_S J_S(t)\, v(t)\, \mathrm{d}S \right\} \mathrm{d}t,\ v \in \mathcal{V}. \end{aligned} \tag{4.54}$$

In the same way as in Section 4.4 we derive an *abstract a posteriori error estimate*:

Theorem 4.6. *Under condition* (4.47), *the error* e_h *satisfies the estimate*

$$\begin{aligned} &\frac{1}{2}(1-c_b)\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\, |e_h(t)|^2_{H^1(\Omega)} \mathrm{d}t \le \\ &\le \ \frac{1}{2} T \sum_{K \in \mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \\ &+ \ \int_0^T (T-t) \left\{ \sum_{K \in \mathcal{T}_h} (h_K \|r_K(t)\|_{L^2(K)})\, (h_K^{-1} \|e_h(t) - v_h(t)\|_{L^2(K)}) + \right. \\ &\left. + \ \sum_{S \in \mathcal{E}_h} (h_S^{1/2} \varepsilon \|J_S(t)\|_{L^2(S)})\, (h_S^{-1/2} \|e_h(t) - v_h(t)\|_{L^2(S)}) \right\} \mathrm{d}t,\ v_h \in \mathcal{V}_h. \end{aligned} \tag{4.55}$$

Proof uses (4.48), (4.50) – (4.15) and is quite analogous to the proof of Theorem 4.2. *Q.E.D.*

Now we can set $v_h := I_h e_h$ in (4.55), where I_h is the interpolation operator defined in Section 4.4. We distinguish again the cases $d = 1$ and $d \geq 2$:

Theorem 4.7. *Let $d = 1$ and (4.47) be satisfied. Then*

$$
\begin{aligned}
(4.56)\qquad & (1-c_b)\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \leq \\
\leq\ & T\sum_{i=1}^N \|e_h(0)\|^2_{L^2(K_i)} + \frac{1}{\pi^2}\int_0^T (T-t)\sum_{i=1}^N \left\| \frac{h_{K_i}}{\sqrt{\varepsilon}}\, r_{K_i}(t)\right\|^2_{L^2(K_i)} \mathrm{d}t.
\end{aligned}
$$

Proof. We use the same notation as in I) from Section 4.4. Substituting $v_h := I_h e_h$ in (4.55), where I_h is the Lagrange interpolation, the jumps J_S disappear. Using (4.34), the Cauchy inequality and Young's inequality, we obtain the estimate

$$
\begin{aligned}
& \frac{1}{2}(1-c_b)\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \leq \\
\leq\ & \frac{1}{2}T\sum_{i=1}^N \|e_h(0)\|^2_{L^2(K_i)} + \frac{1}{2\pi^2}\int_0^T (T-t)\,\frac{1}{\varepsilon}\sum_{i=1}^N \|h_{K_i}\, r_{K_i}(t)\|^2_{L^2(K_i)}\,\mathrm{d}t + \\
& + \frac{\varepsilon}{2}\int_0^T (T-t)\sum_{i=1}^N |e_h(t)|^2_{H^1(K_i)}\,\mathrm{d}t.
\end{aligned}
$$

This already implies (4.56). *Q.E.D.*

In the case $d \geq 2$ we arrive at

Theorem 4.8. *Let $d \geq 2$ and let Assumptions 4.1, 2 and (4.47) be satisfied. Then*

$$
\begin{aligned}
(4.57)\quad & (1-c_b)\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \leq \\
\leq\ & T\sum_{K\in\mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + c^* \int_0^T (T-t)\sum_{K\in\mathcal{T}_h} \|h_K\,\varepsilon^{-1/2} r_K(t)\|^2_{L^2(K)}\,\mathrm{d}t + \\
& + c^{**}\int_0^T (T-t)\sum_{S\in\mathcal{E}_h} \|h_S^{1/2}\,\varepsilon J_S(t)\|^2_{L^2(S)}\,\mathrm{d}t, \quad h\in(0,h_0),
\end{aligned}
$$

where c^*, c^{**} are constants depending on the constants $\hat{c}$, $\tilde{C}$, $\tilde{c}_1$, $\tilde{c}_2$ from Assumptions 4.1, 4.2 and estimates (4.38), *but independent of h and ε.*

Proof. We proceed analogously as in the proof of Theorem 4.4. The abstract estimate (4.55) together with (4.38), Assumptions 4.1, 2 and the Cauchy inequality imply that

$$
\begin{aligned}
&\frac{1}{2}(1-c_b)\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \le \\
\le\ & \frac{1}{2}T \sum_{K\in\mathcal{T}_h} \|e_h(0)\|^2_{L^2(K)} + \left\{ \tilde{C}^{1/2}\tilde{c}_1 \left(\int_0^T (T-t) \sum_{K\in\mathcal{T}_h} \|h_K\, r_K(t)\|^2_{L^2(K)}\,\mathrm{d}t \right)^{1/2} + \right. \\
&+\ \left. (\hat{c}\,\tilde{C}/2)^{1/2}\tilde{c}_2 \left(\int_0^T (T-t) \sum_{S\in\mathcal{E}_h} \|h_S^{1/2}\,\varepsilon\, J_S(t)\|^2_{L^2(S)}\,\mathrm{d}t \right)^{1/2} \right\} \times \\
&\quad \times \left(\int_0^T (T-t) \sum_{K\in\mathcal{T}_h} |e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \right)^{1/2} .
\end{aligned}
$$

Now the use of Young's inequality already yields (4.57). *Q.E.D.*

Let us note that the statement of Remark 4.1 is true for the estimate from Theorem 4.8.

4.6 A Posteriori Error Estimates for Hyperbolic Problems

The approach treated in the previous sections can also be applied to second order hyperbolic equations, namely to the following initial-boundary value problem: Find $u : Q_T \to \mathbb{R}$ such that

$$
\frac{\partial^2 u}{\partial t^2} = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\, \frac{\partial u}{\partial x_j} \right) + f(x,t) \quad \text{in } Q_T, \tag{4.58}
$$

$$
u(x,0) = u^0(t), \quad \frac{\partial u}{\partial t}(x,0) = u^1(x), \quad x \in \Omega, \tag{4.59}
$$

$$
u(x,t) = 0, \quad x \in \partial\Omega,\ t \in (0,T). \tag{4.60}
$$

We use the same notation and assumptions as in Section 4.2. Only, instead of (4.4) we introduce the conditions

$$(4.61)\quad a_{ij} \in C^1(\overline{\Omega}),\ a_{ij} = a_{ji}, \quad c_0 \sum_{i=1}^{d} |\xi_i|^2 \le \sum_{i,j=1}^{d} a_{ij}(x)\, \xi_i\, \xi_j \le c_1 \sum_{i=1}^{d} |\xi_i|^2,$$

$$\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d,\ x \in \overline{\Omega},\ c_0,\ c_1 > 0 \text{ independent of } x,\ \xi,$$
$$f, \partial f/\partial t \in C(\overline{Q}_T),\ u^0 \in V,\ u^1 \in L^2(\Omega)$$

The weak formulation of this problem reads: find $u : Q_T \to \mathbb{R}$ satisfying

$$(4.62)\ \text{a)}\qquad u \in L^2(0,T;V), \quad \frac{\partial u}{\partial t} \in L^2(Q_T),$$

$$\text{b)}\qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} \int_\Omega u(t)\, v \, \mathrm{d}x + \int_\Omega \sum_{i,j=1}^{d} a_{ij} \frac{\partial u(t)}{\partial x_i} \frac{\partial v}{\partial x_j} \, \mathrm{d}x = \int_\Omega f(\cdot,t)\, v \, \mathrm{d}x$$

$$\forall\, v \in V, \quad \text{in the sense of distributions over } (0,T),$$

$$\text{c)}\qquad u(0) = u^0, \quad \frac{\partial u}{\partial t}(0) = u^1.$$

In what follows we shall assume the existence of a weak solution u satisfying (4.62) and

$$(4.63)\qquad \frac{\partial^2 u}{\partial t^2} \in L^2(Q_T), \quad \frac{\partial u}{\partial t} \in L^2(0,T;V).$$

In [59], Chap. 20, the existence and uniqueness of such a solution was proven.

Then from (4.62), b) using (4.12) we obtain the identity

$$(4.64)\quad \int_0^T \int_0^\tau \int_\Omega \frac{\partial^2 u(t)}{\partial t^2} v(t) \, \mathrm{d}x \mathrm{d}t \mathrm{d}\tau + \int_0^T \int_0^\tau \int_\Omega \sum_{i,j=1}^{d} a_{ij} \frac{\partial u(t)}{\partial x_i} \frac{\partial v(t)}{\partial x_j} \, \mathrm{d}x \mathrm{d}t \mathrm{d}\tau$$
$$= \int_0^T (T-t) \int_\Omega f(t)\, v(t) \, \mathrm{d}x \mathrm{d}t, \quad v \in \mathcal{V} = L^2(0,T;V).$$

The space semidiscretization of (4.62) leads to the definition of the *approximate solution*:

$$(4.65)\qquad \text{a)}\qquad u_h \in C^2([0,T];V_h),$$

$$\text{b)}\qquad \int_\Omega \frac{\partial^2 u_h(t)}{\partial t^2} v_h \, \mathrm{d}x + \int_\Omega \sum_{i,j=1}^{d} a_{ij} \frac{\partial u_h(t)}{\partial x_i} \frac{\partial v_h}{\partial x_j} \, \mathrm{d}x$$
$$= \int_\Omega f(t)\, v_h \, \mathrm{d}x \quad \forall\, v_h \in V_h,\ \forall\, t \in (0,T),$$

$$\text{c)}\qquad u_h(0) = u_h^0, \quad \frac{\partial u_h(0)}{\partial t} = u_h^1,$$

where u_h^0 and u_h^1 are V_h-approximations of u^0 and u^1, respectively. The above discrete problem (4.65) is obviously equivalent to an initial value problem for a system of second-order ordinary differential equations, which possesses a unique solution.

From (4.65), b) we obtain

$$\int_0^T \int_0^\tau \int_\Omega \frac{\partial^2 u_h(t)}{\partial t^2}\, v_h(t)\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \int_0^T \int_0^\tau \int_\Omega \sum_{i,j=1}^d a_{ij} \frac{\partial u_h(t)}{\partial x_i} \frac{\partial v_h(t)}{\partial x_j}\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau$$

(4.66)

$$= \int_0^T (T-t) \int_\Omega f(t)\, v_h(t)\,\mathrm{d}x\mathrm{d}t, \quad v_h \in \mathcal{V}_h = L^2(0,T;V_h).$$

In the sequel we shall use the notation

$$\mathcal{B}(u,v) = \int_0^T \int_0^\tau \int_\Omega \frac{\partial^2 u(t)}{\partial t^2}\, v(t)\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \int_0^T \int_0^\tau \int_\Omega \sum_{i,j=1}^d a_{ij} \frac{\partial u(t)}{\partial x_i} \frac{\partial v(t)}{\partial x_j}\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau,$$

(4.67)

$$u \in L^2(0,T;V),\ \frac{\partial^2 u}{\partial t^2} \in L^2(Q_T),\ v \in \mathcal{V},$$

(4.68) $$\mathcal{L}(v) = \int_0^T (T-t) \int_\Omega f(t)\, v(t)\,\mathrm{d}x\mathrm{d}t, \quad v \in \mathcal{V}.$$

Then (4.64) and (4.66) can be written in the form

(4.69) $$\mathcal{B}(u,v) = \mathcal{L}(v), \quad v \in \mathcal{V},$$

and

(4.70) $$\mathcal{B}(u_h,v_h) = \mathcal{L}(v_h), \quad v_h \in \mathcal{V}_h,$$

respectively.

Now we proceed in an analogous way as in Section 4.4. For $v \in \mathcal{V}$,

(4.71) $$\mathcal{B}(u,v) - \mathcal{B}(u_h,v) = \mathcal{L}(v) - \mathcal{B}(u_h,v) =: \mathcal{R}_h(v) = \text{residual}.$$

By (4.70), we have again the orthogonality property

(4.72) $$\mathcal{R}_h(v_h) = 0, \quad v_h \in \mathcal{V}_h.$$

Let us define *local residuals*

(4.73) $$r_K(x,t) = -\frac{\partial^2 u_h(x,t)}{\partial t^2} + \sum_{i,j=1}^d \frac{\partial}{\partial x_i}\left(a_{ij}(x)\,\frac{\partial u_h(x,t)}{\partial x_j}\right) + f(x,t),$$
$$x \in K,\ t \in (0,T),\ K \in \mathcal{T}_h,$$

and *local interelement jumps*:

$$(4.74)\, J_S(x,t) = -\sum_{i,j=1}^{d} a_{ij}(x)\left[\frac{\partial u(x,t)}{\partial x_i}\right]_S n_{Sj}, \quad x \in S,\ t \in (0,T),\ S \in \mathcal{E}_h,$$

where the jump $[\partial u/\partial x_i]_S$ is defined by (4.28). With the aid of Green's theorem and (4.12) we find that the residual $\mathcal{R}_h(v)$ can be expressed as in (4.30):

$$\mathcal{R}_h(v) = \int_0^T (T-t)\left\{\sum_{K\in\mathcal{T}_h}\int_K r_K(t)\,v(t)\,\mathrm{d}x + \sum_{S\in\mathcal{E}_h}\int_S J_S(t)\,v(t)\,\mathrm{d}S\right\}\mathrm{d}t,\ v \in \mathcal{V}. \tag{4.75}$$

In the sequel, we shall use the simplified notation $\dot{u} = \partial u/\partial t$, $\dot{u}_h = \partial u_h/\partial t$, $\ddot{u} = \partial^2 u/\partial t^2$, etc. In virtue of (4.61) and (4.65),

$$(4.76)\quad \begin{aligned} &u_h \in C^3(K\times[0,T]), \quad \dot{r}_K = \partial r_K/\partial t \in C(K\times[0,T]), \quad K \in \mathcal{T}_h,\\ &\dot{J}_S = \partial J_S/\partial t \in C(S\times[0,T]), \quad S \in \mathcal{E}_h. \end{aligned}$$

It follows from (4.62) and (4.63) that

$$(4.77)\qquad u \in C([0,T];V), \quad \partial u/\partial t \in C([0,T];L^2(\Omega)).$$

(The continuity is even absolute.)

Let us derive an *abstract a posteriori error estimate*.

Theorem 4.9. *For the error $e_h = u - u_h$ we have the following bound*:

$$(4.78)\qquad \begin{aligned} &\|\dot{e}_h\|^2_{L^2(Q_T)} + c_0\int_0^T |e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t \le \\ &\le T\|\dot{e}_h(0)\|^2_{L^2(\Omega)} + c_1 T|e_h(0)|^2_{H^1(\Omega)} + \\ &+2T\Big\{\sum_{K\in\mathcal{T}_h}\|h_K r_K(0)\|_{L^2(K)}\|h_K^{-1}(e_h(0)-v_h(0))\|_{L^2(K)} + \\ &+\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}J_S(0)\|_{L^2(S)}\|\,h_S^{-1/2}(e_h(0)-v_h(0))\|_{L^2(S)}\Big\} + \\ &+2\int_0^T\Big\{\sum_{K\in\mathcal{T}_h}\|h_K(r_K(t)-(T-t)\,\dot{r}_K(t))\|_{L^2(K)}\,\|h_K^{-1}(e_h(t)-v_h(t))\|_{L^2(K)} + \\ &+\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}(J_S(t)-(T-t)\,\dot{J}_S(t))\|_{L^2(S)}\,\|h_S^{-1/2}(e_h(t)-v_h(t))\|_{L^2(S)}\Big\}\mathrm{d}t, \\ &\qquad v_h \in C^1([0,T];V_h). \end{aligned}$$

Proof. If $v_h \in C^1([0,T];V_h)$, then $\dot v_h \in \mathcal{V}_h$. Substituting $v := \dot e_h \in C([0,T];L^2(\Omega)) \cap L^2(0,T;V)$ in (4.71) and using (4.72), we obtain the relation

$$\mathcal{B}(e_h, \dot e_h) = \mathcal{R}_h(\dot e_h - \dot v_h). \tag{4.79}$$

By (4.67) and Fubini's theorem,

$$\mathcal{B}(e_h,\dot e_h) = \int_0^T\int_0^\tau\int_\Omega \ddot e_h(t)\,\dot e_h(t)\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau + \int_0^T\int_0^\tau\int_\Omega \sum_{i,j=1}^d a_{ij}\,\frac{\partial e_h(t)}{\partial x_i}\,\frac{\partial \dot e_h(t)}{\partial x_j}\,\mathrm{d}x\mathrm{d}t\mathrm{d}\tau =$$

$$= \frac12\int_\Omega\int_0^T\int_0^\tau \frac{\partial}{\partial t}(\dot e_h(t))^2\,\mathrm{d}t\mathrm{d}\tau\mathrm{d}x + \frac12\int_\Omega\int_0^T\int_0^\tau \sum_{i,j=1}^d a_{ij}\frac{\partial}{\partial t}\left(\frac{\partial e_h(t)}{\partial x_i}\,\frac{\partial e_h(t)}{\partial x_j}\right)\mathrm{d}t\mathrm{d}\tau\mathrm{d}x =$$

$$= \frac12\int_{Q_T}(\dot e_h(t))^2\mathrm{d}x\mathrm{d}t - \frac12 T\int_\Omega(\dot e_h(0))^2\mathrm{d}x + \frac12\int_{Q_T}\sum_{i,j=1}^d a_{ij}\,\frac{\partial e_h(t)}{\partial x_i}\,\frac{\partial e_h(t)}{\partial x_j}\,\mathrm{d}x\mathrm{d}t -$$

$$- \frac12 T\int_\Omega \sum_{i,j=1}^d a_{ij}\,\frac{\partial e_h(0)}{\partial x_i}\,\frac{\partial e_h(0)}{\partial x_j}\,\mathrm{d}x\mathrm{d}t.$$

This and conditions (4.61) imply that

$$\begin{aligned}\mathcal{B}(e_h,\dot e_h) \;\ge\; & \frac12\|\dot e_h\|^2_{L^2(Q_T)} - \frac12 T\|\dot e_h(0)\|^2_{L^2(\Omega)} + \\ & + \frac{c_0}{2}\int_0^T |e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t - \frac{c_1}{2}T|e_h(0)|^2_{H^1(\Omega)}.\end{aligned} \tag{4.80}$$

Now let us estimate the right-hand side in (4.79). Starting from the representation (4.75) and using integration by parts with respect to time, we obtain

$$\begin{aligned}\mathcal{R}_h(\dot e_h - \dot v_h) = & \int_0^T (T-t)\Big\{\sum_{K\in\mathcal{T}_h}\int_K r_K(t)\,(\dot e_h(t)-\dot v_h(t))\,\mathrm{d}x + \\ & + \sum_{S\in\mathcal{E}_h}\int_S J_S(t)\,(\dot e_h(t)-\dot v_h(t))\,\mathrm{d}S\Big\}\mathrm{d}t =\end{aligned} \tag{4.81}$$

$$= -T\Big\{\sum_{K\in\mathcal{T}_h}\int_K r_K(0)\,(e_h(0)-v_h(0))\,\mathrm{d}x + \sum_{S\in\mathcal{E}_h}\int_S J_S(0)\,(e_h(0)-v_h(0))\,\mathrm{d}S\Big\} +$$

$$+ \int_0^T\Big\{\sum_{K\in\mathcal{T}_h}\int_K (r_K(t)-(T-t)\dot r_K(t))\,(e_h(t)-v_h(t))\,\mathrm{d}x +$$

$$+ \sum_{S\in\mathcal{E}_h} \int_S (J_S(t) - (T-t)\,\dot{J}_S(t))\,(e_h(t) - v_h(t))\,\mathrm{d}S\Big\}\mathrm{d}t.$$

The application of the Cauchy inequality to individual integrals in (4.81) yields the estimate

$$\mathcal{R}_h(\dot{e}_h - \dot{v}_n) \le T\Big\{ \sum_{K\in\mathcal{T}_h} \|h_K r_K(0)\|_{L^2(K)} \|h_K^{-1}(e_h(0) - v_h(0))\|_{L^2(K)} +$$

$$+ \sum_{S\in\mathcal{E}_h} \|h_S^{1/2} J_S(0)\|_{L^2(S)}\, \|h_S^{-1/2}(e_h(0) - v_h(0))\|_{L^2(S)}\Big\} +$$

$$+ \int_0^T \Big\{ \sum_{K\in\mathcal{T}_h} \|h_K(r_K(t) - (T-t)\,\dot{r}_K(t))\|_{L^2(K)} \|h_K^{-1}(e_h(t) - v_h(t))\|_{L^2(K)} +$$

$$+ \sum_{S\in\mathcal{E}_h} \|h_S^{1/2}(J_S(t) - (T-t)\,\dot{J}_S(t))\|_{L^2(S)}\, \|h_S^{-1/2}(e_h(t) - v_h(t))\|_{L^2(S)}\Big\}\mathrm{d}t.$$

This and (4.80) immediately give (4.78). *Q.E.D.*

From Theorem 4.9 we obtain the a posteriori error estimates:

Theorem 4.10. *If $d = 1$, then*

$$\|\dot{e}_h\|^2_{L^2(Q_T)} + \frac{c_0}{2}\int_0^T |e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t \le \tag{4.82}$$

$$\le T\|\dot{e}_h(0)\|^2_{L^2(\Omega)} + T(c_1+1)\,|e_h(0)|^2_{H^1(\Omega)} + \frac{T}{\pi^2}\sum_{i=1}^N \|h_{K_i} r_{K_i}(0)\|^2_{L^2(K_i)} +$$

$$+ \frac{4}{\pi^2 c_0}\int_0^T \sum_{i=1}^N \|h_{K_i}(r_{K_i}(t) - (T-t)\,\dot{r}_{K_i}(t))\|^2_{L^2(K_i)}.$$

Proof. For $d = 1$, we have $\sum_{K\in\mathcal{T}_h} = \sum_{i=1}^N$. Let us set $v_h := I_h e_h \in C^1([0,T]; V_h)$ (Lagrange interpolation). Then in virtue of (4.33), the terms in (4.78) with jumps vanish. The use of (4.34) and the Cauchy and Young's inequalities yields the estimates

$$2\sum_{K_i\in\mathcal{T}_h} \|h_{K_i} r_{K_i}(0)\|_{L^2(K_i)}\, \|h_{K_i}^{-1}(e_h(0) - I_h e_h(0))\|_{L^2(K_i)} \le \tag{4.83}$$

$$\le \frac{1}{\pi^2}\sum_{i=1}^N \|h_{k_i} r_{K_i}(0)\|^2_{L^2(K_i)} + \sum_{i=1}^N |e_h(0)|^2_{H^1(K_i)} =$$

$$= \frac{1}{\pi^2}\sum_{i=1}^{N}\|h_{K_i} r_{K_i}(0)\|^2_{L^2(K_i)} + |e_h(0)|^2_{H^1(\Omega)}.$$

Similarly we find that

$$2\int_0^T \sum_{K\in\mathcal{T}_h} \|h_K(r_K(t)-(T-t)\,\dot r_K(t))\|_{L^2(K)}\|h_K^{-1}(e_h(t)-I_h e_h(t))\|_{L^2(K)} \le$$

(4.84)

$$\le \frac{4}{\pi^2 c_0}\int_0^T \sum_{i=1}^{N}\|h_{K_i}(r_{K_i}(t)-(T-t)\,\dot r_{K_i}(t))\|^2_{L^2(K_i)} + \frac{c_0}{2}\int_0^T \sum_{i=1}^{N}|e_h(t)|^2_{H^1(K_i)}\mathrm{d}t.$$

Now (4.78), where $v_h = I_h e_h$, together with (4.83) and (4.84) already imply (4.82). *Q.E.D.*

Theorem 4.11. *Under Assumptions* 4.1 *and* 2, *for* $d \ge 2$, *the following a posteriori error estimate holds:*

(4.85)
$$\|\dot e_h\|^2_{L^2(Q_T)} + \frac{c_0}{2}\int_0^T |e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t \le$$

$$\le T\|\dot e_h(0)\|^2_{L^2(\Omega)} + T(c_1+1)\,|e_h(0)|^2_{H^1(\Omega)} + T\Big\{(c_0 c^*/2)\sum_{K\in\mathcal{T}_h}\|h_K r_K(0)\|^2_{L^2(K)} +$$

$$+(c_0 c^{**}/2)\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}J_S(0)\|^2_{L^2(\Omega)}\Big\}+$$

$$+c^*\int_0^T \sum_{K\in\mathcal{T}_h}\|h_K(r_K(t)-(T-t)\,\dot r_K(t))\|^2_{L^2(K)}\mathrm{d}t+$$

$$+c^{**}\int_0^T \sum_{S\in\mathcal{E}_h}\|h_S^{1/2}(J_S(t)-(T-t)\,\dot J_S(t))\|^2_{L^2(S)}\,\mathrm{d}t,$$

where $c^* = 4\tilde C\tilde c_1^2/c_0$ *and* $c^{**} = 2\hat c\tilde C\tilde c_2^2/c_0$.

Proof. From the abstract estimate (4.78), where we put $v_h := I_h e_h =$ Clement's interpolation, (4.38) and the Cauchy inequality, using the same notation and similar process as in the proof of Theorem 4.4, we obtain

$$\|\dot e_h\|^2_{L^2(Q_T)} + c_0\int_0^T |e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t \le$$

$$\le T\|\dot e_h(0)\|^2_{L^2(\Omega)} + c_1 T|e_h(0)|^2_{H^1(\Omega)} + 2T\Big\{\tilde C^{1/2}\tilde c_1\Big(\sum_{K\in\mathcal{T}_h}\|h_K r_K(0)\|^2_{L^2(K)}\Big)^{1/2}+$$

$$+(\hat{c}\tilde{C}/2)^{1/2}\tilde{c}_2\Big(\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}J_S(0)\|^2_{L^2(S)}\Big)^{1/2}\Big\}\Big(\sum_{S\in\mathcal{E}_h}|e_h(0)|^2_{H^1(K)}\Big)^{1/2}+$$

$$+2\Big\{\tilde{C}^{1/2}\tilde{c}_1\Big(\int_0^T\sum_{K\in\mathcal{T}_h}\|h_K(r_K(t)-(T-t)\,\dot{r}_K(t))\|^2_{L^2(K)}\mathrm{d}t\Big)^{1/2}+$$

$$+(\hat{c}\tilde{C}/2)^{1/2}\tilde{c}_2\Big(\int_0^T\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}(J_S(t)-(T-t)\,\dot{J}_S(t))\|^2_{L^2(S)}\mathrm{d}t\Big)^{1/2}\Big\}\times$$

$$\times\Big(\int_0^T|e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t\Big)^{1/2}\le$$

$$\le T\|\dot{e}_h(0)\|^2_{L^2(\Omega)}+T(c_1+1)|e_h(0)|^2_{H^1(\Omega)}+T\Big\{2\tilde{C}\tilde{c}_1^2\sum_{K\in\mathcal{T}_h}\|h_Kr_K(0)\|^2_{L^2(K)}+$$

$$+\hat{c}\tilde{C}\tilde{c}_2^2\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}J_S(0)\|^2_{L^2(S)}\Big\}+\frac{c_0}{2}\int_0^T|e_h(t)|^2_{H^1(\Omega)}\mathrm{d}t+$$

$$+\frac{4\tilde{C}\tilde{c}_1^2}{c_0}\int_0^T\sum_{K\in\mathcal{T}_h}\|h_K(r_K(t)-(T-t)\,\dot{r}_K(t))\|^2_{L^2(K)}\mathrm{d}t+$$

$$+\frac{2\hat{c}\tilde{C}\tilde{c}_2^2}{c_0}\int_0^T\sum_{S\in\mathcal{E}_h}\|h_S^{1/2}(J_S(t)-(T-t)\,\dot{J}_S(t))\|^2_{L^2(S)}\mathrm{d}t.$$

As we see, (4.85) holds. *Q.E.D.*

Estimate (4.85) is introduced here because of completness. However, its practical applicability is limited in virtue of Remark 4.1.

4.7 Examples and Improved Estimates

The numerical verification of the described a posteriori error estimates is in progress. Here we give several examples of the application of the derived estimates to parabolic problems with one space dimension. For more complex situations, see [5].

I) First we consider the *linear heat conduction problem*

(4.86)
$$\begin{aligned}
&\text{a)}\quad \frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+f(x,t),\quad x\in(0,1),\ t\in(0,T),\\
&\text{b)}\quad u(0,t)=u(1,t)=0,\quad t\in(0,T),\\
&\text{c)}\quad u(x,0)=u^0(x),\qquad x\in(0,1),
\end{aligned}$$

with solution

$$u(x,t) = (1-t)\sin \pi x \tag{4.87}$$

and corresponding right-hand side f and initial condition u^0.

Problem (4.86) was discretized in space by conforming piecewise linear finite elements on a uniform partition of the interval $(0,1)$, formed by $N =$ 10, 20, 50, 100 subintervals. The resulting system of ordinary differential equations was solved by the explicit Euler method with a very small time step $\tau = 10^{-5}$.

In this case the a posteriori error estimate (4.35) was used with $\alpha = 1$. The integrals were evaluated by quadrature formulae. The integration in time uses the trapezoidal rule with step $\tau = 10^{-5}$. In space, the integrals of quadratic functions were evaluated exactly.

Problem (4.86) was solved on the time interval $(0,T)$ with $T = 0.2,\ 0.4,\ 1.0,\ 2.0,\ 5.0$. Always the *effectivity index* σ_{eff} was evaluated:

$$\sigma_{\mathrm{eff}} = E_h / |||e_h|||, \tag{4.88}$$

where E_h is the error estimator:

$$E_h = \left(T\|e_h(0)\|^2_{L^2(\Omega)} + \frac{1}{\alpha\pi^2}\int_0^T (T-t) \sum_{i=1}^{N} \|h_{K_i} r_{K_i}(t)\|^2_{L^2(K_i)}\,\mathrm{d}t \right)^{1/2} \tag{4.89}$$

and

$$|||e_h||| = \left(\|e_h\|^2_{L^2(Q_T)} + \alpha \int_0^T (T-t)\,|e_h|^2_{H^1(\Omega)}\,\mathrm{d}t \right)^{1/2}, \tag{4.90}$$

(cf. (4.35)), where $\alpha = 1$. The following table shows the effectivity index for different cases of T and density of the partition of the interval $(0,1)$.

N \ T	0.2	0.4	1.0	2.0	5.0
10	1.103744	1.104311	1.104689	1.104752	1.104567
20	1.102918	1.103058	1.103150	1.103169	1.103140
50	1.102685	1.102706	1.102718	1.102726	1.102740
100	1.102651	1.102656	1.102656	1.102662	1.102683

Table 4.1: Effectivity index for estimate (4.35)

As we see, the effectivity index remains nearly constant, i.e., independent of the length of time interval and the density of the space partition. These results are quite satisfactory, but they can be still improved.

In fact, the constant C_1 from (4.34) is the eigenvalue of the problem

$$\begin{aligned} -u'' &= \lambda u \quad \text{in } (0,1), \\ u(0) &= u(1) = 0. \end{aligned} \tag{4.91}$$

Estimate (4.34) yields a guaranteed error estimate. The function realizing (4.34) is $v = \sin \frac{\pi}{k} x$ which leads to the present estimate.

Let us derive the asymptotic estimate. Note that the estimate follows from the evaluation of the term $\sum_{i=1}^N \int_{K_i} r_{K_i}(e_h - I_h e_h)\,\mathrm{d}x$. If the solution is smooth, then e_h is also smooth and so is r_h. Hence, asymptotically we can take $r_{K_i} \approx a =$ constant on each K_i and asymptotically, instead of the constant $C_1 = 1/\pi$ in (4.34) we get

$$C_1 = \frac{\left(\int_0^1 x(1-x)\,\mathrm{d}x\right)^{1/2}}{\left(\int_0^1 (1-2x)^2\,\mathrm{d}x\right)^{1/2}} = \left(\frac{1}{12}\right)^{1/2}. \tag{4.92}$$

Then we obtain the a posteriori error estimate in the form

$$|||e_h|||^2 \le T\|e_h(0)\|^2_{L^2(\Omega)} + \frac{1}{12}\int_0^T (T-t)\sum_{i=1}^N \|h_{K_i}\, r_{K_i}(t)\|^2_{L^2(K_i)}\,\mathrm{d}t. \tag{4.93}$$

Similar estimates can be obtained for parabolic convection-diffusion problems and hyperbolic problems. For example, instead of estimate (4.56) we get

$$\begin{aligned} &(1-c_b)\,\|e_h\|^2_{L^2(Q_T)} + \varepsilon\int_0^T (T-t)\,|e_h(t)|^2_{H^1(\Omega)}\,\mathrm{d}t \le \\ &\le T\|e_h(0)\|^2_{L^2(\Omega)} + \frac{1}{12}\int_0^T (T-t)\sum_{i=1}^N \left\|\frac{h_{K_i}}{\sqrt{\varepsilon}}\, r_{K_i}(t)\right\|^2_{L^2(K_i)}\mathrm{d}t. \end{aligned} \tag{4.94}$$

The use of estimate (4.93) leads to results shown in Table 4.2 indicating that estimate (4.93) is asymptotically exact.

N \ T	0.2	0.4	1.0	2.0	5.0
10	1.002099	1.002143	1.002235	1.002194	1.001753
20	1.000517	1.000525	1.000545	1.000538	1.000442
50	1.000069	1.000070	1.000070	1.000074	1.000075
100	1.000005	1.000005	1.000003	1.000007	1.000023

Table 4.2: Effectivity index for estimate (4.93)

II) The second test problem is the *linear convection-diffusion equation*

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial u}{\partial x} + f, \quad x \in (0,1),\ t \in (0,T), \tag{4.95}$$

with constant coefficients $\alpha > 0$, $\beta \in \mathbb{R}$, equipped with conditions (4.86), b) – c) with solution (4.87) and corresponding right-hand side f and initial condition u^0.

This problem is discretized in the same way as (4.86). In the case $\alpha = 1$, $\beta = 1$ (i.e., $\epsilon = \alpha = 1$) estimate (4.56) leads to the effectivity index ≈ 1.1. Table 4.3 contains the results obtained on the basis of estimate (4.94). In Table 4.4 the effectivity index for estimate (4.94) is shown for the case $\alpha = 1$, $\beta = 100$. In order to follow the behaviour of the effectivity index in dependence on the density of the partition of the space interval $(0,1)$, we consider here $N = 10, 20, 50, 100$. In these examples the constant c_b from (4.47) is zero, the error estimator E_h is defined by

$$E_h = \left(T\|e_h(0)\|^2_{L^2(\Omega)} + \frac{1}{12}\int_0^T (T-t) \sum_{i=1}^{N} \left\| \frac{h_{K_i}}{\sqrt{\varepsilon}}\, r_{K_i}(t) \right\|^2_{L^2(K_i)} \mathrm{d}t \right)^{1/2} \tag{4.96}$$

and the norm of the error is given by

$$|||e_h||| = \left(\|e_h\|^2_{L^2(Q_T)} + \varepsilon \int_0^T (T-t)\, |e_h(t)|^2_{H^1(\Omega)}\, \mathrm{d}t \right)^{1/2}. \tag{4.97}$$

N \ T	0.2	0.4	1.0	2.0	5.0
10	1.002561	1.002617	1.002716	1.002674	1.002238
20	1.000634	1.000645	1.000666	1.000659	1.000564
50	1.000088	1.000089	1.000090	1.000093	1.000095
100	1.000010	1.000009	1.000007	1.000012	1.000028

Table 4.3: Effectivity index corresponding to estimate (4.94) for $\alpha = \beta = 1$

N \ T	0.2	0.4	1.0	2.0	5.0
10	3.053230	3.055344	3.056544	3.056646	3.370285
20	1.756602	1.756739	1.756823	1.756806	1.756706
50	1.154948	1.154927	1.154916	1.154912	1.154899
100	1.040894	1.040886	1.040879	1.040882	1.040892

Table 4.4: Effectivity index corresponding to estimate (4.94) for $\alpha = 1, \beta = 100$

From Tables 4.3 and 4.4 it is possible to conclude that estimate (4.94) is asymptotically exact.

Bibliography

[1] S. Adjerid, J. E. Flaherty: A moving finite element method with error estimation and refinement for one-dimensional time dependent partial differential equations. SIAM J. Numer. Anal. 23 (1986), 778–796.

[2] S. Adjerid, J. E. Flaherty, Y. J. Wang: A posteriori error estimation with finite element methods of lines for one-dimensional parabolic systems. Tech. Report 91–1, Troy, NY, Dept. of Computer Science, Rensselar Polytechnic Institute, 1991.

[3] I. Babuška, R. Durán, R. Rodríguez: Analysis of the efficiency of an a posteriori error estmator for linear triangular finite elements. SIAM J. Numer. Anal. 29 (1992), 947-964.

[4] I. Babuška, W. Gui: Basic principles of feedback and adaptive approaches in the finite element method. Comput. Methods Appl. Mech. Engrg. 55 (1986), 27–42.

[5] I. Babuška, S. Ohnimus: A posteriori error estimation for the semidiscrete finite element method of parabolic differential equations. Preprint, TICAM, Univ. of Texas at Austin, 1999.

[6] I. Babuška, W. C. Rheinboldt: A-posteriori error estimates for the finite element method. Internat. J. Numer. Methods Engrg. 12 (1978), 1597–1615.

[7] I. Babuška, T. Strouboulis: *Finite Element Method and its Reliability*. Manuscript for Oxford University Press.

[8] I.Babuška and M.Suri, *The p- and h-p versions of the Finite Element Method*, Computer Methods in Applied Mechanics and Engineering, 80(1990), 5-26, North Holland

[9] R. E. Bank, A. Weiser: Some a posteriori error estimators for elliptic partial differential equations. Math. Comp. 44 (1985), 283–301.

[10] R. Becker, R. Rannacher: A feed-back approach to error control in finite element methods: basic analysis and examples. Preprint 96–52 (SFB 359), Heidelberg, 1996.

[11] M. Beiterman, I. Babuška: An adaptive method of lines with error control for parabolic equations of the reaction-diffusion type. J. Comput. Phys. 63 (1986), 33–66.

[12] M. Brdička, *Continuum mechanics*, ČSAV, Prague, 1959.

[13] P.N.Brown, G.D. Byrne and A.C.Hindmarsh, *VODE, a Variable-Coefficient ODE Solver*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 1038-1051.

[14] P.N.Brown and A.C.Hindmarsh, *Reduced Storage Matrix Methods in Stiff ODE Systems*, J. Appl. Math.&Comput. 31 (1989), pp. 40-91.

[15] J.C.Butcher, *The numerical analysis of ordinary differential equations*, John Wiley and Sons, NY, 1987

[16] C. Carstensen, S. A. Funken: Constants in Clement-interpolation error and residual based a posteriori estimates in finite element methods. Bericht 97-11 (August 1997), Christian-Albrechts-Universität zu Kiel.

[17] F. Chalot amd T.J.R. Hughes, *Analysis of hypersonic flows in thermomechanical equilibrium by application of the Galerkin Least Square formulation*, In R.E.O'Malley, Jr., editor, Proc. 2nd International Conference for Industrial and Applied Mathematics, SIAM, pp. 146-159, 1992.

[18] P. G. Ciarlet: *The Finite Element Method for Elliptic Problems.* North–Holland, Amsterdam, 1979.

[19] P. Clément: Approximation by finite element functions using local regularization. RAIRO Anal. Numér. 9 (1975), 77–84.

[20] A. Coddington and N. Levinson. *Ordinary Differential Equations*, New York, McGraw–Hill, 1955.

[21] L. J. Durlofsky, B. Enquist, and S. Osher, *Triangle based adaptive stencils for the solution of hyperbolic conservation laws.* J. Comput. Phys., 98:64–73, 1992.

[22] B. Einfeldt, *On Godunov-type methods for gas-dynamics*, SIAM J. Numer.Anal, 25 (1988), pp. 357–393.

[23] B. Einfeldt, C.C.Munz, P.L.Roe and B.Sjorgreen, *On Godunov-type methods near low densities*, J. Comput. Phys., 92 (1991), pp. 273–295.

[24] K. Eriksson, C. Johnson: Adaptive finite element methods for linear elliptic problems. Math. Comp. 50 (1988), 361–383.

[25] K. Eriksson, C. Johnson: Adaptive finite element methods for parabolic problems I: A linear model problem. SIAM J. Numer. Anal. 28 (1991), 43–77.

[26] K. Eriksson, C. Johnson: Adaptive finite element methods for parabolic problems IV: Nonlinear problems. SIAM J. Numer. Anal. 32 (1995), 1729–1749.

[27] Miloslav Feistauer, *Mathematical Methods in Fluid Dynamics.* Longman Scientific & Technical, Harlow, 1993.

[28] J. Felcman and P. Šolín. *Construction of the Osher-Solomon Scheme for 3D Euler Equations.* East-West Journal of Numerical Mathematics, Vol. 6, No. 1, 1998.

[29] J. Fořt, K. Kozel and M. Vavřincová: Numerical simulation of transonic flows through cascades using cell centered and cell vertex type schemes, Acta Technica, Vol. 33, pp. 37 – 59, 1993.

[30] J. Fořt, M. Huněk, K. Kozel and M. Vavřincová: Numerical simulation of steady and unsteady flow through plane cascades, Numerical Modelling in Continuous Mechanics-II (M. Feistauer, K. Kozel, and R. Rannacher – Eds), Charles University, 1995, Prague.

[31] R. M. Furzeland, J. G. Verwer, P. A. Zegeling: A numerical study of three moving grid methods for one-dimensional partial differential equations which are based on the method of lines. J. Comput. Phys. 89 (1990), 349–388.

[32] H. Gajewski, K. Gröger, K. Zacharias: *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen.* Akademie–Verlag, Berlin, 1994.

[33] B. Gustafsson and A. Sundström, *Incompletely parabolic systems in fluid dynamics*, SIAM J. Appl. Math., Vol. 35, No. 2, pp. 343-357, 1978

[34] Martin H. Gutknecht, *Lanczos-type solvers for non-symmetric linear systems of equations*, Technical report, Swiss center for scientific computing, ETH Zürich, 1997.

[35] P. Hansbo, *Explicit streamline diffusion finite element methods for the compressible euler equations in conservation variables* , Department of mathematics, Chalmers University of technology, Göteborg, Preprint No. 1991:31/ISSN 0347-2809,

[36] P. Hansbo and C. Johnson, *Adaptive streamline-diffusion methods for compressible flow using conservation variables*, Computer Methods in Applied Mechanics and Engineering, 87: 267-280, 1991

[37] G. Hauke and T.J.R. Hughes, *A unified approach to compressible and incompressible flows*, Computer Methods in Applied Mechanics and Engineering, 113: 389-395, 1994

[38] F.-K. Hebeker, C. Führer, R. Rannacher: An adaptive finite element method for unsteady convection-dominated flows with stiff source terms. Preprint 97 - 15 (SFB 359) März 1997, Interdisciplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg.

[39] A. C. Hindmarsh, *ODEPACK, a systematized collection of ODE solvers.* Scientific Computing (IMACS Transactions on Scientific Computation Vol. 1). Amsterdam, North-Holland 1983, 55-64.

[40] A. C. Hindmarsch: LSODE and LSODI, two new initial value ordinary differential equation solvers. ACMSIGNUM Newsletter 15 (1980), 10–11.

[41] T.J.R. Hughes, L.P.Franca and M. Mallet, *A new finite element formulation for computational fluid dynamics: I.Symmetric forms of the*

compressible Navier-Stokes equations and the second law of thermodynamics, Computer Methods in Applied Mechanics and Engineering, 54: 223-234, 1986.

[42] M. Huněk, K. Kozel and M. Vavřincová: Numerical solution of transonic potential flow in a 2D compressor using multi-grid techniques, Proceedings of IV. GAMM Seminar, 1988, Kiel.

[43] K. Kozel and M. Vavřincová: Finite volume method for the Euler equations, Proceedings of Second ISNA Conference, 1987, Prague"

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

[45] J. Kurzweil, *Ordinary Differential Equations*, SNTL Prague, 1978

[46] J.D.Lambert. *Computational Methods in Ordinary Differential Equations*, London–New York, J. Wiley & Sons, 1973.

[47] Le Veque, R. J., *Numerical Methods for Conservation Laws* BirkhHauser Verlag, Basel – Boston – Berlin, 1990.

[48] J. L. Lions: *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod, Paris, 1969.

[49] M.-S. Liou and C.J.Steffen, *A new flux-splitting scheme*, J. Comput. Phys., 107 (1993), pp. 23–29.

[50] P. K. Moore: A posteriori error estimation with finite element semi- and fully discrete methods for nonlinear parabolic equations in one space dimension. SIAM J. Numer. Anal. 31 (1994), 149–169.

[51] R. H. Nochetto, G. Savaré, C. Verdi: A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations. Report No. 20, 1997/98, Institut Mittag–Leffler.

[52] J. Nordström, *The influence of open boundary conditions on the convergence to steady state for the Navier-Stokes equations*, J. Comp. Phys., Vol 85, pp. 210-244, 1989

[53] J.T.Oden, *Theory and implementation of high-order adaptive h-p methods for the analysis of incompressible viscous flow*, Texas Institute for Computational Mechanics, The University of Texas at Austin

[54] J. T. Oden, L. Demkowicz, W. Rachowicz, T. A. Westermann: Towards a universal h – p adaptive finite element strategy, Part 2. A posteriori error estimation. Comp. Meth. Appl. Mech. Engrg 77 (1989), 113–180.

[55] J. Oliger and A. Sundström, *Theoretical and practical aspects of some initial boundary value problems in fluid dynamics*, SIAM J.Appl.Math., Vol.35, No.3, pp. 419-446, 1978

[56] Osher, S., Solomon, F., *Upwind difference schemes for hyperbolic systems of conservation laws*, Math. Comput. 38 (1982), 339–374.

[57] L. R. Petzold: A description of DDASSL: A differential/algebraic system solver. Sandia Report No. Sand 82-8637, Livermore, CA, Sandia National Laboratory, 1982.

[58] T.J.Poinsot and S.K.Lele, *Boundary conditions for direct simulations of compressible viscous flows*, J. Comp. Phys., Vol 101, pp. 104 - 129, 1992

[59] K. Rektorys: *The Method of Discretization in Time and Partial Differential Equations.* Reidel Publishing Company, Dordrecht – Boston – London, 1982.

[60] H.-G. Roos, M. Stynes, L. Tobiska: *Numerical Methods for Singularly Pertubed Differential Equations (Convection-Diffusion and Flow Problems).* Springer-Verlag, Berlin, 1996.

[61] K. Segeth: Grid adjustment based on a posteriori error estimates. Appl. Math. 38 (1993), 488–504.

[62] K. Segeth: A posteriori error estimates for parabolic differential systems solved by the finite element method of lines. Appl. Math. 39 (1994), 415–443.

[63] K. Segeth: A posteriori error estimation with the finite element method of lines for a nonlinear parabolic equation in one space dimension. Preprint, Mathematical Institute of the Academy of Sciences, Prague, 1998.

[64] G. Sewell, *The numerical solution of ordinary and partial differential equations*, Academic Press, Inc., Harcourt Brace Jovanovich, Publishers.

[65] F. Shakib, T.J.R. Hughes and Z. Johan, *A new finite element formulation for computational fluid dynamics: X. The compressible Euler and Navier-Stokes equations*, Computer Methods in Applied Mechanics and Engineering, 89: 141-219, 1991

[66] G. A. Sod, *A survey of several finite difference methods for system of nonlinear hyperbolic conservation laws*, J. Comput. Phys., 27:1–31, 1978.

[67] S.P.Spekreijse. *Multigrid Solution of the Steady Euler Equations.* PhD Thesis, Center for Mathematics and Computer Science, Amsterdam, 1988.

[68] Steger, J. L., Warming, R. F., *Flux vector splitting of the inviscid gas dynamics equations with applications to finite difference methods*, J. Comput. Phys. 40 (1981), 263–293.

[69] J. Stiller, *Zum Randwertproblem fr̈ die kompressiblen Navier-Stokes Gleichungen*, Technical Report I 1505, Technische Universität Dresden, Institut für Strömungsmechanik, März 1995

[70] J.C. Strikwerda, *Initial Boundary Value Problems for Incompletely Parabolic Systems*, Comm. Pure Appl. Math., Vol. 30, pp. 797-822

[71] T. Strouboulis, J. T. Oden: A posteriori error estimation of finite element approximations in fluid mechanics. Comp. Meth. Appl. Mech. Engrg 78 (1990), 201–242.

[72] V. Thomée: *Galerkin Finite Element Methods for Parabolic Problems.* Springer-Verlag, Berlin, 1984.

[73] P. Tichy, *Behaviour of some BiCG and CGS algorithms*, Master thesis, Charles University, Prague, 1997.

[74] J. W. Van der Burg, *Numerical Techniques for Transonic Flow Computations*, University of Twente, 1992.

[75] R. Verfürth: *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* Willey – Teubner, Chichester, 1996.

[76] R. Verfürth: A posteriori error estimate for nonlinear problems: $L^r(0,T;W^{1,\rho}(\Omega))-$ error estimates for finite element discretizations

of parabolic equations. Numerical Methods Partial Differential Eq 14 (1998), 487–518.

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

[78] Y. Wada and M.-S. Liou, *An accurate and robust flux-splitting scheme for shock and contact discontinuities*, SIAM J. Sci. Comput., 18 (1997), pp. 633-657.

[79] J. Z. Zhu, O. C. Zienkiewicz: Adaptive techniques in the finite element method. Commun. Appl. Numer. Methods 4 (1988), 197–204.