

**Raport Badawczy**  
**Research Report**

**RB/16/2017**

**Drag minimization for the obstacle  
in compressible flow using shape  
derivatives and finite volume -  
finite element method**

**A. Kaźmierczak, J. Sokołowski  
A. Żochowski**

**Instytut Badań Systemowych**  
**Polska Akademia Nauk**

**Systems Research Institute**  
**Polish Academy of Sciences**



# **POLSKA AKADEMIA NAUK**

## **Instytut Badań Systemowych**

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:  
Prof. dr hab. inż. Antoni Żochowski

Warszawa 2016

# DRAG MINIMIZATION FOR THE OBSTACLE IN COMPRESSIBLE FLOW USING SHAPE DERIVATIVES AND FINITE VOLUME – FINITE ELEMENT METHOD

ANNA KAŻMIERCZAK

Faculty of Mathematics and Computer Science  
University of Lodz, Poland

JAN SOKOŁOWSKI

Institut Élie Cartan  
Laboratoire de Mathématiques  
Université Henri Poincaré Nancy 1

ANTONI ŻOCHOWSKI

Systems Research Institute  
Polish Academy of Sciences  
Warsaw, Poland

**ABSTRACT.** In the paper the shape optimization problem for the static, compressible Navier–Stokes equations is analyzed. The drag minimizing of an obstacle immersed in the gas stream is considered. The continuous gradient of the drag is obtained by application of the sensitivity formulas derived in the works of one of the co-authors. The numerical approximation scheme uses mixed Finite Volume – Finite Element formulation. The novelty of our numerical method is a particular choice of the regularizing term for the non-homogeneous Stokes boundary value problem, which may be tuned to obtain the best accuracy. The convergence of the discrete solutions for the model under considerations is proved. The non-linearity of the model is treated by means of the fixed point procedure. The numerical example of an optimal shape is given.

**1. Introduction.** One of the main applications of the theory of compressible viscous flows [27] is the optimal shape design in aerodynamics. The classical example is the minimization problem of the drag of an airfoil travelling in the atmosphere with uniform speed.

In the paper we present numerical computations for a simple geometry in the plane. The theory leading to the formulas for the shape derivatives is based on a series of papers by P.I. Plotnikov and J. Sokolowski [23]–[26], while computational scheme is based on [28]. For the approximation of solutions the method of Finite Volumes is used, what required supplementing the results of [10],[7] with some additional analysis. This analysis constitutes the main novelty of the paper.

---

2010 *Mathematics Subject Classification.* Primary: 49Q10, 35Q30; Secondary: 74S10.

*Key words and phrases.* shape optimization, compressible Navier-Stokes, drag minimization, Finite Volume Method, Finite Element Method .

The optimal control of compressible flows is still poorly investigated, in contrast to incompressible flows. We can cite here see e.g. [6], [29] or a more recent work [5] where the dynamic version of the problem was considered.

**2. Mathematical model.** Let  $B \subset \mathbb{R}^2$  be a doubly connected, bounded *hold all domain* with a smooth boundary  $\Sigma = \partial B$ . We denote two connected components of  $\partial B$  by  $\Sigma_{in}$  and  $\Sigma_{out}$ ,  $\Sigma = \Sigma_{in} \cup \Sigma_{out}$ . Suppose that  $B$  contains an obstacle  $S$ , which is a compact set with a sufficiently smooth boundary  $\Gamma$ . Suppose also that the volume of  $S$  is given and small with respect to the volume of  $B$  and the obstacle is far from the boundary of  $B$ . We define the computational domain as  $\Omega = B \setminus S$ . Thus the boundary of  $\Omega$  consists of three parts:  $\partial\Omega = \Sigma_{in} \cup \Sigma_{out} \cup \Gamma$ . We assume that a viscous, compressible fluid is flowing across  $\Omega$ . Fluid enters  $\Omega$  through  $\Sigma_{in}$  and leaves through  $\Sigma_{out}$ . The computational domain is presented in Fig. 1.

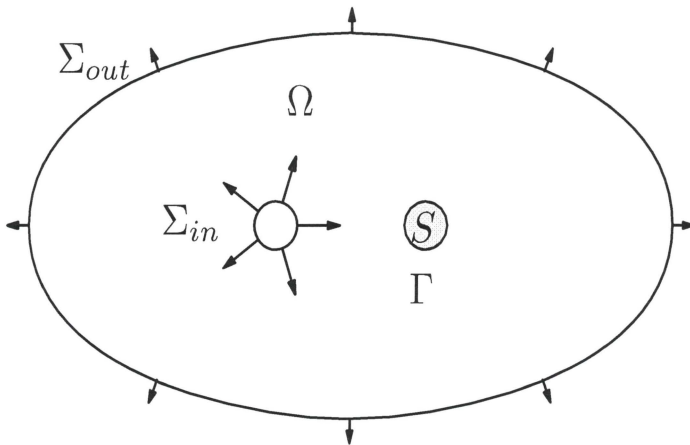


Figure 1. Computational domain  $\Omega = B \setminus S$

We restrict our consideration to stationary flows in the framework of the existence theory established in [27], Theorem 11.2.6 on page 311. In fact we need the existence of unique strong solutions to the model under considerations in order to assure the weak differentiability of solutions with respect to small parameter.

The dynamics of the fluid is determined by three functions: the velocity field  $\mathbf{u}$ , the density  $\varrho$  and the pressure  $p(\varrho)$ , where  $p = p(\varrho)$  is a smooth, strictly monotone function of the density with  $p(0) = 0$ . In computational experiments we set  $p = p_0(\frac{\varrho}{\varrho_0})^\theta$ , with  $p_0 = 1$  and  $\theta = \frac{5}{3}$ . The quantities:  $\text{Ma}$ ,  $\text{Re}$  and  $\lambda$  are physical

constants, denoting respectively the Mach number (defined as a proportion of the speed of a fluid to the speed of sound), the Reynolds number (which is a proportion of fluid's inertial forces to its friction forces), and the viscosity coefficient. The quantity  $\varrho_0$ , present in the boundary condition for  $\varrho$ , is a constant, in computations we set  $\varrho_0 = 1.2$ . Finally,  $\mathbf{U} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a given smooth vector field, satisfying certain compatibility conditions, see Section 9.2.

**2.1. General theory in three spatial dimensions.** The stationary boundary value problem for compressible Navier-Stokes equations in three spatial dimensions can be formulated as follows.

**Problem.** *Assume that a gas occupies the flow domain  $\Omega = B \setminus S$ , where  $B \subset \mathbb{R}^n$ ,  $n = 3$  in general case, is a bounded hold-all domain, with an obstacle  $S$  in its interior.*

*Find a couple  $(\mathbf{u}, \varrho)$  such that*

$$\Delta \mathbf{u} + \lambda \nabla \operatorname{div} \mathbf{u} = \operatorname{Re} \varrho \mathbf{u} \cdot \nabla \mathbf{u} + \frac{\operatorname{Re}}{\operatorname{Ma}^2} \nabla p(\varrho) \quad \text{in } \Omega, \quad (1a)$$

$$\operatorname{div}(\varrho \mathbf{u}) = 0 \quad \text{in } \Omega, \quad (1b)$$

$$\mathbf{u} = \mathbf{U} \quad \text{on } \Sigma, \quad \mathbf{u} = 0 \quad \text{on } \partial S, \quad (1c)$$

$$\varrho = \varrho_0 \quad \text{on } \Sigma_{\text{in}}. \quad (1d)$$

where  $\mathbf{U}$  is a given vector field and  $\Sigma = \partial B$ . The inlet  $\Sigma_{\text{in}}$  and the outlet  $\Sigma_{\text{out}}$  are defined by

$$\Sigma_{\text{in}} = \{x \in \Sigma : \mathbf{U} \cdot \mathbf{n} < 0\}, \quad \Sigma_{\text{out}} = \{x \in \Sigma : \mathbf{U} \cdot \mathbf{n} > 0\},$$

respectively. Here  $\mathbf{n}$  stands for the outward normal to  $\partial(B \setminus S) = \Sigma \cup \partial S$ . For the sake of simplicity we assume that the mass force equals zero, and  $\varrho_0$  is a given positive constant.

**2.2. Change of variables in Navier-Stokes equations.** We consider the general case, the results apply as well in two space dimensions. In order to perform the shape sensitivity analysis in three space dimensions we choose a vector field  $\mathbf{T} \in C^2(\mathbb{R}^3)^3$  vanishing in a neighborhood of  $\Sigma$ , and introduce the mapping  $\Phi_\varepsilon(x) = x + \varepsilon \mathbf{T}(x)$ , which defines the perturbation  $S_\varepsilon = \Phi_\varepsilon(S)$  of the obstacle  $S$ . For small  $\varepsilon$ , the mapping  $\Phi_\varepsilon$  takes diffeomorphically the flow region  $\Omega = B \setminus S$  onto  $\Omega_\varepsilon = B \setminus S_\varepsilon$ . In the perturbed domain Problem 2.1 reads

**Problem.** *Find a solution  $(\bar{\mathbf{u}}_\varepsilon, \bar{\varrho}_\varepsilon)$ , to the following boundary value problem posed in the variable domain  $\Omega_\varepsilon = B \setminus S_\varepsilon$ , for the shape parameter  $\varepsilon \in (-\delta, \delta)$  with  $\delta > 0$ :*

$$\Delta \bar{\mathbf{u}}_\varepsilon + \lambda \nabla \operatorname{div} \bar{\mathbf{u}}_\varepsilon = \operatorname{Re} \bar{\varrho}_\varepsilon \bar{\mathbf{u}}_\varepsilon \cdot \nabla \bar{\mathbf{u}}_\varepsilon + \frac{\operatorname{Re}}{\operatorname{Ma}^2} \nabla p(\bar{\varrho}_\varepsilon) \quad \text{in } B \setminus S_\varepsilon, \quad (2a)$$

$$\operatorname{div}(\bar{\varrho}_\varepsilon \bar{\mathbf{u}}_\varepsilon) = 0 \quad \text{in } B \setminus S_\varepsilon, \quad (2b)$$

$$\bar{\mathbf{u}}_\varepsilon = \mathbf{U} \quad \text{on } \Sigma, \quad \bar{\mathbf{u}}_\varepsilon = 0 \quad \text{on } \partial S_\varepsilon, \quad (2c)$$

$$\bar{\varrho}_\varepsilon = \varrho_0 \quad \text{on } \Sigma_{\text{in}}. \quad (2d)$$

Now, we perform a change of variables in equations (2) in order to reduce Problem 2.2 in the variable domain  $\Omega_\varepsilon$ , depending on a small parameter  $\varepsilon$ , to a problem in the fixed domain  $\Omega$ , named the *reference domain*. Denote by  $\mathbb{M}$  the Jacobi matrix of the mapping  $\Phi_\varepsilon$  and by  $\mathbf{g}$  the determinant of  $\mathbb{M}$ , i.e.,

$$\mathbb{M}(x) = \mathbb{I} + \varepsilon D \mathbf{T}(x), \quad \mathbf{g}(x) = \det(\mathbb{I} + \varepsilon D \mathbf{T}(x)). \quad (3)$$

Further the notation  $\mathbf{N}$  stands for the adjugate matrix

$$\mathbf{N} = (\det \mathbb{M}) \mathbb{M}^{-1} \text{ with } \det \mathbf{N} = \mathbf{g}^2. \quad (4)$$

**Definition 2.1.**

- In shape sensitivity analysis the *material derivatives* of solutions to stationary Navier-Stokes equations are obtained by the differentiation of the mapping

$$\varepsilon \rightarrow (\bar{\mathbf{u}}_\varepsilon(x + \varepsilon \mathbf{T}(x)), \bar{\varrho}_\varepsilon(x + \varepsilon \mathbf{T}(x))), \quad x \in \Omega.$$

- The *shape derivatives* of solutions to stationary Navier-Stokes equations are defined by derivatives of the mapping

$$\varepsilon \rightarrow (\bar{\mathbf{u}}_\varepsilon(y), \bar{\varrho}_\varepsilon(y)), \quad y \in \Omega_\varepsilon.$$

- The *shape gradient* of the drag functional for stationary Navier-Stokes equations is given by the derivative of the function

$$\varepsilon \mapsto J(\Omega_\varepsilon).$$

Introduce the functions

$$\mathbf{u}_\varepsilon(x) = \mathbf{N}(x) \bar{\mathbf{u}}_\varepsilon(x + \varepsilon \mathbf{T}(x)), \quad \varrho_\varepsilon(x) = \bar{\varrho}_\varepsilon(x + \varepsilon \mathbf{T}(x)), \quad x \in \Omega. \quad (5)$$

The function  $\mathbf{u}_\varepsilon$  is called the *Piola transformation* of  $\bar{\mathbf{u}}_\varepsilon$ .

**Lemma 2.2.** *Let  $(\bar{\mathbf{u}}_\varepsilon(y), \bar{\varrho}_\varepsilon(y))$  be a solution to Problem 2.2. Then the couple  $(\mathbf{u}_\varepsilon(x), \varrho_\varepsilon(x))$  defined by (5) satisfies*

$$\begin{aligned} \Delta \mathbf{u}_\varepsilon + \nabla \left( \lambda \mathbf{g}^{-1} \operatorname{div} \mathbf{u}_\varepsilon - \frac{\operatorname{Re}}{\operatorname{Ma}^2} p(\varrho_\varepsilon) \right) \\ = \mathcal{A}(\mathbf{u}_\varepsilon) + \operatorname{Re} \mathcal{B}(\varrho_\varepsilon, \mathbf{u}_\varepsilon, \mathbf{u}_\varepsilon) \quad \text{in } \Omega, \end{aligned} \quad (6a)$$

$$\operatorname{div}(\varrho_\varepsilon \mathbf{u}_\varepsilon) = 0 \quad \text{in } \Omega, \quad (6b)$$

$$\mathbf{u}_\varepsilon = \mathbf{U} \quad \text{on } \Sigma, \quad \mathbf{u}_\varepsilon = 0 \quad \text{on } \partial S, \quad (6c)$$

$$\varrho_\varepsilon = \varrho_0 \quad \text{on } \Sigma_{\text{in}}. \quad (6d)$$

Here, the linear operator  $\mathcal{A}$  and the trilinear form  $\mathcal{B}$  are defined by

$$\begin{aligned} \mathcal{A}(\mathbf{u}) &= \Delta \mathbf{u} - \mathbf{N}^{-\top} \operatorname{div}(\mathbf{g}^{-1} \mathbf{N} \mathbf{N}^\top \nabla(\mathbf{N}^{-1} \mathbf{u})), \\ \mathcal{B}(\varrho, \mathbf{u}, \mathbf{w}) &= \varrho \mathbf{N}^{-\top} (\mathbf{u} \nabla(\mathbf{N}^{-1} \mathbf{w})). \end{aligned} \quad (7)$$

Thus we come to the following problem:

**Problem.** *Find  $(\mathbf{u}_\varepsilon, \varrho_\varepsilon)$  such that*

$$\Delta \mathbf{u}_\varepsilon + \nabla \left( \lambda \mathbf{g}^{-1} \operatorname{div} \mathbf{u}_\varepsilon - \frac{\operatorname{Re}}{\operatorname{Ma}^2} p(\varrho_\varepsilon) \right) = \mathcal{A}(\mathbf{u}_\varepsilon) + \operatorname{Re} \mathcal{B}(\varrho_\varepsilon, \mathbf{u}_\varepsilon, \mathbf{u}_\varepsilon) \quad \text{in } \Omega, \quad (8a)$$

$$\operatorname{div}(\varrho_\varepsilon \mathbf{u}_\varepsilon) = 0 \quad \text{in } \Omega, \quad (8b)$$

$$\mathbf{u}_\varepsilon = \mathbf{U} \quad \text{on } \Sigma, \quad \mathbf{u}_\varepsilon = 0 \quad \text{on } \partial S, \quad (8c)$$

$$\varrho_\varepsilon = \varrho_0 \quad \text{on } \Sigma_{\text{in}}. \quad (8d)$$

Here the operators  $\mathcal{A}$  and  $\mathcal{B}$  are defined by (7).

Observe that equations (8) depend only on the matrix  $\mathbf{N}$  and do not depend on  $\varepsilon$  directly. In this framework, the fact that  $\mathbf{N}$  is the adjugate matrix to the Jacobi matrix  $\mathbb{I} + \varepsilon D \mathbf{T}$  is not so important. Hence our task is to show that Problem 2.2 is well-posed and its solution is differentiable with respect to  $\mathbf{N}$ . We begin by analysing the well-posedness.

**2.3. Solution scheme for stationary problem in reference domain.** In [27] we use the simple algebraic scheme proposed in [22]. The basic idea is to introduce the *effective viscous pressure*  $q_\varepsilon$  and a large parameter  $\sigma_0$ ,

$$q_\varepsilon := \frac{\text{Re}}{\text{Ma}^2} p(\varrho_\varepsilon) - \lambda g^{-1} \operatorname{div} \mathbf{U}_\varepsilon, \quad \sigma_0 = \frac{\text{Re}}{\lambda \text{Ma}^2}, \quad (9)$$

and write equations (8) in the form

$$\Delta \mathbf{U}_\varepsilon - \nabla q_\varepsilon = \mathcal{A}(\mathbf{U}_\varepsilon) + \operatorname{Re} \mathcal{B}(\varrho_\varepsilon, \mathbf{U}_\varepsilon, \mathbf{U}_\varepsilon) \quad \text{in } \Omega, \quad (10a)$$

$$\operatorname{div} \mathbf{U}_\varepsilon = g \sigma_0 p(\varrho_\varepsilon) - \frac{g q}{\lambda} \quad \text{in } \Omega, \quad (10b)$$

$$\mathbf{U}_\varepsilon \cdot \nabla \varrho_\varepsilon + g \sigma_0 p(\varrho_\varepsilon) \varrho_\varepsilon = \frac{g q_\varepsilon}{\lambda} \varrho_\varepsilon \quad \text{in } \Omega, \quad (10c)$$

$$\mathbf{U}_\varepsilon = \mathbf{U}i \quad \text{on } \Sigma, \quad \mathbf{U}_\varepsilon = 0 \quad \text{on } \partial S, \quad (10d)$$

$$\varrho_\varepsilon = \varrho_0 \quad \text{on } \Sigma_{\text{in}}, \quad (10e)$$

System (10) consists of perturbed Stokes equations (10a)–(10b) for  $(\mathbf{u}_\varepsilon, q_\varepsilon)$  and a perturbed transport equation (10c) for  $\varrho_\varepsilon$ . It is easy to see that equations (10b) and (10c) are equivalent to (9) and the mass balance equation  $\operatorname{div}(\varrho_\varepsilon \mathbf{u}_\varepsilon) = 0$ . Therefore, equations (10) are equivalent to (8). It was shown in [22] and [23] that problem (10) has a strong solution if the parameters satisfy the conditions

$$\lambda \gg 1, \quad \text{Re} \ll 1, \quad \sigma_0 \gg 1,$$

which are more restrictive than the standard conditions  $\text{Re} \ll 1$ ,  $\text{Ma}^2 \ll 1$  adopted in [2], [16], and [24].

**Perturbations.** We will look for a local solution to problem (8) which is close to an approximate solution  $(\mathbf{u}_\star, q_\star, \varrho_\star)$ . In order to define such a solution notice that for small Mach and Reynolds numbers equations (10a)–(10b) are close to Stokes equations and the density  $\varrho_\varepsilon$  is close to  $\varrho_0$ . Thus we take

$$\varrho_\star = \varrho_0 \quad (11)$$

and choose  $(\mathbf{u}_\star, q_\star)$  as the solution to the following boundary problem for the Stokes equation:

$$\begin{aligned} \Delta \mathbf{u}_\star - \nabla q_\star &= 0, \quad \operatorname{div} \mathbf{u}_\star = 0 \quad \text{in } \Omega, \\ \mathbf{u}_\star &= \mathbf{U} \quad \text{on } \Sigma, \quad \mathbf{u}_\star = 0 \quad \text{on } \partial S, \quad \Pi q_\star = q_\star, \end{aligned} \quad (12)$$

where

$$\Pi q := q - \frac{1}{\operatorname{meas} \Omega} \int_\Omega q \, dx. \quad (13)$$

The solution to problem (8) is decomposed into the approximate solution  $(\mathbf{u}_\star, q_\star, \varrho_\star)$  and small perturbations  $(\mathbf{v}_\varepsilon, \pi_\varepsilon, \varphi_\varepsilon, m_\varepsilon)$ ,

$$\mathbf{u}_\varepsilon = \mathbf{u}_\star + \mathbf{v}_\varepsilon, \quad \varrho_\varepsilon = \varrho_\star + \varphi_\varepsilon, \quad q_\varepsilon = q_\star + \lambda \sigma_0 p(\varrho_\star) + \pi_\varepsilon + \lambda m_\varepsilon, \quad (14)$$

with some unknown functions  $\vartheta = (\mathbf{v}_\varepsilon, \pi_\varepsilon, \varphi_\varepsilon)$  and an unknown constant  $m_\varepsilon$ .

We point out that our decomposition (14) holds exclusively in the reference domain  $\Omega = B \setminus S$ . The unknown functions  $\vartheta$  and the unknown constant  $m_\varepsilon$  contain the necessary and sufficient information on the shape dependence of solutions to the equations (8) upon the shape parameter  $\varepsilon \rightarrow 0$ . This information is used in order to determine the material derivatives of the solutions to the model (2) with respect to the shape perturbation. However, the Stokes boundary value problem (1) defined in the reference domain  $\Omega$  is meaningless for the model (2) defined in the variable domain  $\Omega_\varepsilon$ .

Substituting decomposition (14) into (10) we obtain the following boundary value problem for  $\vartheta = (\mathbf{v}_\varepsilon, \pi_\varepsilon, \varphi_\varepsilon)$ :

$$\Delta \mathbf{v}_\varepsilon - \nabla \pi_\varepsilon = \mathcal{A}(\mathbf{u}_\varepsilon) + \operatorname{Re} \mathcal{B}(\varrho_\varepsilon, \mathbf{u}_\varepsilon, \mathbf{u}_\varepsilon) \quad \text{in } \Omega, \quad (15a)$$

$$\operatorname{div} \mathbf{v}_\varepsilon = \mathfrak{g} \left( \frac{\sigma}{\varrho_\star} \varphi_\varepsilon - \Psi[\vartheta] - m_\varepsilon \right) \quad \text{in } \Omega, \quad (15b)$$

$$\mathbf{u}_\varepsilon \cdot \nabla \varphi_\varepsilon + \sigma \varphi_\varepsilon = \Psi_1[\vartheta] + m_\varepsilon \mathfrak{g} \varrho_\varepsilon \quad \text{in } \Omega, \quad (15c)$$

$$\mathbf{v}_\varepsilon = 0 \quad \text{on } \partial\Omega, \quad \varphi_\varepsilon = 0 \quad \text{on } \Sigma_{\text{in}}, \quad \Pi \pi_\varepsilon = \pi_\varepsilon, \quad (15d)$$

where

$$\Psi[\vartheta] = \frac{q_\star + \pi_\varepsilon}{\lambda} - \frac{\sigma}{p'(\varrho_\star)\varrho_\star} H(\varphi_\varepsilon), \quad \Psi_1[\vartheta] = \mathfrak{g} \left( \varrho_\varepsilon \Psi[\vartheta] - \frac{\sigma}{\varrho_\star} \varphi_\varepsilon^2 \right) + \sigma \varphi_\varepsilon (1 - \mathfrak{g}), \quad (15e)$$

$$\sigma = \sigma_0 p'(\varrho_\star) \varrho_\star, \quad H(\varphi_\varepsilon) = p(\varrho_\star + \varphi_\varepsilon) - p(\varrho_\star) - p'(\varrho_\star) \varphi_\varepsilon, \quad (15f)$$

and the vector field  $\mathbf{u}_\varepsilon$  and the function  $\varrho_\varepsilon$  are given by (14).

**2.4. Function spaces.** To establish existence results for problem (15), we begin with the definition of an admissible class of solutions to this problem.

Notice that we are looking for strong solutions, which means that in equations (15) the unknown functions and their derivatives are integrable, i.e., these equations are satisfied a.e. in  $\Omega$ . To this end it suffices to assume that  $\mathbf{v} \in W^{r,2,2}(\Omega)$  and  $(\pi, \varphi, \zeta) \in W^{1,2}(\Omega)$ . Next, we are looking for solutions such that  $(\pi, \varphi, \zeta)$  are continuous and  $\mathbf{v}$  is continuously differentiable in  $\Omega$ . In order to satisfy this demand it suffices to take  $\mathbf{v} \in W^{s+1,r}(\Omega)$  and  $(\pi, \varphi, \zeta) \in W^{s,r}(\Omega)$  with  $sr > 3$ . To address both issues we introduce the Banach spaces

$$\begin{aligned} X^{s,r} &= W^{s,r}(\Omega) \cap W^{1,2}(\Omega), \\ Y^{s,r} &= W^{s+1,r}(\Omega) \cap W^{2,2}(\Omega), \quad 0 < s < 1 < r < \infty, \end{aligned} \quad (16)$$

equipped with the norms

$$\|u\|_{X^{s,r}} = \|u\|_{W^{s,r}(\Omega)} + \|u\|_{W^{1,2}(\Omega)}, \quad \|u\|_{Y^{s,r}} = \|u\|_{W^{s+1,r}(\Omega)} + \|u\|_{W^{2,2}(\Omega)}.$$

Throughout of the rest of this chapter we intensively exploit the fact that the spaces  $X^{s,r}$  and  $Y^{s,r}$  are reflexive. The proof of this fact is nontrivial and follows the line given below.

It is well known (see [1, Thm. 3.5]) that for any bounded domain  $\Omega$  with  $C^1$  boundary, any integer  $m \geq 0$  and  $1 < r < \infty$ , the Sobolev space  $W^{m,r}(\Omega)$  is reflexive. The space  $W_0^{1,r}(\Omega)$  is obviously reflexive as a closed subspace of a reflexive Banach space (see [1, Thm. 1.21]).

**Lemma 2.3.** *Let  $\Omega$  be a bounded domain with  $C^1$  boundary and  $0 < s < 1 < r < \infty$ . Then the spaces  $W^{s,r}(\Omega)$  and  $W^{s+1,r}(\Omega)$  are reflexive.*

**Lemma 2.4.** *For any functional  $f \in (X^{s,r})'$  there are continuous functionals  $g \in (W^{s,r}(\Omega))'$  and  $h \in (W^{1,2}(\Omega))'$  such that*

$$\langle f, \varphi \rangle = \langle g, \varphi \rangle + \langle h, \varphi \rangle \quad \text{for all } \varphi \in X^{s,r}.$$

**Lemma 2.5.** *Let  $\Omega$  be a bounded domain with  $C^1$  boundary and  $0 < s < 1 < r < \infty$ . Then the spaces  $X^{s,r}$  and  $Y^{s,r}$  are reflexive.*

While  $W^{s,r}(\Omega)$  and  $W^{1,2}(\Omega)$  are separable, we cannot conclude from this that  $X^{s,r}$  is separable. However, the dual spaces  $(X^{s,r})'$  and  $(Y^{s,r})'$  are separable:



**Lemma 2.6.** *Let  $\Omega \subset \mathbb{R}^2$  be a bounded domain with  $C^1$  boundary and  $0 < s < 1 < r < \infty$ . Then  $(X^{s,r})'$  and  $(Y^{s,r})'$  are separable.*

Existence theory. Define the Banach space

$$E = (Y^{s,r})^3 \times (X^{s,r})^2, \quad (17)$$

and denote by  $\mathcal{B}_\tau \subset E$  the closed ball of radius  $\tau$  centered at 0. Next, note that for  $r > 3$ , elements of the ball  $\mathcal{B}_\tau$  satisfy the inequality

$$\|\mathbf{v}\|_{C^1(\Omega)} + \|\pi\|_{C(\Omega)} + \|\varphi\|_{C(\Omega)} \leq c_\varepsilon(r, s, \Omega) \|\vartheta\|_E \leq c_\varepsilon \tau, \quad (18)$$

where the norm in  $E$  is defined by

$$\|\vartheta\|_E = \|\mathbf{v}\|_{Y^{s,r}} + \|\pi\|_{X^{s,r}} + \|\varphi\|_{X^{s,r}}.$$

Further,  $c$  denotes generic constants, which are different in different places and depend only on  $\Omega$ ,  $\mathbf{U}$ ,  $\sigma$  and  $r, s$ . We assume that the flow domain and the given data satisfy the following conditions.

**Condition 1.**

- $\partial\Omega$  is a closed surface of class  $C^3$  and the set  $\Gamma = \text{cl } \Sigma_{\text{in}} \cap \Sigma \setminus \Sigma_{\text{in}}$  is a closed  $C^3$  one-dimensional manifold such that  $\Sigma = \Sigma_{\text{in}} \cup \Gamma \cup \Sigma_{\text{out}}$ .
- The vector field  $\mathbf{U} \in C^3(\partial\Omega)$  satisfies

$$\int_{\partial\Omega} \mathbf{U} \cdot \mathbf{n} \, d\Sigma = 0.$$

Moreover, we can assume that it is extended to  $\mathbb{R}^3$  in such a way that the extension  $\mathbf{U} \in C^3(\mathbb{R}^3)$  vanishes in a neighborhood of  $S$ .

- There is a positive constant  $c$  such that

$$\mathbf{U}i \cdot \nabla(\mathbf{U}i \cdot \mathbf{n}) > c > 0 \quad \text{on } \Gamma.$$

Since the vector field  $\mathbf{U}i$  is tangent to  $\partial\Omega$  on  $\Gamma$ , the left hand side is well defined.

Let us consider the augmented problem of the following form:

**Problem.** *Given  $(\mathbf{u}_\star, q_\star, \varrho_\star)$ , find  $(\mathbf{v}_\varepsilon, \pi_\varepsilon, \varphi_\varepsilon)$  and a constant  $m_\varepsilon$ , depending on  $\varepsilon$ , such that*

$$\begin{aligned} \Delta \mathbf{v}_\varepsilon - \nabla \pi_\varepsilon &= \mathcal{A}(\mathbf{u}_\varepsilon) + \text{Re } \mathcal{B}(\varrho_\varepsilon, \mathbf{u}_\varepsilon, \mathbf{u}_\varepsilon) \quad \text{in } \Omega, \\ \text{div } \mathbf{v}_\varepsilon &= \mathfrak{g} \left( \frac{\sigma}{\varrho_\star} \varphi_\varepsilon - \Psi[\vartheta] - m_\varepsilon \right) \quad \text{in } \Omega, \\ \mathbf{u}_\varepsilon \cdot \nabla \varphi_\varepsilon + \sigma \varphi_\varepsilon &= \Psi_1[\vartheta] + m_\varepsilon \mathfrak{g} \varrho_\varepsilon \quad \text{in } \Omega, \\ \mathbf{v}_\varepsilon &= 0 \quad \text{on } \partial\Omega, \quad \varphi_\varepsilon = 0 \quad \text{on } \Sigma_{\text{in}}, \quad \Pi \pi_\varepsilon = \pi_\varepsilon. \end{aligned} \quad (19a)$$

The constant  $m_\varepsilon$  is determined from the following relations:

$$m_\varepsilon = \varkappa \int_{\Omega} (\varrho_\star^{-1} \Psi_1[\vartheta] \zeta_\varepsilon - \mathfrak{g} \Psi[\vartheta]) \, dx, \quad \varkappa_\varepsilon = \left( \int_{\Omega} \mathfrak{g} (1 - \zeta_\varepsilon - \varrho_\star^{-1} \zeta_\varepsilon \varphi_\varepsilon) \, dx \right)^{-1}, \quad (19b)$$

where the auxiliary function  $\zeta_\varepsilon$  is a solution to the adjoint boundary value problem

$$-\text{div}(\mathbf{u}_\varepsilon \zeta_\varepsilon) + \sigma \zeta_\varepsilon = \sigma \mathfrak{g} \quad \text{in } \Omega, \quad \zeta_\varepsilon = 0 \quad \text{on } \Sigma_{\text{out}}. \quad (19c)$$

The following existence theorem is the main result of this section.

**Theorem 2.7.** *Assume that the surface  $\Sigma$  and the given vector field  $\mathbf{U}$  satisfy Condition 1. Furthermore, let numbers  $r$  and  $s$  satisfy*

$$0 < s < 1, \quad 1 < r < \infty, \quad 2s - 3r^{-1} < 1, \quad sr > 3. \quad (20)$$

*Then there exists  $\sigma^* > 1$ , depending only on  $\mathbf{U}$ ,  $\Omega$  and  $s$ ,  $r$ , with the following property. For every  $\sigma > \sigma^*$  there are positive  $\tau^*$  and  $c$ , depending only on  $\mathbf{U}$ ,  $\Omega$ ,  $r$ ,  $s$ , and  $\sigma$ , such that whenever*

$$\tau \in (0, \tau^*], \quad \lambda^{-1}, \operatorname{Re} \in (0, \tau^2], \quad \|\mathbf{N} - \mathbb{I}\|_{C^2(\Omega)} \leq \tau^2, \quad (21)$$

*Problem 2.4 has a solution  $\vartheta \in \mathcal{B}_\tau$ ,  $\zeta_\varepsilon \in X^{s,r}$ ,  $m_\varepsilon \in \mathbb{R}$ . The auxiliary function  $\zeta_\varepsilon$  and the constants  $\varkappa_\varepsilon, m_\varepsilon$  admit the estimates*

$$\|\zeta_\varepsilon\|_{X^{s,r}} + |\varkappa_\varepsilon| \leq c, \quad |m_\varepsilon| \leq c\tau < 1. \quad (22)$$

*Let  $\Theta_{\mathbf{N}} \subset \mathcal{B}_\tau \times W^{s,r}(\Omega) \times \mathbb{R}$  be the set of all solutions  $(\vartheta, \zeta_\varepsilon, m_\varepsilon) = (\mathbf{v}_\varepsilon, \pi_\varepsilon, \varphi_\varepsilon, \zeta_\varepsilon, m_\varepsilon)$  corresponding to a matrix-valued function  $\mathbf{N}$ . Then for every  $\mathbf{N}$  in the ball  $B(\tau^2) = \{\mathbf{N} : \|\mathbb{I} - \mathbf{N}\|_{C^2(\Omega)} \leq \tau^2\}$  there is a nonempty subset  $\Theta'_{\mathbf{N}} \subset \Theta_{\mathbf{N}}$  such that  $\bigcup_{\mathbf{N} \in B(\tau^2)} \Theta'_{\mathbf{N}}$  is relatively compact in  $W^{s+1,r}(\Omega) \times W^{s,r}(\Omega)^3 \times \mathbb{R}$ .*

**3. Minimization of the shape functional.** Now we are going to adopt the general theory to the specific problem in two spatial dimensions. Our goal is to propose the numerical scheme for drag minimization and obtain numerical results. We refer the reader to e.g., [13], [14], [30], [31], [32] for incompressible modeling and numerical methods. The drag minimization is considered also in [12].

**3.1. Shape functional.** It is known that the theory of compressible fluids has many essential applications in aerodynamics. For example, it is used in aircraft construction. Thus the obstacle  $S$  can be identified with an airfoil. We work in the reference frame with the fixed obstacle  $S$ . (see Fig.1). Aerodynamical force, acting on the obstacle  $S$  can be expressed as

$$\mathbf{F}_a(\Gamma) = - \int_\Gamma (\nabla \mathbf{u} + (\nabla \mathbf{u})^T + (\lambda - 1) \operatorname{div} \mathbf{u} \mathbb{I} - \frac{\operatorname{Re}}{M_a^2} p \mathbb{I}) \cdot \mathbf{n} dS, \quad (23)$$

where  $\mathbf{n}$  denotes the unit normal field defined on  $\Gamma$ , pointing outward from  $S$ , and  $\mathbb{I}$  stands for the unit matrix. Similarly as in the case of an airfoil moving in the horizontal direction, the drag functional  $J$  can be defined as the horizontal component of the aerodynamical force  $\mathbf{F}_a$ . For the horizontal unit vector  $\mathbf{e}_1 = [1, 0]$ , it reads:

$$J(\Gamma) = \mathbf{e}_1 \cdot \mathbf{F}_a(\Gamma). \quad (24)$$

Then, applying any smooth function  $\eta$  (in the program  $\eta$  is a solution to the Laplace equation) satisfying suitable boundary conditions, the aerodynamical force can be expressed as an integral over  $\Omega$ . For  $\eta \equiv 1$  on  $\Gamma$  and  $\eta = 0$  on  $\Sigma$ , integrating (23) by parts gives:

$$\begin{aligned} \mathbf{F}_a(\Omega) = & - \int_\Omega (\nabla \mathbf{u} + \nabla \mathbf{u}^T - \operatorname{div}(\mathbf{u}) \mathbb{I} - q \mathbb{I} - \operatorname{Re} \varrho \mathbf{u} \otimes \mathbf{u}) \cdot \nabla \eta \, dx \\ & - \int_\Omega \eta \varrho \mathbf{u} \cdot \nabla \mathbf{u} \, dx. \end{aligned} \quad (25)$$

In consequence, the drag functional is:

$$\begin{aligned}
 J(\Omega) = & -\mathbf{e}_1 \cdot \int_{\Omega} (\nabla \mathbf{u} + \nabla \mathbf{u}^T - \operatorname{div}(\mathbf{u})\mathbb{I} - q\mathbb{I} - \operatorname{Re} \rho \mathbf{u} \otimes \mathbf{u}) \cdot \nabla \eta \, dx \\
 & - \mathbf{e}_1 \cdot \int_{\Omega} \eta \rho \mathbf{u} \cdot \nabla \mathbf{u} \, dx.
 \end{aligned} \tag{26}$$

**3.2. Discretization of continuous shape gradient.** The existence of shape gradient of the drag functional is shown for compressible model in [27]. In the numerical procedure we are going to use the material derivatives of the solutions to the governing equations. The obtained continuous shape gradient is discretized for the purposes of numerical method. We point out that the shape derivatives of the continuous compressible model also lead to the shape gradient of the cost [26]. The shape derivatives are obtained by the singular limit procedure in the boundary value problem that is formulated for the material derivatives. Formally, the same result can be derived by an application of the boundary variations technique to the strong formulation of the stationary compressible Navier-Stokes equations. However, it is important to remember that even the existence of the weak solutions to the stationary model with nonhomogenous boundary conditions is a difficult problem in the physical range of parameters. Therefore, the formal considerations for compressible model are not in general justified, and are performed in the framework of experimental mathematics. The results obtained here are proved for almost incompressible model.

**3.3. Deformations of  $\Gamma$  and  $\Omega$ .** We seek a deformation of  $\Gamma$ , such that the value of the drag functional  $J$  is minimal relatively to a certain family of admissible shapes. We introduce the family of admissible shapes by defining the set of admissible deformations of the initial shape  $\Gamma$ . The set of admissible deformations of  $\Gamma$  consists of linear combinations of basic transformations of  $\Gamma$ , defined as bell-shaped smooth hat functions (see (109)) depending on the discretization of  $\Gamma$ . It is assumed that  $\mathbf{h} = [h_1, h_2]$  is a smooth vector field normal to  $\Gamma$ , that is a deformation of  $\Gamma$  in the normal direction. It defines a transformation  $\Phi_\varepsilon(x) = x + \varepsilon \mathbf{T}(x)$  of  $\Omega$  into  $\Omega_\varepsilon$ , depending on the vector field  $\mathbf{T}$  constituting the unique solution to the Laplace equation:

$$\Delta \mathbf{T} = 0 \quad \text{in } \Omega, \quad \mathbf{T} = 0 \quad \text{on } \Sigma, \quad \mathbf{T} = \mathbf{h} \quad \text{on } \Gamma. \tag{27}$$

Observe that, for  $\mathbf{h} = 0$  and consequently  $\mathbf{T} = 0$ , this transformation reduces to identity. Since both boundaries of  $\Omega$  and boundary data for  $h_1$  and  $h_2$  are smooth, such  $\mathbf{T}$  is a vector field on  $\Omega$  of class  $C^\infty(\bar{\Omega})$ , normal on  $\Gamma$  and vanishing on  $\Sigma$ .

The derivative of the functional  $J(\Omega)$  in the direction  $\mathbf{T}$  is given by the limit

$$dJ(\Omega; \mathbf{T}) = \lim_{\varepsilon \rightarrow 0^+} \frac{J(\Omega_\varepsilon) - J(\Omega)}{\varepsilon}$$

We use  $\mathbf{T}$  instead of  $\mathbf{h}$  in the definition of this derivative because such a formulation could be used for arbitrary field  $\mathbf{T}$  as well.

By applying the same substitutions as in Section 2.2, the shape functional, defined previously as an integral over the domain  $\Omega_\varepsilon$ , can be written as an integral over  $\Omega$ :

$$\begin{aligned}
J_\epsilon(\Omega) &= -\operatorname{Re} \mathbf{e}_1 \cdot \int_{\Omega} \varrho^\epsilon \mathbf{u}^\epsilon \nabla (\mathbb{N}^{-1} \mathbf{u}^\epsilon) \eta dx - \\
&\mathbf{e}_1 \cdot \int_{\Omega} [\mathbf{g}^{-1} (\mathbb{N}^T \nabla (\mathbb{N}^{-1} \mathbf{u}^\epsilon) + \nabla (\mathbb{N}^{-1} \mathbf{u}^\epsilon)^T \mathbb{N} - \operatorname{div} \mathbf{u}^\epsilon) - q^\epsilon \mathbb{I}] \mathbb{N}^T \nabla \eta dx
\end{aligned} \tag{28}$$

**4. Shape derivative.** Differentiating formally the integrand in (28), we obtain the formula for the derivative of  $J$ , which can be symbolically written as the sum:

$$\frac{d}{d\epsilon} J_\epsilon(\Omega) \Big|_{\epsilon=0} = L_e(\mathbf{T}) + L_u(\delta \mathbf{u}, \delta \pi, \delta \varphi), \tag{29}$$

where the linear forms  $L_e$  and  $L_u$  are given by the formulas [25]

$$\begin{aligned}
L_e(\mathbf{T}) &= \mathbf{e}_1 \cdot \int_{\Omega} \operatorname{Tr} D\mathbf{T} (\nabla \mathbf{u} + (\nabla \mathbf{u})^T - \operatorname{div} \mathbf{u} \mathbb{I}) \nabla \eta dx - \\
&\mathbf{e}_1 \cdot \int_{\Omega} [\mathbb{D}^T \nabla \mathbf{u} + (\nabla \mathbf{u})^T \mathbb{D} - \nabla (\mathbb{D} \mathbf{u}) - (\nabla (\mathbb{D} \mathbf{u}))^T] \nabla \eta dx - \\
&\mathbf{e}_1 \cdot \int_{\Omega} [\nabla \mathbf{u} + (\nabla \mathbf{u})^T - \operatorname{div} \mathbf{u} - q \mathbb{I}] \mathbb{D}^T \nabla \eta dx + \\
&\operatorname{Re} \mathbf{e}_1 \cdot \int_{\Omega} \varrho \mathbf{u} \nabla (\mathbb{D} \mathbf{u}) \eta \cdot dx,
\end{aligned} \tag{30}$$

$$\begin{aligned}
L_u(\delta \mathbf{u}, \delta \pi, \delta \varphi) &= -\mathbf{e}_1 \cdot \int_{\Omega} [(\nabla \delta \mathbf{u} + (\nabla \delta \mathbf{u})^T - \operatorname{div} \delta \mathbf{u} - \delta \pi) \nabla \eta \\
&+ \operatorname{Re} \mathcal{C}(\delta \mathbf{u}, \delta \varphi) \eta] dx,
\end{aligned} \tag{31}$$

where

$$\frac{d\mathbf{g}}{d\epsilon} \Big|_{\epsilon=0} = \operatorname{Tr} D\mathbf{T} \equiv \operatorname{div} \mathbf{T} \tag{32}$$

and  $(\delta \mathbf{v}, \delta \pi, \delta \varphi, \delta \zeta, \delta m)$  denote the material derivatives  $\frac{d}{d\epsilon} (\mathbf{v}_\epsilon, \pi_\epsilon, \varphi_\epsilon, \zeta_\epsilon, m_\epsilon) \Big|_{\epsilon=0}$  for the solution to the correction system. Since by definition (1)  $(\mathbf{u}_\star, q_\star, \rho_\star)$  do not vary then, from (14), we have  $(\delta \mathbf{u}, \delta q, \delta \varrho) = (\delta \mathbf{v}, \delta \pi, \delta \varphi)$ . The material derivatives  $(\delta \mathbf{v}, \delta \pi, \delta \varphi, \delta \zeta, \delta m)$  can be derived by formal differentiation of the correction system (10), which gives the following system for derivatives:

$$\Delta \delta \mathbf{v} - \nabla \delta \pi = \mathcal{C}(\delta \varphi, \delta \mathbf{v}) + \mathcal{D}(\mathbb{D}), \tag{33}$$

$$\operatorname{div} \delta \mathbf{v} = b_{21} \delta \varphi + b_{22} \delta \pi + b_{23} \delta m + b_{20} \operatorname{Tr} D\mathbf{T}, \tag{34}$$

$$\mathbf{u} \nabla \delta \varphi + \sigma_1 \delta \varphi = -\delta \mathbf{v} \nabla \varphi + b_{11} \delta \varphi + b_{12} \delta \pi + b_{13} \delta m + b_{10} \operatorname{Tr} D\mathbf{T}, \tag{35}$$

$$-\operatorname{div}(\delta \zeta \mathbf{u}) + \sigma_1 \delta \zeta = \operatorname{div}(\zeta \delta \mathbf{v}) + \sigma_1 \operatorname{Tr} D\mathbf{T}, \tag{36}$$

$$\delta m = \chi \int_{\Omega} (b_{31} \delta \varphi + b_{32} \delta \pi + b_{34} \delta \zeta b_{30} \operatorname{Tr} D\mathbf{T}) dx, \tag{37}$$

with the boundary conditions:

$$\begin{aligned}
\delta \mathbf{v} &= 0 \text{ on } \partial \Omega, \quad \delta \varphi = 0 \text{ on } \Sigma_{in}, \quad \delta \zeta = 0 \text{ on } \Sigma_{out} \\
\Pi(\delta \pi) &= \delta \pi,
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
\mathcal{D}(\mathbb{D}) &= \operatorname{div} [((\operatorname{Tr} D\mathbf{T}) \mathbb{I} - \mathbb{D} - \mathbb{D}^T) \nabla \mathbf{u}] + \mathbb{D}^T \Delta \mathbf{u} + \Delta (\mathbb{D} \mathbf{u}) - \operatorname{Re} \varrho \mathbb{D}^T (\mathbf{u} \nabla \mathbf{u}) - \operatorname{Re} \varrho \mathbf{u} \nabla (\mathbb{D} \mathbf{u}), \\
\mathcal{C}(\delta \varphi, \delta \mathbf{v}) &= \operatorname{Re} \delta \varphi \mathbf{u} \cdot \nabla \mathbf{u} + \operatorname{Re} \varrho \mathbf{u} \cdot \nabla \delta \mathbf{v} + \operatorname{Re} \varrho \delta \mathbf{v} \cdot \nabla \mathbf{u},
\end{aligned}$$

and the coefficients  $b_{ij}$  equal:

$$\begin{aligned}
b_{11} &= \Psi + m - \frac{2\sigma}{\varrho_0}\varphi - \frac{\varrho\sigma}{p'(\varrho_0)}Hl(\varphi), \\
b_{12} &= \frac{\varrho}{\lambda}, \\
b_{13} &= \varrho, \\
b_{10} &= \varrho(\Psi + m) - \frac{\sigma}{\varrho_0}\varphi^2 - \sigma\varphi, \\
b_{21} &= \frac{\sigma}{\varrho_0} + \frac{\sigma}{p'(\varrho_0)\varrho_0}Hl(\varphi), \\
b_{22} &= -\frac{1}{\lambda}, \\
b_{23} &= -1, \\
b_{20} &= \frac{\sigma}{\varrho_0}\varphi - \Psi - m \\
b_{31} &= \frac{\zeta}{\varrho_0}\left(\left(\Psi + m\right) - \frac{2\sigma}{\varrho_0}\varphi\right) + \frac{\sigma}{p'(\varrho_0)\varrho_0}\left(1 - \frac{\varrho\zeta}{\varrho_0}\right)Hl(\varphi) \\
b_{32} &= \frac{1}{\lambda}\left(\frac{\varrho}{\varrho_0}\zeta - 1\right), \\
b_{34} &= \frac{1}{\varrho_0}\left(\varrho(\Psi + m) - \frac{\sigma}{\varrho_0}\varphi^2\right), \\
b_{30} &= \left(\frac{\varrho}{\varrho_0}\zeta - 1\right)(\Psi + m) - \frac{\sigma\varrho}{\varrho_0^2}\varphi\zeta
\end{aligned}$$

It is easy to see that the left hand sides of (19a) and (33)-(37) are identical, which will simplify further computations.

**5. Computational scheme.** In order to make an optimization step, we need to calculate the shape derivative, which is given by the formula (29). Thus we have to find the unique solution of the system (19a) and also the material derivatives  $(\delta\mathbf{v}, \delta\pi, \delta\varphi, \delta\zeta, \delta m)$ , that are the solution to the system (33)-(37). Both these systems can be written in the common form:

$$\begin{aligned}
\Delta z_1^1 - \partial_1 z^2 &= f_1^1(\vartheta), \\
\Delta z_2^1 - \partial_2 z^2 &= f_2^1(\vartheta) \\
\partial_1 z_1^1 + \partial_2 z_2^1 &= f^2(\vartheta), \\
\mathbf{u}^T \nabla z^3 + \sigma_1 z^3 &= f^3(\vartheta), \\
-\mathbf{u}^T \nabla z^4 + \sigma_2 z^4 &= f^4(\vartheta), \\
z^5 &= \int_{\Omega} f^5(\vartheta) dx,
\end{aligned} \tag{39}$$

where  $\vartheta = (z^1, z^2, z^3, z^4, z^5) \in (H_0^1(\Omega))^2 \times L^2(\Omega) \times H^1(\Omega) \times H^1(\Omega) \times L^2(\Omega)$  and  $\sigma_2 = \sigma_1 - \operatorname{div} \mathbf{u}$ , with the boundary conditions:

$$\begin{aligned}
z^1 &= 0 \quad \text{on} \quad \partial\Omega, \\
M(z^2) &= 0, \\
z^3 &= 0 \quad \text{on} \quad \Sigma_{in}, \\
z^4 &= 0 \quad \text{on} \quad \Sigma_{out}.
\end{aligned}$$

We rewrite the above system in the compact form:

$$\mathcal{L}(\mathbf{u}(\vartheta); \vartheta) = \mathcal{R}(\vartheta), \quad (40)$$

where the coefficients of (39) depend on  $\mathbf{u}$  that is computed in the previous step ( $\mathbf{u}$ , in turn, depends on  $\vartheta$  and  $\mathbf{u}_0$  through (14)). Denote by

$$\xi = \mathcal{S}(\mathbf{u}(\vartheta); \vartheta)$$

the inverse of the linear operator  $\xi \mapsto \mathcal{L}(\mathbf{u}(\vartheta); \xi)$  such that  $\mathcal{L}(\mathbf{u}(\vartheta); \xi) = \mathcal{R}(\vartheta)$ .

In order to find the approximate solution [9], in which an iteration step is defined by the formula:

$$\vartheta_{n+1} := \alpha_n \vartheta_{n+\frac{1}{2}} + (1 - \alpha_n) \vartheta_n, \quad (41)$$

where  $\vartheta_n$  is a value of solution from the previous step, and

$$\vartheta_{n+\frac{1}{2}} := S(\mathbf{u}(\vartheta_n); \vartheta_n). \quad (42)$$

We take  $\vartheta_0 = 0$ . Elements of the sequence  $\alpha_n$  (in the program we set:  $\alpha_n = \frac{1}{2+n}$ ,  $n = 0, 1, \dots$ ) have to satisfy the following conditions:

$$0 \leq \alpha_n \leq 1, \quad \lim_{n \rightarrow \infty} \alpha_n = 0, \quad \sum_{n=0}^{\infty} \alpha_n = \infty.$$

In every iteration of the fixed point method we find the solution (42) of the system (40) with fixed right-hand sides  $f_1^1, f_2^1, f^2, \dots, f^5 \in L^2(\Omega)$ . Let us notice that if we substitute in the system (40) the null operator for  $\mathcal{R}$  and consider only the first three equations, we obtain the Stokes system (it has to be solved once in every iteration). To solve (40) numerically in every iteration, we apply the modification of the Finite Element (FE), used in [28], replacing it by the mixture of Finite Volume and Finite Element (which we will denote shortly by FV). The latter one was used in [10] and [7] for homogeneous systems of the type (1).

Similarly as FE, the FV method serves for solving systems of differential equations in the weak form. Let us then multiply the first 5 equations of (40) by the test functions  $t \in C_0^\infty(\Omega)$  and integrate them over  $\Omega$ . Next, applying integration by parts, we get:

$$\begin{aligned} \int_{\Omega} \nabla t \nabla z_1^1 dx - \int_{\Omega} (\partial_1 t) z^2 dx &= \int_{\Omega} f_1^1 t dx, \\ \int_{\Omega} \nabla t \nabla z_2^1 dx - \int_{\Omega} (\partial_2 t) z^2 dx &= \int_{\Omega} f_2^1 t dx, \\ \int_{\Omega} t \partial_1 z_1^1 dx + \int_{\Omega} t \partial_2 z_2^1 dx &= \int_{\Omega} f^2 t dx, \\ \int_{\Omega} u^T \nabla z^3 t dx + \int_{\Omega} \sigma_1 z^3 t dx &= \int_{\Omega} f^3 t dx, \\ - \int_{\Omega} u^T \nabla z^4 t + \int_{\Omega} \sigma_2 z^4 t &= \int_{\Omega} f^4 t \\ z^5 &= \int_{\Omega} f^5 dx. \end{aligned} \quad (43)$$

The first three equations of (43) constitute a system which is hardest to solve (the last three equations are independent and can be solved separately). The boundary conditions will be discussed later on (cf. (60)).

**6. Approximation method.** Assume that there is a given triangulation of the domain  $\Omega$ , so  $\Omega \approx \sum_{T \subset \mathcal{T}} T$ . Denote by  $h$  the diameter of the biggest triangle in that mesh (that is the diameter of the circumscribed circle). We will denote the set of all triangle edges from  $\mathcal{T}$  by  $E$ , and the sets of all inner edges and outer edges by  $E_{int}$  and  $E_{ext}$ , respectively. We will approximate the functions appearing in the system (39) by functions from certain finitely dimensional spaces (the same spaces as in [10] and [7]). The approximations of the functions  $z_1^1, z_2^1, z^3, z^4$  and  $t$ , except from the test functions from (43/3), are the elements of the space  $V_h$  :

$$V_h =$$

$$\{k_h \in L^2(\Omega) : \forall_{t_i \in \mathcal{T}} k_h|_{t_i} \in P_1(t_i), \forall_{\substack{t_i, t_j \in \mathcal{T} \\ e = t_i | t_j}} \int_e \tilde{k}_h|_{cl(t_i)} dS = \int_e \tilde{k}_h|_{cl(t_j)} dS\}, \quad (44)$$

where  $t_i | t_j$  denotes the common edge of triangles  $t_i$  and  $t_j$ ,  $P_1(t_i)$  - the space of first order polynomials defined on a triangle  $t_i$ , and  $\tilde{k}_h|_{cl(t_i)}$  is a continuous extension of  $k_h|_{t_i}$  on the closure of  $t_i$ . Let  $n_e$  be the number of triangle edges in a mesh  $\mathcal{T}$ . The elements of the space  $V_h$ , which are linear functions on interiors of triangles, are uniquely determined by their values in the centres of edges. Thus this space is of dimension  $n_e$  and its basis consists of the functions:

$$\phi_i = \begin{cases} 1 & \text{on the } i\text{-th edge,} \\ 0 & \text{at the centres of other edges of triangles adjacent to } e_i, \end{cases} \quad (45)$$

$i = 1 \dots n_e$ . On the other hand,  $z^2$ , the right-hand sides  $f_1^1, f_2^1, f^2, \dots, f^5$  and the test functions for the equation (43/3) are approximated by piecewise constant functions  $l_h$  over the triangles,

$$L_h = \{l_h \in L^2(\Omega) : l_h|_t \text{ is a constant for } t \in \mathcal{T}\}, \quad (46)$$

with the basis of characteristic functions over the triangles from  $\mathcal{T}$ :

$$\chi_k, \quad k = 1, \dots, n_t, \quad (47)$$

where  $n_t$  denotes the number of triangles in the mesh. Thus, approximations of functions from (43) in the above bases will have the form:

$$\begin{aligned} z_1^1 &\approx \sum_{i=1}^{n_e} z_{1,i}^1 \phi_i, & z_2^1 &\approx \sum_{i=1}^{n_e} z_{2,i}^1 \phi_i, \\ z^2 &\approx \sum_{k=1}^{n_k} z_k^2 \chi_k, & z^3 &\approx \sum_{i=1}^{n_e} z_i^3 \phi_i, & z^4 &\approx \sum_{i=1}^{n_e} z_i^4 \phi_i, \\ t &\approx \sum_{i=1}^{n_e} k_i \phi_i \quad \text{or} \quad t \approx \sum_{k=1}^{n_k} l_k \chi_k \quad \text{in (43/3),} & & & & (48) \\ f_1^1 &\approx \sum_{k=1}^{n_t} f_{1,k}^1 \chi_k, & f_2^1 &\approx \sum_{k=1}^{n_t} f_{2,k}^1 \chi_k, & f^2 &\approx \sum_{k=1}^{n_t} f_k^2 \chi_k, \\ f^3 &\approx \sum_{k=1}^{n_t} f_k^3 \chi_k, & f^4 &\approx \sum_{k=1}^{n_t} f_k^4 \chi_k, & f^5 &\approx \sum_{k=1}^{n_t} f_k^5 \chi_k, \end{aligned}$$

with real coefficients  $z_{1,i}^1, z_{2,i}^1, z_i^2, z_i^3, z_i^4, k_i, l_k$  and  $f_{1,k}^1, f_{2,k}^1, f_k^2, f_k^3, f_k^4$ . Substituting above sums to the system (43), we obtain a system of algebraic equations in the

matrix form:

$$\begin{aligned}
S z_1^1 - W_1 z^2 &= F_1^1 \\
S z_2^1 - W_2 z^2 &= F_2^1 \\
W_1^T z_1^1 + W_2^T z_2^1 &= F^2 \\
(D + \sigma_1 M) z^3 &= F^3 \\
(-D + \sigma_2 M) z^4 &= F^4 \\
z^5 &= \sum_{k=1}^{n_t} \int_{t_k} f_k^5 \chi_k dx.
\end{aligned} \tag{49}$$

where  $\sigma_2 = \sigma_1 - \operatorname{div} \mathbf{u}_h \in L_h$  (in turn,  $\mathbf{u}_h \in (V_h)^2$  is a discretization of  $\mathbf{u}$ ) and  $G_{1,2}^k$  are vectors such that

$$G_1^k \cdot [u_{e1}^1, u_{e2}^1, u_{e3}^1]^T = \partial_1 u^1|_{t_k}, \quad G_2^k \cdot [u_{e1}^2, u_{e2}^2, u_{e3}^2]^T = \partial_2 u^2|_{t_k}.$$

Here  $[u_{e1}^1, u_{e2}^1, u_{e3}^1]$  and  $[u_{e1}^2, u_{e2}^2, u_{e3}^2]$  are the values of the first and the second velocity component on the edges of a given triangle  $t_k$ . The matrices used above are defined in the following way:

$$\begin{aligned}
S &= \left[ \sum_{k=1}^{n_t} \int_{t_k} \nabla \phi_i \nabla \phi_j dx \right]_{i=1 \dots n_e \times j=1 \dots n_e} \\
W_1 &= \left[ \int_{t_k} (\chi_k \partial_1 \phi_i) dx \right]_{i=1 \dots n_e \times k=1 \dots n_t} & W_2 &= \left[ \int_{t_k} (\chi_k \partial_2 \phi_i) dx \right]_{i=1 \dots n_e \times k=1 \dots n_t} \\
D &= \left[ \sum_{k=1}^{n_t} \int_{t_k} \phi_i \phi_j (G_1 [u_{e1}^1, u_{e2}^1, u_{e3}^1]^T + G_2 [u_{e1}^2, u_{e2}^2, u_{e3}^2]^T) dx \right]_{i=1 \dots n_e \times j=1 \dots n_e}, \\
M &= \left[ \sum_{k=1}^{n_t} \int_{t_k} \phi_i \phi_j dx \right]_{i=1 \dots n_e \times j=1 \dots n_e}
\end{aligned} \tag{50}$$

while the right-hand side vectors are:

$$\begin{aligned}
F_1^1 &= \left[ \sum_{k=1}^{n_t} \int_{t_k} f_{1,k}^1 \chi_k \phi_i dx \right]_{i=1 \dots n_e}, & F_2^1 &= \left[ \sum_{k=1}^{n_t} \int_{t_k} f_{2,k}^1 \chi_k \phi_i dx \right]_{i=1 \dots n_e}, \\
F^2 &= \left[ \sum_{k=1}^{n_t} \int_{t_k} f_k^2 \chi_k \phi_i dx \right]_{i=1 \dots n_e}, & F^3 &= \left[ \sum_{k=1}^{n_t} \int_{t_k} f_k^3 \chi_k \phi_i dx \right]_{i=1 \dots n_e}, \\
F^4 &= \left[ \sum_{k=1}^{n_t} \int_{t_k} f_k^4 \chi_k \phi_i dx \right]_{i=1 \dots n_e}
\end{aligned} \tag{51}$$

The elements of the matrices, as well as the components of the right-hand side vectors can be computed using elementary calculations, like the formula for the gradient of linear function. In order to improve numerical properties of the system (49), we add to its third equation the stabilizing terms  $T_{STAB,1}$  and  $T_{STAB,2}$ . The term  $T_{STAB,3}$  is added to the next two equations. These terms are defined as follows.



Let  $\beta > 0$ ,  $\gamma > 0$ , then

$$T_{STAB,1} = h^\beta \begin{bmatrix} |t_1| & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & |t_2| & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & |t_{n_t}| \end{bmatrix}, \quad (52)$$

where  $|t_k|$  denotes the area of  $k$ -th triangle. Next,

$$T_{STAB,2} = h^\gamma [b_{ij}]_{i=1\dots n_t \times j=1\dots n_t}, \quad (53)$$

with

$$b_{ij} = \begin{cases} |e_{ij_1}|g_{ij_1} + \dots + |e_{ij_n}|g_{ij_n} & \text{for } t_{j_1}, \dots, t_{j_n} \in \mathcal{N}(t_i), \text{ if } i = j \\ -|e_{ij}|g_{ij} & \text{if } i \neq j \text{ and } t_j \in \mathcal{N}(t_i) \\ 0 & \text{in other cases,} \end{cases} \quad (54)$$

Here  $e_{ij} = t_i|t_j$  ( $t_i|t_j$  is a common edge of triangles  $t_i$  and  $t_j$ ),  $\mathcal{N}(t_i)$  denotes the set of neighbours of  $t_i$ , and  $g_{ij} = (s_i h_i + s_j h_j)$ , where  $h_i, h_j$  are heights from  $t_i$  and  $t_j$  which are falling on  $e_{ij}$ .

Let  $t_k \in \mathcal{T}$ . Then:

$$s_k = \begin{cases} \frac{1}{2} & \text{if } t_k \text{ has 1 neighbour,} \\ \frac{1}{4} & \text{if } t_k \text{ has 2 neighbours,} \\ \frac{1}{6} & \text{if } t_k \text{ has 3 neighbours,} \end{cases} \quad (55)$$

The factor  $s_i h_i |e_{ij}|$  corresponds to the  $1$ ,  $\frac{1}{2}$ , or  $\frac{1}{3}$  part of the area of triangle  $t_i$ . Thus the penalty for discontinuity between neighbouring triangles  $t_i$  and  $t_j$  is proportional to their areas. The last term is

$$T_{STAB,3} = \kappa S, \quad (56)$$

and the value of small  $\kappa > 0$  different places of the program; for the definition of  $S$  see formulas (50). Owing to the first term, the system (49) becomes invertible (it corresponds to the condition  $\Pi(\pi) = \pi$  in (19)).  $T_{STAB,2}$  is a regularizing term and  $T_{STAB,3}$  is a term that stabilizes the discretized transport equations, equivalent to adding on the left-hand side the Laplace operator multiplied by  $\kappa$ . Thus the final form of the discretized scheme is the following:

$$\begin{aligned} Sz_1^1 - W_1 z^2 &= F_1^1 \\ Sz_2^1 - W_2 z^2 &= F_2^1 \\ W_1^T z_1^1 + W_2^T z_2^1 + (T_{STAB,1} + T_{STAB,2})z^2 &= F^2 \\ (D + \sigma_1 M + T_{STAB,3})z^3 &= F^3 \\ (-D + \sigma_2 M + T_{STAB,3})z^4 &= F^4 \\ z^5 &= \sum_{k=1}^{n_t} \int_{t_k} f_k^5 \chi_k dx. \end{aligned} \quad (57)$$

**7. Correctness of the approximation method.** For the convergence of numerical method we are going to establish:

1. The existence of solutions to the discrete scheme (57) by a fixed point argument.

2. The convergence of the sequence of discrete solutions to the solution (43) of the model, taking into account appropriate a priori estimates for discrete solutions.

First of all, the linear Stokes system (1) with the nonhomogeneous Dirichlet boundary conditions is transformed by translation to the equivalent system with homogeneous Dirichlet boundary conditions. Recall that the Stokes system under considerations takes the form:

$$\begin{aligned}\Delta \mathbf{u}_0 - \nabla q_0 &= \mathbf{f}^1 \text{ in } \Omega \\ \operatorname{div}(\mathbf{u}_0) &= f^2 \text{ in } \Omega \\ \mathbf{u}_0 &= \mathbf{U} \text{ on } \Sigma, \mathbf{u}_0 = 0 \text{ on } \Gamma, M(q_0) = 0,\end{aligned}$$

Let  $\mathbf{u}'$  satisfy the system:

$$\begin{aligned}\Delta \mathbf{u}' &= 0 \text{ in } \Omega \\ \mathbf{u}' &= \mathbf{U} \text{ on } \Sigma, \mathbf{u}' = 0 \text{ on } \Gamma\end{aligned}\tag{58}$$

Substituting  $\mathbf{w} = \mathbf{u} - \mathbf{u}'$ , we get:

$$\begin{aligned}\Delta \mathbf{w} - \nabla q &= \mathbf{f}^1 \text{ in } \Omega \\ \operatorname{div} \mathbf{w} &= f^2 + \operatorname{div} \mathbf{u}' \text{ in } \Omega \\ \mathbf{w} &= 0 \text{ on } \partial\Omega\end{aligned}\tag{59}$$

Of course, the second equation satisfies the compatibility condition if and only if

$$\int_{\Sigma} \mathbf{U} \cdot \mathbf{n} dS = 0,$$

this condition is imposed in the subsection 8.2.

It means that we can apply the scheme (40) with homogeneous Dirichlet boundary conditions to the whole systems which we solve, including (1). Thus, we replace the approximation space  $V_h$  (defined in (44)) by the space  $V_h^0$ :

$$V_h^0 = \{k_h \in V_h : \forall_{e \in E_{ext}} \int_e \tilde{k}_h dS = 0\},\tag{60}$$

corresponding to zero boundary conditions.

Denote by  $\mathcal{T}_h$  a triangulation  $\mathcal{T}$  of the diameter  $h$ , and by  $(\mathbf{z}_h^1, z_h^2)$  - the sequence of solutions for (61) in the space  $(V_h^0)^2 \times L_h$  corresponding to the triangulation  $\mathcal{T}_h$ . The last three equations of the system (57) are solved, after regularization, in a standard way. Therefore we will concentrate only on its first three equations having the form (59). We rewrite them in the slightly different, but analogical form:  $\forall \mathbf{k}_h \in (V_h^0)^2, \forall t_i \in \mathcal{T}_h$

$$\begin{aligned}\int_{\Omega, b} \nabla \mathbf{z}_h^1 \cdot \nabla \mathbf{k}_h dx - \int_{\Omega, b} z_h^2 \operatorname{div} \mathbf{k}_h dx &= \int_{\Omega, b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \\ \sum_{e=t_i|t_j} v_{e,h} + h^\beta |t_i| z_{t_i,h}^2 + h^\gamma \sum_{e=t_i|t_j} |e|(s_i h_i + s_j h_j)(z_{t_i,h}^2 - z_{t_j,h}^2) &= |t_i| f_{t_i,h}^2,\end{aligned}\tag{61}$$

where  $v_{e,h} = |e| \mathbf{z}_{e,h}^1 \cdot \mathbf{n}_{t_i}$ ,  $\mathbf{n}_{t_i}$  is a normal vector to the edge  $e$ , pointing outwards the triangle  $t_i$  and  $\int_{\Omega, b}$  denotes the following sum of integrals:  $\sum_{t_i \in \mathcal{T}_h} \int_{t_i}$ .

By the symbol  $\|\cdot\|_b$  we will denote the norm of the space  $(V_h^0)^2$ , given by the formula:  $\|\mathbf{z}^1\|_b^2 = \int_{\Omega, b} \nabla \mathbf{z}_h^1 : \nabla \mathbf{z}_h^1 dx$ .

**7.1. Existence of a solution to the discrete scheme.** By an application of the Leray-Schauder Fixed Point Theorem (cf. [11]) we can show that the discrete scheme (61) admits a solution. We recall the theorem.

**Theorem (Leray-Schauder) 1.** Let  $\mathcal{B}$  be a Banach space,  $T$  - compact mapping from  $\mathcal{B} \times [0, 1]$  to  $\mathcal{B}$ , such that  $T(X, 0) = 0$  for all  $X \in \mathcal{B}$ . Assume that there exists a constant  $M$  such that:

$$\|x\|_{\mathcal{B}} \leq M \quad (62)$$

for all  $(x, \lambda) \in \mathcal{B} \times [0, 1]$  satisfying  $x = T(x, \lambda)$ . Then the mapping  $T_1 : \mathcal{B} \rightarrow \mathcal{B}$  of the form  $T_1 X = T(X, 1)$  has a fixed point.

Set  $\mathcal{B} = (V_h^0)^2 \times L_h$  and denote  $X = (\mathbf{A}, B) \in \mathcal{B}$ . Take  $T : \mathcal{B} \times [0, 1] \rightarrow \mathcal{B}$  such that  $T(X, \lambda) = x \in \mathcal{B} : x = (z_h^1, z_h^2)$  satisfies the following system:

$$\begin{aligned} \forall \mathbf{k}_h \in (V_h^0)^2, \quad \forall t_i \in \mathcal{T}_h \\ \int_{\Omega, b} \nabla z_h^1 \cdot \nabla \mathbf{k}_h dx - \lambda \int_{\Omega, b} B \operatorname{div} \mathbf{k}_h dx = \lambda \int_{\Omega, b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \\ \lambda \sum_{e=t_i|t_j} A_{e,h} + h^\beta |t_i| z_{t_i,h}^2 + \\ \lambda h^\gamma \sum_{e=t_i|t_j} (s_k h_k + s_l h_l) |e| (z_{t_i,h}^2 - z_{t_j,h}^2) = \lambda |t_i| f_{t_i,h}^2 \end{aligned} \quad (63)$$

where  $A_{e,h}$  is defined in the same manner as  $v_{e,h}$ . We need to show that the mapping  $T$  satisfies all the assumptions. For  $\lambda = 0$  it has the form:

$$\begin{aligned} \forall \mathbf{k}_h \in (V_h^0)^2 \quad \int_{\Omega, b} \nabla z_h^1 \cdot \nabla \mathbf{k}_h dx = 0 \\ \forall t_i \in \mathcal{T}_h \quad h^\beta |t_i| z_{t_i,h}^2 = 0 \end{aligned} \quad (64)$$

Obviously, the only solution to the above system is the trivial one. Thus  $T$  satisfies the condition:  $T((\mathbf{A}, B), 0) = 0 \forall (\mathbf{A}, B) \in \mathcal{B}$ . The zero solution obviously meets the estimate (62) too. Furthermore,  $T$  - as a linear operator, acting in finite dimensional space - is compact. In order to prove the estimate (62) for the solution to the system:  $\forall \mathbf{k}_h \in (V_h^0)^2, \forall t_i \in \mathcal{T}_h$

$$\begin{aligned} \int_{\Omega, b} \nabla z_h^1 \cdot \nabla \mathbf{k}_h dx - \lambda \int_{\Omega, b} z_h^2 \operatorname{div} \mathbf{k}_h dx = \lambda \int_{\Omega, b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \\ \lambda \sum_{e=t_i|t_j} v_{e,h} + h^\beta |t_i| z_{t_i,h}^2 + \\ \lambda h^\gamma \sum_{e=t_i|t_j} (s_k h_k + s_l h_l) |e| (z_{t_i,h}^2 - z_{t_j,h}^2) = \lambda |t_i| f_{t_i,h}^2 \end{aligned} \quad (65)$$

we can use estimates derived for the scheme (61) (which is equal to (65) for  $\lambda = 1$ ), that are presented further in this paper (see 7.2.1). Since, employing the analogical technique as for (61), we get the following inequality:

$$\begin{aligned} \|z_h^1\|_b \leq \\ \frac{1}{c_2} \sqrt{\lambda c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|f_h^2\|_{L^2(\Omega)} + \lambda c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} + \frac{1}{c_2} \|f_h^2\|_{L^2(\Omega)}}, \end{aligned} \quad (66)$$

(symbols  $c_2, c_1$  denote certain positive constants, described later - see 7.2.1). In turn, estimates for  $z_h^2$  can be derived from the equation (65/2). Let us multiply it by  $z_h^2$  and sum over all triangles from  $\mathcal{T}_h$ :

$$\begin{aligned} & \lambda \int_{\Omega, b} z_h^2 \operatorname{div} \mathbf{z}_h^1 dx + h^\beta \int_{\Omega, b} (z_{t_i, h}^2)^2 dx + \\ & \lambda h^\gamma \sum_{e \in E} |e| (s_i h_i + s_j h_j) (z_{t_i, h}^2 - z_{t_j, h}^2)^2 = \lambda \int_{\Omega, b} f_h^2 \cdot z_h^2 dx. \end{aligned} \quad (67)$$

After substituting for  $\lambda \int_{\Omega, b} z_h^2 \operatorname{div} \mathbf{z}_h^1$  the identity from (65/1) (with  $\mathbf{k}_h = \mathbf{z}_h^1$ ), we get (from definition of norm  $\|\cdot\|_b$  in the space  $(V_h^0)^2$ ):

$$\begin{aligned} \|\mathbf{z}_h^1\|_b^2 + h^\beta \int_{\Omega, b} (z_{t_i, h}^2)^2 dx + \lambda h^\gamma \sum_{e \in E} |e| (s_i h_i + s_j h_j) (z_{t_i, h}^2 - z_{t_j, h}^2)^2 = \\ \lambda \int_{\Omega, b} \mathbf{f}_h^1 \cdot \mathbf{z}_h^1 dx + \lambda \int_{\Omega, b} f_h^2 \cdot z_h^2 dx. \end{aligned} \quad (68)$$

Since the first and the third term on the left are non-negative, the following inequality must hold:

$$h^\beta \int_{\Omega, b} (z_{t_i, h}^2)^2 dx \leq \lambda \int_{\Omega, b} \mathbf{f}_h^1 \cdot \mathbf{z}_h^1 dx + \lambda \int_{\Omega, b} f_h^2 \cdot z_h^2 dx, \quad (69)$$

and after applying definition of the norm in  $L^2(\Omega)$  and Hölder's inequality, we have:

$$h^\beta \|z_h^2\|_{L^2(\Omega)}^2 \leq c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b + \|f_h^2\|_{L^2(\Omega)} \|z_h^2\|_{L^2(\Omega)}, \quad (70)$$

then:

$$\|z_h^2\|_{L^2(\Omega)} (h^\beta \|z_h^2\|_{L^2(\Omega)} - \lambda \|f_h^2\|_{L^2(\Omega)}) \leq \lambda c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b, \quad (71)$$

and in consequence:

$$\sqrt{h^\beta} \|z_h^2\|_{L^2(\Omega)} - \frac{\lambda}{\sqrt{h^\beta}} \|f_h^2\|_{L^2(\Omega)} \leq \lambda c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b, \quad (72)$$

if only

$$\|z_h^2\|_{L^2(\Omega)} - \frac{\lambda}{h^\beta} \|f_h^2\|_{L^2(\Omega)} > 0.$$

Otherwise it holds:

$$\|z_h^2\|_{L^2(\Omega)} \leq \frac{\lambda}{h^\beta} \|f_h^2\|_{L^2(\Omega)},$$

which already gives an estimate for  $\|z_h^2\|_{L^2(\Omega)}$ . From (72) results:

$$\|z_h^2\|_{L^2(\Omega)} \leq \frac{\sqrt{\lambda c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b}}{\sqrt{h^\beta}} + \frac{\lambda}{h^\beta} \|f_h^2\|_{L^2(\Omega)} \quad (73)$$

In the presence of inequalities (66), (73) and the fact that  $\lambda \in (0, 1)$ , we have:

$$\|\mathbf{z}_h^1\|_b \leq \frac{1}{c_2} \sqrt{c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|f_h^2\|_{L^2(\Omega)}} + c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} + \frac{1}{c_2} \|f_h^2\|_{L^2(\Omega)}, \quad (74)$$

and

$$\|z_h^2\|_{L^2(\Omega)} \leq \frac{1}{\sqrt{h^\beta}} \sqrt{c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b} + \frac{1}{h^\beta} \|f_h^2\|_{L^2(\Omega)} \quad (75)$$

The right-hand sides of the above inequalities are given and independent of  $\lambda$ , so also the assumption (62) is satisfied. That ensures the existence of a solution to the system (61).

**7.2. Convergence of discrete solutions to the solution of the model.** We will show that the sequence  $(\mathbf{z}_h^1, z_h^2)$  has an estimate in  $V_0^1(\Omega)^2 \times L^2(\Omega)$  independent of the diameter  $h$  of the mesh. This makes possible to select a weakly convergent subsequence. Furthermore, we have to know that this subsequence converges to the model's solution when the mesh diameter  $h \rightarrow 0$ .

**7.2.1. Estimates for discrete solutions.** Multiplying the second equation of the scheme (57) by  $z_h^2$  and summing it over all triangles from  $\mathcal{T}_h$ , we get:

$$\int_{\Omega,b} z_h^2 \operatorname{div} \mathbf{z}_h^1 dx + h^\beta \int_{\Omega,b} (z_{t_i,h}^2)^2 dx + h^\gamma \sum_{e \in E} |e| (s_i h_i + s_j h_j) (z_{t_i,h}^2 - z_{t_j,h}^2)^2 = \int_{\Omega,b} f_h^2 \cdot z_h^2 dx. \quad (76)$$

Next, substituting for  $\int_{\Omega,b} z_h^2 \operatorname{div} \mathbf{z}_h^1 dx$  the identity (61/1) with  $\mathbf{k}_h = \mathbf{z}_h^1$ , we get from definition of the norm  $\|\cdot\|_b$  in the space  $V_h^0$ :

$$\|\mathbf{z}_h^1\|_h^2 + h^\beta \int_{\Omega,b} z_{t_i,h}^2 dx + h^\gamma \sum_{e \in E} |e| (s_i h_i + s_j h_j) (z_{t_i,h}^2 - z_{t_j,h}^2)^2 = \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{z}_h^1 dx + \int_{\Omega,b} f_h^2 \cdot z_h^2 dx. \quad (77)$$

Removing the last two terms on the left-hand side, we obtain:

$$\|\mathbf{z}_h^1\|_h^2 \leq \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{z}_h^1 dx + \int_{\Omega,b} f_h^2 \cdot z_h^2 dx. \quad (78)$$

Employing Hölder's inequality to the right-hand side gives:

$$\|\mathbf{z}_h^1\|_h^2 \leq \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_{L^2(\Omega)^2} + \|f_h^2\|_{L^2(\Omega)} \|z_h^2\|_{L^2(\Omega)} \leq c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{z}_h^1\|_b + \|f_h^2\|_{L^2(\Omega)} \|z_h^2\|_{L^2(\Omega)}, \quad (79)$$

wherein the last inequality results from Friedrichs-Poincaré inequality for the space  $(V_h^0)^2$ , having a form:

$$\int_{\Omega} f^2(x) dx \leq c_1 \sum_{t_i \in \mathcal{T}_h} \int_{t_i} |\nabla f(x)|^2 dx \quad \forall f \in (V_h^0)^2, \forall h \geq 0, \quad (80)$$

where the constant  $c_1$  depends only on  $\Omega$  and the regularity of the mesh. In consequence:

$$\|z_h^2\|_{L^2(\Omega)} \geq \frac{\|\mathbf{z}_h^1\|_b (\|\mathbf{z}_h^1\|_b - c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2})}{\|f_h^2\|_{L^2(\Omega)}}, \quad (81)$$

assuming that  $\|f_h^2\|_{L^2(\Omega)} > 0$  (otherwise the inequality (78) simplifies). Now we can use the inf-sup stability of the pair of spaces  $((V_h^0)^2, L_h)$  (see [10],[7]):

$$\sup_{\mathbf{k}_h \in (V_h^0)^2} \frac{\int_{\Omega,b} z_h^2 \operatorname{div}(\mathbf{k}_h) dx}{\|\mathbf{k}_h\|_h} \geq c_2 \|z_h^2 - M(z_h^2)\|_{L^2(\Omega)}, \quad (82)$$

where the constant  $c_2 > 0$  is independent of the size of triangulation. Substituting for  $\int_{\Omega,b} z_h^2 \operatorname{div}(\mathbf{k}_h) dx$  the expression (61/1), we obtain:

$$\sup_{\mathbf{k}_h \in (V_h^0)^2} \frac{\int_{\Omega,b} \nabla z_h^1 \cdot \nabla \mathbf{k}_h dx - \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx}{\|\mathbf{k}_h\|_h} \geq c_2 \|z_h^2 - M(z_h^2)\|_{L^2(\Omega)}. \quad (83)$$

But, since  $M(z_h^2) = 0$ , we have  $\|z_h^2 - M(z_h^2)\|_{L^2(\Omega)} = \|z_h^2\|_{L^2(\Omega)}$ . Next, the following sequence of estimates holds:

$$\begin{aligned} & \int_{\Omega,b} \nabla \mathbf{z}_h^1 \cdot \nabla \mathbf{k}_h dx - \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \\ & \leq \left| \int_{\Omega,b} \nabla \mathbf{z}_h^1 \cdot \nabla \mathbf{k}_h dx \right| + \left| \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \right| \leq \\ & \|\mathbf{z}_h^1\|_b \|\mathbf{k}_h\|_b + \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{k}_h\|_{L^2(\Omega)} \leq \\ & \|\mathbf{z}_h^1\|_b \|\mathbf{k}_h\|_b + c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|\mathbf{k}_h\|_b. \end{aligned} \quad (84)$$

Then, substituting the estimate from (84) to inequality (83), we get:

$$\|\mathbf{z}_h^1\|_b + c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \geq c_2 \|z_h^2\|_{L^2(\Omega)}. \quad (85)$$

Employing (81), we obtain the following relation leading directly to the estimate for  $\mathbf{z}_h^1$ :

$$\|\mathbf{z}_h^1\|_b + c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \geq c_2 \frac{\|\mathbf{z}_h^1\|_b (\|\mathbf{z}_h^1\|_b - c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2})}{\|f_h^2\|_{L^2(\Omega)}}. \quad (86)$$

Above inequality gives:

$$\begin{aligned} c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|f_h^2\|_{L^2(\Omega)} & \geq \\ & \|\mathbf{z}_h^1\|_b (c_2 \|\mathbf{z}_h^1\|_b - c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} - \|f_h^2\|_{L^2(\Omega)}). \end{aligned} \quad (87)$$

Of course, the right-hand side of that inequality can be estimated from below by the expression:

$$c_2 (\|\mathbf{z}_h^1\|_b - c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} - \frac{1}{c_2} \|f_h^2\|_{L^2(\Omega)})^2, \quad (88)$$

if only it is non-negative; otherwise we arrive in the estimate of the form:

$$\|\mathbf{z}_h^1\|_b \leq c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} + \frac{1}{c_2} \|f_h^2\|_{L^2(\Omega)}. \quad (89)$$

Thus:

$$\sqrt{c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|f_h^2\|_{L^2(\Omega)}} \geq c_2 \|\mathbf{z}_h^1\|_b - c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} - \|f_h^2\|_{L^2(\Omega)}, \quad (90)$$

so finally we have:

$$\|\mathbf{z}_h^1\|_b \leq \frac{1}{c_2} \sqrt{c_2 c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} \|f_h^2\|_{L^2(\Omega)}} + c_1 \|\mathbf{f}_h^1\|_{L^2(\Omega)^2} + \frac{1}{c_2} \|f_h^2\|_{L^2(\Omega)}. \quad (91)$$

Simultaneously, in the presence of (85) and (91), we easily obtain a bound for  $\|z_h^2\|_{L^2(\Omega)}$ :

$$\|z_h^2\|_{L^2(\Omega)} \leq \frac{1}{c_2} \|\mathbf{z}_h^1\|_b + \frac{c_1}{c_2} \|\mathbf{f}_h^1\|_{L^2(\Omega)^2}. \quad (92)$$

For further computations it will be also necessary to estimate the expression  $|z_h^2|_{T_h}$ , defined in the space  $L_h$  in the following way:

$$|z_h^2|_{T_h}^2 = \sum_{e \in E_{int}} (h_i + h_j) |e| (z_{t_i,h}^2 - z_{t_j,h}^2)^2. \quad (93)$$

From the equality (77) follows that:

$$h^\gamma \sum_{e \in E} |e| (s_i h_i + s_j h_j) (z_{t_i,h}^2 - z_{t_j,h}^2)^2 \leq \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{z}_h^1 dx + \int_{\Omega,b} f_h^2 \cdot p dx. \quad (94)$$

Furthermore, as  $s_i \geq \frac{1}{6}$ , then  $6(s_i h_i + s_j h_j) \geq (h_i + h_j)$  and:

$$h^\gamma \sum_{e \in E} |e|(h_i + h_j)(z_{t_i,h}^2 - z_{t_j,h}^2)^2 \leq 6h^\gamma \sum_{e \in E} |e|(s_i h_i + s_j h_j)(z_{t_i,h}^2 - z_{t_j,h}^2)^2. \quad (95)$$

Thus, an estimate for  $|z_h^2|_{\mathcal{T}_h}$  can be obtained on the basis of the estimate of the right-hand side of (94).

By Theorem 3.3 in [10] it follows that the estimates for discrete solutions  $(\mathbf{z}_h^1, z_h^2)$  in  $V_0^1(\Omega)^2 \times L^2(\Omega)$  which are independent of the mesh size imply the strong convergence of the sequence  $\mathbf{z}_h^1$  in  $L^2(\Omega)^2$  to the limit  $\mathbf{z}^1$  from  $H_0^1(\Omega)^2$  and the weak convergence of  $z_h^2$  in  $L^2(\Omega)$  (these convergences hold for subsequences only).

*7.2.2. Convergence of discrete schemes to the model.* We want to show that the scheme:  $\forall \mathbf{k}_h \in (V_h)^2 \quad \forall t_i \in \mathcal{T}_h$

$$\begin{aligned} \int_{\Omega,b} \nabla \mathbf{z}_h^1 \cdot \nabla \mathbf{k}_h dx - \int_{\Omega,b} z_h^2 \operatorname{div} \mathbf{k}_h dx &= \int_{\Omega,b} \mathbf{f}_h^1 \cdot \mathbf{k}_h dx \\ \sum_{e=t_i|t_j} v_{e,h} + h^\beta |t_i| z_{t_i,h}^2 + h^\gamma \sum_{e=t_i|t_j} |e|(s_i h_i + s_j h_j)(z_{t_i,h}^2 - z_{t_j,h}^2) &= |t_i| f_{t_i,h}^2 \end{aligned} \quad (96)$$

converges to the system of the form:

$$\begin{aligned} \int_{\Omega} \nabla \mathbf{z}^1 \cdot \nabla \mathbf{k} dx - \int_{\Omega} z^2 \operatorname{div} \mathbf{k} &= \int_{\Omega} \mathbf{f}^1 \cdot \mathbf{k} dx \quad \forall \mathbf{k} \in C_0^\infty(\Omega)^2 \\ \int_{\Omega} \operatorname{div}(\mathbf{z}^1) l dx &= \int_{\Omega} f^2 l dx \quad \forall l \in C_0^\infty(\Omega) \end{aligned} \quad (97)$$

Since the proof of convergence of the first equation in the scheme (96) to the first one in (97) is considered in the paper [10], we can direct the reader to that source. Regarding the convergence of the second one, we do not have a ready result but we may apply some ideas from [10]. First, multiplying (96/2) by the approximation  $l_h \in L^h$  of a smooth function  $l \in C_0^\infty(\Omega)$  and summing over all triangles from  $\mathcal{T}_h$ , we obtain:

$$T_1 + T_2 + T_3 = \sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 l_{t_i,h}, \quad (98)$$

denoting:

$$\begin{aligned} T_1 &= \sum_{t_i \in \mathcal{T}_h} \sum_{e=t_i|t_j} v_{e,h} l_{t_i,h} \\ T_2 &= \sum_{t_i \in \mathcal{T}} h^\beta |t_i| z_{t_i,h}^2 l_{t_i,h} \\ T_3 &= \sum_{t_i \in \mathcal{T}} h^\gamma \sum_{e=t_i|t_j} |e|(s_i h_i + s_j h_j)(z_{t_i,h}^2 - z_{t_j,h}^2) l_{t_i,h}, \end{aligned} \quad (99)$$

where  $l_{t_i,h}$  is the mean value of  $l$  on the triangle  $t_i$ . We will show that the first term converges to  $\int_{\Omega} \operatorname{div} \mathbf{z}^1 l dx$  and the other two tend to zero when  $h \rightarrow 0$ . At the beginning, one can notice that the following equalities hold:

$$T_1 = \int_{\Omega,b} \operatorname{div} \mathbf{z}_h^1 l_h dx = \int_{\Omega,b} \operatorname{div} \mathbf{z}_h^1 (l_h - l) dx + \int_{\Omega,b} \operatorname{div} \mathbf{z}_h^1 l dx, \quad (100)$$

since:  $l_h \rightarrow l$  in  $L^2(\Omega)$  (according to *Lemma 2.1(3)* from [10]), and  $\operatorname{div} \mathbf{z}_h^1$  is bounded in  $L^2(\Omega)$  on every triangle  $t_i$ , as:

$$\begin{aligned} \int_{t_i} (\operatorname{div} \mathbf{z}_h^1)^2 dx &= \int_{t_i} \left( \frac{\partial \mathbf{z}_{1,h}^1}{\partial x_1} dx + \frac{\partial \mathbf{z}_{2,h}^1}{\partial x_2} \right)^2 dx = \\ & \int_{t_i} \left( \frac{\partial \mathbf{z}_{1,h}^1}{\partial x_1}^2 + 2 \frac{\partial \mathbf{z}_{1,h}^1}{\partial x_1} \frac{\partial \mathbf{z}_{2,h}^1}{\partial x_2} + \frac{\partial \mathbf{z}_{2,h}^1}{\partial x_2}^2 \right) dx \leq \\ & 2 \int_{t_i} \left( \frac{\partial \mathbf{z}_{1,h}^1}{\partial x_1}^2 + \frac{\partial \mathbf{z}_{2,h}^1}{\partial x_2}^2 \right) dx \leq 2 \int_{t_i} \nabla \mathbf{z}_h^1 : \nabla \mathbf{z}_h^1 = 2 \| \mathbf{z}_h^1 \|_b. \end{aligned} \quad (101)$$

Next, integrating the second term by parts, we get

$$\int_{\Omega, b} \operatorname{div} \mathbf{z}_h^1 l dx = - \int_{\Omega, b} \mathbf{z}_h^1 \nabla l dx \xrightarrow{h \rightarrow 0} - \int_{\Omega} \mathbf{z}^1 \nabla l dx,$$

because it was already demonstrated that  $\mathbf{z}_h^1 \rightarrow \mathbf{z}^1$  in  $L^2(\Omega)^2$  to a function from  $H_0^1(\Omega)^2$ . Applying integration by parts one more time, we obtain the needed result for the first term. Obviously, the second term  $T_2$  tends to zero when  $h$  tends to zero since, as it was already shown, the expression  $\| z_h^2 \|_{L^2(\Omega)}$  is bounded. As for the term  $T_3$ , from the boundedness of  $|z_h^2|_{\mathcal{T}_h}$  we get  $T_3 \rightarrow 0$ , owing to:

$$\begin{aligned} T_3 &= \sum_{t_i \in \mathcal{T}} h^\gamma \sum_{e=t_i|t_j} |e|(s_i h_i + s_j h_j)(z_{t_i,h}^2 - z_{t_j,h}^2) l_{t_i} \leq \\ & \leq \sum_{t_i \in \mathcal{T}} h^\gamma \sum_{e=t_i|t_j} |e|(h_i + h_j)(z_{t_i,h}^2 - z_{t_j,h}^2) l_{t_i} \leq h^\gamma |z_h^2|_{\mathcal{T}_h} \|l_h\|_{C_0^\infty(\Omega)} \xrightarrow{h \rightarrow 0} 0 \end{aligned} \quad (102)$$

Finally, again from *Lemma 2.1(3)* from [10], it is obvious that:

$$\begin{aligned} \sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 l_{t_i,h} &= \sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 (l_{t_i,h} - l) + \sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 l \\ & \rightarrow \sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 l, \end{aligned} \quad (103)$$

and from the weak convergence of  $f_{t_i,h}^2$  in  $L^2(\Omega)$  we have:

$$\sum_{t_i \in \mathcal{T}_h} |t_i| f_{t_i,h}^2 l \rightarrow \int_{\Omega} f^2 l dx. \quad (104)$$

This, with convergence of the terms  $T_1$ ,  $T_2$  and  $T_3$  from the left hand side of (98), gives the result.

The novelty of the analysis performed above consists in considering the system with  $f^2 \neq 0$ . The reasoning is based on [10], [7], where a nonlinear system was considered, but had to be supplemented by the elements taking into account this non-homogeneity.

## 8. Example of obstacle shape optimization.

**8.1. Shape and size of the computational domain  $\Omega$ .** We define the computational domain  $\Omega$  as an ellipse with two small circular holes inside (one of them is the inlet, and the other one - the initial shape of the obstacle, see Fig.1), situated symmetrically with respect to its centre. The radii of the considered ellipse are 30 and 16 units. The centres of the circles lie in the horizontal axis of the ellipse, each



in the distance of 3,5 from its centre. The length of the radii of both circles equals 1.

**8.2. Boundary conditions - vector field  $\mathbf{U}$ .** Vector field  $\mathbf{U}$ , that appears in the boundary condition (2c) is chosen in such a way, that it satisfies the compatibility condition for the equation (1), which reads:

$$\int_{\Sigma_{in}} \mathbf{U} \cdot \mathbf{n} \, ds + \int_{\Sigma_{out}} \mathbf{U} \cdot \mathbf{n} \, ds = 0.$$

It corresponds to the physical condition that the same amount of the fluid enters and leaves  $\Omega$ . We assume that the fluid passes through the boundary  $\partial\Omega$  in normal direction. We take  $\mathbf{U} = a \frac{(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|^2}$  ( $a \in \mathbb{R}$ ), which is the velocity field in the solution to the system (116) defined for  $\mathbf{x}_0$ , the source of the gas, taken as the coordinates of the centre of  $\Sigma_{in}$ .

**8.3. Optimization step.** Using the gradient method, we want to find a transformation of  $\Omega$  acting in the direction of the steepest descent of  $J$ . For a fixed discretization of  $\Gamma$ , the vector field  $\mathbf{T}$  is a linear combination of basic fields  $\mathbf{T}_i$  related to the boundary points  $i = 1..n_b$ , namely:

$$\mathbf{T} = t_1 \mathbf{T}_1 + \dots + t_{n_b} \mathbf{T}_{n_b}, \quad (105)$$

where  $\mathbf{t} = [t_i]$ ,  $i \in 1..n_b$  are real coefficients. Thus the gradient of the functional  $J$  contains the derivatives with respect to the basic fields

$$\mathbf{g} = [g^1, \dots, g^{n_b}] = \frac{\partial J}{\partial \mathbf{t}} = \left[ \frac{\partial J}{\partial t_1}, \dots, \frac{\partial J}{\partial t_{n_b}} \right] \quad (106)$$

taken at  $t_i = 0$ ,  $i = 1..n_b$ .

In order to construct these fields, we approximate the boundary  $\Gamma$  by a closed spline curve  $\xi$  of class  $C^2$ , passing through all the points of the discretized boundary (which are the vertices of triangles from  $\mathcal{T}$  that touch  $\Gamma$ ) and parametrized by the arc length  $s$ :

$$\xi = \xi(s), \quad s \in [0, L], \quad \xi(s_k) = \mathbf{p}_k, \quad k = 1..n_b, \quad (107)$$

where  $L$  is the length of  $\Gamma$ . Since we want to construct normal movements of the boundary, for every point  $\mathbf{p}_k$  we need a formula for the outer normal vector. It has the form:

$$\mathbf{n}_k = \frac{N \xi'(s_k)}{\|\xi'(s_k)\|}, \quad N = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (108)$$

Then we define the hat function around a given boundary discretization point of  $\Gamma$  by the following formula:

$$b_k(s) = \exp\left[-\left(\frac{\text{dist}(s, s_k)}{w_0}\right)^2\right], \quad (109)$$

where  $\text{dist}(\cdot, s_k)$  is a minimal distance of  $s$  from  $s_k$ :

$$\text{dist}(s, s_k) = \min\{|s - s_k|, L - |s - s_k|\} \quad (110)$$

and the constant  $w_0$  determines the "width" of the function  $b_k(s)$ . To obtain a normal shift of every boundary point  $\mathbf{p}_j$ , we multiply  $b_k(s_j)$  by the normal vector  $\mathbf{n}_j$ :

$$\mathbf{h}_k(\mathbf{p}_j) = b_k(s_j) \mathbf{n}_j, \quad j = 1..n_b \quad (111)$$

Applying the formula (27), all vector fields  $\mathbf{T}_k$  on  $\Omega$  can be evaluated.

During our computations we impose two obvious constraints. The first is the constancy of the volume, since otherwise the optimal obstacle would reduce to the single point. If we assume, that the boundary nodal points of  $S_h$  move by the vector  $\sum_{k \in 1..n_b} \zeta^k \mathbf{h}_k$ , then the requirement of fixed volume reduces to

$$\sum_{k \in 1..n_b} \zeta^k \left( \int_{\Omega_h} \operatorname{div} \mathbf{T}_k dx \right) = \sum_{k \in 1..n_b} \zeta^k d_k = 0$$

where  $d_k = \int_{\Omega_h} \operatorname{div} \mathbf{T}_k dx$ .

The second constraint concerns the position of the mass centre of the boundary of the obstacle. The speed of the flow decreases with the distance from the inlet, therefore, in order to reduce the drag, the obstacle would move as far to the right as it is possible, see Fig. 1. We want to prevent it, because we are interested in the optimal shape at the specified position.

The condition for the mass center to be fixed at the origin reads

$$\int_{\partial S_h} x_j ds = 0, \quad j = 1, 2. \quad (112)$$

This may be transformed as follows:

$$\int_{\partial S_h} x_j ds = \sum_{k=1}^{n_b} l_k x_j^k \approx l_0 \sum_{k=1}^{n_b} x_j^k,$$

Here we have assumed that the lengths of boundary edges  $l_k$  are approximately the same and equal  $l_0$ . This condition holds at the beginning of each optimization step, see next paragraph. Therefore, in view of (112),

$$\sum_{k=1}^{n_b} x_j^k = 0, \quad j = 1, 2. \quad (113)$$

Now if we move each nodal point by  $\sum_{k=1}^{n_b} \zeta^k \mathbf{h}_k(\mathbf{p}_i)$ ,  $i = 1, \dots, n_b$ , the above condition takes on the form

$$\sum_{i=1}^{n_b} \left[ x_j^i + \sum_{k=1}^{n_b} \zeta^k h_k^j(\mathbf{p}_i) \right] = 0, \quad j = 1, 2.$$

Because of (113), the condition for keeping mass centre of  $\Gamma_h$  in place thus becomes

$$\sum_{k=1}^{n_b} \zeta^k \left[ \sum_{i=1}^{n_b} h_k^j(\mathbf{p}_i) \right] = 0, \quad j = 1, 2. \quad (114)$$

Let us denote

$$\mathbf{d} = [d_1, \dots, d_{n_b}]^T,$$

$$\tilde{\mathbf{h}}^j = [\tilde{h}_1^j, \dots, \tilde{h}_{n_b}^j]^T, \quad \text{where} \quad \tilde{h}_k^j = \sum_{i=1}^{n_b} h_k^j(\mathbf{p}_i), \quad j = 1, 2.$$

As a result we must project the gradient  $\mathbf{g}$  on the intersection of three hyperplanes

$$\sum_{k=1}^{n_b} \zeta^k d_k = 0, \quad \sum_{k=1}^{n_b} \zeta^k \tilde{h}_k^j = 0, \quad j = 1, 2.$$

It is easily done by taking

$$\mathbf{g}_1 = \mathbf{g} - \alpha_0 \mathbf{d} - \alpha_1 \tilde{\mathbf{h}}^1 - \alpha_2 \tilde{\mathbf{h}}^2$$

and finding  $\alpha_0, \alpha_1, \alpha_2$  from equations

$$\mathbf{g}_1^T \mathbf{d} = 0, \quad \mathbf{g}_1^T \tilde{\mathbf{h}}^1 = 0, \quad \mathbf{g}_1^T \tilde{\mathbf{h}}^2 = 0.$$

As it was mentioned above, after each optimization step we distribute  $s_k$  approximately uniformly along  $\xi(s)$ , see (107). This simplifies the condition for keeping the obstacle in place, but also prevents other undesirable behaviour. In places of higher curvature the nodal points on  $\Gamma_h$  may get nearer and nearer in the process of optimization. Keeping the distances between these points equal prevents them from overlapping during the consecutive optimization steps. The results of the computations are shown in Fig.2.

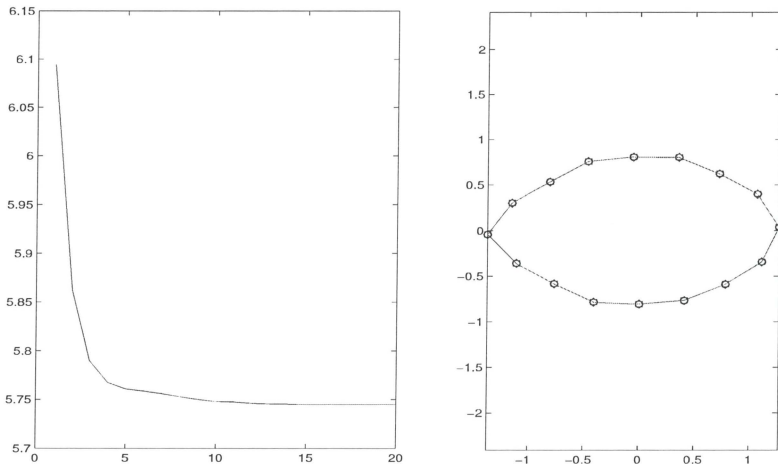


Figure 2. On the left: the plot of  $J$  versus number of steps; on the right: the final shape of the obstacle  $\Gamma$  after minimization of the drag

**8.4. Experimental computation of the rates of convergence of FE and FV.** To investigate the rates of convergence of both methods, we use the Stokes system:

$$\begin{aligned} \Delta \mathbf{u} - \nabla p &= 0 \text{ in } \Omega \\ \operatorname{div}(\mathbf{u}) &= 0 \text{ in } \Omega \\ \mathbf{u} &= \mathbf{U} \text{ on } \partial\Omega \end{aligned} \tag{115}$$

with the field  $\mathbf{U}$  generated by a fundamental solution (Stoklet) to the system:

$$\begin{aligned} \Delta \mathbf{U}_1 - \partial_1 P &= \delta(\mathbf{x} - \mathbf{x}_0) \text{ in } \mathbb{R}^2, \\ \Delta \mathbf{U}_2 - \partial_2 P &= 0 \text{ in } \mathbb{R}^2, \\ \operatorname{div}(\mathbf{U}) &= 0 \text{ in } \mathbb{R}^2, \end{aligned} \tag{116}$$

where  $\mathbf{x}_0 \notin \Omega$ . Thus, the solution to (115) is given by the formula:

$$\begin{aligned} u_1 &= -0.5 \log(\|\mathbf{x} - \mathbf{x}_0\|^2) + \frac{x_1^0 - x_1}{\|\mathbf{x} - \mathbf{x}^0\|^2} \\ u_2 &= \frac{(x_1 - x_1^0)(x_2 - x_2^0)}{\|\mathbf{x} - \mathbf{x}^0\|^2}, \\ p &= \frac{2(x_1 - x_1^0)}{\|\mathbf{x} - \mathbf{x}^0\|^2}, \end{aligned} \quad (117)$$

In order to compute the rates of convergence of  $\mathbf{u}$  and  $p$ , we make 3 subsequent mesh condensations (every time we take  $h := \frac{1}{2}h$ ). Regularity of the mesh is assured by Matlab's algorithm. Denote by  $u_1err$ ,  $u_2err$  and  $perr$  errors of the velocity components and the pressure, computed in the norm of  $L^2(\Omega)$ . We take the total error of the velocity as:

$$uerr = \sqrt{u_1err^2 + u_2err^2}. \quad (118)$$

In both methods the same scheme (57) and the same stabilizing terms (52), (53), (56) have been employed. The values of  $\beta$  and  $\gamma$  are chosen in such a way ( $\beta = 1, 2$ ,  $\gamma = 0, 6$ ) that they minimize the errors of the Stoklet solution. The same values were used in the computations done during the optimization. The convergence rates of the velocity and pressure obtained from comparison with exact solutions are 3.0714 and 0.9158 for FE compared to 2.5763 and 1.5480 for FV. One can remark that error values themselves are not actually important, only the convergence rate - since for the same triangulation the number of discrete variables for FV is approximately 3 times bigger than for FE. The obtained results are presented on the plot below:

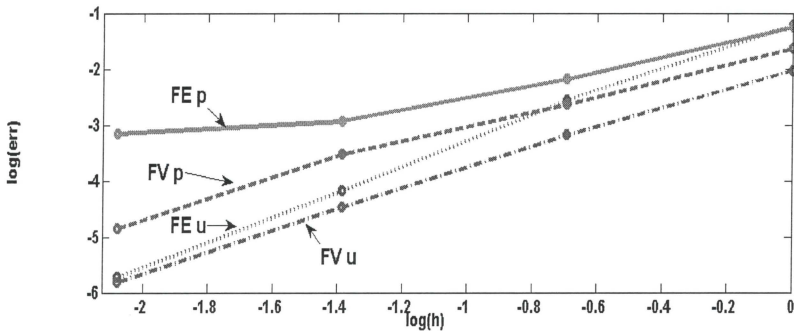


Figure 3.

It can be concluded that FV approximates the pressure more accurately, whereas its advantage in accuracy for the velocity decreases with subsequent mesh condensation steps. However, the accuracy of the pressure is crucial in this case.

#### REFERENCES

- [1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
- [2] H. Beirão da Veiga, Stationary motions and the incompressible limit for compressible viscous fluids, *Houston J. Math.* 13 (1987), 527–544.

- [3] H. Beirão da Veiga, An  $L^p$ -theory for the  $n$ -dimensional, stationary, compressible Navier-Stokes equations, and the incompressible limit for compressible fluids. The equilibrium solutions, *Comm. Math. Phys.* **109** (1987), 229–248.
- [4] H. Beirão da Veiga, Existence results in Sobolev spaces for a transport equation, *Ricerca Mat.* **36** (1987), suppl., 173–184.
- [5] C. Brandenburg, F. Lindemann, M. Ulbrich, S. Ulbrich, Advanced numerical methods for PDE constrained optimization with application to optimal design in Navier Stokes flow, *Constrained optimization and optimal control for partial differential equations, in series: Internat. Ser. Numer. Math.*, **160**, Birkhäuser/Springer Basel AG, Basel, (2012), 257–275.
- [6] E. Casas An optimal control problem governed by the evolution Navier-Stokes equations. In S. S. Sretharan, editor, Optimal control of viscous flows, *Frontiers in applied mathematics*, SIAM, Philadelphia (1993).
- [7] R. Eymard, T. Gallouet, R. Herbin, J. Latche, A convergent finite element-finite volume scheme for the compressible Stokes problem; part II - The isentropic case, *Mathematics of Computation*, **79**, No.270 (2010), 649–675
- [8] E. Feireisl, T. Karper, M. Pokorný, *Mathematical Theory of Compressible Viscous Fluids*, Birkhäuser Basel, 2016
- [9] V. Berinde, *Iterative Approximation of Fixed Points*, Lecture Notes in Mathematics 1912, Springer Verlag 2007
- [10] T. Gallouet, R. Herbin, J. Latche, A convergent finite element-finite volume scheme for the compressible Stokes problem; part I - The isothermal case, *Mathematics of Computation*, **78**, No.267, (2009), 1333–1352
- [11] D. Gilbarg, N. Trudinger *Elliptic partial differential equations of second order*, Springer, 2001
- [12] T. Kondoh, T. Matsumori, A. Kawamoto, Drag minimization and lift maximization in laminar flows via topology optimization employing simple objective function expressions based on body force integration, *Struct. Multidiscip. Optim.*, **45**, (2012), 693–701.
- [13] B. Mohammadi, O. Pironneau, Shape optimization in fluid mechanics, *Ann. Rev. Fluid Mech.* **36** (2004), 255–279.
- [14] B. Mohammadi, O. Pironneau, *Applied Shape Optimization for Fluids*, 2<sup>nd</sup> edition, Oxford Univ. Press, Oxford, 2010.
- [15] S. Novo, Compressible Navier-Stokes model with inflow-outflow boundary conditions, *J. Math. Fluid Mech.* **7** (2005), 485–514.
- [16] A. Novotný, About steady transport equation. I.  $L^p$ -approach in domains with smooth boundaries, *Comment. Math. Univ. Carolin.* **37** (1996), 43–89.
- [17] A. Novotný, About steady transport equation. II. Schauder estimates in domains with smooth boundaries, *Portugal. Math.* **54** (1997), 317–333.
- [18] A. Novotný, M. Padula, Existence and uniqueness of stationary solutions for viscous compressible heat conductive fluid with large potential and small non-potential external forces, *Sibirsk. Mat. Zh.* **34** no. 5, (1993), 120–146 (in Russian).
- [19] A. Novotný, M. Padula,  $L^p$ -approach to steady flows of viscous compressible fluids in exterior domains, *Arch. Ration. Mech. Anal.* **126** (1994), 243–297.
- [20] A. Novotný, M. Padula, Physically reasonable solutions to steady compressible Navier-Stokes equations in 3D-exterior domains ( $v_\infty \neq 0$ ), *Math. Ann.* **308** (1997), 439–489.
- [21] A. Novotný, I. Straškraba, *Introduction to the Mathematical Theory of Compressible Flow*, Oxford Univ. Press, Oxford, 2004.
- [22] M. Padula, Existence and uniqueness for viscous steady compressible motions, *Arch. Ration. Mech. Anal.* **97** (1987), 89–102.
- [23] P. I. Plotnikov, E. V. Ruban, J. Sokolowski, Inhomogeneous boundary value problems for compressible Navier-Stokes equations: well-posedness and sensitivity analysis, *SIAM J. Math. Anal.* **40** (2008), 1152–1200.
- [24] P. I. Plotnikov, E. V. Ruban, J. Sokolowski, Inhomogeneous boundary value problems for compressible Navier-Stokes and transport equations, *J. Math. Pures Appl.* **92** (2009), 113–162.
- [25] P. Plotnikov, J. Sokolowski, Domain dependence of solutions to compressible Navier-Stokes equations, *SIAM Journal on Control and Optimization*, **45** (2006), 1165–1197
- [26] P. Plotnikov, J. Sokolowski, Shape derivative of drag functional, *SIAM J. Control Optim.*, **48** no.7 (2010), 4680–4706
- [27] P. Plotnikov, J. Sokolowski, Compressible Navier-Stokes Equations. Theory and Shape Optimization, *Monografie Matematyczne*, **73**, Birkhauser, (2012)

- [28] P. Plotnikov, J. Sokolowski, A. Zochowski, Numerical experiments in drag minimization for compressible Navier-Stokes flows in bounded domains, *14th International IEEE/IFAC Conference on Methods and Models in Automation and Robotics MMAR'09* (2009), 4 pages.
- [29] S. Schmidt, V. Schulz, Shape derivatives for general objective functions and the incompressible Navier-Stokes equations, *Control and Cybernetics*, **39**, (2010), 677–713.
- [30] J. Simon, Domain variation for drag in Stokes flow, in: *Control Theory of Distributed Parameter Systems and Applications* (Shanghai, 1990), *Lecture Notes in Control and Inform. Sci.* **159**, Springer, Berlin, (1991), 28–42.
- [31] T. Slawig, A formula for the derivative with respect to domain variations in Navier-Stokes flow based on an embedding domain method, *SIAM J. Control Optim.* **42** (2003), 495–512.
- [32] T. Slawig, An explicit formula for the derivative of a class of cost functionals with respect to domain variations in Stokes flow, *SIAM J. Control Optim.* **39** (2000), 141–158.

Received xxxx 20xx; revised xxxx 20xx.

*E-mail address:* akaz@math.uni.lodz.pl

*E-mail address:* jan.sokolowski@univ-lorraine.fr

*E-mail address:* zochowsk@ibspan.waw.pl



