Quasi-static poroelastic equations as a symmetric positive system and its numerical approximation

by

Mohammad H Akanda

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2015

Keywords: Poroelasticity, Friedrich's System, Least square finite element method

Approved by

A. J. Meir, Chair, Professor of Mathematics and Statistics
Yanzhao Cao, Co-chair, Professor of Mathematics and Statistics
Paul Schmidt, Professor of Mathematics and Statistics
Dmitry Glotov, Associate Professor of Mathematics and Statistics

Abstract

This dissertation is concerned with the equations of linear poroelasticity and numerical simulation in the framework of symmetric positive systems. Physical systems arising in geomechanics, hydrology, soil mechanics, reservoir engineering, biomedical engineering etc. are modeled with linear poroelasticity equations. The purpose of this dissertation is to present well-posedness results and numerical analysis techniques using the framework of symmetric positive systems for the variants of poroelasticity equations.

Symmetric positive system, commonly known as Friedrich's system is a system of first order partial differential equations (PDEs) with symmetry and positivity properties. A PDE, that can be written in this framework is well-posed and such PDE can be numerically solved easily. We will exploit these properties of Friedrich's systems in our model problem of poroelasticity.

We consider a quasi-static poroelasticity model with two sets of different base variables. First we consider fluid content (η), rotation variables (w_{ij}) and pressure gradients (p_{x_i}). With those variables, the original PDE (or its arbitrary purturbation) is written in a symmetric positive form and subsequently a least square finite element analysis, followed by numerical simulation results is presented. In the second case, we choose stress components (σ_{ij}), displacement variables (u_i), pressure (p), pressure gradients (p_{x_i}). A scaling technique is used after semi-descretization in order to ensure the sufficient condition for positivity. We have successfully applied this technique for a wide varieties of rocks. Finally, a least square finite element method has been employed to find its numerical solutions.

Acknowledgments

I would like to express my gratitude and appreciation to my advisers Dr. Meir and Dr. Cao for their research guidance, support and patience. I tremendously benefited from each of them in different areas of my research. Dr. Meir introduced me to an interesting topic of research and provided important insight throughout. I am especially thankful for his gracious personality and financial support. Besides scholarly advice and continuous encouragement, I also appreciate Dr. Cao for giving seminars to understand different mathematical techniques, much needed for my research.

I would also like to thank Dr. Paul Schimdt and Dr. Glotov for serving on my committee and providing important comments and suggestions. Dr. Paul Schimdt has been a particularly inspirational teacher to me throughout my years at Auburn University. I would like to express my sincere gratitude to Dr. Frank Ulhig who has always been a great source of encouragement, self-esteem and enthusiasm. I am particularly thankful to Dr. Wlodzimierz Kuperberg who helped me to redefine my definition of mathematics, from engineering perspective to real mathematical perspective. Lastly, I am especially grateful to my parents, wife and kids for their unconditional love and support.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Poroelasticity, first coined by J. Geertsma in 1966, describes the physical phenomenon where fluid flows into a deformable porous medium under the assumption of relatively small deformation. The deformable porous medium is generally described as solid. When fluid flows into solid structure, the external load due to the fluid flow causes the deformation of solid matrix, which in turn affects the fluid or pore pressure. Two inherent couplings can be easily identified, solid-to-fluid coupling which refers to change of pore pressure due to change in applied stress and fluid-to-solid coupling which refers to solid deformation due to change in pore pressure. Modeling a poroelastic system requires recognition of these couplings. The mathematical description which accounts for these couplings between solid and fluid is simply a set of linear constitutive equations. More precisely, Darcy's law relates fluid velocity and pressure in a solid matrix, and another law relate fluid-to-solid interactions by introducing pressure term in stress field. The earliest work incorporating the couplings between solid and fluid dates back to 1923 when Terzaghi [1] and others proposed one dimensional consolidation of clay soils. Although three dimensional generalization of consolidation was proposed by Rendulic [2] in 1936, the most comprehensive mathematical formulation of isothermal linear poroelasticity is given by Biot [3, 4] in 1935 and 1941. Since then, Biot [5, 6, 7, 8] reformulated the theory for different specialized circumstances, also Rice and Cleary [9] explained the asymtotic poroelastic behaviour of geological entity, especially for rocks and soils. Later on, Barry and Mercer [15], Coussy [16] came up with a few analytical solutions of simplified (axis-symmetric or one dimensional problems) poroelastic system. These analytical solutions are of little importance because the corresponding modeling is too simple to represent any real life problem. With the continuous advent of enormous computational power, numerical solution of poroelastic equations is now possible. Numerical solution with higher accuracy for large engineering system can be achieved through using the tremendous computational power available. Hence, developing numerical algorithm for solving poroelastic equations has great prospect in future.

This chapter starts with the motivation behind our work where a few applications of poroelasticity in various discipline will be described. It effectively gives an idea that the application fields of poroelasticity keeps on increasing day by day by adding new areas, which were completely unknown before. Later on, mathematical modeling of poroelasticity will be derived from the very basic principles, such as mass, and momentum conservation and also necessary assumptions will be stated in order to get it quasi static form. Plan of this dissertation will be listed at the end.

1.1 Motivation

The mathematical theory of poroelasticity concerns the mechanics of porous elastic solids with fluid-filled pores. It was first applied to solve several geological problems such as consolidation of saturated soil under a uniform load, dynamic wave propagation problems in geomechanics etc. Since then the theory has grown to cover many and varied applications in many disciplines such as reservoir engineering, earthquake engineering, environmental engineering, biomedical engineering etc. We have broadly classified the application of poroelasticity as geophysical and biological, as described below.

1.1.1 Geophysical applications

The theories of poroelasticity are essential in many geophysical applications, where pore-filling materials are of interest such as the seepage of liquid waste disposed of underground, borehole damage, soil consolidation and glaciers dynamics, gas-hydrate detection, oil and gas exploration, seismic monitoring of CO_2 storage, wave propagation in the earth, hydrogeology, etc. It is worthwhile to mention a couple of historical examples in this regard. F.H.

King (1892) reported that water level in a well near the train station went up as trains approached and went down as train left the station. The following Figure 1.1 [17] shows the fluctuations. Other historical examples are vertical subsidence (due to oil or gas exploration or ground water removal) and earthquakes [17]. In 1926, massive vertical subsidence occurred due to 100 million barrels of fluid and sand being extracted from the Goose Creek oil field near Galveston, Texas. Subsidence related to exploitation of oil and gas fields also include Wilmington, California; Lake Maracaibo, Venezuela; Niigata, Japan; and the Po Delta in Italy. The areas of major subsidence related to ground-water withdrawal include areas in Japan; Mexico City, Mexico; Texas, Arizona, Nevada, and California [18]. Some examples [19] which show excessive ground water withdrawal caused huge environmental problems such as sinking locality. On the other hand, in 1935, a number of small earthquakes occurred beneath Lake Mead in Colorado due to the newly constructed Hoover Dam along the Colorado river. This dam and flowing water were stressing faults to the failure point of the lake and as a result, the earthquake occurred. The consideration of subsidence in oil industry is of great importance. Reservoir engineers often recommend drilling at a certain location to maximize the oil/gas recovery and of course, drilling induced subsidence phenomena must be considered beforehand to construct drilling platforms and other facilities. Another problem in reservoir engineering is the borehole fracture due to subsurface

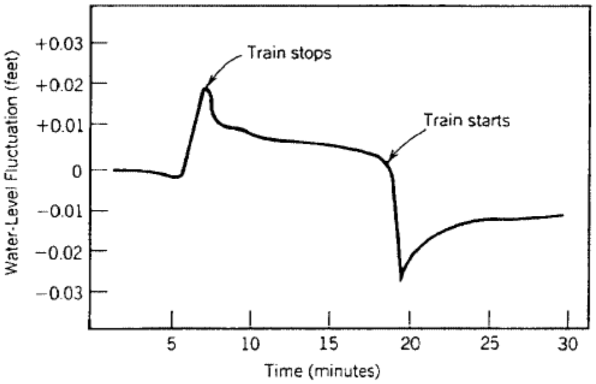


Figure 1.1: Water level fluctuations due to passing train [17]

shifting. The poroelastic model can also be applied to different waste disposal and seepage flow control problem, which have important application in environmental engineering. In earthquake engineering, earthquake liquefaction describes a phenomenon whereby a saturated or partially saturated soil substantially loses strength and stiffness in response to a earthquake shaking, causing it to behave like a liquid. Some of the earthquake liquefaction occurs in Alaska, USA, 1964, Niigata, Japan, 1964, Loma Prieta, USA, 1989, Kobe, Japan, 1995 (see Figure 1.2). Poroelastic model can be used for prediction and prevention of this type of catastrophe.

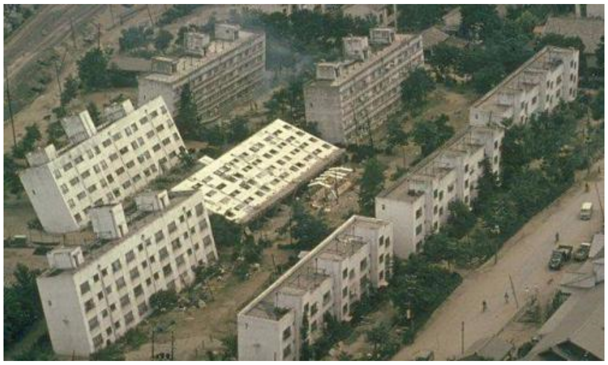


Figure 1.2: Effects of earthquake liquefaction in Niigata, Japan, 1964

1.1.2 Biological applications

Poroelastic models of bone were first reported around 45 years ago [10, 11, 12, 13, 14]. A survey of the application of poroelasticity in bone mechanics has been given by Cowin (1999). In this paper, there is a detailed review of the literature related to the application of poroelasticity to bone saturated fluid. It also describes the specific physical and modeling considerations that establish poroelasticity as an effective and useful model for deformation-driven bone fluid movement in bone tissue. Several models [20, 21, 22] have been proposed to investigate the biomechanics of soft biological tissue based on Biot's poroelastic model. Later on, the poroelastic model of soft biological tissue has been refined by considering the effect of the micro-mechanics of cells on the macro-mechanics of tissue in [23]. A poroelastic model for interstitial pressure in tumors was proposed in [24]. In 2003, Roose et al. [25] proposed a linear poroelasticity model to estimate tumor-induced stress in confined environments such as the brain. The mathematical model can essentially be used to estimate tumor growth in the brain and finally conclude tumor cell size could be a direct indicator of solid stress level inside the tumors and hence provided assistance in a clinical diagnostic setting. In 2009, Li et al. [26] proposed three dimensional poroelastic model of brain edema, a consequences of serious head injury due to the enhancement of water content and thus the increased brain volume. A detailed discussion of a poroelastic model of the cerebrospinal fluid (a water-like liquid inside the brain) system in the human brain can be found in [27].

Depending on the characteristic Stokes length $L_s = \sqrt{\nu\tau}$, where ν is the kinematic viscosity of the interstitial liquid and τ is the time-scale of the motion, we can classify the applications of poroelasticity, as found in the following table [28].

	Geometry	
	Infinite medium	Finite medium
$L_s \sim$ pore size	High-frequency acoustic wave propagation in saturated rock	Sound absorption High-frequency vibrating gels/ biological tissues
$L_s \geq$ pore size	Low-frequency acoustic wave propagation in saturated rock Consolidation and settling phenomena	Low-frequency vibrating gels/ biological tissues Bone mechanics Cartilage deformation Dynamics of poroelastic filaments

Table 1.1: Applications of poroelasticity

1.2 Mathematical modeling

The equations of linear poroelasticity are in fact momentum and mass conservation equations at macroscopic level [4, 29].

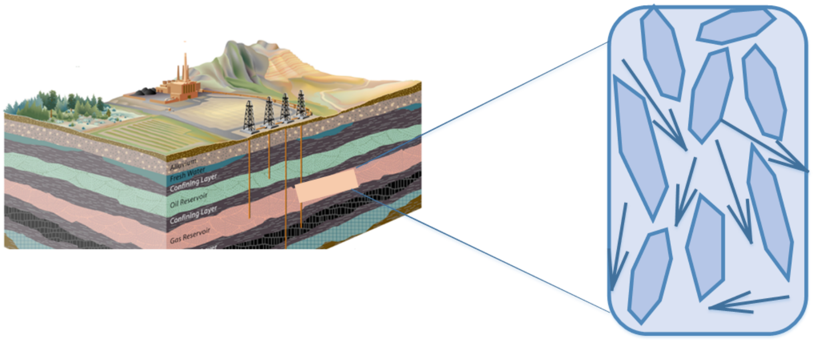


Figure 1.3: Macroscopic scales in linear poroelasticity

Here, We consider its 3-dimensional formulation, where the axes are $x_1(1)$, $x_2(2)$, and $x_3(3)$. We also consider a general situation where poroelastic system is not in static equilibrium, such as phenomenon during seismic wave propagation. The governing equation is the equation of motion, which is found by applying the law of conservation of linear momentum (second law of Newton). To formulate the law of motion, let $B \subset \Omega$ be an arbitrary finite open set with the boundary ∂B. Different forces acting on B are

$$x_1 - \text{component of inertia} = \iiint\limits_B \rho \frac{\partial^2 u_1}{\partial t^2}\, dV,$$

$$x_1 - \text{component of force due to surface tractions} = \iint\limits_{\partial B} (\sigma_{11} n_1 + \sigma_{21} n_2 + \sigma_{31} n_3)\, dA,$$

$$x_1 - \text{component of body force} = \iiint\limits_B f_1\, dV,$$

where ρ is the density, u_1 is x_1-component of the displacement vector, σ's are stress components and f_1 is the x_1-component of the body force per unit volume. Force balance gives

$$\iiint\limits_B \rho \frac{\partial^2 u_1}{\partial t^2}\, dV = \iint\limits_{\partial B} (\sigma_{11} n_1 + \sigma_{21} n_2 + \sigma_{31} n_3)\, dA + \iiint\limits_B f_1\, dV.$$

Invoking the divergence theorem,

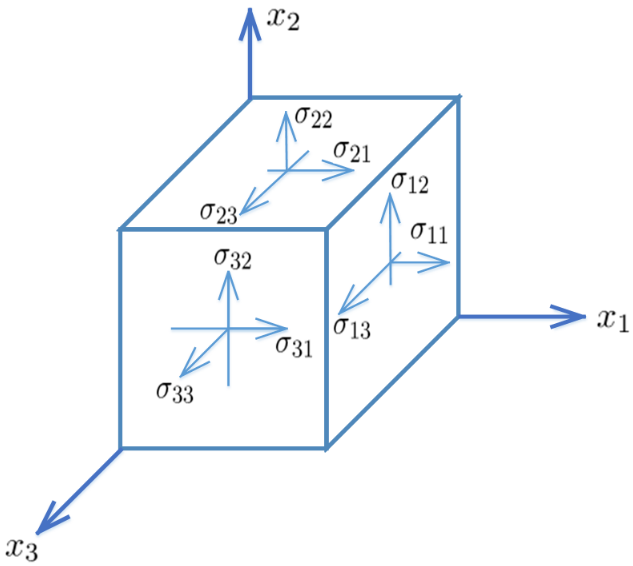


Figure 1.4: Stress tensor on a volume element

$$\iiint\limits_B \left[\rho\frac{\partial^2 u_1}{\partial t^2} - \left(\frac{\partial\sigma_{11}}{\partial x_1} + \frac{\partial\sigma_{21}}{\partial x_2} + \frac{\partial\sigma_{31}}{\partial x_3}\right)\right] dV = \iiint\limits_B f_1\, dV.$$

As the equation is valid for every $B \subset \Omega$, it follows that

$$\rho\frac{\partial^2 u_1}{\partial t^2} - \left(\frac{\partial\sigma_{11}}{\partial x_1} + \frac{\partial\sigma_{21}}{\partial x_2} + \frac{\partial\sigma_{31}}{\partial x_3}\right) = f_1.$$

Similar application of conservation of momentum in x_2 and x_2 direction gives

$$\rho\frac{\partial^2 u_2}{\partial t^2} - \left(\frac{\partial\sigma_{12}}{\partial x_1} + \frac{\partial\sigma_{22}}{\partial x_2} + \frac{\partial\sigma_{32}}{\partial x_3}\right) = f_2,$$

$$\rho\frac{\partial^2 u_3}{\partial t^2} - \left(\frac{\partial\sigma_{13}}{\partial x_1} + \frac{\partial\sigma_{23}}{\partial x_2} + \frac{\partial\sigma_{33}}{\partial x_3}\right) = f_3.$$

Using Einstein's summation convention, we can write

$$\rho\frac{\partial^2 u_j}{\partial t^2} - \frac{\partial\sigma_{ij}}{\partial x_i} = f_j,\ 1 \le j \le 3. \tag{1.1}$$

In vector-tensor notation,

$$\rho \frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla \cdot \sigma(\mathbf{x},t) = f(\mathbf{x},t), \tag{1.2}$$

where u is the displacement vector, σ is stress tensor and f is a volume-distributed external force. In the poroelastic system, the total stress σ must account for effective stress σ_e due to the deformation of the solid matrix according to the Hooke's law and the stress σ_p due to the fluid pressure inside the porous body, described as follows

$$\sigma = \sigma_e + \sigma_p. \tag{1.3}$$

Effective stress σ_e is given by Hooke's law

$$\sigma_e = 2\mu\epsilon + \lambda tr(\epsilon) I, \tag{1.4}$$

where $\epsilon = \frac{1}{2}\left(\nabla u + \nabla u^T\right)$, $tr(\epsilon)$ = Trace of $\epsilon = \epsilon_{ii}$, I is the identity matrix, and λ (Dilatation modulus), μ (Shear modulus) are the Lamé constants. The Lamé constants can be found [17] from the properties of solid matrix, Young's modulus or modulus of elasticity E and Poisson's ratio ν are

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \; \mu = \frac{E}{2+2\nu}.$$

Young's modulus E measures the force (per unit area) that is needed to stretch (or compress) a material sample. On the other hand, Poisson's ratio ν is the ratio of transverse contraction strain to longitudinal extension strain in the direction of stretching force. When a material is compressed in one direction, it usually induces expansion in the other two directions perpendicular to the direction of compression. This effect is known as Poisson effect, which is measured by Poisson's ratio ν.

In order to find σ_p, we recall the stress-strain relationship for poroelastic system as in [17]

$$\begin{aligned}
\epsilon_{11} &= \frac{1}{2\mu}\left[\sigma_{11} - \frac{\nu}{1+\nu}\sigma_{kk}\right] + \frac{\alpha}{3K}p, \\
\epsilon_{22} &= \frac{1}{2\mu}\left[\sigma_{22} - \frac{\nu}{1+\nu}\sigma_{kk}\right] + \frac{\alpha}{3K}p, \\
\epsilon_{33} &= \frac{1}{2\mu}\left[\sigma_{33} - \frac{\nu}{1+\nu}\sigma_{kk}\right] + \frac{\alpha}{3K}p, \\
\epsilon_{12} &= \frac{1}{2\mu}\sigma_{12}, \\
\epsilon_{23} &= \frac{1}{2\mu}\sigma_{23}, \\
\epsilon_{31} &= \frac{1}{2\mu}\sigma_{31}.
\end{aligned}$$

For the last three equations, there is no term containing pore pressure as the change in pore pressure does not induces shear strain. Here, $\alpha = \frac{K}{H}$ is the Biot-Willis constant where

$$\text{Compressibility of material under drained condition } \frac{1}{K} = \left.\frac{\delta\epsilon}{\delta\sigma}\right|_{p=0},$$

$$\text{Poroelastic expansion coefficient } \frac{1}{H} = \left.\frac{\delta\epsilon}{\delta p}\right|_{\sigma=0},$$

where $\epsilon = \frac{\delta V}{V}$ is the volumetric strain and σ is the isotropic applied stress field. α has the following bound

$$0 < \alpha \leq 1.$$

$\alpha \approx 1$ corresponds to an incompressible solid matrix. The system can be written in index notion

$$\underbrace{\epsilon_{ij}}_{\text{Total Strain}} = \underbrace{\frac{1}{2\mu}\left[\sigma_{ij} - \frac{\nu}{1+\nu}\sigma_{kk}\delta_{ij}\right]}_{\text{Poroelastic Strain}} + \underbrace{\frac{\alpha}{3K}p\delta_{ij}}_{\text{Free Strain}}, \tag{1.5}$$

where δ_{ij} is the Kronecker delta as defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

In (1.5), we have seen that pore pressure has an effect on the total strain. In absence of pore pressure, this equation is just a stress-strain relationship for nonporous material. Using (1.5), we can easily find the stress

$$\sigma_{12} = 2\mu\epsilon_{12},$$

$$\sigma_{23} = 2\mu\epsilon_{23},$$

$$\sigma_{31} = 2\mu\epsilon_{31}.$$

In order to find other stress

$$\begin{bmatrix} 1 & -\nu & -\nu \\ -\nu & 1 & -\nu \\ -\nu & -\nu & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \end{bmatrix} = \begin{bmatrix} 2\mu(1+\nu)(\epsilon_{11} - \frac{\alpha}{3K}p) \\ 2\mu(1+\nu)(\epsilon_{22} - \frac{\alpha}{3K}p) \\ 2\mu(1+\nu)(\epsilon_{33} - \frac{\alpha}{3K}p) \end{bmatrix}. \tag{1.6}$$

Inverting the matrix,

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \end{bmatrix} = \frac{1}{(2\nu-1)(1+\nu)} \begin{bmatrix} \nu-1 & -\nu & -\nu \\ -\nu & \nu-1 & -\nu \\ -\nu & -\nu & \nu-1 \end{bmatrix} \begin{bmatrix} (\epsilon_{11} - \frac{\alpha}{3K}p) \\ (\epsilon_{22} - \frac{\alpha}{3K}p) \\ (\epsilon_{33} - \frac{\alpha}{3K}p) \end{bmatrix} 2\mu(1+\nu). \tag{1.7}$$

For σ_{11}

$$\sigma_{11} = \frac{2\mu}{2\nu-1}\left[(\nu-1)(\epsilon_{11}-\frac{\alpha}{3K}p)-\nu(\epsilon_{22}-\frac{\alpha}{3K}p)-\nu(\epsilon_{33}-\frac{\alpha}{3K}p)\right],$$
$$\Rightarrow \sigma_{11} = \frac{2\mu}{1-2\nu}\left[(1-\nu)(\epsilon_{11}-\frac{\alpha}{3K}p)+\nu(\epsilon_{22}-\frac{\alpha}{3K}p)+\nu(\epsilon_{33}-\frac{\alpha}{3K}p)\right],$$
$$\Rightarrow \sigma_{11} = \frac{2\mu}{1-2\nu}\left[\epsilon_{11}-\frac{\alpha}{3K}p\nu-\nu\epsilon_{11}+\frac{\alpha}{3K}p\nu+\nu\epsilon_{22}-\frac{\alpha}{3K}p\nu+\nu\epsilon_{33}-\frac{\alpha}{3K}p\nu\right],$$
$$\Rightarrow \sigma_{11} = \frac{2\mu}{1-2\nu}\left[\epsilon_{11}-2\nu\epsilon_{11}+\nu(\epsilon_{11}+\epsilon_{22}+\epsilon_{33})-\frac{\alpha}{3K}p-\frac{\alpha}{3K}p\nu\right],$$
$$\Rightarrow \sigma_{11} = \frac{2\mu}{1-2\nu}\left[\epsilon_{11}(1-2\nu)+\nu\epsilon_{kk}-\frac{\alpha}{3K}p(1+\nu)\right],$$
$$\Rightarrow \sigma_{11} = 2\mu\epsilon_{11}+\frac{2\mu\nu}{1-2\nu}\epsilon_{kk}-\frac{\alpha p}{3K}\cdot\frac{2\mu(1+\nu)}{(1-2\nu)},$$
$$\Rightarrow \sigma_{11} = 2\mu\epsilon_{11}+\lambda\epsilon_{kk}-\alpha p.$$

The last equation is found using $\lambda = \frac{2\mu\nu}{1-2\nu}$ and $\mu = \frac{3K(1-2\nu)}{2(1+\nu)}$. Similarly, we have

$$\sigma_{22} = 2\mu\epsilon_{22}+\lambda\epsilon_{kk}-\alpha p,$$
$$\sigma_{33} = 2\mu\epsilon_{33}+\lambda\epsilon_{kk}-\alpha p.$$

The stress field can be written as

$$\sigma = 2\mu\epsilon+\lambda tr(\epsilon)I-\alpha pI. \tag{1.8}$$

Comparing with Equation (1.3), we have

$$\sigma_p = -\alpha pI\cdot \tag{1.9}$$

Using the σ of (1.8) in (1.1), we have

$$\rho\frac{\partial^2 u_1}{\partial t^2} - (\lambda+\mu)\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}\right) - \mu\left(\frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_1}{\partial x_2^2} + \frac{\partial^2 u_1}{\partial x_3^2}\right) + \alpha\frac{\partial p}{\partial x_1} = f_1,$$
$$\rho\frac{\partial^2 u_2}{\partial t^2} - (\lambda+\mu)\frac{\partial}{\partial x_2}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}\right) - \mu\left(\frac{\partial^2 u_2}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_2^2} + \frac{\partial^2 u_2}{\partial x_3^2}\right) + \alpha\frac{\partial p}{\partial x_2} = f_2,$$
$$\rho\frac{\partial^2 u_3}{\partial t^2} - (\lambda+\mu)\frac{\partial}{\partial x_3}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}\right) - \mu\left(\frac{\partial^2 u_3}{\partial x_1^2} + \frac{\partial^2 u_3}{\partial x_2^2} + \frac{\partial^2 u_3}{\partial x_3^2}\right) + \alpha\frac{\partial p}{\partial x_3} = f_3.$$

In vector-tensor notation

$$\rho\frac{\partial^2 u}{\partial t^2} - (\lambda+\mu)\nabla(\nabla\cdot u) - \mu\Delta u + \alpha\nabla p = f. \tag{1.10}$$

For the mass conservation equation, corresponding variables are increment of fluid content (or just fluid content) η, the fluid flux q, and external volumetric fluid source h. To formulate the law of mass conservation, let $B \subset \Omega$ be an arbitrary finite open set with the boundary ∂B. Mass balance on the control volume B

$$\text{Rate of change of fluid volume} = \frac{\partial}{\partial t}\iiint\limits_B \eta\, dV,$$

$$\text{Fluid input rate} = \iiint\limits_B h\, dV,$$

$$\text{Fluid output rate} = \iint\limits_{\partial B} q.n\, dA.$$

Mass conservation leads to

$$\frac{\partial}{\partial t}\iiint\limits_B \eta\, dV = \iiint\limits_B h\, dV - \iint\limits_{\partial B} q.n\, dA,$$

$$\Rightarrow \iiint\limits_B \frac{\partial \eta}{\partial t}\, dV = \iiint\limits_B h\, dV - \iint\limits_{\partial B} (q_1 n_1 + q_2 n_2 + q_3 n_3)\, dA.$$

Invoking the divergence theorem,

$$\iiint\limits_B \frac{\partial \eta}{\partial t}\, dV + \iiint\limits_B \left(\frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3} \right) dV = \iiint\limits_B h\, dV,$$

$$\Rightarrow \iiint\limits_B \left[\frac{\partial \eta}{\partial t} + \left(\frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3} \right) \right] dV = \iiint\limits_B h\, dV,$$

$$\Rightarrow \iiint\limits_B \left[\frac{\partial \eta}{\partial t} + \nabla \cdot q \right] dV = \iiint\limits_B h\, dV.$$

As the equation is valid for every $B \subset \Omega$, it follows

$$\frac{\partial \eta(\mathbf{x}, t)}{\partial t} + \nabla \cdot q(\mathbf{x}, t) = h(\mathbf{x}, t). \tag{1.11}$$

Before deriving the expression for fluid content η, the following definitions of parameters are needed. The unconstrained specific storage coefficient S_σ is the change of fluid volume in storage per unit control volume per unit change in pressure at constant stress

$$S_\sigma = \left. \frac{\delta \eta}{\delta p} \right|_{\sigma = 0} = \frac{1}{R}. \tag{1.12}$$

Another specific storage coefficient, S_ϵ is defined as the volume of fluid released from the storage per unit control volume per unit pressure decline holding the control volume constant

$$S_\epsilon = \left. \frac{\delta \eta}{\delta p} \right|_{\epsilon = 0} = \frac{1}{M}. \tag{1.13}$$

Skempton's coefficient B is defined as the ratio of the induced pore pressure to the change in applied stress while no fluid is allowed to move in or out of the control volume.

$$B = - \left. \frac{\delta p}{\delta \sigma} \right|_{\eta = 0} \text{with} \qquad 0 \le B \le 1. \tag{1.14}$$

This coefficient indicates how the applied stress is distributed over the solid matrix and the fluid. $B \approx 0$ for pores filled with gas because the load is supported by the solid matrix. On the other hand, $B \approx 1$ for saturated soil because the load is supported by the fluid. The interrelationships between these coefficients are as follows

$$S_\sigma = \frac{1}{BH} = \frac{\alpha}{KB}, \tag{1.15}$$

$$S_\epsilon = S_\sigma - \frac{K}{H^2} = S_\sigma - \frac{\alpha^2}{K} = S_\sigma \left(1 - \alpha B\right). \tag{1.16}$$

The fluid content is given by [29]

$$\eta = S_\epsilon p + \alpha \nabla \cdot u = S_\epsilon p + \alpha \epsilon_{kk}. \tag{1.17}$$

Here, $S_\epsilon p$ accounts for the fluid content that can be injected into the fixed volume storage by pressure and $\alpha \nabla \cdot u$ accounts for the fluid that can be squeezed out. Using (1.8)

$$\begin{aligned} \sigma_{11} + \sigma_{22} + \sigma_{33} = 2\mu \left(\epsilon_{11} + \epsilon_{22} + \epsilon_{33}\right) + 3\lambda \epsilon_{kk} - 3\alpha p, \\ \Rightarrow \sigma_{kk} = \left(2\mu + 3\lambda\right) \epsilon_{kk} - 3\alpha p, \\ \Rightarrow \epsilon_{kk} = \frac{\sigma_{kk} + 3\alpha p}{2\mu + 3\lambda}, \\ \Rightarrow \nabla \cdot u = \epsilon_{kk} = \frac{\sigma_{kk}}{3K} + \frac{\alpha}{K} p, \text{ using } K = \lambda + \frac{2}{3}\mu. \end{aligned}$$

Using this expression

$$\begin{aligned} \eta = S_\epsilon p + \alpha \nabla \cdot u, \\ \Rightarrow \eta = S_\epsilon p + \alpha \left(\frac{\sigma_{kk}}{3K} + \frac{\alpha}{K} p\right), \\ \Rightarrow \eta = \left(S_\epsilon + \frac{\alpha^2}{K}\right) p + \frac{\alpha}{3K} \sigma_{kk}. \end{aligned}$$

So,

$$\eta = S_\sigma p + \frac{\alpha}{3K}\sigma_{kk}, \tag{1.18}$$

$$\eta = \frac{\alpha}{KB}p + \frac{\alpha}{3K}\sigma_{kk}. \tag{1.19}$$

On the other hand, the flux q is given by the Darcy's Law for the diffusive flow through the porous medium, as follows

$$q = -\frac{k_s}{\mu_f}\nabla p, \tag{1.20}$$

where μ_f is the fluid viscosity and k_s is the permeability of solid matrix. Defining hydraulic diffusivity or mobility $k = \frac{k_s}{\mu_f}$, we have Darcy's equation

$$q = -k\nabla p. \tag{1.21}$$

Using (1.17) ,(1.19), and (1.21), the mass conservation equation (1.11) becomes

$$S_\epsilon \frac{\partial p}{\partial t} + \alpha \frac{\partial\left(\nabla \cdot u\right)}{\partial t} - \nabla \cdot (k\nabla p) = h, \tag{1.22}$$

$$\frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{3K}\frac{\partial \sigma_{kk}}{\partial t} - \nabla \cdot (k\nabla p) = h. \tag{1.23}$$

If the hydraulic conductivity k is not a function of the spatial variables (such as in homogeneous and isotropic medium, the permeability and viscosity are constant), then we have

$$S_\epsilon \frac{\partial p}{\partial t} + \alpha \frac{\partial\left(\nabla \cdot u\right)}{\partial t} - k\nabla^2 p = h, \tag{1.24}$$

$$\frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{3K}\frac{\partial \sigma_{kk}}{\partial t} - k\nabla^2 p = h. \tag{1.25}$$

Parameters	SI Unit	Description
λ,μ	N/m^2	Positive Lamé constants
S_ϵ or M^{-1}	m^2/N	Constrained specific storage coefficient
S_σ or R^{-1}	m^2/N	Unconstrained specific storage coefficient
α	$-$	Biot-Willis constant
μ_f	$(N \cdot s)/m^2$	Viscosity of fluid
k_s	m^2	Permeability of solid matrix
k	$m^4/(N \cdot s)$	Hydraulic conductivity
B	$-$	Skempton's coefficient
K^{-1}	m^2/N	Drained compressibility coefficient
H^{-1}	m^2/N	Expansion coefficient at constant stress
ν	$-$	Poisson constant
E	m^2/N	Young's modulus
ϵ	$-$	Volumetric strain

Table 1.2: Different Physical Parameters

The system of partial differential equations for poroelastic model,

In $\sigma - u - p$ formulation

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha pI = 0, \\ \rho\frac{\partial^2 u}{\partial t^2} - \nabla \cdot \sigma = f, \\ S_\epsilon\frac{\partial p}{\partial t} + \alpha\frac{\partial(\nabla \cdot u)}{\partial t} - k\nabla^2 p = h, \\ \text{or } \frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{3K}\frac{\partial \sigma_{kk}}{\partial t} - k\nabla^2 p = h. \end{cases} \tag{1.26}$$

In $u - p$ formulation

$$\begin{cases} \rho\frac{\partial^2 u}{\partial t^2} - (\lambda + \mu)\nabla(\nabla \cdot u) - \mu\Delta u + \alpha\nabla p = f, \\ S_\epsilon\frac{\partial p}{\partial t} + \alpha\frac{\partial(\nabla \cdot u)}{\partial t} - k\nabla^2 p = h, \\ \text{or } \frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{3K}\frac{\partial \sigma_{kk}}{\partial t} - k\nabla^2 p = h. \end{cases} \tag{1.27}$$

We assume the deformation of solid matrix is much slower than the fluid flow rate. With this quasi-static assumption, the term $\rho\frac{\partial^2 u}{\partial t^2}$ in (1.26) and (1.27) can be ignored. So, we restrict our attention to linear quasi-static flow in a deformable porous medium. The PDEs for quasi-static poroelasticity are listed below.

In the $\sigma - u - p$ formulation

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha pI = 0, \\ -\nabla \cdot \sigma = f, \\ S_\epsilon\frac{\partial p}{\partial t} + \alpha\frac{\partial(\nabla \cdot u)}{\partial t} - k\nabla^2 p = h, \\ \text{or } \frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{3K}\frac{\partial \sigma_{kk}}{\partial t} - k\nabla^2 p = h. \end{cases} \tag{1.28}$$

In the $u - p$ formulation

$$\begin{cases} -\left(\lambda + \mu\right) \nabla \left(\nabla \cdot u\right) - \mu \Delta u + \alpha \nabla p = f, \\ S_\epsilon \frac{\partial p}{\partial t} + \alpha \frac{\partial (\nabla \cdot u)}{\partial t} - k \nabla^2 p = h, \\ \text{or } \frac{\alpha}{KB} \frac{\partial p}{\partial t} + \frac{\alpha}{3K} \frac{\partial \sigma_{kk}}{\partial t} - k \nabla^2 p = h. \end{cases} \tag{1.29}$$

In order to have well-possedness results, we must supplement the equations by suitable boundary conditions. The following boundary conditions will be enforced along with the initial condition.

1. Dirichlet boundary condition:

$$u = u_1, \; p = p_1 \text{ on } \Omega.$$

2. Neumann boundary condition:

$$\sigma \cdot n = s, \; p = p_1 \text{ on } \Omega.$$

3. Mixed boundary condition:Let us partition the boundary as $\partial \Omega = \bar{\Gamma}_c \cup \bar{\Gamma}_t$ with $\Gamma_c \cap \Gamma_t = \emptyset$ where Γ_c and Γ_t are regular open sets of Ω. Γ_c and Γ_t are called clamped (Dirichlet form of boundary condition) and traction (Neumann form of boundary condition) boundary. We impose

$$u = u_1 \text{ on } \Gamma_c,$$
$$\sigma \cdot n = s \text{ on } \Gamma_t,$$
$$p = p_1 \text{ on } \Omega.$$

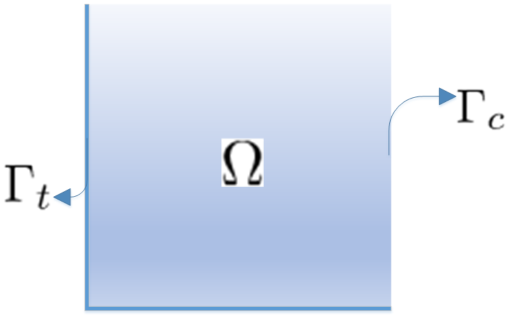


Figure 1.5: Partition of boundary, $\partial\Omega = \bar{\Gamma}_c \cup \bar{\Gamma}_t$

1.3 Notations and assumptions

We consider the system (1.26)-(1.28) on an open bounded subset Ω of $\mathbb{R}^d$ $(d = 2$ or $3)$. To understand the corresponding functional setting, the following Sobolev spaces are defined [42, 43, 44].

Let α be an multi-index

$$\alpha = (\alpha_1,\, \alpha_2,\, \alpha_3,\, \cdots,\, \alpha_d) \in \mathbb{N}^d.$$

Define

$$|\alpha| = \sum_{i=1}^{d} \alpha_i \text{ and } x^\alpha = (x_1,\, x_2, \cdots,\, x_d)^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d},$$

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}.$$

For k be a non-negative integer and $p \in [1,\, \infty]$, the Sobolev space

$$W^{k,p} = \{f \in L^1_{loc} |\ D^\alpha \in L^p(\Omega)\}.$$

The space $W^{k,p}$ is a Banach space with the norm,

for $1 \leq p < 1$

$$\|v\|_{k,\,p,\,\Omega} = \left(\sum_{|\alpha| \leq k} \|D^\alpha v\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}},$$

for $p = \infty$

$$\|v\|_{k,\,\infty,\,\Omega} = \max_{|\alpha| \leq k} \|D^\alpha v\|^p_{L^\infty(\Omega)}.$$

The subscript Ω is often omitted in the norm when there is no confusion. The case $p = 2$ is of great importance as it leads to a Hilbert space. In such case, $W^{k,\,2} = H^k$ is a Banach space with $\|\cdot\|_k = \|\cdot\|_{k,\,2,\,\Omega}$. Moreover, it is a Hilbert space with following inner product

$$(u,\, v)_k = \sum_{|\alpha| \leq k} \int_\Omega D^\alpha u\, D^\alpha v\; d\Omega.$$

For $k = 0$, $W^{0,\,2} = H^0 = L^2$. We also denote $C_0^\infty(\Omega)$ the space of infinitely continuously differentiable function with compact support in Ω. The space $W_0^{k,p}(\Omega)$ is the closure of $C_0^\infty(\Omega)$ in $W^{k,p}(\Omega)$. Similarly, for $p = 2$, $W_0^{k,\,2} = H_0^k$.

We can also define Sobolev space for non-negative non-integer order. For $s = k + \sigma$ with $k \geq 0$ an integer and $\sigma \in (0,\, 1)$, $1 \leq p < \infty$, $W^{s,p}$ is the collection of all function $v \in W^{k,p}(\Omega)$ such that

$$\frac{|D^\alpha v(x) - D^\alpha v(y)|}{\|x - y\|^{\sigma + \frac{d}{p}}} \in L^p(\Omega \times \Omega) \text{ for all } |\alpha| = k,$$

with the given norm (that makes it Banach space)

$$\|v\|_{s,\,p,\,\Omega} = \left(\|v\|^p_{k,\,p,\,\Omega} + \sum_{|\alpha|=k} \int_{\Omega\times\Omega} \frac{|D^\alpha v(x) - D^\alpha v(y)|^p}{\|x-y\|^{\sigma p + d}}\, dx\, dy \right)^{\frac{1}{p}}.$$

For $p = 2$, $H^s(\Omega) = W^{s,\,2}(\Omega)$ is again a Hilbert space with the following inner product

$$(u,\, v)_{s,\,\Omega} = (u,\, v)_{k,\,\Omega} + \sum_{|\alpha|=k} \int_{\Omega\times\Omega} \frac{(D^\alpha u(x) - D^\alpha u(y))\,(D^\alpha v(x) - D^\alpha v(y))}{\|x-y\|^{2\sigma + d}}\, dx\, dy.$$

A Sobolev space of negative order is in fact dual space of positive order Sobolev space. As before $W_0^{s,\,p}(\Omega)$ is the closure of the space $C_0^\infty(\Omega)$ in $W^{s,\,p}(\Omega)$. When $p = 2$, we have the Hilbert space $H_0^s(\Omega) = W_0^{s,\,2}(\Omega)$. For $s \geq 0$, $p \in [1,\, \infty)$, $q = \frac{p}{p-1}$, we define $W^{-s,\,q}(\Omega)$ to be the dual space of $W_0^{s,\,p}(\Omega)$. As before, for $p = 2$, $H^{-s}(\Omega) = W^{-s,\,2}(\Omega)$. Clearly, these spaces are Banach space with appropriate norm. All the spaces defined above can be easily extended for vector valued function as follows

$$\mathbf{H}^k(\Omega) = \left(H^k(\Omega)\right)^d,\ \mathbf{H}_0^k(\Omega) = \left(H_0^k(\Omega)\right)^d,\ \mathbf{L}^2(\Omega) = \left(L^2(\Omega)\right)^d.$$

We also define spaces involving time. In this work, we define $I = (0,\, T]$ where T is a finite real number. Let X be any Banach space, then the space $L^p(0,\, T,\, X)$ consists of all measurable function $u : [0,\, T] \to X$ with
for $1 \leq p < \infty$

$$\|u\|_{L^p(0,\,T,\,X)} = \left(\int_0^T \|u(t)\|^p\, dt \right)^{\frac{1}{p}} < \infty,$$

for $p = \infty$

$$\|u\|_{L^{\infty}(0,T,X)} = \operatorname*{ess\,sup}_{0 \le t \le T} \|u\|_X < \infty.$$

With these norm, $L^p(0,\, T,\, X)$ are Banach space. X is usually $W^{s,p}(\Omega)$ for some s and p.

1.4 Plan of dissertation

Chapter 1 is concerned with introduction to model problem and its mathematical formulation. Our model problem is poroelastic system, a few motivation for studying the poroelastic system were presented, followed by the mathematical modeling of fluid saturated porous medium. Necessary assumptions were made to get quasi-static form of poroelastic system, which will be studied in the subsequent chapters. Physical parameters used in the modeling were listed and a brief description of mathematical foundations were also discussed.

Chapter 2 deals with basic foundation of the analysis tools used in this dissertation and its numerical approximation technique i.e. symmetric positive system and least square finite element method. A review of symmetric positive system is discussed, followed by a couple of examples relevant to this dissertation. The numerical tool, LSFEM is discussed briefly at the end.

Chapter 3 commences with fluid content-rotation-pressure content formulation of quasi-static poroelastic system. For simplicity, two dimensional formulation will be presented. With necessary perturbation of the PDE system, the system is a symmetric positive system. LSFEM adapts this framework and numerical analysis with its approximate solution is presented.

In Chapter 4, a new practically useful formulation namely stress-displacement-pressure formulation is developed. Although it can be generalized for any dimension, again for simplicity, two dimensional formulation is presented. With necessary perturbation of the PDE system and later on with no perturbation, the system is a symmetric positive system. LS-FEM for system of first order PDE easily accommodates this framework and numerical analysis with its approximate solution is presented.

Chapter 5 deals with conclusions and remarks. This dissertation is concluded with a brief discussion of future work.

Chapter 2

Symmetric positive system and least square finite element method

2.1 Symmetric positive system

2.1.1 Introduction

Although solutions of elliptic, hyperbolic and parabolic partial differential equations have very different properties, it may seem very difficult to have a unified treatment of these equations. An attempt at such unified treatment has been made by Friedrich (1958) [30]. He introduced a class of boundary value problems, named symmetric positive systems, encompassing a variety of elliptic, parabolic and hyperbolic problems. Also, his unified approach provided a framework for a successful treatment of some equations of mixed type, such as the Tricomi equation, $y\frac{\partial^2\phi}{\partial x^2} - \frac{\partial^2\phi}{\partial y^2} = 0$ and transonic flow problems. The very basic idea is to write any partial differential equation as a first order systems that satisfies some algebraic properties which will be sufficient to ensure its well-posedness provided with suitable boundary conditions. This approach also allows us to enforce different boundary conditions using different boundary operators satisfying certain properties. Such boundary conditions are called admissible boundary conditions. Once a linear partial differential equation can be written as a symmetric positive system, the existence and uniqueness of the solution is immediate. However, writing a linear partial differential equation as a symmetric positive system depends on proper choice of variables and multiplier or transformations which are neither unique nor straightforward.

2.1.2 Mathematical formulation

Friedrich's systems are systems of first-order PDE's endowed with a symmetry and positivity property. Let Ω be a bounded, open, Lipschitz domain in $\mathbb{R}^d$ and let m be a positive integer that corresponds to the number of scalar-valued PDEs in the system. Let $(d+1)$ $\mathbb{R}^{m,m}$ -valued fields defined in the domain Ω, say $\mathbb{A}^0$, $\mathbb{A}^1, \ldots,$ $\mathbb{A}^d$ and set $\mathbb{X} = \sum_{k=1}^{d} \partial_k \mathbb{A}^k$. The assumption on the fields $\mathbb{A}^0$, $\mathbb{A}^1, \ldots,$ $\mathbb{A}^d$ are

- For all $k \in \{0, 1, \ldots, d\}$, $\mathbb{A}^k \in [L^\infty(\Omega)]^{m,m}$ and $\mathbb{X} \in [L^\infty(\Omega)]^{m,m}$.
- For all $k \in \{1, \ldots, d\}$, $\mathbb{A}^k = (\mathbb{A}^k)^T$ a.e. in Ω (symmetry).
- There exists $\mu_0 > 0$, $\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \mathbb{X} \geq 2\mu_0 \mathbb{I}_m$ a.e. in Ω (positivity).

Let $\mathbf{L} = [L^2(\Omega)]^m$ be equipped with its natural scalar product

$$(f,\, g)_L = \int_\Omega f^T g, \tag{2.1}$$

and the associated norm $\|.\|_{\mathbf{L}}$.

We are interested in the following differential operator

$$A : [C^1(\overline{\Omega})]^m \ni z \longmapsto Az := A_{(0)}z + A_{(1)}z \in \mathbf{L}, \tag{2.2}$$

where

$$A_{(0)}z := \mathbb{A}^0 z, A_{(1)}z := \sum_{k=1}^{d} \mathbb{A}^k \partial_k z.$$

We also consider the following differential operator which is the formal adjoint of A

$$\overline{A} : [C^1(\overline{\Omega})]^m \ni z \longmapsto \overline{A}z := \overline{A}_{(0)}z - A_{(1)}z \in \mathbf{L}, \tag{2.3}$$

where

$$\overline{A}_{(0)}z := ((\mathbb{A}^0)^t - \mathbb{X})z.$$

Note that

$$\forall \phi, \psi \in [C_0^\infty(\Omega)]^m,\ (A\phi,\, \psi)_{\mathbf{L}} = (\phi,\, \overline{A}\psi)_{\mathbf{L}}.$$

Assuming the fields $\mathbb{A}^1,\, \mathbb{A}^2, \ldots,\, \mathbb{A}^d$ are smooth enough to be defined on the boundary $\partial\Omega$. Introduce a boundary field $D : \partial\Omega \to \mathbb{R}^m$ such that, for a.e. $x \in \partial\Omega$, $D := \sum_{k=1}^{d} n_k \mathbb{A}^k$ where $n = (n_1,\, n_2, \ldots,\, n_d)$ is the outward unit normal to Ω.

Assume, there is a boundary field $M : \partial\Omega \to \mathbb{R}^m$, such that, for a.e. $x \in \partial\Omega$, the following conditions hold

$$M \geq 0,\ \forall \xi \in \mathbb{R}^m, (\xi)^t M \xi \geq 0, \tag{2.4}$$

$$\mathbb{R}^m = \mathrm{Ker}(D - M) + \mathrm{Ker}(D + M). \tag{2.5}$$

Note that D is symmetric by construction and by varying the boundary field M, we can enforce different boundary conditions. Let $f \in \mathbf{L}$ and consider the following differential equation

$$Az = f \quad \text{in} \quad \Omega, \tag{2.6}$$

$$(D - M)z = 0 \quad \text{on} \quad \partial\Omega. \tag{2.7}$$

Friedrich showed in [30] the uniqueness of the strong solution $z \in [C^1(\overline{\Omega})]^m$of the above mentioned boundary value problem. He also showed that the existence of a so-called weak solution $z \in \mathbf{L}$ such that $(z,\, \overline{A}y)_{\mathbf{L}} = (f,\, y)_{\mathbf{L}}$ for all $y \in [C^1(\overline{\Omega})]^m$ such that $(D + M^t)y = 0$ on $\partial\Omega$. The following two theorems are due to [30].

Theorem 2.1 (Uniqueness of strong solution). *Let $u \in C^1(\overline{\Omega})$, $(D - M)u = 0$ on $\partial\Omega$ and $v \in C^1(\overline{\Omega})$, $(D + M^T)v = 0$ on $\partial\Omega$, then there exist c_1 and c_2 such that $c_1\|u\| \leq \|Au\|$ and $c_2\|v\| \leq \|\overline{A}v\|$*

Theorem 2.2 (Existence of weak solution). *Let $f \in \mathbf{L}$, then there exists a weak solution of (2.6)-(2.7).*

Let us define the graph space of operator A, $W \subset \mathbf{L}$ as $W = \{u \in \mathbf{L};\, Au \in \mathbf{L}\}$. This is a Hilbert space with the following inner product

$$(u,\, v)_W = (u,\, v)_{\mathbf{L}} + (Au,\, Av)_{\mathbf{L}}\,. \tag{2.8}$$

Theorem 2.3. *Let $V = \{u \in W;\, (D - M)\, u = 0 \text{ on } \partial\Omega\}$, then the operator $A : V \to \mathbf{L}$ is an isomorphism.*

Proof. See [33] □

2.1.3 Classical applications

Soon after the concept of symmetric positive systems in 1958 as a unification tool for general PDEs, it has been successfully applied to a vast number of model PDEs. A few of them are

- Advection-reaction [32], $\mu z + \beta \cdot \nabla z = f$,
- Diffusion-advection-reaction [32, 34], $-\nabla \cdot (A \nabla u) + \beta \cdot \nabla u + \mu u = f$,
- Linear elasticity (static/dynamic) [31, 32],
- The curl-curl problem [32],
- Maxwell's equation [33],
- Darcy's equation [33],
- Wave equation [34], $u_{tt} - \gamma^2 u_{xx} = f$,
- Second order linear ODE [34], $-(p(x)u^{'}(x))^{'} + q(x)u(x) = f(x)$,
- Tricomi equation [30], $y\frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \phi}{\partial y^2} = 0$,
- Hyperbolic equation [30], $\frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial x^2} = h(x,t)$,

- Parabolic equation [30], $\frac{\partial \phi}{\partial t} - \frac{\partial^2 \phi}{\partial x^2} = h(x, t)$,
- Non-homogeneous Laplace equation [30], $\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = h(x, t)$,
- Embedding [30], $Lu \pm k\overline{u} = f$,
- Ultrahyperbolic equation [30], $\frac{\partial^2 \phi}{\partial x_1^2} + \frac{\partial^2 \phi}{\partial x_2^2} - \frac{\partial^2 \phi}{\partial x_3^2} - \frac{\partial^2 \phi}{\partial x_4^2} = 0$,
- Cauchy-riemann equation [30], $(\frac{\partial}{\partial x_1} + i\frac{\partial}{\partial x_2})u = 0$,
- Standard hyperbolic and elliptic equation of second order [30],
- Linear second order equation with variable co-efficient [34], $(\alpha(x, y)u_x)_x + (\beta(x, y)u_y)_y + \gamma(x, y) = f(x, y)$,
- The generalized heat equation [35], $\partial_t u - \operatorname{div}(A\nabla u) + b \cdot \nabla u + cu = f$.

2.1.4 Examples

Based on the relevance to our work, a couple of PDEs are chosen to show they can be formulated as symmetric positive systems. We have chosen the linear elasticity equation, the heat equation and the ultrahyperbolic equations for this purpose.

Linear elasticity equation

We consider the following linear elasticity equations

$$\begin{aligned} \sigma - \lambda(\nabla \cdot u)I_d - \mu(\nabla u + \nabla u^T) &= 0, \\ -\frac{1}{2}\nabla \cdot (\sigma + \sigma^T) + \beta u &= f. \end{aligned} \tag{2.9}$$

Remark 2.1. *Linear elasticity equation can be formulated as symmetric positive system as in [32]. But we have slightly different approach in here based on how to prove the positivity of $\mathcal{B}$.*

One dimensional case:

In one dimension, (2.9) will be as follows

$$\begin{aligned} \sigma - \lambda\frac{\partial u}{\partial x} - 2\mu\frac{\partial u}{\partial x} &= 0, \\ -\frac{\partial \sigma}{\partial x} + \beta u &= f. \end{aligned} \tag{2.10}$$

With further manipulation gives,

$$\begin{aligned} \frac{2\mu}{\lambda + 2\mu}\sigma - 2\mu\frac{\partial u}{\partial x} &= 0, \\ -2\mu\frac{\partial \sigma}{\partial x} + 2\mu\beta u &= f_1. \end{aligned} \tag{2.11}$$

In matrix form,

$$\begin{bmatrix} \frac{2\mu}{\lambda+2\mu} & -2\mu\frac{\partial}{\partial x} \\ & \\ -2\mu\frac{\partial}{\partial x} & 2\mu\beta \end{bmatrix} \begin{bmatrix} \sigma \\ \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ \\ f_1 \end{bmatrix}. \tag{2.12}$$

Theorem 2.4. *The PDE system* (2.10) *can be formulated as a symmetric positive system with some unknown variables.*

Proof. Referring to the equivalent formulation (2.12) of (2.10), the corresponding matrices are

$$\mathcal{A}_x = \begin{bmatrix} 0 & -2\mu \\ & \\ -2\mu & 0 \end{bmatrix}, \ \mathbb{A}^0 = \begin{bmatrix} \frac{2\mu}{\lambda+2\mu} & 0 \\ & \\ 0 & 2\mu\beta \end{bmatrix}, \text{ and } \mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \sum_{k=1}^{d} \partial_k \mathbb{A}^k = \begin{bmatrix} \frac{4\mu}{\lambda+2\mu} & 0 \\ & \\ 0 & 4\mu\beta \end{bmatrix}.$$

Note that $\mathcal{A}_x$ is symmetric and $\mathcal{B}$ is positive definite, so by the definition the PDE system (2.10) is symmetric positive. □

In order to generalize the symmetric positiveness of linear elasticity equations (2.9) for $d = 2$ or $d = 3$, we note

$$\sigma = \sigma^T,$$
$$\sigma_{kk} = \lambda d(\nabla \cdot u) + 2\mu(\nabla \cdot u),$$
$$\Rightarrow \nabla \cdot u = \tfrac{\sigma_{kk}}{\lambda d + 2\mu}.$$

So,

$$\begin{aligned} \sigma - a(\sigma_{kk})I_d - \mu(\nabla u + \nabla u^T) &= 0, \\ -\nabla \cdot \sigma + \beta u &= f, \end{aligned} \tag{2.13}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$. Equation (2.9) can be written

$$\begin{aligned} \sigma_{i,j} - a(\sigma_{kk})\delta_{i,j} - \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) = 0 \quad \forall i, j \in \{1, 2, \cdots, d\}, \\ -\nabla \cdot \sigma + \beta u = f. \end{aligned} \tag{2.14}$$

Eliminating the Kronecker delta

$$\begin{aligned} b\sigma_{i,i} - a\sum_{k \neq i} \sigma_{kk} - 2\mu\frac{\partial u_i}{\partial x_i} = 0 \quad \forall i \in \{1, 2, \cdots, d\}, \\ \sigma_{i,j} - \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) = 0 \quad \forall i, j \in \{1, 2, \cdots, d\} \text{ with } i \neq j, \\ -\frac{\partial \sigma_{ij}}{\partial x_i} + \beta u_j = f_j \, \forall j \in \{1, 2, \cdots, d\}, \end{aligned} \tag{2.15}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$ and $b = 1 - a$ and note that a and b are positive constants.
Generalization of Theorem 2.4 can be listed as

Theorem 2.5. *The PDE system* (2.9) *can be formulated as a symmetric positive system for* $d = 2$ *or* $d = 3$ *with some unknown variables. Moreover, admissible boundary conditions can be achieved for the system.*

Proof. Without loss of generality, we will prove it for $d = 2$. Referring to (2.15), the system is as follows

$$\begin{aligned}
b\sigma_{11} - a\sigma_{22} - 2\mu\frac{\partial u_1}{\partial x_1} &= 0,\\
b\sigma_{22} - a\sigma_{11} - 2\mu\frac{\partial u_2}{\partial x_2} &= 0,\\
\sigma_{12} - \mu\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) &= 0,\\
\sigma_{21} - \mu\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) &= 0,\\
-\frac{\partial \sigma_{11}}{\partial x_1} - \frac{\partial \sigma_{21}}{\partial x_2} + \beta u_1 &= f_1,\\
-\frac{\partial \sigma_{12}}{\partial x_1} - \frac{\partial \sigma_{22}}{\partial x_2} + \beta u_2 &= f_2.
\end{aligned} \tag{2.16}$$

Using the symmetry of σ, the system

$$\begin{aligned}
b\sigma_{11} - a\sigma_{22} - 2\mu\frac{\partial u_1}{\partial x_1} &= 0,\\
\sigma_{21} - \mu\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) &= 0,\\
\sigma_{12} - \mu\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) &= 0,\\
b\sigma_{22} - a\sigma_{11} - 2\mu\frac{\partial u_2}{\partial x_2} &= 0,\\
-2\mu\frac{\partial \sigma_{11}}{\partial x_1} - \mu\frac{\partial \sigma_{21}}{\partial x_2} - \mu\frac{\partial \sigma_{12}}{\partial x_2} + 2\mu\beta u_1 &= f_1^*,\\
-\mu\frac{\partial \sigma_{21}}{\partial x_1} - \mu\frac{\partial \sigma_{12}}{\partial x_1} - \mu\frac{\partial \sigma_{22}}{\partial x_2} + 2\mu\beta u_2 &= f_2^*.
\end{aligned} \tag{2.17}$$

In matrix form,

$$\begin{bmatrix} b & 0 & 0 & -a & -2\mu\frac{\partial}{\partial x_1} & 0 \\ 0 & 1 & 0 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & 0 & 1 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -a & 0 & 0 & b & 0 & -2\mu\frac{\partial}{\partial x_2} \\ -2\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta & 0 \\ 0 & -\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_1} & -2\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \\ u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ f_1^* \\ f_2^* \end{bmatrix}.$$

The corresponding matrices are

$$\mathcal{A}_{x_1} = \begin{bmatrix} 0 & 0 & 0 & 0 & -2\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & -\mu \\ 0 & 0 & 0 & 0 & 0 & -\mu \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & -\mu & -\mu & 0 & 0 & 0 \end{bmatrix},\ \mathcal{A}_{x_2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 0 & 0 & -\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & -2\mu \\ 0 & -\mu & -\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\mu & 0 & 0 \end{bmatrix},$$

$$\mathbb{A}^0 = \begin{bmatrix} b & 0 & 0 & -a & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -a & 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu\beta & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu\beta \end{bmatrix},$$

$$\text{and } \mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \sum_{k=1}^{d} \partial_k \mathbb{A}^k = 2 \begin{bmatrix} b & 0 & 0 & -a & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -a & 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu\beta & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu\beta \end{bmatrix}.$$

As $\frac{b}{a} = \frac{1-a}{a} = \frac{1}{a} - 1 = \frac{\lambda d + 2\mu}{\lambda} - 1 = d - 1 + \frac{2\mu}{\lambda} = 1 + \frac{2\mu}{\lambda} > 1$, $\mathcal{B}$ is diagonally dominant and hence positive definite. On the other hand, $\mathcal{A}_{x_1}$ and $\mathcal{A}_{x_2}$ are clearly symmetric matrices. So, the PDE system (2.9) is symmetric positive.

For admissible boundary condition

$$D = \sum_{k=1}^{d} n_k \mathcal{A}^k = \begin{bmatrix} 0 & 0 & 0 & 0 & -2\mu n_1 & 0 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & 0 & -2\mu n_2 \\ -2\mu n_1 & -\mu n_2 & -\mu n_2 & 0 & 0 & 0 \\ 0 & -\mu n_1 & -\mu n_1 & -2\mu n_2 & 0 & 0 \end{bmatrix}.$$

- Admissible boundary condition 1:

$$\text{Consider} \quad M = \begin{bmatrix} 0 & 0 & 0 & 0 & 2\mu n_1 & 0 \\ 0 & 0 & 0 & 0 & \mu n_2 & \mu n_1 \\ 0 & 0 & 0 & 0 & \mu n_2 & \mu n_1 \\ 0 & 0 & 0 & 0 & 0 & 2\mu n_2 \\ -2\mu n_1 & -\mu n_2 & -\mu n_2 & 0 & 0 & 0 \\ 0 & -\mu n_1 & -\mu n_1 & -2\mu n_2 & 0 & 0 \end{bmatrix}.$$

$$\text{So,} \quad D - M = 2\begin{bmatrix} 0 & 0 & 0 & 0 & -2\mu n_1 & 0 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & 0 & -2\mu n_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and $(D - M)v = 0$ is equivalent to $u_1 = u_2 = 0$ on the boundary. M satisfied the required conditions, $M \geq 0$ and $\mathbb{R}^m = Ker(D - M) + Ker(D + M)$.

- Admissible boundary condition 2:

$$\text{Consider} \quad M = \begin{bmatrix} 0 & 0 & 0 & 0 & -2\mu n_1 & 0 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & -\mu n_2 & -\mu n_1 \\ 0 & 0 & 0 & 0 & 0 & -2\mu n_2 \\ 2\mu n_1 & \mu n_2 & \mu n_2 & 0 & 0 & 0 \\ 0 & \mu n_1 & \mu n_1 & 2\mu n_2 & 0 & 0 \end{bmatrix}.$$

$$\text{So,} \quad D - M = 2\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2\mu n_1 & -\mu n_2 & -\mu n_2 & 0 & 0 & 0 \\ 0 & -\mu n_1 & -\mu n_1 & -2\mu n_2 & 0 & 0 \end{bmatrix},$$

and $(D - M)v = 0$ is equivalent to $2n_1\sigma_{11} + n_2\sigma_{21} + n_2\sigma_{12} = 0$ and $n_1\sigma_{21} + n_1\sigma_{12} + 2n_2\sigma_{22} = 0$ on the boundary. As $\sigma_{21} = \sigma_{12}$, it implies $\sigma \cdot n = 0$ on the boundary. M satisfied the required conditions, $M \geq 0$ and $\mathbb{R}^m = Ker(D - M) + Ker(D + M)$.

□

The Heat equation

Consider the heat equation

$$\frac{\partial p}{\partial t} - k\nabla^2 p = h \text{ on } \Omega \times (0, T], \tag{2.19}$$

where Ω is a rectangle.

Remark 2.2. *The Heat equation can be formulated as symmetric positive system as in [30]. But we have slightly different approach in here based on the choice of unknown variables.*

Theorem 2.6. *The PDE* (2.19) *can be formulated as a symmetric positive system for* $d = 2$ *or* $d = 3$ *with some unknown variables. Moreover, admissible boundary condition can be given for the system.*

Proof. Without loss of generality, we will prove it for $d = 2$. In order to write as a first order system, define $p_{x_1} = \frac{\partial p}{\partial x_1}$ and $p_{x_2} = \frac{\partial p}{\partial x_2}$. The equation (2.19) can be written in the following matrix form

$$\begin{bmatrix} e^{-t}\frac{\partial}{\partial t} & -ke^{-t}\frac{\partial}{\partial x_1} & -ke^{-t}\frac{\partial}{\partial x_2} \\ -ke^{-t}\frac{\partial}{\partial x_1} & ke^{-t} & 0 \\ -ke^{-t}\frac{\partial}{\partial x_2} & 0 & ke^{-t} \end{bmatrix} \begin{bmatrix} p \\ p_{x_1} \\ p_{x_2} \end{bmatrix} = \begin{bmatrix} h_1 \\ 0 \\ 0 \end{bmatrix}. \tag{2.20}$$

So, the corresponding matrices are $\mathcal{A}_t = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$,

$$\mathcal{A}_{x_1} = \begin{bmatrix} 0 & -ke^{-t} & 0 \\ -ke^{-t} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \mathcal{A}_{x_2} = \begin{bmatrix} 0 & 0 & -ke^{-t} \\ 0 & 0 & 0 \\ -ke^{-t} & 0 & 0 \end{bmatrix}, \text{ and } \ \mathbb{A}^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & ke^{-t} & 0 \\ 0 & 0 & ke^{-t} \end{bmatrix}.$$

It is easy to see $\mathcal{A}_t$, $\mathcal{A}_{x_1}$ and $\mathcal{A}_{x_2}$ are symmetric and $\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_t \mathcal{A}_t - \partial_{x_1}\mathcal{A}_{x_1} - \partial_{x_2}\mathcal{A}_{x_2} = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & 2ke^{-t} & 0 \\ 0 & 0 & 2ke^{-t} \end{bmatrix}$ which is clearly positive definite. So, the PDE (2.19) is symmetric positive. The arguments can be generalized for any higher dimension.

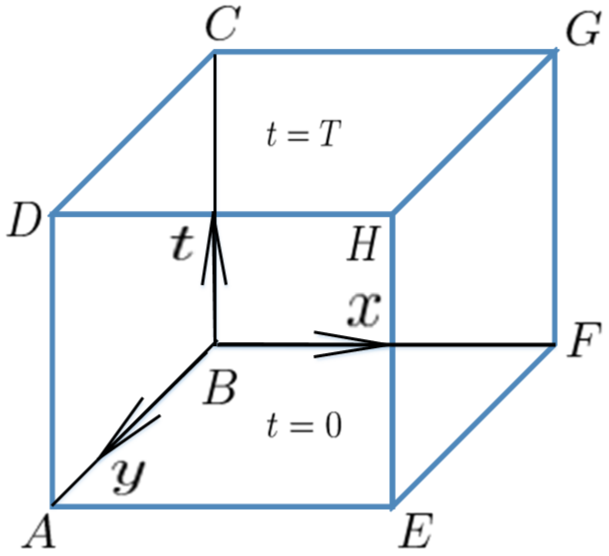


Figure 2.1: Domain $\Omega \times (0,\, T]$

$$\text{Here} \quad D = e^{-t} \begin{bmatrix} n_t & -kn_x & -kn_y \\ -kn_x & 0 & 0 \\ -kn_y & 0 & 0 \end{bmatrix},.$$

We can describe admissible boundary condition as follows

- On AEFB, $n_t = -1$, $n_x = n_y = 0$. So, $D = e^{-t} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and set $M = -D$.

  Now, $D - M = 2D$ which implies $p = 0$ on AEFB.

- On DHGC, $n_t = 1$, $n_x = n_y = 0$. So, $D = e^{-t} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and set $M = D$.

  Now, $D - M = 0$ which implies nothing to impose on DHGC.

- On ABCD, $n_t = 0$, $n_x = -1$, $n_y = 0$. So, $D = e^{-t} \begin{bmatrix} 0 & k & 0 \\ k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and set $M = e^{-t} \begin{bmatrix} 0 & k & 0 \\ -k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

  Now, $D - M = e^{-t} \begin{bmatrix} 0 & 0 & 0 \\ 2k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ which implies $p = 0$ on ABCD.

- On EFGH, $n_t = n_y = 0$, $n_x = 1$. So, $D = e^{-t} \begin{bmatrix} 0 & -k & 0 \\ -k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and set $M = e^{-t} \begin{bmatrix} 0 & -k & 0 \\ k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

  Now, $D - M = e^{-t} \begin{bmatrix} 0 & 0 & 0 \\ -2k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ which implies $p = 0$ on EFGH.

- On AEHD, $n_t = n_x = 0$, $n_y = -1$. So, $D = e^{-t} \begin{bmatrix} 0 & 0 & k \\ 0 & 0 & 0 \\ k & 0 & 0 \end{bmatrix}$ and set $M = e^{-t} \begin{bmatrix} 0 & 0 & k \\ 0 & 0 & 0 \\ -k & 0 & 0 \end{bmatrix}$.

  Now, $D - M = e^{-t} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2k & 0 & 0 \end{bmatrix}$ which implies $p = 0$ on AEHD.

- On BFGC, $n_t = n_x = 0$, $n_y = 1$. So, $D = e^{-t}\begin{bmatrix} 0 & 0 & -k \\ 0 & 0 & 0 \\ -k & 0 & 0 \end{bmatrix}$ and set $M = e^{-t}\begin{bmatrix} 0 & 0 & -k \\ 0 & 0 & 0 \\ k & 0 & 0 \end{bmatrix}$.

  Now, $D - M = e^{-t}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -2k & 0 & 0 \end{bmatrix}$ which implies $p = 0$ on BFGC.

In each case M satisfied the required conditions, $M \geq 0$ and $\mathbb{R}^m = Ker(D-M)+Ker(D+M)$. □

Sometime discretization in one variable is conducted and the resulted PDE is cast into symmetric positive framework. This approach is specially important for numerical computation. We can discretize the equation (2.19) as follows

$$\frac{p^n - p^{n-1}}{\Delta t} - k\frac{\partial^2 p^n}{\partial x_1^2} - k\frac{\partial^2 p^n}{\partial x_2^2} = h^n. \tag{2.21}$$

After manipulation

$$p^n - k\Delta t\frac{\partial^2 p^n}{\partial x_1^2} - k\Delta t\frac{\partial^2 p^n}{\partial x_2^2} = h_1^n. \tag{2.22}$$

In matrix form,

$$\begin{bmatrix} 1 & -k_1\frac{\partial}{\partial x_1} & -k_1\frac{\partial}{\partial x_2} \\ -k_1\frac{\partial}{\partial x_1} & k_1 & 0 \\ -k_1\frac{\partial}{\partial x_2} & 0 & k_1 \end{bmatrix}\begin{bmatrix} p^n \\ p^n_{x_1} \\ p^n_{x_2} \end{bmatrix} = \begin{bmatrix} h_1^n \\ 0 \\ 0 \end{bmatrix}, \tag{2.23}$$

where $k_1 = k\Delta t$. The following theorem shows that the equation (2.22) is indeed symmetric positive.

Theorem 2.7. *The PDE* (2.22) *can be formulated as a symmetric positive system with some unknown variables. Moreover, admissible boundary condition can be given for the system.*

Proof. Equation (2.22) can be written as a system of first order (2.23). So, the corresponding matrices are

$$\mathcal{A}_{x_1} = \begin{bmatrix} 0 & -k_1 & 0 \\ -k_1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; \mathcal{A}_{x_2} = \begin{bmatrix} 0 & 0 & -k_1 \\ 0 & 0 & 0 \\ -k_1 & 0 & 0 \end{bmatrix}, \text{ and } \mathbb{A}^0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & k_1 & 0 \\ 0 & 0 & k_1 \end{bmatrix}.$$

Now $\mathcal{A}_{x_1}$ and $\mathcal{A}_{x_2}$ are symmetric, and

$$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_{x_1}\mathcal{A}_{x_1} - \partial_{x_2}\mathcal{A}_{x_2} = 2\begin{bmatrix} 1 & 0 & 0 \\ 0 & k_1 & 0 \\ 0 & 0 & k_1 \end{bmatrix},$$

which is clearly positive definite. The claim follows.

For admissible boundary condition, we note

$$D = \begin{bmatrix} 0 & -k_1 n_{x_1} & -k_1 n_{x_2} \\ -k_1 n_{x_1} & 0 & 0 \\ -k_1 n_{x_2} & 0 & 0 \end{bmatrix}, \text{ and set } M = \begin{bmatrix} 0 & -k_1 n_{x_1} & -k_1 n_{x_2} \\ k_1 n_{x_1} & 0 & 0 \\ k_1 n_{x_2} & 0 & 0 \end{bmatrix}.$$

. Clearly, M satisfied the required conditions, $M \geq 0$ and $\mathbb{R}^m = Ker(D-M) + Ker(D+M)$. So,

$$D - M = \begin{bmatrix} 0 & 0 & 0 \\ -2k_1 n_{x_1} & 0 & 0 \\ -2k_1 n_{x_2} & 0 & 0 \end{bmatrix},$$

which implies $p^n = 0$ on the boundary. □

Remark 2.3. *We note, $p^{(n)}$ can be found once we have known $p^{(n-1)} = 0$. Thus, we can solve the time marching problem.*

The Ultrahyperbolic equation

We consider

$$\frac{\partial^2 \phi}{\partial x_1^2} + \frac{\partial^2 \phi}{\partial x_2^2} - \frac{\partial^2 \phi}{\partial x_3^2} - \frac{\partial^2 \phi}{\partial x_4^2} = 0, \tag{2.24}$$

which can be written as

$$\begin{aligned} \frac{\partial}{\partial x_1}(\frac{\partial \phi}{\partial x_1}) + \frac{\partial}{\partial x_2}(\frac{\partial \phi}{\partial x_2}) - \frac{\partial}{\partial x_3}(\frac{\partial \phi}{\partial x_3}) - \frac{\partial}{\partial x_4}(\frac{\partial \phi}{\partial x_4}) &= 0, \\ \frac{\partial}{\partial x_2}(\frac{\partial \phi}{\partial x_1}) - \frac{\partial}{\partial x_1}(\frac{\partial \phi}{\partial x_2}) &= 0, \\ -\frac{\partial}{\partial x_3}(\frac{\partial \phi}{\partial x_1}) + \frac{\partial}{\partial x_1}(\frac{\partial \phi}{\partial x_3}) &= 0, \\ -\frac{\partial}{\partial x_4}(\frac{\partial \phi}{\partial x_1}) + \frac{\partial}{\partial x_1}(\frac{\partial \phi}{\partial x_4}) &= 0. \end{aligned} \tag{2.25}$$

In matrix form,

$$\begin{bmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & -\frac{\partial}{\partial x_3} & -\frac{\partial}{\partial x_4} \\ \frac{\partial}{\partial x_2} & -\frac{\partial}{\partial x_1} & 0 & 0 \\ -\frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} & 0 \\ -\frac{\partial}{\partial x_4} & 0 & 0 & \frac{\partial}{\partial x_1} \end{bmatrix} \begin{bmatrix} \frac{\partial \phi}{\partial x_1} \\ \frac{\partial \phi}{\partial x_2} \\ \frac{\partial \phi}{\partial x_3} \\ \frac{\partial \phi}{\partial x_4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.26}$$

.

The Ultrahyperbolic equation (2.24) is symmetric positive as shown in the following theorem [30].

Theorem 2.8. *There is a transformation of PDE* (2.24) *or its equivalent form* (2.26) *such that the resulting first order system can be formulated as a symmetric positive system.*

Proof. It suffices to consider the following system as indicated in [30]

$$\begin{bmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & -\frac{\partial}{\partial x_3} & -\frac{\partial}{\partial x_4} \\ \frac{\partial}{\partial x_2} & -\frac{\partial}{\partial x_1} & 0 & 0 \\ -\frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} & 0 \\ -\frac{\partial}{\partial x_4} & 0 & 0 & \frac{\partial}{\partial x_1} \end{bmatrix} \begin{bmatrix} \frac{\partial\phi}{\partial x_1} \\ \frac{\partial\phi}{\partial x_2} \\ \frac{\partial\phi}{\partial x_3} \\ \frac{\partial\phi}{\partial x_4} \end{bmatrix} + 2k \begin{bmatrix} 0 \\ \frac{\partial\phi}{\partial x_2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{2.27}$$

where k is an arbitrarily small positive constant. Now, consider the transformation

$$v = e^{-kx_1}u \quad \text{with} \quad u = \left(\frac{\partial\phi}{\partial x_1} \quad \frac{\partial\phi}{\partial x_2} \quad \frac{\partial\phi}{\partial x_3} \quad \frac{\partial\phi}{\partial x_4} \right)^T.$$

By the transformation, we can write

$$\begin{bmatrix} e^{kx_1}(\frac{\partial}{\partial x_1} + k) & e^{kx_1}\frac{\partial}{\partial x_2} & -e^{kx_1}\frac{\partial}{\partial x_3} & -e^{kx_1}\frac{\partial}{\partial x_4} \\ e^{kx_1}\frac{\partial}{\partial x_2} & -e^{kx_1}(\frac{\partial}{\partial x_1} - k) & 0 & 0 \\ -e^{kx_1}\frac{\partial}{\partial x_3} & 0 & e^{kx_1}(\frac{\partial}{\partial x_1} + k) & 0 \\ -e^{kx_1}\frac{\partial}{\partial x_4} & 0 & 0 & e^{kx_1}(\frac{\partial}{\partial x_1} + k) \end{bmatrix} \begin{bmatrix} e^{-kx_1}\frac{\partial\phi}{\partial x_1} \\ e^{-kx_1}\frac{\partial\phi}{\partial x_2} \\ e^{-kx_1}\frac{\partial\phi}{\partial x_3} \\ e^{-kx_1}\frac{\partial\phi}{\partial x_4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.28}$$

The corresponding matrices are

$$\mathcal{A}_{x_1} = \begin{bmatrix} e^{kx_1} & 0 & 0 & 0 \\ 0 & -e^{kx_1} & 0 & 0 \\ 0 & 0 & e^{kx_1} & 0 \\ 0 & 0 & 0 & e^{kx_1} \end{bmatrix}, \ \mathcal{A}_{x_2} = \begin{bmatrix} 0 & e^{kx_1} & 0 & 0 \\ e^{kx_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{A}_{x_3} = \begin{bmatrix} 0 & 0 & -e^{kx_1} & 0 \\ 0 & 0 & 0 & 0 \\ -e^{kx_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathcal{A}_{x_4} = \begin{bmatrix} 0 & 0 & 0 & -e^{kx_1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -e^{kx_1} & 0 & 0 & 0 \end{bmatrix}, \ \mathbb{A}^0 = \begin{bmatrix} ke^{kx_1} & 0 & 0 & 0 \\ 0 & ke^{kx_1} & 0 & 0 \\ 0 & 0 & ke^{kx_1} & 0 \\ 0 & 0 & 0 & ke^{kx_1} \end{bmatrix}.$$

So,

$$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_{x_1}\mathcal{A}_{x_1} - \partial_{x_2}\mathcal{A}_{x_2} - \partial_{x_3}\mathcal{A}_{x_3} - \partial_{x_4}\mathcal{A}_{x_4} = e^{kx_1} \begin{bmatrix} 2k-k^2 & 0 & 0 & 0 \\ 0 & 2k-k^2 & 0 & 0 \\ 0 & 0 & 2k-k^2 & 0 \\ 0 & 0 & 0 & 2k-k^2 \end{bmatrix}.$$

Now $\mathcal{A}_{x_1}$, $\mathcal{A}_{x_2}$, $\mathcal{A}_{x_3}$ and $\mathcal{A}_{x_4}$ are clearly symmetric. On the other hand, for $0 < k < 2$, $\mathcal{B}$ is positive definite. The claim follows. □

2.2 Least square finite element method

Given a system of partial differential equations, the least square finite element method (LSFEM) defines an unconstrained minimization problem and based on that, a finite element method can be developed in a variational setting. The very basic concept of LSFEM is to define the least square functional as the residuals measured in some suitable norm, often challenging to find, in Hilbert space. Assuming the well-posedness of original PDE along with suitable boundary conditions, the least square functional will have a unique minimizer and thus, the corresponding variational formulation has a unique solution. Also, by construction of the least square functional, the bilinear form associated with the corresponding variational formulation is symmetric positive definite. So, in the case of the discrete formulation, it leads to a symmetric positive linear system, relatively easy to solve numerically. Additionally, if the induced energy norm is norm-equivalent to some norm in a suitable Hilbert space, optimal properties [40] of the resulting least-squares method can be achieved. One of the main difference between conventional Galerkin method and LSFEM is to require different regularity of finite element space. In the former case, regularity requirements are weakened by the integration by parts whereas in LSFEM, it requires higher regularity of the finite element space.

Application of LSFEM to solve problems dates back to early seventies [47, 48]. Soon after that LSFEM suffered a couple of disadvantages that greatly limited its appeal. In many problems, discretization requires C^1 or better finite element spaces, which lead to assembly matrix with high condition number and hence difficult to solve. Later on, this problem of having higher regularity has been greatly optimized by first transforming the PDE into first order system and then using LSFEM. Because of that, in the last few decades, the LSFEM has been receiving increasing attention in both the engineering and mathematics communities [38, 39, 45]. The increased attention is also due to the fact that LSFEM offers significant analytic and computational advantages over conventional finite element method. The main advantages include LSFEM is not subject to the inf-sup stability condition, boundary conditions can be enforced weakly, resulting assembly matrix is symmetric positive leading to efficient computation etc. One of the great example of LSFEM application is to use this method for the numerical approximation of Navier-Stokes equation instead of using conventional mixed Galerkin method, needing to satisfy Ladyzenskaja-Babuska-Brezzi(LBB) condition at discretization level. Using LSFEM in this case offers easy discretization process, same finite element space for all unknowns, important information about physically important variables such as vorticity, reasonable condition number for the discrete problem. Other applications of LSFEM are stationary incompressible flow, time dependent incompressible flow, convection-diffusion problems, purely hyperbolic problems etc.

Basic strategies of LSFEM include transformation of the PDE and identification of proper subspace. Before any numerical treatment, it is advantageous to transform the original PDE into a system of first order PDE. This process often offers discretization by C^0 finite element and approximations of physically important fields, such as vorticity, fluxes, stresses etc. On the other hand, identification of proper subspace ensures some priori estimate holds or the original operator is bounded below. A number of advantages is derived from this step. First of all, it provides existence and uniqueness of the minimizers. For numerical treatment, it ensures stability of discretization, avoiding any inf-sup type condition. The linear systems resulting from discretization process are symmetric positive definite matrices, which can be solved by robust iterative methods (such as preconditioned conjugate gradient methods).

2.2.1 Applications of LSFEM

LSFEM is particularly used in engineering communities now a days. The followings are typical applications of this method as found in [46].

- Div-Curl system,
- Div-Curl-Grad system,
- Incompressible irrotational flow,
- Subsonic compressible irrotational flow,
- The Stokes flows,
- The Navier-Stokes equations,
- Natural convection,
- Rayleigh Benard convection cells,
- Doubly diffusive convective flows,

- Surface-Tension driven convection,
- Convective transport equations,
- Thermally stratified flows,
- Incompressible Euler equations,
- The compressible Navier-Stokes equations,
- Flows over a backward facing steps,
- Two fluid flows,
- High-speed compressible flows,
- Maxwell equations,
- Transient scattering waves,
- Coupled Stokes-Darcy flow,
- Poisson-Boltzmann equations,
- Domain decomposition based LSFEM for large scale parallel computations.

2.2.2 Mathematical formulation

Consider the following boundary value problem

$$\begin{aligned} Lu = f \quad &\text{in} \quad \Omega, \\ Bu = g \quad &\text{on} \quad \partial\Omega, \end{aligned} \tag{2.29}$$

where L is a first order differential operator as follows

$$Lu = \sum_{j=1}^{d} L_j \frac{\partial u}{\partial x_j} + L_0 u. \tag{2.30}$$

Here Ω is an open bounded subset of $\mathbb{R}^d$ with sufficiently smooth boundary $\partial\Omega$ and $u^T = (u_1\, u_2\, u_3 \cdots u_m)$, a vector of m unknown functions of $x = (x_1\, x_2\, x_3 \cdots x_d)$. For this setting, L_j and L_0 are $d \times m$ matrices, f and g are given vector valued $(d \times 1)$ functions, and B is a boundary operator. The fundamental principle of the least squares variational method is the minimization of the mean squared error in the equations over the problem domain. In particular, the objective is to find an approximation that satisfies equation (2.29). Without loss of generality we may assume $g = 0$, and we now consider the following boundary value problem

$$\begin{aligned} Lu = f \quad &\text{in} \quad \Omega, \\ Bu = 0 \quad &\text{on} \quad \partial\Omega. \end{aligned} \tag{2.31}$$

For the proper functional setting, we suppose $f \in \mathbf{L}(\Omega)$ and an appropriate Hilbert space $W = \{v \in \mathbf{L}(\Omega)\,;\; Bv = 0 \text{ on } \partial\Omega\} \subset \mathbf{L}(\Omega)$. We consider the operator L maps W into $\mathbf{L}(\Omega)$ as follows

$$L : W \to \mathbf{L}(\Omega)\,. \tag{2.32}$$

Define the residual function

$$E(u) = Lu - f \quad \text{for all } u \in W, \tag{2.33}$$

and the least square quadratic functional

$$I(u) = \|Lu - f\|^2_{\mathbf{L}} = (Lu - f,\, Lu - f) = (E(u),\, E(u))\,. \tag{2.34}$$

A solution u to the problem (2.31) can be interpreted as $u \in W$ that minimizes $E(u)$ i.e.

$$0 = I(u) \leq I(v) \quad \text{for all } v \in W. \tag{2.35}$$

So, we are seeking a minimizer of the quadratic functional I in W. A necessary condition that $u \in W$ be a minimizer of the functional I in W is that its first variation vanishes at that point. It follows

$$\begin{aligned} \lim_{t\to 0} \frac{d}{dt} I\left(u+tv\right) &= 0, \\ \lim_{t\to 0} \frac{d}{dt} \|L(u+tv)-f\|_0^2 &= 0, \\ \lim_{t\to 0} \frac{d}{dt} \int\limits_{\Omega} \left(L(u+tv)-f\right)^2\ d\Omega &= 0, \\ \lim_{t\to 0} \frac{d}{dt} \int\limits_{\Omega} \big(LuLv - 2fLu + 2tLuLv - 2ftLv + t^2 LvLv\big)\ d\Omega &= 0, \\ \lim_{t\to 0} \int\limits_{\Omega} \left(2LuLv - 2fLv + tLvLv\right)\ d\Omega &= 0, \\ \int\limits_{\Omega} \left(2LuLv - 2fLv\right)\ d\Omega = 0 \Rightarrow 2\int\limits_{\Omega} Lv\left(Lu-f\right)\ d\Omega &= 0, \end{aligned} \tag{2.36}$$

for all $v \in W$. This leads to the following variational formulation.

Find $u \in W$ such that

$$a(u,\, v) = F(v) \quad \text{for all} \;\; v \in W, \tag{2.37}$$

where

$$\begin{aligned} a(u,\, v) &= \left(Lu,\, Lv\right), \\ F(v) &= \left(f,\, Av\right). \end{aligned} \tag{2.38}$$

In the finite element approximation, we partition the domain into a finite number of elements, characterized by a discretization parameter h and then introduce an appropriate finite element basis. Let N denote the number of nodes for one element and ϕ_j denote the element shape functions. Assuming the same finite element is used for all unknown variables, we have

$$u_h^e(x) = \sum_{j=1}^{N} \phi_j(x)\left(u_1\ u_2 \cdots u_m\right)_j^T, \tag{2.39}$$

where $\left(u_1\ u_2 \cdots u_m\right)_j$ are nodal values at the jth node. Introducing the finite element approximation defined in (2.39) into the variational formulation (2.37), we have the following linear system

$$\begin{aligned}\mathbf{AU} &= \mathbf{F},\\ \mathbf{A}_e &= \int\limits_{\Omega_e} (L\phi_1 \; L\phi_2 \; \cdots L\phi_N \;)^T \; (L\phi_1 \; L\phi_2 \; \cdots L\phi_N \;) \; d\Omega,\\ \mathbf{F}_e &= \int\limits_{\Omega_e} (L\phi_1 \; L\phi_2 \; \cdots L\phi_N \;)^T f \, d\Omega,\end{aligned} \tag{2.40}$$

where $\mathbf{U}$ is global vector of nodal values and $\Omega_e \subset \Omega$ is the domain of the e-th element. The global matrix $\mathbf{A}$ is assembled from $\mathbf{A_e}$s and $\mathbf{F}$ is assembled from $\mathbf{F_e}$s.

In LSFEM, boundary condition can be enforced weakly. In the mathematical formulation, boundary conditions are imposed on W. If we want to use boundary conditions weakly, we may do so by adding boundary terms in the least square functional as below

$$I(u) = \|Lu - f\|^2_{\mathbf{L},\Omega} + \|Bu - g\|^2_{\mathbf{L},\partial\Omega}. \tag{2.41}$$

We summarize this section by stating the following theorem [41] without proof.

Theorem 2.9. *Assume* (2.29) *is well-posed and a finite dimensional space $W^h \subset W$. Then*

1. *The bilinear form $a(\cdot, \cdot)$ defined in* (2.38) *is continuous, symmetric, and coercive.*

2. *The linear functional $F(\cdot)$ defined in* (2.38) *is continuous.*

3. *The variational formulation defined in* (2.37) *has a unique solution $u \in W$ that is also the unique solution of the minimization problem* (2.35).

4. *The corresponding discrete problem of* (2.37) *has unique solution $u_h \in W^h$.*

5. *The matrix $\mathbf{A}$ is symmetric and positive definite.*

6. *There is a constant $C > 0$ such that u and u_h satisfy the error estimate*

$$\|u - u_h\|_W \leq C \inf_{w^h \in W^h} \|u - w_h\|_W. \tag{2.42}$$

2.2.3 An example of LSFEM

This example [46] is intended to show why LSFEM is advantageous in comparison with other numerical methods. Consider the following first order differential equation

$$u'(x) = \frac{e^{-\frac{1-x}{\eta}}}{\eta\left(1 - e^{-\eta^{-1}}\right)} \quad \text{for } x \in [0,1], \qquad u(0) = 0, \tag{2.43}$$

with $0 < \eta < 1$. The exact solution of this differential equation is

$$u(x) = 1 - \left(1 - e^{-\eta^{-1}}\right)^{-1}\left(1 - e^{-\frac{1-x}{\eta}}\right). \tag{2.44}$$

The LSFEM is performed with equidistant 10 linear elements and Simpson quadrature rule. The same setting has been used for Galerkin method. The Galerkin solution wildly oscillates over the domain and it is far from the exact solution. On the other hand, the LSFEM solution is very close to the exact solution with any oscillation. The comparison is illustrated in the following figure for $\eta = 0.05$ [46].

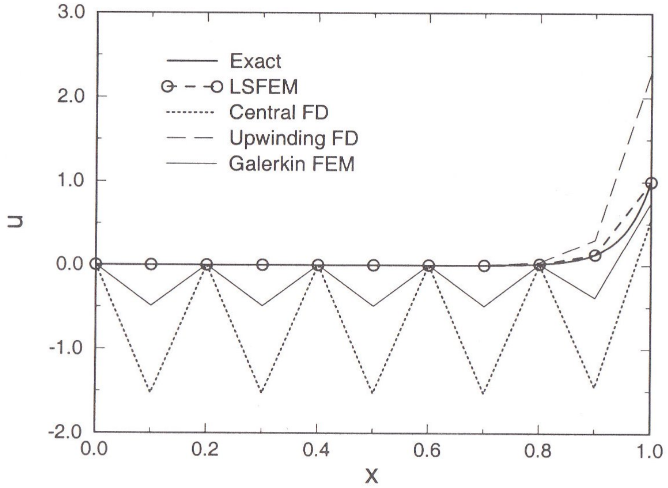


Figure 2.2: Solution of the model differential equation (2.43)

Chapter 3

Fluid content-rotation-pressure gradient formulation

We consider the following partial differential equations

$$\begin{cases} -\left(\lambda+\mu\right)\nabla\left(\nabla\cdot u\right)-\mu\Delta u+\alpha\nabla p=f, \\ \frac{\partial}{\partial t}\left[S_{\epsilon}p+\alpha\left(\nabla\cdot u\right)\right]-k\nabla^{2}p=h. \end{cases} \tag{3.1}$$

Although one dimensional formulation for symmetric positive system is simple, it is different in terms of unknown variables than that of higher dimensional case. Thus, we start with the one dimensional case.

3.1 One dimensional case

One dimensional formulation for (3.1)

$$\begin{aligned} -(\lambda+\mu)\frac{\partial^2 u}{\partial x^2}-\mu\frac{\partial^2 u}{\partial x^2}+\alpha\frac{\partial p}{\partial x}&=f, \\ \frac{\partial}{\partial t}\left[S_{\epsilon}p+\alpha\nabla.u\right]-k\frac{\partial^2 p}{\partial x^2}&=h. \end{aligned} \tag{3.2}$$

Define two positive constants $\alpha_2=S_{\epsilon}k+\frac{\alpha^2 k}{\lambda+2\mu}$ and $\eta=S_{\epsilon}p+\alpha\nabla\cdot u$.

The first equation

$$\begin{aligned} -(\lambda+2\mu)\frac{\partial^2 u}{\partial x^2}+\alpha\frac{\partial p}{\partial x}&=f, \\ \Rightarrow\frac{\partial}{\partial x}\left((\lambda+2\mu)\frac{\partial u}{\partial x}-\alpha p\right)&=-f, \\ \Rightarrow\frac{\partial}{\partial x}\left(\alpha\frac{\partial u}{\partial x}-\frac{\alpha^2}{(\lambda+2\mu)}p\right)&=-\frac{f\alpha}{(\lambda+2\mu)}, \\ \Rightarrow\frac{\partial}{\partial x}\left(\alpha\frac{\partial u}{\partial x}+S_{\epsilon}p-S_{\epsilon}p-\frac{\alpha^2}{(\lambda+2\mu)}p\right)&=-\frac{f\alpha}{(\lambda+2\mu)}, \end{aligned}$$

$$
\begin{aligned}
\Rightarrow \frac{\partial}{\partial x}\left(S_\epsilon p + \alpha\frac{\partial u}{\partial x}\right) - \left(S_\epsilon + \frac{\alpha^2}{(\lambda+2\mu)}\right)\frac{\partial p}{\partial x} &= -\frac{f\alpha}{(\lambda+2\mu)},\\
\Rightarrow -k\frac{\partial \eta}{\partial x} + k\left(S_\epsilon + \frac{\alpha^2}{(\lambda+2\mu)}\right)\frac{\partial p}{\partial x} &= \frac{f\alpha k}{(\lambda+2\mu)},\\
\Rightarrow -k\frac{\partial \eta}{\partial x} + \alpha_2\frac{\partial p}{\partial x} &= f^*.
\end{aligned}
$$

So, the system of first order equations

$$
\begin{aligned}
\frac{\partial \eta}{\partial t} - k\frac{\partial}{\partial x}\left(\frac{\partial p}{\partial x}\right) &= h,\\
-k\frac{\partial \eta}{\partial x} + \alpha_2\frac{\partial p}{\partial x} &= f^*.
\end{aligned}
\tag{3.3}
$$

In matrix form,

$$
\begin{bmatrix} \frac{\partial}{\partial t} & -k\frac{\partial}{\partial x} \\ -k\frac{\partial}{\partial x} & \alpha_2 \end{bmatrix}
\begin{bmatrix} \eta \\ \frac{\partial p}{\partial x} \end{bmatrix}
=
\begin{bmatrix} h \\ f^* \end{bmatrix}. \tag{3.4}
$$

Theorem 3.1. *There is a transformation of PDE* (3.2) *such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

Proof. As equation (3.2) and (3.4) are equivalent, we can work with (3.4). Consider the transformation, $v = e^{-\xi t}\left(\eta\ ,\ \frac{\partial p}{\partial x}\right)^T$ where ξ is some positive constant. With this transformation equation (3.4) becomes

$$
\begin{bmatrix} \frac{\partial}{\partial t} + \xi & -k\frac{\partial}{\partial x} \\ -k\frac{\partial}{\partial x} & \alpha_2 \end{bmatrix}
\begin{bmatrix} \eta e^{-\xi t} \\ \frac{\partial p}{\partial x} e^{-\xi t} \end{bmatrix}
=
\begin{bmatrix} h_1 \\ f^{**} \end{bmatrix}. \tag{3.5}
$$

The corresponding matrices are

$$\mathcal{A}_t = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \mathcal{A}_x = \begin{bmatrix} 0 & -k \\ -k & 0 \end{bmatrix}, \text{ and } \mathbb{A}^0 = \begin{bmatrix} \xi & 0 \\ 0 & \alpha_2 \end{bmatrix}$$

So,

$$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_t \mathcal{A}_t - \partial_x \mathcal{A}_x = 2 \begin{bmatrix} \xi & 0 \\ 0 & \alpha_2 \end{bmatrix}.$$

Now $\mathcal{A}_t$, $\mathcal{A}_x$ are clearly symmetric. On the other hand, as $\xi > 0$ and $\alpha_2 > 0$, $\mathcal{B}$ is positive definite. So, the system is symmetric positive. Consider the domain of PDE is $\Omega \times (0, T]$ where open subset $\Omega \subset \mathbf{R}$.

We can implement admissible boundary condition as follows

$$\text{Here} \quad D = \begin{bmatrix} n_t & -kn_x \\ -kn_x & 0 \end{bmatrix}.$$

- On $\Omega \times \{0\}$, $n_t = -1$, $n_x = 0$. So, $D = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$, and set $M = -D$.

  Now, $D - M = 2D$ which implies $\eta = 0$ on $\Omega \times \{0\}$.

- On $\Omega \times \{T\}$, $n_t = 1$, $n_x = 0$. So, $D = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and set $M = D$.

  Now, $D - M = 0$ which implies nothing on $\Omega \times \{T\}$.

- On $\partial\Omega \times (0, T]$, $n_t = 0$, $n_x = \pm 1$. So, $D = \begin{bmatrix} 0 & \mp k \\ \mp k & 0 \end{bmatrix}$, and set $M = \begin{bmatrix} 0 & \mp k \\ \pm k & 0 \end{bmatrix}$.

  Now, $D - M = \begin{bmatrix} 0 & 0 \\ \mp 2k & 0 \end{bmatrix}$ which implies $\eta = 0$ on $\partial\Omega \times (0, T]$.

In each cases, M satisfied the required conditions, $M \geq 0$ and $\mathbb{R}^m = Ker(D - M) + Ker(D + M)$. So, the claim follows. □

3.2 Two dimensional case

For $d = 2$, we have the equation (3.1)

$$\begin{aligned}
-(\lambda+\mu)\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\left(\frac{\partial^2 u_1}{\partial x_1^2}+\frac{\partial^2 u_1}{\partial x_2^2}\right)+\alpha\frac{\partial p}{\partial x_1} &= f_1,\\
-(\lambda+\mu)\frac{\partial}{\partial x_2}\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\left(\frac{\partial^2 u_2}{\partial x_1^2}+\frac{\partial^2 u_2}{\partial x_2^2}\right)+\alpha\frac{\partial p}{\partial x_2} &= f_2,\\
\frac{\partial}{\partial t}\left[S_\epsilon p+\alpha\nabla.u\right]-k\nabla^2 p &= h.
\end{aligned}\tag{3.6}$$

Define $\alpha_2 = \frac{2\alpha\mu k}{\lambda+2\mu}$, $\alpha_3 = S_\epsilon k + \frac{\alpha^2 k}{\lambda+2\mu}$, $w_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right)$ and $\eta = S_\epsilon p + \alpha\nabla.u$.

The first equation

$$\begin{aligned}
-(\lambda+\mu)\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\left(\frac{\partial^2 u_1}{\partial x_1^2}+\frac{\partial^2 u_1}{\partial x_2^2}\right)+\alpha\frac{\partial p}{\partial x_1} &= f_1,\\
\Rightarrow \frac{\partial}{\partial x_1}\left((\lambda+\mu)\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)+\mu\frac{\partial u_1}{\partial x_1}-\alpha p\right)+\mu\frac{\partial^2 u_1}{\partial x_2^2} &= -f_1,\\
\Rightarrow \frac{\partial}{\partial x_1}\left((\lambda+\mu)\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)+\mu\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\frac{\partial u_2}{\partial x_2}-\alpha p\right)+\mu\frac{\partial^2 u_1}{\partial x_2^2} &= -f_1,\\
\Rightarrow \frac{\partial}{\partial x_1}\left((\lambda+2\mu)\nabla.u-\alpha p\right)+\mu\frac{\partial^2 u_1}{\partial x_2^2}-\mu\frac{\partial^2 u_2}{\partial x_1\partial x_2} &= -f_1,\\
\Rightarrow \frac{\partial}{\partial x_1}\left((\lambda+2\mu)\nabla.u-\alpha p\right)+\mu\frac{\partial}{\partial x_2}\left(\frac{\partial u_1}{\partial x_2}-\frac{\partial u_2}{\partial x_1}\right) &= -f_1,\\
\Rightarrow \frac{\partial}{\partial x_1}\left(\alpha\nabla.u-\frac{\alpha^2}{\lambda+2\mu}p\right)+\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_2} &= -\frac{f_1\alpha}{\lambda+2\mu},\\
\Rightarrow \frac{\partial}{\partial x_1}\left(\alpha\nabla.u+S_\epsilon p-S_\epsilon p-\frac{\alpha^2}{\lambda+2\mu}p\right)+\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_2} &= -\frac{f_1\alpha}{\lambda+2\mu},\\
\Rightarrow \frac{\partial\eta}{\partial x_1}-\left(S_\epsilon+\frac{\alpha^2}{\lambda+2\mu}\right)\frac{\partial p}{\partial x_1}+\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_2} &= -\frac{f_1\alpha}{\lambda+2\mu},\\
\Rightarrow -k\frac{\partial\eta}{\partial x_1}+\left(S_\epsilon+\frac{\alpha^2}{\lambda+2\mu}\right)k\frac{\partial p}{\partial x_1}-\frac{2\alpha\mu k}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_2} &= \frac{f_1\alpha k}{\lambda+2\mu},\\
\Rightarrow -k\frac{\partial\eta}{\partial x_1}-\alpha_2\frac{\partial w_{12}}{\partial x_2}+\alpha_3\frac{\partial p}{\partial x_1} &= f_1^*.
\end{aligned}$$

The second equation can be written as

$$
\begin{aligned}
-(\lambda+\mu)\frac{\partial}{\partial x_2}\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\left(\frac{\partial^2 u_2}{\partial x_1^2}+\frac{\partial^2 u_2}{\partial x_2^2}\right)+\alpha\frac{\partial p}{\partial x_2} &= f_2,\\
\Rightarrow \frac{\partial}{\partial x_2}\left((\lambda+\mu)\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)+\mu\frac{\partial u_2}{\partial x_2}-\alpha p\right)+\mu\frac{\partial^2 u_2}{\partial x_2^2} &= -f_2,\\
\Rightarrow \frac{\partial}{\partial x_2}\left((\lambda+\mu)\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)+\mu\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)-\mu\frac{\partial u_1}{\partial x_1}-\alpha p\right)+\mu\frac{\partial^2 u_2}{\partial x_1^2} &= -f_2,\\
\Rightarrow \frac{\partial}{\partial x_2}\left((\lambda+2\mu)\nabla.u-\alpha p\right)+\mu\frac{\partial^2 u_2}{\partial x_1^2}-\mu\frac{\partial^2 u_1}{\partial x_1\partial x_2} &= -f_2,\\
\Rightarrow \frac{\partial}{\partial x_2}\left((\lambda+2\mu)\nabla.u-\alpha p\right)-\mu\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_2}-\frac{\partial u_2}{\partial x_1}\right) &= -f_2,\\
\Rightarrow \frac{\partial}{\partial x_2}\left(\alpha\nabla.u-\frac{\alpha^2}{\lambda+2\mu}p\right)-\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_1} &= -\frac{f_2\alpha}{\lambda+2\mu},\\
\Rightarrow \frac{\partial}{\partial x_2}\left(\alpha\nabla.u+S_\epsilon p-S_\epsilon p-\frac{\alpha^2}{\lambda+2\mu}p\right)-\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_1} &= -\frac{f_2\alpha}{\lambda+2\mu},\\
\Rightarrow \frac{\partial\eta}{\partial x_2}-\left(S_\epsilon+\frac{\alpha^2}{\lambda+2\mu}\right)\frac{\partial p}{\partial x_2}-\frac{2\alpha\mu}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_1} &= -\frac{f_2\alpha}{\lambda+2\mu},\\
\Rightarrow -k\frac{\partial\eta}{\partial x_2}+\left(S_\epsilon+\frac{\alpha^2}{\lambda+2\mu}\right)k\frac{\partial p}{\partial x_2}+\frac{2\alpha\mu k}{\lambda+2\mu}\frac{\partial w_{12}}{\partial x_1} &= \frac{f_2\alpha k}{\lambda+2\mu},\\
\Rightarrow -k\frac{\partial\eta}{\partial x_2}+\alpha_2\frac{\partial w_{12}}{\partial x_1}+\alpha_3\frac{\partial p}{\partial x_2} &= f_2^*.
\end{aligned}
$$

So, the system of first order equations

$$
\begin{aligned}
\frac{\partial\eta}{\partial t}-k\frac{\partial}{\partial x_1}\left(\frac{\partial p}{\partial x_1}\right)-k\frac{\partial}{\partial x_2}\left(\frac{\partial p}{\partial x_2}\right) &= h,\\
-\alpha_2\frac{\partial}{\partial x_2}\left(\frac{\partial p}{\partial x_1}\right)+\alpha_2\frac{\partial}{\partial x_1}\left(\frac{\partial p}{\partial x_2}\right) &= 0,\\
-k\frac{\partial\eta}{\partial x_1}-\alpha_2\frac{\partial w_{12}}{\partial x_2}+\alpha_3\frac{\partial p}{\partial x_1} &= f_1^*,\\
-k\frac{\partial\eta}{\partial x_2}+\alpha_2\frac{\partial w_{12}}{\partial x_1}+\alpha_3\frac{\partial p}{\partial x_2} &= f_2^*,
\end{aligned}
\tag{3.7}
$$

with unknown variables η, w_{12}, $\frac{\partial p}{\partial x_1}$ and $\frac{\partial p}{\partial x_2}$.

In matrix form,

$$\begin{bmatrix} \frac{\partial}{\partial t} & 0 & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & 0 & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} \\ -k\frac{\partial}{\partial x_1} & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_3 & 0 \\ -k\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} & 0 & \alpha_3 \end{bmatrix} \begin{bmatrix} \eta \\ w_{12} \\ \frac{\partial p}{\partial x_1} \\ \frac{\partial p}{\partial x_2} \end{bmatrix} = \begin{bmatrix} h \\ 0 \\ f_1^* \\ f_2^* \end{bmatrix}. \tag{3.8}$$

With respect to symmetric positive formulation, it suffices to consider the following perturbed system

$$\begin{bmatrix} \frac{\partial}{\partial t} & 0 & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & 0 & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} \\ -k\frac{\partial}{\partial x_1} & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_3 & 0 \\ -k\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} & 0 & \alpha_3 \end{bmatrix} \begin{bmatrix} \eta \\ w_{12} \\ \frac{\partial p}{\partial x_1} \\ \frac{\partial p}{\partial x_2} \end{bmatrix} + \epsilon \begin{bmatrix} 0 \\ w_{12} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} h \\ 0 \\ f_1^* \\ f_2^* \end{bmatrix}, \tag{3.9}$$

for arbitrarily small $\epsilon > 0$. So, the PDE system is

$$\begin{bmatrix} \frac{\partial}{\partial t} & 0 & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & \epsilon & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} \\ -k\frac{\partial}{\partial x_1} & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_3 & 0 \\ -k\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} & 0 & \alpha_3 \end{bmatrix} \begin{bmatrix} \eta \\ w_{12} \\ \frac{\partial p}{\partial x_1} \\ \frac{\partial p}{\partial x_2} \end{bmatrix} = \begin{bmatrix} h \\ 0 \\ f_1^* \\ f_2^* \end{bmatrix}. \tag{3.10}$$

Theorem 3.2. *Consider the system of PDE* (3.10) *in* $\Omega \times (0, T]$ *where* Ω *is an open subset of* $\mathbf{R}^2$ *with Lipschitz boundary. Then, there is a transformation of the PDE system such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

Proof. Consider the transformation, $v = e^{-\xi t}\left(\eta,\ w_{12},\ \frac{\partial p}{\partial x_1},\ \frac{\partial p}{\partial x_2}\right)^T$ where ξ is some positive constant. With this transformation equation (3.10) becomes

$$\begin{bmatrix} \frac{\partial}{\partial t}+\xi & 0 & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & \epsilon & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} \\ -k\frac{\partial}{\partial x_1} & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_3 & 0 \\ -k\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} & 0 & \alpha_3 \end{bmatrix} \begin{bmatrix} \eta e^{-\xi t} \\ w_{12}e^{-\xi t} \\ \frac{\partial p}{\partial x_1}e^{-\xi t} \\ \frac{\partial p}{\partial x_2}e^{-\xi t} \end{bmatrix} = \begin{bmatrix} h \\ 0 \\ f_1^* \\ f_2^* \end{bmatrix}. \tag{3.11}$$

The corresponding matrices are

$$\mathcal{A}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ \mathcal{A}_{x_1} = \begin{bmatrix} 0 & 0 & -k & 0 \\ 0 & 0 & 0 & \alpha_2 \\ -k & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \end{bmatrix},\ \mathcal{A}_{x_2} = \begin{bmatrix} 0 & 0 & 0 & -k \\ 0 & 0 & -\alpha_2 & 0 \\ 0 & -\alpha_2 & 0 & 0 \\ -k & \alpha_2 & 0 & 0 \end{bmatrix},$$

and

$$\mathbb{A}^0 = \begin{bmatrix} \xi & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_3 \end{bmatrix}.$$

So,

$$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_t\mathcal{A}_t - \partial_{x_1}\mathcal{A}_{x_1} - \partial_{x_2}\mathcal{A}_{x_2} = 2\begin{bmatrix} \xi & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_3 \end{bmatrix}.$$

Now $\mathcal{A}_t,\ \mathcal{A}_{x_1}$ and $\mathcal{A}_{x_2}$ are clearly symmetric. On the other hand, as $\xi,\ \epsilon,\ \alpha_3 > 0$ and $\mathcal{B}$ is positive definite. So, the system is symmetric positive.

An admissible boundary condition is as follows.

Here

$$D = \begin{bmatrix} n_t & 0 & -kn_x & -kn_y \\ 0 & 0 & -\alpha_2 n_y & \alpha_2 n_x \\ -kn_x & -\alpha_2 n_y & 0 & 0 \\ -kn_y & \alpha_2 n_x & 0 & 0 \end{bmatrix}.$$

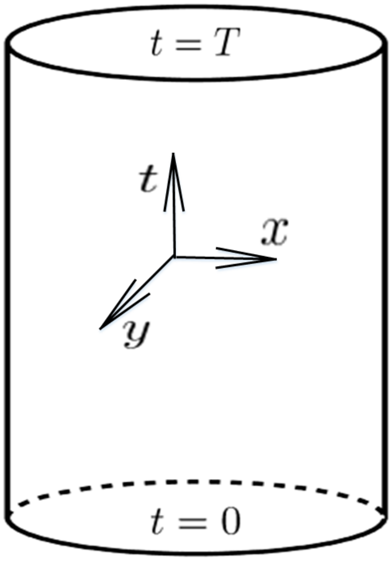


Figure 3.1: A typical domain for equation (3.10)

- On bottom surface $\Omega \times \{0\}$, $n_t = -1$, $n_x = n_y = 0$. So, $D = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, and

  set $M = D$.

  Now, $D - M = 2D$ which implies $\eta = 0$ on bottom surface.

- On top surface $\Omega \times \{T\}$, $n_t = 1$, $n_x = n_y = 0$. So, $D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, and

  set $M = D$. Now, $D - M = 0$ which implies nothing to impose on top surface.

- On curved surface $\partial\Omega \times (0, T]$, $n_t = 0$, $n_x^2 + n_y^2 = 1$. So, $D = \begin{bmatrix} 0 & 0 & -kn_x & -kn_y \\ 0 & 0 & -\alpha_2 n_y & \alpha_2 n_x \\ -kn_x & -\alpha_2 n_y & 0 & 0 \\ -kn_y & \alpha_2 n_x & 0 & 0 \end{bmatrix}$,

  and set $M = \begin{bmatrix} 0 & 0 & -kn_x & -kn_y \\ 0 & 0 & -\alpha_2 n_y & \alpha_2 n_x \\ kn_x & \alpha_2 n_y & 0 & 0 \\ kn_y & -\alpha_2 n_x & 0 & 0 \end{bmatrix}$. Now, $D - M = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2kn_x & -2\alpha_2 n_y & 0 & 0 \\ -2kn_y & 2\alpha_2 n_x & 0 & 0 \end{bmatrix}$,

  which implies

$$\begin{aligned} kn_x\eta + \alpha_2 n_y w_{12} &= 0, \\ kn_y\eta - \alpha_2 n_x w_{12} &= 0. \end{aligned} \tag{3.12}$$

or

$$\begin{aligned} kn_x^2\eta + \alpha_2 n_x n_y w_{12} &= 0, \\ kn_y^2\eta - \alpha_2 n_x n_y w_{12} &= 0. \end{aligned} \tag{3.13}$$

or $\eta k(n_x^2 + n_y^2) = 0 \Rightarrow k\eta = 0 \Rightarrow \eta = 0$ and so, $w_{12} = 0$ on the curved surface.

□

Remark 3.1. *The two dimensional formulation can be easily generalized for higher dimensional case allowing more unknown variables. We note in higher dimension increasing number of pressure gradient ($\frac{\partial p}{\partial x_i}$) and also the rotation vector (w_{ij}). As an example, the three* *dimensional formulation is as follows*

$$
\begin{aligned}
\frac{\partial \eta}{\partial t} - k\frac{\partial}{\partial x_1}\left(\frac{\partial p}{\partial x_1}\right) - k\frac{\partial}{\partial x_2}\left(\frac{\partial p}{\partial x_2}\right) - k\frac{\partial}{\partial x_3}\left(\frac{\partial p}{\partial x_3}\right) &= h,\\
-\alpha_1\frac{\partial}{\partial x_2}\left(\frac{\partial p}{\partial x_1}\right) + \alpha_1\frac{\partial}{\partial x_1}\left(\frac{\partial p}{\partial x_2}\right) &= 0,\\
-\alpha_1\frac{\partial}{\partial x_3}\left(\frac{\partial p}{\partial x_2}\right) + \alpha_1\frac{\partial}{\partial x_2}\left(\frac{\partial p}{\partial x_3}\right) &= 0,\\
-\alpha_1\frac{\partial}{\partial x_1}\left(\frac{\partial p}{\partial x_3}\right) + \alpha_1\frac{\partial}{\partial x_3}\left(\frac{\partial p}{\partial x_1}\right) &= 0,\\
-k\frac{\partial \eta}{\partial x_1} - \alpha_1\left(\frac{\partial w_{12}}{\partial x_2} - \frac{\partial w_{31}}{\partial x_3}\right) + \alpha_2\frac{\partial p}{\partial x_1} &= f_1^*,\\
-k\frac{\partial \eta}{\partial x_2} - \alpha_1\left(\frac{\partial w_{23}}{\partial x_3} - \frac{\partial w_{12}}{\partial x_1}\right) + \alpha_2\frac{\partial p}{\partial x_2} &= f_2^*,\\
-k\frac{\partial \eta}{\partial x_3} - \alpha_1\left(\frac{\partial w_{31}}{\partial x_1} - \frac{\partial w_{23}}{\partial x_2}\right) + \alpha_2\frac{\partial p}{\partial x_3} &= f_3^*,
\end{aligned}
\tag{3.14}
$$

with unknown variables η, w_{12}, w_{23}, w_{31}, $\frac{\partial p}{\partial x_1}$, $\frac{\partial p}{\partial x_2}$ *and* $\frac{\partial p}{\partial x_3}$.

Remark 3.2. *For $d-$dimensional case, we have the number of unknown variables*

- *1 for* η
- $(d-1)+(d-2)+\cdots+1 = \frac{d(d-1)}{2}$ *for* w_{ij}
- d *for* $\frac{\partial p}{\partial x_i}$

So, the total number of unknown is $1+\frac{d(d-1)}{2}+d$, *which is* $1+\frac{d(d+1)}{2}$. *On the other hand, we have one mass conservation equation, $C(d,2)$ pressure gradient equations and d force balance equations, all together* $1+C(d,2)+d = \frac{d(d+1)}{2}$, *same as the number of unknown.*

3.3 Time discretization formulation

The time discretization of equation (3.10) can be readily realized by using finite difference methods. Backward-Euler scheme is used in the following formulation

$$\begin{aligned}\frac{\eta^n - \eta^{n-1}}{\Delta t} - k\frac{\partial p_1^n}{\partial x_1} - k\frac{\partial p_2^n}{\partial x_2} &= h^n,\\ \epsilon w_{12}^n - \alpha_2\frac{\partial p_1^n}{\partial x_2} + \alpha_2\frac{\partial p_2^n}{\partial x_1} &= 0,\\ -k\frac{\partial \eta^n}{\partial x_1} - \alpha_2\frac{\partial w_{12}^n}{\partial x_2} + \alpha_3 p_1^n &= f_1^n,\\ -k\frac{\partial \eta^n}{\partial x_2} + \alpha_2\frac{\partial w_{12}^n}{\partial x_1} + \alpha_3 p_2^n &= f_2^n,\end{aligned} \tag{3.15}$$

where $\Delta t = t^n - t^{n-1}$, $p_1 = \frac{\partial p}{\partial x_1}$ and $p_2 = \frac{\partial p}{\partial x_2}$. We can write it in the form of $Lu = f$ as

$$u = \begin{bmatrix} \eta^n \\ w_{12}^n \\ p_1^n \\ p_2^n \end{bmatrix}, \; Lu = \begin{bmatrix} \eta^n - k\Delta t\frac{\partial p_1^n}{\partial x_1} - k\Delta t\frac{\partial p_2^n}{\partial x_2} \\ \epsilon w_{12}^n - \alpha_2\frac{\partial p_1^n}{\partial x_2} + \alpha_2\frac{\partial p_2^n}{\partial x_1} \\ -k\frac{\partial \eta^n}{\partial x_1} - \alpha_2\frac{\partial w_{12}^n}{\partial x_2} + \alpha_3 p_1^n \\ -k\frac{\partial \eta^n}{\partial x_2} + \alpha_2\frac{\partial w_{12}^n}{\partial x_1} + \alpha_3 p_2^n \end{bmatrix}, \text{ and } f = \begin{bmatrix} h^n\Delta t + \eta^{n-1} \\ 0 \\ f_{11}^n \\ f_{22}^n \end{bmatrix}. \tag{3.16}$$

In matrix form,

$$\begin{bmatrix} 1 & 0 & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & \epsilon & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} \\ -k\frac{\partial}{\partial x_1} & -\alpha_2\frac{\partial}{\partial x_2} & \alpha_3 & 0 \\ -k\frac{\partial}{\partial x_2} & \alpha_2\frac{\partial}{\partial x_1} & 0 & \alpha_3 \end{bmatrix} \begin{bmatrix} \eta^n \\ w_{12}^n \\ p_1^n \\ p_2^n \end{bmatrix} = \begin{bmatrix} h^n\Delta t + \eta^{n-1} \\ 0 \\ f_{11}^n \\ f_{22}^n \end{bmatrix}. \tag{3.17}$$

Theorem 3.3. *Consider the system of PDE* (3.17) *in* $\Omega \times (0, T]$ *where* Ω *is an open subset of* $\mathbb{R}^2$ *with Lipschitz boundary. Then, there is a transformation of the PDE system such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

Proof. Observing the equation (3.17), the corresponding matrices are

$$\mathcal{A}_{x_1} = \begin{bmatrix} 0 & 0 & -k & 0 \\ 0 & 0 & 0 & \alpha_2 \\ -k & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \end{bmatrix}, \ \mathcal{A}_{x_2} = \begin{bmatrix} 0 & 0 & 0 & -k \\ 0 & 0 & -\alpha_2 & 0 \\ 0 & -\alpha_2 & 0 & 0 \\ -k & \alpha_2 & 0 & 0 \end{bmatrix}, \ \mathbb{A}^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_3 \end{bmatrix}.$$

So,

$$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_{x_1}\mathcal{A}_{x_1} - \partial_{x_2}\mathcal{A}_{x_2} = 2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_3 \end{bmatrix}.$$

Now $\mathcal{A}_{x_1}$ and $\mathcal{A}_{x_2}$ are clearly symmetric. On the other hand, as ϵ, $\alpha_3 > 0$, and $\mathcal{B}$ is positive definite, the system is symmetric positive.

An admissible boundary condition is as follows.

Here

$$D = \begin{bmatrix} 0 & 0 & -kn_x & -kn_y \\ 0 & 0 & -\alpha_2 n_y & \alpha_2 n_x \\ -kn_x & -\alpha_2 n_y & 0 & 0 \\ -kn_y & \alpha_2 n_x & 0 & 0 \end{bmatrix}, \text{ set } M = \begin{bmatrix} 0 & 0 & -kn_x & -kn_y \\ 0 & 0 & -\alpha_2 n_y & \alpha_2 n_x \\ kn_x & \alpha_2 n_y & 0 & 0 \\ kn_y & -\alpha_2 n_x & 0 & 0 \end{bmatrix}.$$

So,

$$D - M = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -2kn_x & -2\alpha_2 n_y & 0 & 0 \\ -2kn_y & 2\alpha_2 n_x & 0 & 0 \end{bmatrix},$$

which implies $\eta^n = w_{12}^n = 0$ on the boundary. □

3.4 Error analysis

Theorem 3.4. *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with sufficiently smooth boundary $\partial\Omega$. The boundary value problem* (3.17) *$Lu = f$ in Ω with $\eta^n = w_{12}^n = 0$ on $\partial\Omega$ has a unique solution $u \in [L^2(\Omega)]^4$ for every $f \in [L^2(\Omega)]^4$.*

Proof. By theorem 3.3, L is a symmetric positive operator with an admissible boundary condition $\eta^n = w_{12}^n = 0$ on $\partial\Omega$. Moreover, by theorem 2.3, L is an isomorphism between the proper graph space and $[L^2(\Omega)]^4$. So, the claim follows. □

Theorem 3.5. *Consider the PDE system* (3.17)*. There is a normed subspace $V \subset \mathbf{L} = L^2(\Omega)]^4$ and two positive constants c_1 and c_2 such that $c_1\|u\|_V \leq \|Lu\|_{\mathbf{L}} \leq c_2\|u\|_V$ for every $u \in V$.*

Proof. Define $V = \{u \in \mathbf{L};\, Lu \in \mathbf{L},\, \eta^n = w_{12}^n = 0 \text{ on } \partial\Omega\}$ with inner product $(u,\, v)_V = (u,\, v)_{\mathbf{L}} + (Lu,\, Lv)_{\mathbf{L}}$ for every $u,\, v \in V$. Equipped with this inner product V is a Hilbert space. The induced norm is $\|u\|_V^2 = \|u\|_{\mathbf{L}}^2 + \|Lu\|_{\mathbf{L}}^2$ for all $u \in V$. Now consider the operator

$$L : V \to \mathbf{L}$$

and for all $u \in V$

$$\|u\|_V^2 = \|u\|_{\mathbf{L}}^2 + \|Lu\|_{\mathbf{L}}^2 \Rightarrow \|u\|_V^2 \geq \|Lu\|_{\mathbf{L}}^2 \Rightarrow \|Lu\|_{\mathbf{L}} \leq \|u\|_V$$

So, we can choose $c_2 = 1$. Moreover, as L is a symmetric positive operator, by theorem 2.3, L is an isomorphism between V and L. So, L^{-1} is continuous and hence there is a c_1 such that $c_1\|u\|_V \leq \|Lu\|_{\mathbf{L}}$ □

Theorem 3.6. *Consider the PDE system* (3.17) *and the least square problem, find $u \in V$ such that $(Lu,\, Lv)_{\mathbf{L}} = (f,\, Lv)_{\mathbf{L}}$ for every $v \in V$. Then, the problem has a unique solution.*

Proof. Define a bilinear form, $\tilde{a} : V \times V \to \mathbb{R}$ as

$$\tilde{a}(u,\, v) = (Lu,\, Lv)_{\mathbf{L}} \quad \text{for every} \quad u, v \in V.$$

Due to the last theorem 3.5, $\tilde{a}$ is coercive and continuous. Since, $f \in \mathbf{L}$ and L is a bounded linear operator, $(f,\, L\cdot)_{\mathbf{L}}$ is a continuous form on V. So, the conclusion follows from the Lax-Milgram Lemma. As discussed in LSFEM formulation in Chapter 2, the unique solution minimizes the quadratic functional $E(v) = \|Lv - f\|_{\mathbf{L}}$ for $v \in V$. □

Theorem 3.7. *Consider the PDE system* (3.17)*,* $V_h \subset V$ *be a finite dimensional space and the least square finite element problem, find* $u_h \in V_h$ *such that* $(Lu_h,\, Lv_h)_{\mathbf{L}} = (f,\, Lv_h)_{\mathbf{L}}$ *for every* $v_h \in V_h$*. Then, the problem has a unique solution.*

Proof. Consider the bilinear form, $\tilde{a_h} : V_h \times V_h \to \mathbb{R}$ as $\tilde{a_h}(u_h,\, v_h) = (Lu_h,\, Lv_h)_{\mathbf{L}}$ for every $u_h, v_h \in V_h$. Due to the conformity $V_h \subset V$, $\tilde{a_h}$ is coercive and continuous. Since, $f \in \mathbf{L}$ and L is a bounded linear operator, $(f,\, L\cdot)_{\mathbf{L}}$ is a continuous form on V_h. The conclusion follows from the Lax-Milgram Lemma. □

Consider two problems as follows, referring to PDE system (3.17) and V the graph space of the linear operator L
find $u \in V$ such that

$$(Lu,\, Lv)_{\mathbf{L}} = (f,\, Lv)_{\mathbf{L}} \quad \text{for every} \quad v \in V. \tag{3.18}$$

Let $V_h \subset V$ be a finite dimensional space,
find $u_h \in V_h$ such that

$$(Lu_h,\, Lv_h)_{\mathbf{L}} = (f,\, Lv_h)_{\mathbf{L}} \quad \text{for every} \quad v_h \in V_h. \tag{3.19}$$

Based on this setting, we can state the following theorem

Theorem 3.8. *Let u and u_h be solution of problem* (3.18) *and* (3.19) *respectively. Assume also, $u \in H^{m+1}(\Omega)$ for some integer $m \geq 1$. Then, there is a $c > 0$ such that for every $h > 0$,*

$$\|\eta-\eta_h\|_1+\|w_{12}-w_{12h}\|_1+\|p_1-p_{1h}\|_1+\|p_2-p_{2h}\|_1 \leq ch^m(\|\eta\|_{m+1}+\|w_{12}\|_{m+1}+\|p_1\|_{m+1}+\|p_2\|_{m+1}).$$

Proof. Since $v_h \subset V$

$$(Lu,\, Lv_h) = (f,\, Lv_h) \quad \text{for all} \quad v_h \in V_h,$$

$$(Lu_h,\, Lv_h) = (f,\, Lv_h) \quad \text{for all} \quad v_h \in V_h.$$

Upon subtraction

$$(L(u-u_h),\, Lv_h) = 0 \quad \text{for all} \quad v_h \in V_h.$$

Let $\Pi_h u \in V_h$ be a equal order interpolant of u. Then,

$$\begin{aligned}\|L(u-u_h)\|_{\mathbf{L}}^2 &= (L(u-u_h),\, L(u-u_h)) \\ &= (L(u-u_h),\, L(u-\Pi_h u)) + (L(u-u_h),\, L(\Pi_h u - u_h)) \\ &= (L(u-u_h),\, L(u-\Pi_h u)) \\ &\leq \|L(u-u_h)\|_{\mathbf{L}}\, \|L(u-\Pi_h u)\|_{\mathbf{L}}.\end{aligned}$$

So, $\|L(u-u_h)\|_{\mathbf{L}} \leq \|L(u-\Pi_h u)\|_{\mathbf{L}}$.

By theorem 3.5

$$c_1\|u-u_h\|_V \leq \|L(u-u_h)\|_{\mathbf{L}} \leq \|L(u-u_h)\|_{\mathbf{L}} \leq c_2\|u-\Pi_h u\|_V. \tag{3.20}$$

Choosing $\Pi_h u \in V_h$ such that

$$\|u-\Pi_h u\|_V \leq h^m\left(\|\eta\|_{m+1} + \|w_{12}\|_{m+1} + \|p_1\|_{m+1} + \|p_2\|_{m+1}\right). \tag{3.21}$$

The claim follows from equation (3.20) and (3.21). □

3.5 Numerical solution

The explicit weak form of the least square formulation is as follows.

Find $u \in V$ such that

$$(Lu,\, Lv)_{\mathbf{L}} = (f,\, Lv)_{\mathbf{L}} \quad \text{for all} \quad v \in V,$$

with $u = (\eta^n \;\; w_{12}^n \;\; p_1^n \;\; p_2^n),\;\; v = (\tilde{\eta}^n \;\; \tilde{w}_{12}^n \;\; \tilde{p}_1^n \;\; \tilde{p}_2^n)$ and using (3.16), we have

$$(Lu,\, Lv)_{\mathbf{L}} = \int_\Omega (I + II + III + IV)\, d\Omega,$$

where

$$I = \left(\eta^n - k\Delta t \frac{\partial p_1^n}{\partial x_1} - k\Delta t \frac{\partial p_2^n}{\partial x_2}\right)\left(\tilde{\eta}^n - k\Delta t \frac{\partial \tilde{p}_1^n}{\partial x_1} - k\Delta t \frac{\partial \tilde{p}_2^n}{\partial x_2}\right),$$

$$II = \left(\epsilon w_{12}^n - \alpha_2 \frac{\partial p_1^n}{\partial x_2} + \alpha_2 \frac{\partial p_2^n}{\partial x_1}\right)\left(\epsilon \tilde{w}_{12}^n - \alpha_2 \frac{\partial \tilde{p}_1^n}{\partial x_2} + \alpha_2 \frac{\partial \tilde{p}_2^n}{\partial x_1}\right),$$

$$III = \left(-k\frac{\partial \eta^n}{\partial x_1} - \alpha_2 \frac{\partial w_{12}^n}{\partial x_2} + \alpha_3 p_1^n\right)\left(-k\frac{\partial \tilde{\eta}^n}{\partial x_1} - \alpha_2 \frac{\partial \tilde{w}_{12}^n}{\partial x_2} + \alpha_3 \tilde{p}_1^n\right),$$

$$IV = \left(-k\frac{\partial \eta^n}{\partial x_2} + \alpha_2 \frac{\partial w_{12}^n}{\partial x_1} + \alpha_3 p_2^n\right)\left(-k\frac{\partial \tilde{\eta}^n}{\partial x_2} + \alpha_2 \frac{\partial \tilde{w}_{12}^n}{\partial x_1} + \alpha_3 \tilde{p}_2^n\right).$$

Also,

$$(f,\, Lv)_L = \int_\Omega (A + B + C + D)\, d\Omega,$$

where

$$A = \left(h^n \Delta t + \eta^{n-1}\right)\left(\tilde{\eta}^n - k\Delta t \frac{\partial \tilde{p}_1^n}{\partial x_1} - k\Delta t \frac{\partial \tilde{p}_2^n}{\partial x_2}\right),$$

$$B = 0,$$

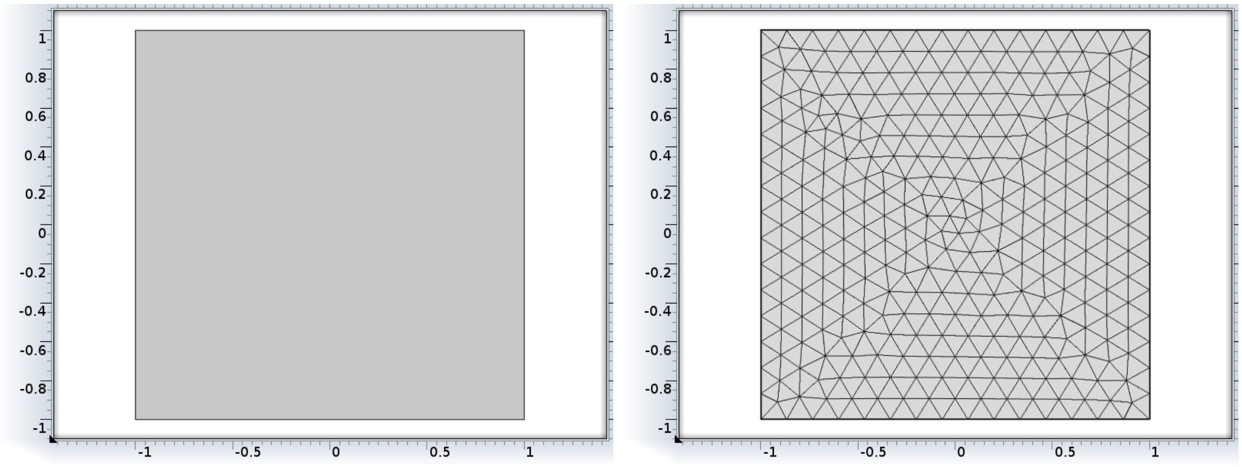
$$C = f_{11}^n \left(-k\frac{\partial \tilde{\eta}^n}{\partial x_1} - \alpha_2 \frac{\partial \tilde{w}_{12}^n}{\partial x_2} + \alpha_3 \tilde{p}_1^n\right),$$

$$D = f_{22}^n \left(-k\frac{\partial \tilde{\eta}^n}{\partial x_2} + \alpha_2 \frac{\partial \tilde{w}_{12}^n}{\partial x_1} + \alpha_3 \tilde{p}_2^n\right).$$

The PDE domain is a square of side 2 with center at (0, 0). The PDEs are supplemented by boundary condition ($\eta = w_{12} = 0$). We also set $\epsilon = 10^{-200}$. We choose $\Delta t = 0.05$, 0.1, 1.0. COMSOL 4.3 Weak form PDE console is used to implement the corresponding weak formulation. In this finite element implementation, 578 elements, 4868 degrees of freedom, Lagrange shape functions with quadratic element order are used.

Parameters	Value	Parameters	Value
$\lambda,\ \mu$	1	f_{11}^n	xy
$k,\ S_\epsilon$	1	f_{22}^n	1
α	0.6	h_1^n	1

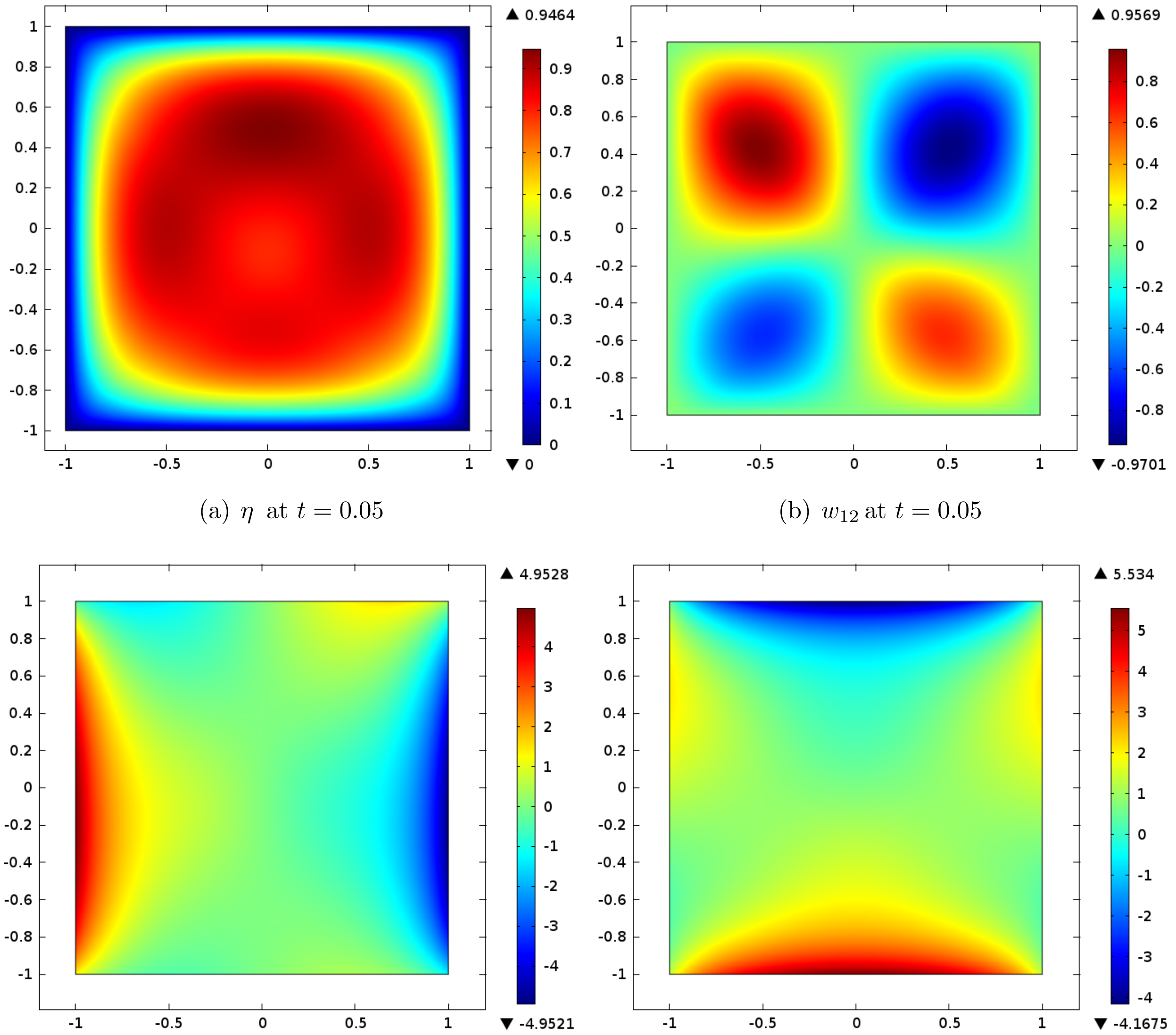
Table 3.1: Different parameters for COMSOL for the first formulation



(a) Physical domain

(b) Meshed domain

Figure 3.2: Domain and its meshing for the first formulation



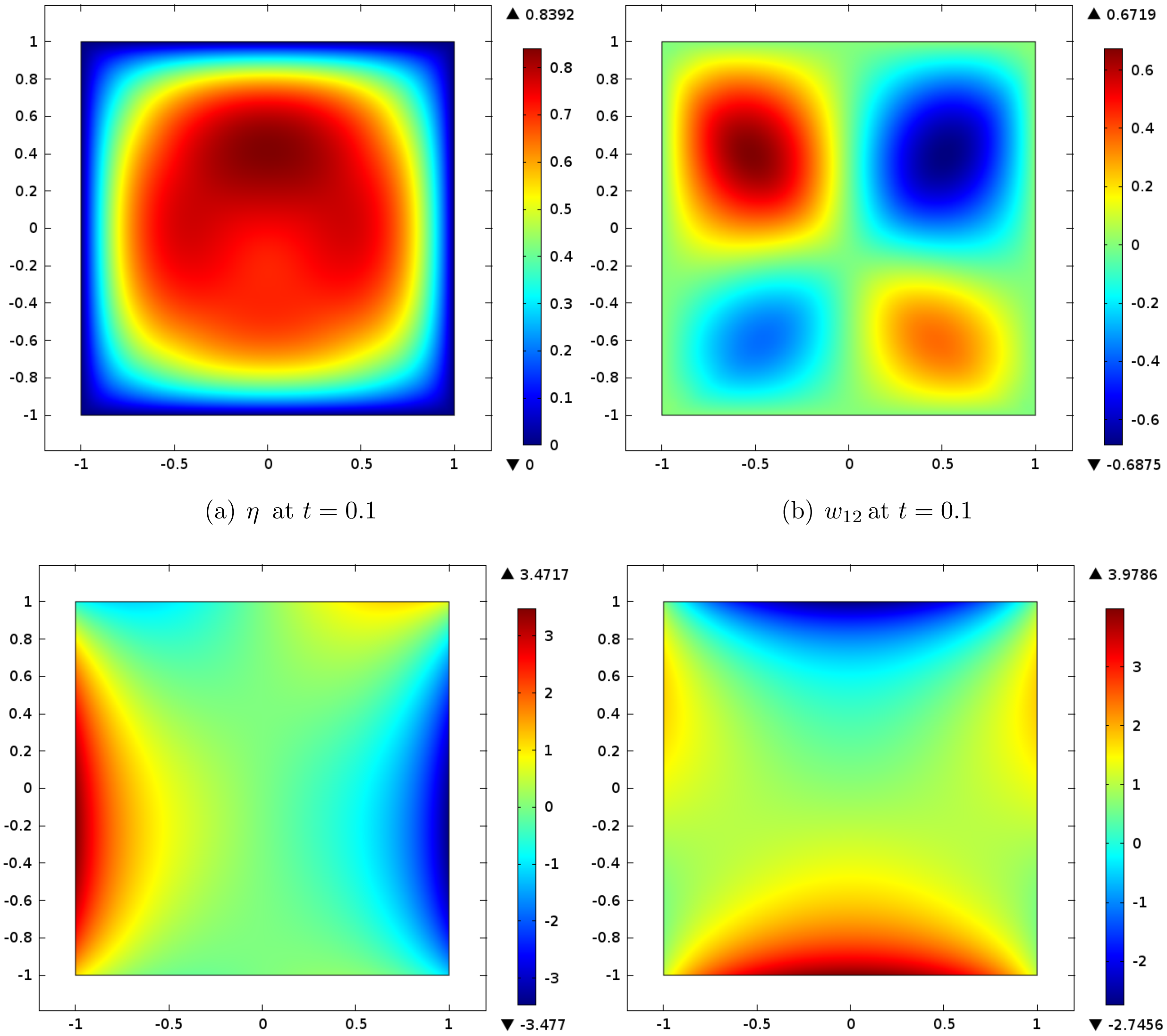
(a) η at $t = 0.05$

(b) w_{12} at $t = 0.05$

(c) p_1 at $t = 0.05$

(d) p_2 at $t = 0.05$

Figure 3.3: Different variables at $t = 0.05$



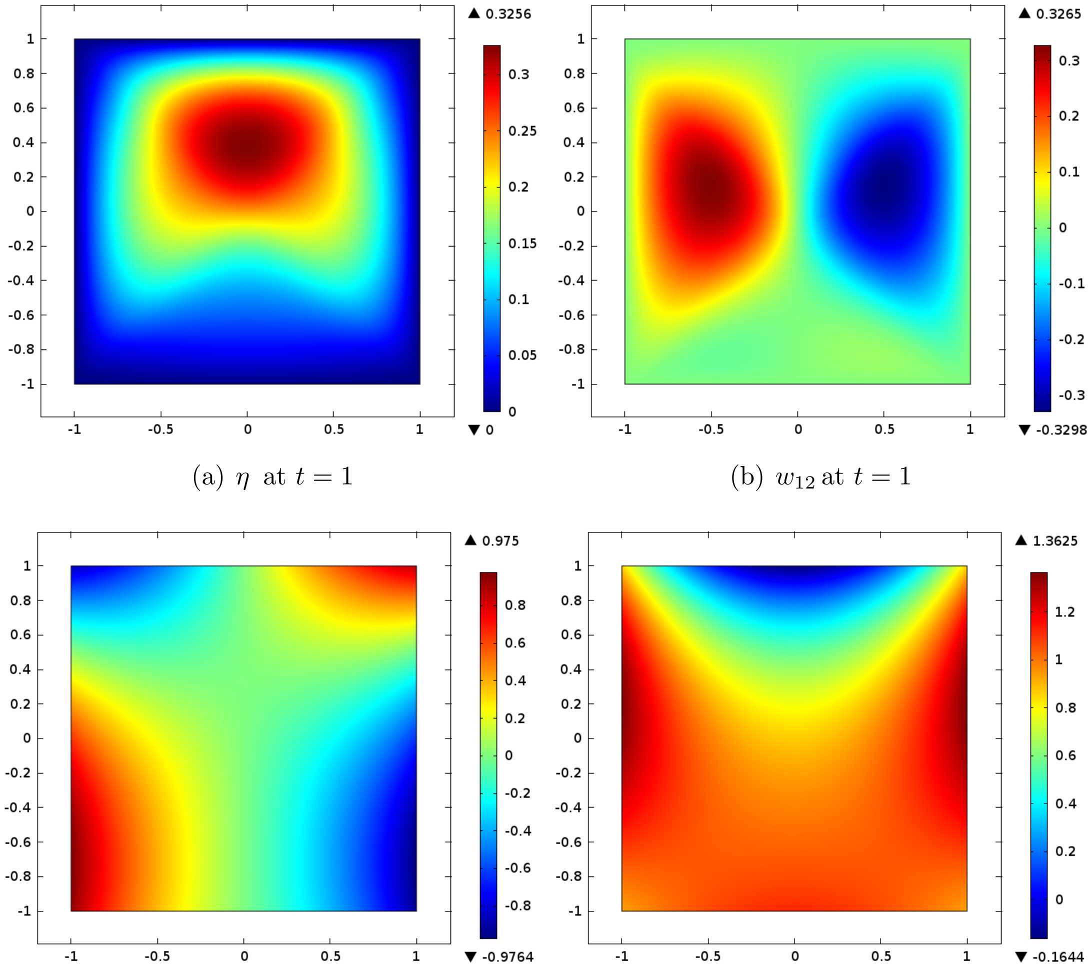
(a) η at $t = 0.1$

(b) w_{12} at $t = 0.1$

(c) p_1 at $t = 0.1$

(d) p_2 at $t = 0.1$

Figure 3.4: Different variables at $t = 0.1$



(a) η at $t = 1$

(b) w_{12} at $t = 1$

(c) p_1 at $t = 1$

(d) p_2 at $t = 1$

Figure 3.5: Different variables at $t = 1$

Chapter 4

Stress-displacement-pressure formulation

We consider the following PDE system

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha p I = 0, \\ -\nabla \cdot \sigma = f, \\ \frac{\partial}{\partial t}\left[S_\epsilon p + \alpha\left(\nabla \cdot u\right)\right] - k\nabla^2 p = h. \end{cases} \tag{4.1}$$

or its equivalent form. Poroelastic equations can be considered as a coupling between linear elasticity equations and heat equation in some sense. In chapter 2, linear elasticity equations and heat equation are discussed. Now, we introduce coupling terms and discuss how the system is a symmetric positive system. It suffices to consider the second equation in (4.1) as in the following form [30, 32]

$$-\nabla \cdot \sigma + \beta u = f.$$

4.1 Completely decoupled system

Consider the following system

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I = 0, \\ -\nabla \cdot \sigma + \beta u = f, \\ \frac{\partial}{\partial t}\left[S_\epsilon p\right] - k\nabla^2 p = h. \end{cases} \tag{4.2}$$

As linear elasticity equations and heat equation are symmetric positive system as proved in chapter 2, and (4.2) is completely decoupled, we have proved the following theorem.

Theorem 4.1. *Consider the system of PDE* (4.2) *in* $\Omega \times (0, T]$ *where* Ω *is an open subset of* $\mathbb{R}^2$ *with Lipschitz boundary. Then, there is a transformation of the PDE system such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

4.2 Simplified coupled system

Consider the following system

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha pI = 0, \\ -\nabla \cdot \sigma + \beta u = f, \\ \frac{\partial}{\partial t}\left[S_\epsilon p\right] - k\nabla^2 p = h. \end{cases} \tag{4.3}$$

This system has some practical importance as noted in [17]. It explains the problems of fluid flow and mechanics are uncoupled when a highly compressible fluid (e.g. air) fills the pore space. In such case, the coupling term σ_{kk} or $\nabla \cdot u$ term approaches zero. Without loss of generality, we set $S_\epsilon = 1$. As in chapter 2, we can write

$$b\sigma_{i,i} - a\sum_{k\neq i}\sigma_{kk} - 2\mu\frac{\partial u_i}{\partial x_i} + p\alpha_1 = 0 \quad \forall i \in \{1,\, 2, \cdots, d\}, \tag{4.4a}$$

$$\sigma_{i,j} - \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) = 0 \quad \forall i,\, j \in \{1,\, 2, \cdots, d\} \text{ with } i \neq j, \tag{4.4b}$$

$$-\frac{\partial \sigma_{ij}}{\partial x_i} + \beta u_j = f_j \quad \forall j \in \{1,\, 2, \cdots, d\}, \tag{4.4c}$$

$$\frac{\partial p}{\partial t} - k\nabla^2 p = h, \tag{4.4d}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$, $b = 1 - a$, $\alpha_1 = \alpha - a\alpha d$ and note that a, b and α_1 are positive constants with $\frac{b}{a} = (d-1) + \frac{2\mu}{\lambda}$ and $\alpha_1 = \frac{2\mu\alpha}{\lambda d + 2\mu}$.

The system for $d = 2$

$$\begin{aligned}
b\sigma_{11} - a\sigma_{22} - 2\mu\frac{\partial u_1}{\partial x_1} + p\alpha_1 &= 0,\\
\sigma_{21} - \mu\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) &= 0,\\
\sigma_{12} - \mu\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) &= 0,\\
b\sigma_{22} - a\sigma_{11} - 2\mu\frac{\partial u_2}{\partial x_2} + p\alpha_1 &= 0,\\
-\frac{\partial\sigma_{11}}{\partial x_1} - \frac{1}{2}\frac{\partial\sigma_{21}}{\partial x_2} - \frac{1}{2}\frac{\partial\sigma_{12}}{\partial x_2} + \beta u_1 &= f_1,\\
-\frac{1}{2}\frac{\partial\sigma_{21}}{\partial x_1} - \frac{1}{2}\frac{\partial\sigma_{12}}{\partial x_1} - \frac{\partial\sigma_{22}}{\partial x_2} + \beta u_2 &= f_2,\\
e^{-t}\frac{\partial p}{\partial t} - ke^{-t}\frac{\partial p_{x_1}}{\partial x_1} - ke^{-t}\frac{\partial p_{x_2}}{\partial x_2} &= h_1,\\
-ke^{-t}\frac{\partial p}{\partial x_1} + ke^{-t}p_{x_1} &= 0,\\
-ke^{-t}\frac{\partial p}{\partial x_2} + ke^{-t}p_{x_2} &= 0.
\end{aligned} \tag{4.5}$$

As in the form of $L\mathbf{u} = \mathbf{f}$

$$L = \left[\begin{array}{c|c|c} B & U_2 & U_3 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} & L_3 \end{array}\right], \quad \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \ \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} p \\ p_{x_1} \\ p_{x_2} \end{bmatrix}. \tag{4.6}$$

Also,

$$B = \begin{bmatrix} b & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & 0 & 0 & b \end{bmatrix}, \ U_2 = \begin{bmatrix} -2\mu\frac{\partial}{\partial x_1} & 0 \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & -2\mu\frac{\partial}{\partial x_2} \end{bmatrix}, \ U_3 = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_1 & 0 & 0 \end{bmatrix}, \tag{4.7}$$

$$L_3 = \begin{bmatrix} e^{-t}\frac{\partial}{\partial t} & -ke^{-t}\frac{\partial}{\partial x_1} & -ke^{-t}\frac{\partial}{\partial x_2} \\ -ke^{-t}\frac{\partial}{\partial x_1} & ke^{-t} & 0 \\ -ke^{-t}\frac{\partial}{\partial x_2} & 0 & ke^{-t} \end{bmatrix}, \ M_2 = \begin{bmatrix} 2\mu\beta & 0 \\ 0 & 2\mu\beta \end{bmatrix}. \tag{4.8}$$

As $\mathbf{f}$ has nothing to do with symmetric positivity, it is often skipped unless numerical results are needed. Here $B = B^T$, $L_3 = L_3^T$ and there is no derivative term in U_3. So, the system is clearly symmetric. Positivity of the system can be found in the following theorem.

Theorem 4.2. *Consider the system of PDE* (4.5) *in* $\Omega \times (0, T]$ *where* Ω *is an open subset of* $\mathbb{R}^2$ *with Lipschitz boundary. Then, there is a transformation of the PDE system such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

Proof. With simple transformation, the system can be written as $L\mathbf{u} = \mathbf{f}$ with

$$L = \left[\begin{array}{c|c|c} B & U_2 & U_3 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} & L_3 \end{array}\right], \ \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \ \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} \frac{p}{\epsilon_1} \\ p_{x_1} \\ p_{x_2} \end{bmatrix}.$$

Also,

$$B = \begin{bmatrix} b & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & 0 & 0 & b \end{bmatrix}, \ U_2 = \begin{bmatrix} -2\mu\frac{\partial}{\partial x_1} & 0 \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & -2\mu\frac{\partial}{\partial x_2} \end{bmatrix}, \ U_3 = \begin{bmatrix} \epsilon_1\alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \epsilon_1\alpha_1 & 0 & 0 \end{bmatrix},$$

$$L_3 = \begin{bmatrix} \epsilon_1^2\epsilon_2 e^{-t}\frac{\partial}{\partial t} & -\epsilon_1\epsilon_2 k e^{-t}\frac{\partial}{\partial x_1} & -\epsilon_1\epsilon_2 k e^{-t}\frac{\partial}{\partial x_2} \\ -\epsilon_1\epsilon_2 k e^{-t}\frac{\partial}{\partial x_1} & \epsilon_2 k e^{-t} & 0 \\ -\epsilon_1\epsilon_2 k e^{-t}\frac{\partial}{\partial x_2} & 0 & \epsilon_2 k e^{-t} \end{bmatrix}, \ M_2 = \begin{bmatrix} 2\mu\beta & 0 \\ 0 & 2\mu\beta \end{bmatrix},$$

where ϵ_1, ϵ_2 are two positive constants. The system is clearly symmetric. The matrix

$\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_t \mathbb{A}^t - \sum_{k=1}^{2} \partial_k \mathbb{A}^k$ as follows

$$\mathcal{B} = \left[\begin{array}{c|c|c} 2B & \mathbb{O} & U_3 \\ \hline \mathbb{O}^T & 2M_2 & \mathbb{O} \\ \hline U_3^T & \mathbb{O}^T & L_4 \end{array}\right], \text{ where } L_4 = \begin{bmatrix} \epsilon_1^2 \epsilon_2 e^{-t} & 0 & 0 \\ 0 & 2\epsilon_2 k e^{-t} & 0 \\ 0 & 0 & 2\epsilon_2 k e^{-t} \end{bmatrix}.$$

Sufficient conditions for $\mathcal{B}$ being positive definite, we need to have $\epsilon_1 \epsilon_2 e^{-T} > 2\alpha_1$ and $2b > 2a + \alpha_1 \epsilon_1$. As $\frac{b}{a} = (d-1) + \frac{2\mu}{\lambda}$ and hence $b > a$, we can easily choose some small ϵ_1 and large ϵ_2 satisfying the conditions. For such ϵ_1 and ϵ_2, the system is clearly symmetric positive. The following admissible boundary condition can be enforced as in the same way for linear elasticity and heat equation in chapter 2.

$$\begin{aligned} u_1 = u_2 = p = 0 \quad &\text{on} \quad \Omega \times \{0\}, \\ u_1 = u_2 = 0 \quad &\text{on} \quad \Omega \times \{T\}, \\ u_1 = u_2 = p = 0 \quad &\text{on} \quad \partial\Omega \times (0, T]. \end{aligned}$$

□

4.3 Time discretization of simplified coupled system

Time discretization of equation (4.3)

$$\begin{aligned} \sigma^n - 2\mu\epsilon^n - \lambda tr(\epsilon^n) I + \alpha p^n I = 0, \\ -\nabla \cdot \sigma^n + \beta u^n = f^n, \\ \frac{S_\epsilon p^n - S_\epsilon p^{n-1}}{\Delta t} - k\nabla^2 p^n = h^n. \end{aligned} \tag{4.9}$$

Without loss of generality, set $S_\epsilon = 1$ and drop the time index, then we have

$$\begin{aligned} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha p I &= 0, \\ -\nabla \cdot \sigma + \beta u &= f, \\ p - k_1 \nabla^2 p &= h_1. \end{aligned} \tag{4.10}$$

The system for $d = 2$

$$\begin{aligned} b\sigma_{11} - a\sigma_{22} - 2\mu \frac{\partial u_1}{\partial x_1} + p\alpha_1 &= 0, \\ \sigma_{21} - \mu \left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2} \right) &= 0, \\ \sigma_{12} - \mu \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) &= 0, \\ b\sigma_{22} - a\sigma_{11} - 2\mu \frac{\partial u_2}{\partial x_2} + p\alpha_1 &= 0, \\ -\frac{\partial \sigma_{11}}{\partial x_1} - \frac{1}{2}\frac{\partial \sigma_{21}}{\partial x_2} - \frac{1}{2}\frac{\partial \sigma_{12}}{\partial x_2} + \beta u_1 &= f_1, \\ -\frac{1}{2}\frac{\partial \sigma_{21}}{\partial x_1} - \frac{1}{2}\frac{\partial \sigma_{12}}{\partial x_1} - \frac{\partial \sigma_{22}}{\partial x_2} + \beta u_2 &= f_2, \\ p - k_1 \frac{\partial p_{x_1}}{\partial x_1} - k_1 \frac{\partial p_{x_2}}{\partial x_2} &= h_1, \\ -k_1 \frac{\partial p}{\partial x_1} + k_1 p_{x_1} &= 0, \\ -k_1 \frac{\partial p}{\partial x_2} + k_1 p_{x_2} &= 0. \end{aligned} \tag{4.11}$$

As in the form of $L\mathbf{u} = \mathbf{f}$ with

$$L = \left[\begin{array}{c|c|c} B & U_2 & U_3 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} & L_3 \end{array} \right], \ \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \ \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} p \\ p_{x_1} \\ p_{x_2} \end{bmatrix}.$$

Also,

$$B = \begin{bmatrix} b & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & 0 & 0 & b \end{bmatrix}, \; U_2 = \begin{bmatrix} -2\mu\frac{\partial}{\partial x_1} & 0 \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & -2\mu\frac{\partial}{\partial x_2} \end{bmatrix}, \; U_3 = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_1 & 0 & 0 \end{bmatrix},$$

$$L_3 = \begin{bmatrix} 1 & -k_1\frac{\partial}{\partial x_1} & -k_1\frac{\partial}{\partial x_2} \\ -k_1\frac{\partial}{\partial x_1} & k_1 & 0 \\ -k_1\frac{\partial}{\partial x_2} & 0 & k_1 \end{bmatrix}, \; M_2 = \begin{bmatrix} 2\mu\beta & 0 \\ 0 & 2\mu\beta \end{bmatrix}.$$

Here $B = B^T$, $L_3 = L_3^T$ and there is no derivative term in U_3. So, the system is clearly symmetric. Positivity of the system can be found in the following theorem.

Theorem 4.3. *Consider the system of PDEs* (4.11) *in* Ω *where* Ω *is an open subset of* $\mathbb{R}^2$ *with Lipschitz boundary. Then, there is a transformation of the PDE system such that the resulting first order system can be formulated as a symmetric positive system. Moreover, there is at least one admissible boundary condition.*

Proof. With simple transformation, the system can be written as $L\mathbf{u} = \mathbf{f}$ with

$$L = \left[\begin{array}{c|c|c} B & U_2 & U_3 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} & L_3 \end{array}\right], \; \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \; \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} \frac{p}{\epsilon_1} \\ p_{x_1} \\ p_{x_2} \end{bmatrix}.$$

Also,

$$B = \begin{bmatrix} b & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & 0 & 0 & b \end{bmatrix}, \; U_2 = \begin{bmatrix} -2\mu\frac{\partial}{\partial x_1} & 0 \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & -2\mu\frac{\partial}{\partial x_2} \end{bmatrix}, \; U_3 = \begin{bmatrix} \epsilon_1\alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \epsilon_1\alpha_1 & 0 & 0 \end{bmatrix},$$

$$L_3 = \begin{bmatrix} \epsilon_1^2\epsilon_2 & -\epsilon_1\epsilon_2 k_1\frac{\partial}{\partial x_1} & -\epsilon_1\epsilon_2 k_1\frac{\partial}{\partial x_2} \\ -\epsilon_1\epsilon_2 k_1\frac{\partial}{\partial x_1} & \epsilon_2 k_1 & 0 \\ -\epsilon_1\epsilon_2 k_1\frac{\partial}{\partial x_2} & 0 & \epsilon_2 k_1 \end{bmatrix}, \ M_2 = \begin{bmatrix} 2\mu\beta & 0 \\ 0 & 2\mu\beta \end{bmatrix},$$

where ϵ_1, ϵ_2 are two positive constants. The system is clearly symmetric. The matrix $\mathcal{B} = \mathbb{A}^0 + (\mathbb{A}^0)^T - \partial_t\mathbb{A}^t - \sum_{k=1}^{2}\partial_k\mathbb{A}^k$ as follows

$$\mathcal{B} = \left[\begin{array}{c|c|c} 2B & \mathbb{O} & U_3 \\ \hline \mathbb{O}^T & 2M_2 & \mathbb{O} \\ \hline U_3^T & \mathbb{O}^T & L_4 \end{array}\right], \text{ where } L_4 = \begin{bmatrix} 2\epsilon_1^2\epsilon_2 & 0 & 0 \\ 0 & 2\epsilon_2 k_1 & 0 \\ 0 & 0 & 2\epsilon_2 k_1 \end{bmatrix}.$$

Sufficient conditions for $\mathcal{B}$ being positive definite, we need to have $\epsilon_1\epsilon_2 > \alpha_1$ and $2b > 2a + \alpha_1\epsilon_1$. As $\frac{b}{a} = (d-1) + \frac{2\mu}{\lambda}$ and hence $b > a$, we can easily choose some small ϵ_1 and large ϵ_2 satisfying the conditions. For such ϵ_1 and ϵ_2, the system is clearly symmetric positive. The following admissible boundary condition can be enforced as in the same way for linear elasticity and heat equation in chapter 2.

$$u_1 = u_2 = p = 0 \quad \text{on} \quad \partial\Omega.$$

□

4.4 Completely coupled system

Consider the following equation

$$\begin{cases} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha p I = 0, \\ -\nabla\cdot\sigma + \beta u = f, \\ \frac{\alpha}{KB}\frac{\partial p}{\partial t} + \frac{\alpha}{dK}\frac{\partial\sigma_{kk}}{\partial t} - k\nabla^2 p = h. \end{cases} \tag{4.12}$$

The system can be written as in chapter 2

$$b\sigma_{i,i} - a\sum_{k\neq i}\sigma_{kk} - 2\mu\frac{\partial u_i}{\partial x_i} + p\alpha_1 = 0 \quad \forall i \in \{1,\, 2, \cdots,\, d\}, \tag{4.13a}$$

$$\sigma_{i,j} - \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) = 0 \quad \forall i,\, j \in \{1,\, 2, \cdots,\, d\} \text{ with } i \neq j, \tag{4.13b}$$

$$-\frac{\partial \sigma_{ij}}{\partial x_i} + \beta u_j = f_j \quad \forall j \in \{1,\, 2, \cdots,\, d\}, \tag{4.13c}$$

$$\frac{\partial}{\partial t}\left[c_1 p + \theta\sigma_{kk}\right] - k\nabla^2 p = h, \tag{4.13d}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$, $b = 1 - a$, $\alpha_1 = \alpha - a\alpha d$, $\frac{b}{a} = (d-1) + \frac{2\mu}{\lambda}$, $\alpha_1 = \frac{2\mu\alpha}{\lambda d + 2\mu}$, $c_1 = \frac{\alpha}{KB}$, $\theta = \frac{\alpha}{dK}$. Here, a, b, and α_1 are positive constants.

4.4.1 Difficulty with time derivative

For $d = 2$, the system in the form of $L\mathbf{u} = \mathbf{f}$ with

$$L = \begin{bmatrix} \begin{array}{c|c|c} B & U_2 & U_3 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline L_1 & \mathbb{O} & L_3 \end{array} \end{bmatrix}, \; \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \; \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} p \\ p_{x_1} \\ p_{x_2} \end{bmatrix}.$$

Also,

$$B = \begin{bmatrix} b & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & 0 & 0 & b \end{bmatrix}, \; U_2 = \begin{bmatrix} -2\mu\frac{\partial}{\partial x_1} & 0 \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} \\ 0 & -2\mu\frac{\partial}{\partial x_2} \end{bmatrix}, \; U_3 = \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_1 & 0 & 0 \end{bmatrix},$$

$$L_1 = \begin{bmatrix} \theta\frac{\partial}{\partial t} & 0 & 0 & \theta\frac{\partial}{\partial t} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ L_3 = \begin{bmatrix} c_1\frac{\partial}{\partial t} & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ -k\frac{\partial}{\partial x_1} & k & 0 \\ -k\frac{\partial}{\partial x_2} & 0 & k \end{bmatrix},\ M_2 = \begin{bmatrix} 2\mu\beta & 0 \\ 0 & 2\mu\beta \end{bmatrix}.$$

In complete matrix form, we can write

$$\begin{bmatrix} b & 0 & 0 & -a & -2\mu\frac{\partial}{\partial x_1} & 0 & \alpha_1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} & 0 & 0 & 0 \\ -a & 0 & 0 & b & 0 & -2\mu\frac{\partial}{\partial x_2} & \alpha_1 & 0 & 0 \\ -2\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta & 0 & 0 & 0 & 0 \\ 0 & -\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_1} & -2\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta & 0 & 0 & 0 \\ \theta\frac{\partial}{\partial t} & 0 & 0 & \theta\frac{\partial}{\partial t} & 0 & 0 & c_1\frac{\partial}{\partial t} & -k\frac{\partial}{\partial x_1} & -k\frac{\partial}{\partial x_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & -k\frac{\partial}{\partial x_1} & k & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k\frac{\partial}{\partial x_2} & 0 & k \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \\ u_1 \\ u_2 \\ p \\ p_{x_1} \\ p_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ f_1 \\ f_2 \\ h \\ 0 \\ 0 \end{bmatrix}.$$

So, the system is not symmetric due to the presence of $\theta\frac{\partial}{\partial t}$ terms in the coefficient matrix. Even with row or column manipulation, it seems very difficult to make it symmetric. There is a chance to make it symmetric if we can get rid of time derivative in the matrix. So, we turn into time discretization formulation to make it symmetric and hopefully positive later on.

4.4.2 Method without scaling

Time discretization of equation (4.12)

$$\begin{aligned} \sigma^n - 2\mu\epsilon^n - \lambda tr(\epsilon^n)I + \alpha p^n I = 0, \\ -\nabla \cdot \sigma^n + \beta u^n = f^n, \\ \frac{1}{\Delta t}\left[\frac{\alpha}{KB}(p^n - p^{n-1}) + \frac{\alpha}{dK}(\sigma_{kk}^n - \sigma_{kk}^{n-1})\right] - k\nabla^2 p^n = h^n. \end{aligned} \tag{4.14}$$

For notational simplicity, we drop the time index and after algebraic manipulation, then we have

$$\begin{aligned} \sigma - 2\mu\epsilon - \lambda tr(\epsilon)I + \alpha p I = 0, \\ -\nabla \cdot \sigma + \beta u = f, \\ \frac{\alpha}{B}p + \frac{\alpha}{d}\sigma_{kk} - k_2\nabla^2 p = h_2. \end{aligned} \tag{4.15}$$

The system for $d = 2$

$$\begin{aligned} b\sigma_{11} - a\sigma_{22} - 2\mu\frac{\partial u_1}{\partial x_1} + p\alpha_1 = 0, \\ \sigma_{21} - \mu\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) = 0, \\ \sigma_{12} - \mu\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) = 0, \\ b\sigma_{22} - a\sigma_{11} - 2\mu\frac{\partial u_2}{\partial x_2} + p\alpha_1 = 0, \\ -\frac{\partial \sigma_{11}}{\partial x_1} - \frac{1}{2}\frac{\partial \sigma_{21}}{\partial x_2} - \frac{1}{2}\frac{\partial \sigma_{12}}{\partial x_2} + \beta u_1 = f_1, \\ -\frac{1}{2}\frac{\partial \sigma_{21}}{\partial x_1} - \frac{1}{2}\frac{\partial \sigma_{12}}{\partial x_1} - \frac{\partial \sigma_{22}}{\partial x_2} + \beta u_2 = f_2, \\ \frac{\alpha}{B}p + \frac{\alpha}{d}\sigma_{kk} - k_2\frac{\partial p_{x_1}}{\partial x_1} - k_2\frac{\partial p_{x_2}}{\partial x_2} = h_2, \\ -k_2\frac{\partial p}{\partial x_1} + k_2 p_{x_1} = 0, \\ -k_2\frac{\partial p}{\partial x_2} + k_2 p_{x_2} = 0. \end{aligned} \tag{4.16}$$

As in the form of $L\mathbf{u} = \mathbf{f}$, we can write

$$\begin{bmatrix} b & 0 & 0 & -a & -2\mu\frac{\partial}{\partial x_1} & 0 & \alpha_1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_1} & 0 & 0 & 0 \\ -a & 0 & 0 & b & 0 & -2\mu\frac{\partial}{\partial x_2} & \alpha_1 & 0 & 0 \\ -2\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_2} & -\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta & 0 & 0 & 0 & 0 \\ 0 & -\mu\frac{\partial}{\partial x_1} & -\mu\frac{\partial}{\partial x_1} & -2\mu\frac{\partial}{\partial x_2} & 0 & 2\mu\beta & 0 & 0 & 0 \\ \frac{\alpha}{2} & 0 & 0 & \frac{\alpha}{2} & 0 & 0 & \frac{\alpha}{B} & -k_2\frac{\partial}{\partial x_1} & -k_2\frac{\partial}{\partial x_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_2\frac{\partial}{\partial x_1} & k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_2\frac{\partial}{\partial x_2} & 0 & k_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \\ u_1 \\ u_2 \\ p \\ p_{x_1} \\ p_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ f_1 \\ f_2 \\ h \\ 0 \\ 0 \end{bmatrix}.$$

The system is clearly symmetric. The matrix $\mathcal{B}$ is as follows

$$\mathcal{B} = \begin{bmatrix} 2b & 0 & 0 & -2a & 0 & 0 & \alpha_1 + \frac{\alpha}{2} & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2a & 0 & 0 & 2b & 0 & 0 & \alpha_1 + \frac{\alpha}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 4\mu\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4\mu\beta & 0 & 0 & 0 \\ \alpha_1 + \frac{\alpha}{2} & 0 & 0 & \alpha_1 + \frac{\alpha}{2} & 0 & 0 & 2\frac{\alpha}{B} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2k_2 \end{bmatrix}.$$

Sufficient condition for $\mathcal{B}$ to be a positive definite matrix

$$\begin{aligned} 2b &> 2a + \frac{\mu\alpha}{\lambda+\mu} + \frac{\alpha}{2}, \\ \frac{\alpha}{B} &> \frac{\mu\alpha}{\lambda+\mu} + \frac{\alpha}{2}. \end{aligned} \tag{4.17}$$

For the d dimensional case, similar analysis indicates $\mathcal{B}$ will be positive definite if

$$\begin{aligned} 2b &> 2a(d-1) + \alpha\frac{2\mu}{\lambda d + 2\mu} + \frac{\alpha}{d}, \\ 2\frac{\alpha}{B} &> d\left(\alpha\frac{2\mu}{\lambda d + 2\mu} + \frac{\alpha}{d}\right). \end{aligned} \tag{4.18}$$

Algebraic manipulation of (4.18) gives

$$\begin{aligned} 2 &> \alpha\left(1 + \frac{1}{d} + \frac{\lambda}{2\mu}\right), \\ \frac{1}{B} &> \frac{\mu d}{\lambda d + 2\mu} + \frac{1}{2}. \end{aligned} \tag{4.19}$$

Now, we can state the following theorem

Theorem 4.4. *Consider the system of PDE* (4.16) *or its equivalent three dimensional formulation in* Ω *where* Ω *is an open subset of* $\mathbb{R}^2$ *or* $\mathbb{R}^3$ *with Lipschitz boundary. Then, the PDE system is symmetric positive if equation* (4.19) *holds. Moreover, there is at least one admissible boundary condition.*

Proof. Sufficient conditions (4.19) have already been established. It is easy to see $u_1 = u_2 = p = 0$ on $\partial\Omega$ is an admissible boundary condition. □

We wish to check if equation (4.19) is satisfied for more realistic situation (three dimensional case) such as poroelastic system in geophysical application. For $d = 3$, equation (4.19) becomes

$$\begin{aligned} 2 &> \alpha\left(1 + \frac{1}{3} + \frac{\lambda}{2\mu}\right), \\ \frac{1}{B} &> \frac{3\mu}{3\lambda + 2\mu} + \frac{1}{2}. \end{aligned} \tag{4.20}$$

The following table lists different physical properties of some rocks. With those data, we try to check whether equation (4.20) is satisfied or not.

Rock	μ (GPa)	K (GPa)	α	B	λ (GPa)
Berea sandstone 1	6.0	8.0	0.79	0.62	4.0
Boise sandstone	4.2	4.6	0.85	0.5	1.8
Ohio sandstone	6.8	8.4	0.74	0.5	3.867
Pecos sandstone	5.9	6.7	0.83	0.61	2.767
Ruhr sandstone	13	13	0.65	0.88	4.333
Weber sandstone	12	13	0.64	0.73	5
Tennessee marble	24	40	0.19	0.51	24
Charcoal granite	19	35	0.27	0.55	22.333
Westerly granite	15	25	0.47	0.85	15
Berea sandstone 2	5.6	6.6	0.77	0.75	2.867
Indiana limestone	12.1	21.2	0.71	0.46	13.133

Table 4.1: Physical constants of Rocks as found in [17]

We find λ as $\lambda = K - \frac{2\mu}{3}$ as in [17].

The following table summarizes the results.

Rock	$\alpha(1+\frac{1}{3}+\frac{\lambda}{2\mu})$	$\frac{3\mu}{3\lambda+2\mu}+\frac{1}{2}$	$\frac{1}{B}$	Positive?
Berea sandstone	1.317	1.25	1.613	Yes
Boise sandstone	1.315	1.413	2	Yes
Ohio sandstone	1.197	1.309	2	Yes
Pecos sandstone	1.301	1.381	1.639	Yes
Ruhr sandstone	0.975	1.5	1.136	No
Weber sandstone	0.987	1.423	1.369	No
Tennessee marble	0.348	1.1	1.961	Yes
Charcoal granite	0.519	1.043	1.818	Yes
Westerly granite	0.862	1.1	1.176	Yes
Berea sandstone 2	1.224	1.348	1.333	No
Indiana limestone	1.332	1.071	2.174	Yes

Table 4.2: Positivity of different poroelastic system without scaling

Here, positivity means being the system as symmetric positive. So, Yes on positive indicates our system is symmetric positive satisfying equation (4.20).

4.4.3 Methods using scaling

We do not have symmetric positive system for Ruhr sandstone, Weber sandstone and Berea sandstone 2. In order to have more control and flexibility, we may scale (row multiplication or column multiplication) some of the equations of (4.16)

For $d = 2$, our system as in the form of $L\mathbf{u} = \mathbf{f}$ $(B, U_2, M_2$ as in (4.3) and (4.3))

$$L = \left[\begin{array}{c|c|c} B & U_2 & U_4 \\ \hline U_2^T & M_2 & \mathbb{O} \\ \hline L_1 & \mathbb{O} & L_3 \end{array}\right], \ \mathbf{u} = \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{u} \\ \boldsymbol{P} \end{bmatrix}, \text{ where } \boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}, \ \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \text{ and } \boldsymbol{P} = \begin{bmatrix} \frac{p}{\epsilon_1} \\ \frac{p_{x_1}}{\epsilon_1} \\ \frac{p_{x_2}}{\epsilon_1} \end{bmatrix},$$

$$U_4 = \begin{bmatrix} \alpha_1\epsilon_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_1\epsilon_1 & 0 & 0 \end{bmatrix}, \ L_3 = \begin{bmatrix} \epsilon_1\epsilon_2\frac{\alpha}{B} & -\epsilon_1\epsilon_2 k_2\frac{\partial}{\partial x_1} & -\epsilon_1\epsilon_2 k_2\frac{\partial}{\partial x_2} \\ -\epsilon_1\epsilon_2 k_2\frac{\partial}{\partial x_1} & \epsilon_1\epsilon_2 k_2 & 0 \\ -\epsilon_1\epsilon_2 k_2\frac{\partial}{\partial x_2} & 0 & \epsilon_1\epsilon_2 k_2 \end{bmatrix}, \ L_1 = \begin{bmatrix} \epsilon_2\frac{\alpha}{2} & 0 & 0 & \epsilon_2\frac{\alpha}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here $B = B^T$, $L_3 = L_3^T$, $M_2 = M_2^T$ and there is no derivative term in U_4 and L_1. So, the system is clearly symmetric. The system is positive if the following matrix is positive definite.

$$\mathcal{B} = \begin{bmatrix} 2b & 0 & 0 & -2a & 0 & 0 & \alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2a & 0 & 0 & 2b & 0 & 0 & \alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4\mu\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4\mu\beta & 0 & 0 & 0 \\ \alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2 & 0 & 0 & \alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2 & 0 & 0 & 2\epsilon_1\epsilon_2\frac{\alpha}{B} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\epsilon_1\epsilon_2 k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\epsilon_1\epsilon_2 k_2 \end{bmatrix}.$$

Sufficient conditions for being positive definite are as follows

$$\begin{aligned} 2b &> 2a + \alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2, \\ 2\epsilon_1\epsilon_2\frac{\alpha}{B} &> 2(\alpha_1\epsilon_1 + \frac{\alpha}{2}\epsilon_2). \end{aligned} \tag{4.21}$$

Setting the values of a, b and α_1 in terms of λ, μ and α, we need to have

$$\begin{aligned} 2 &> \alpha\left[\epsilon_1 + \epsilon_2(\frac{1}{2} + \frac{\lambda}{2\mu})\right], \\ \frac{1}{B} &> \frac{\mu}{\epsilon_2(\lambda+\mu)} + \frac{1}{2\epsilon_1}. \end{aligned} \tag{4.22}$$

Now, we can extend this to the $d-$dimensional case. For $d-$dimensions, sufficient conditions are

$$\begin{aligned} 2b &> 2a(d-1) + \alpha_1\epsilon_1 + \frac{\alpha}{d}\epsilon_2, \\ 2\epsilon_1\epsilon_2\frac{\alpha}{B} &> d\left(\alpha_1\epsilon_1 + \frac{\alpha}{d}\epsilon_2\right). \end{aligned} \tag{4.23}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$, $b = 1 - a$ and $\alpha_1 = \frac{2\mu}{\lambda d + 2\mu}$. After algebraic manipulation, sufficient conditions are

$$\begin{aligned} 2 &> \alpha\left[\epsilon_1 + \epsilon_2(\frac{1}{d} + \frac{\lambda}{2\mu})\right], \\ \frac{1}{B} &> \frac{\mu d}{\epsilon_2(\lambda d+2\mu)} + \frac{1}{2\epsilon_1}. \end{aligned} \tag{4.24}$$

So, we have proved the following theorem.

Theorem 4.5. *Consider the system of PDE* (4.16) *or its equivalent three dimensional formulation in* Ω *where* Ω *is an open subset of* $\mathbb{R}^2$ *or* $\mathbb{R}^3$ *with Lipschitz boundary. Then, the PDE system is symmetric positive if equation* (4.24) *holds for some* $\epsilon_1,\ \epsilon_2 > 0$. *Moreover, there is at least one admissible boundary condition.*

Now, there are two controlling parameters $\epsilon_1,\ \epsilon_2$ to choose, so that equation (4.24) is satisfied. In fact, equation (4.19) is a special case of equation (4.24) with $\epsilon_1 = \epsilon_2 = 1$. Although $\epsilon_1,\ \epsilon_2$ are competing in nature, still there are many different options. Based on the previous geophysical data, we can write the equations as a symmetric positive system. Using scaling technique, the systems corresponding to Ruhr sandstone, Weber sandstone are symmetric positive, whereas without scaling they are not, as shown in Table 4.2. In the three dimensional case, equation (4.24) becomes

$$\begin{aligned} 2 &> \alpha\left[\epsilon_1 + \epsilon_2(\frac{1}{3} + \frac{\lambda}{2\mu})\right], \\ \frac{1}{B} &> \frac{3\mu}{\epsilon_2(3\lambda + 2\mu)} + \frac{1}{2\epsilon_1}. \end{aligned} \tag{4.25}$$

The following table shows the results of the scaling method.

Rock	ϵ_1	ϵ_2	$\alpha[\epsilon_1 + \epsilon_2(\frac{1}{3} + \frac{\lambda}{2\mu})]$	$\frac{3\mu}{\epsilon_2(3\lambda+2\mu)} + \frac{1}{2\epsilon_1}$	$\frac{1}{B}$	Positive?
Berea sandstone 1	1	1	1.317	1.25	1.613	Yes
Boise sandstone	1	1	1.315	1.413	2	Yes
Ohio sandstone	1	1	1.197	1.309	2	Yes
Pecos sandstone	1	1	1.301	1.381	1.639	Yes
Ruhr sandstone	2	2	1.95	0.75	1.136	Yes
Weber sandstone	2.5	1	1.945	1.123	1.369	Yes
Tennessee marble	1	1	0.348	1.1	1.961	Yes
Charcoal granite	1	1	0.519	1.043	1.818	Yes
Westerly granite	1	1	0.862	1.1	1.176	Yes
Berea sandstone 2	2	1	1.994	1.098	1.333	Yes
Indiana limestone	1	1	1.332	1.071	2.174	Yes

Table 4.3: Positivity of different poroelastic systems with scaling

4.5 Least square formulation

Consider the following two dimensional system

$$\begin{aligned} \sigma - \lambda(\nabla.u)1_d - \mu(\nabla u + \nabla u^T) + \alpha(1_d)p = 0, \\ -\frac{1}{2}\nabla \cdot (\sigma + \sigma^T) + \beta u = f, \\ \frac{\partial}{\partial t}\left[\frac{\alpha}{K}\frac{\sigma_{kk}}{2} + \frac{\alpha p}{KB}\right] - k\nabla^2 p = h. \end{aligned} \tag{4.26}$$

where $\beta_1,\ \beta_2 > 0$. After algebraic manipulation, we end up with the system as

$$\begin{aligned} b\sigma_{i,i} - a\sum_{k\neq i}\sigma_{kk} - 2\mu\frac{\partial u_i}{\partial x_i} + p\alpha_1 = 0 \quad \forall i \in \{1,\, 2, \cdots,\, d\}, \\ \sigma_{i,j} - \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) = 0 \quad \forall i,\, j \in \{1,\, 2, \cdots,\, d\} \text{ with } i \neq j, \\ -\frac{\partial \sigma_{ij}}{\partial x_i} + \beta u_j = f_j \quad \forall j \in \{1,\, 2, \cdots,\, d\}, \\ \frac{\partial}{\partial t}\left[\frac{\alpha}{K}\frac{\sigma_{kk}}{2} + \frac{\alpha p}{KB}\right] - k\nabla^2 p = h, \end{aligned} \tag{4.27}$$

where $a = \frac{\lambda}{\lambda d + 2\mu}$, $b = 1 - a = \frac{\lambda(d-1)+2\mu}{\lambda d + 2\mu}$, $\alpha_1 = \alpha - a\alpha d = \frac{2\mu\alpha}{\lambda d + 2\mu}$, and $d = 2$. The time discretization can be realized by using the Backward-Euler scheme.

$$\begin{aligned} b\sigma_{11}^n - a\sigma_{22}^n - 2\mu\frac{\partial u_1^n}{\partial x_1} + \alpha_1 p^n = 0, \\ \sigma_{21}^n - \mu(\frac{\partial u_1^n}{\partial x_2} + \frac{\partial u_2^n}{\partial x_1}) = 0, \\ \sigma_{12}^n - \mu(\frac{\partial u_2^n}{\partial x_1} + \frac{\partial u_1^n}{\partial x_2}) = 0, \\ b\sigma_{22}^n - a\sigma_{11}^n - 2\mu\frac{\partial u_2^n}{\partial x_2} + \alpha_1 p^n = 0, \\ -\frac{\partial \sigma_{11}^n}{\partial x_1} - \frac{\partial \sigma_{21}^n}{\partial x_2} + \beta u_1^n = f_1^n, \\ -\frac{\partial \sigma_{12}^n}{\partial x_1} - \frac{\partial \sigma_{22}^n}{\partial x_2} + \beta u_2^n = f_2^n, \end{aligned}$$

$$
\begin{aligned}
\frac{1}{\Delta t}\left[\frac{\alpha}{2}(\sigma_{11}^n+\sigma_{22}^n)+\frac{\alpha p^n}{B}-\frac{\alpha}{2}(\sigma_{11}^{n-1}+\sigma_{22}^{n-1})-\frac{\alpha p^{n-1}}{B}\right]-kK\frac{\partial p_1^n}{\partial x_1}-kK\frac{\partial p_2^n}{\partial x_2}&=h_1^n,\\
-\frac{\partial p^n}{x_1}+p_1^n&=0,\\
-\frac{\partial p^n}{x_2}+p_2^n&=0,
\end{aligned}
$$

where $\Delta t = t^n - t^{n-1}$.

We can write as in $Lu = f$ as follows where the unknown variables are σ_{11}, σ_{21}, σ_{12}, σ_{22}, u_1, u_2, p, p_1, p_2 and

$$
u=\begin{bmatrix}\sigma_{11}^n\\ \sigma_{21}^n\\ \sigma_{12}^n\\ \sigma_{22}^n\\ u_1^n\\ u_2^n\\ p^n\\ p_1^n\\ p_2^n\end{bmatrix},\ Lu=\begin{bmatrix}b\sigma_{11}^n-a\sigma_{22}^n-2\mu\frac{\partial u_1^n}{\partial x_1}+\alpha_1 p^n\\ \sigma_{21}^n-\mu(\frac{\partial u_1^n}{\partial x_2}+\frac{\partial u_2^n}{\partial x_1})\\ \sigma_{12}^n-\mu(\frac{\partial u_2^n}{\partial x_1}+\frac{\partial u_1^n}{\partial x_2})\\ b\sigma_{22}^n-a\sigma_{11}^n-2\mu\frac{\partial u_2^n}{\partial x_2}+\alpha_1 p^n\\ -\frac{\partial \sigma_{11}^n}{\partial x_1}-\frac{\partial \sigma_{21}^n}{\partial x_2}+\beta u_1^n\\ -\frac{\partial \sigma_{12}^n}{\partial x_1}-\frac{\partial \sigma_{22}^n}{\partial x_2}+\beta u_2^n\\ \frac{\alpha}{2}(\sigma_{11}^n+\sigma_{22}^n)+\frac{\alpha p^n}{B}-kK\Delta t\frac{\partial p_1^n}{\partial x_1}-kK\Delta t\frac{\partial p_2^n}{\partial x_2}\\ -\frac{\partial p^n}{x_1}+p_1^n\\ -\frac{\partial p^n}{x_2}+p_2^n\end{bmatrix},\ \text{and } f=\begin{bmatrix}0\\0\\0\\0\\f_1^n\\f_2^n\\h_2^n\\0\\0\end{bmatrix}, \quad (4.28)
$$

where $h_2^n = h^n K\Delta t+\frac{\alpha}{2}(\sigma_{11}^{n-1}+\sigma_{22}^{n-1})+\frac{\alpha p^{n-1}}{B}$. The Backward-Euler method is unconditionally stable with first order accuracy $\mathcal{O}(\Delta t)$ and thus can be used to find solution with any time step.

For the least square formulation

$$(Lu,\, Lv)_{\mathbf{L}} = (f,\, Lv)_{\mathbf{L}}$$

where

$$u = (\sigma \quad u_1 \quad u_2 \quad p \quad p_1 \quad p_2)^T \quad \text{and} \quad v = (\tilde{\sigma} \quad \tilde{u}_1 \quad \tilde{u}_2 \quad \tilde{p} \quad \tilde{p}_1 \quad \tilde{p}_2)^T$$

Using (4.28), we have

$$(Lu,\, Lv)_{\mathbf{L}} = \int_\Omega (B_1 + B_2 + B_3 + B_4 + B_5 + B_6 + B_7 + B_8 + B_9)\, d\Omega,$$

where

$$B_1 = \left(b\sigma_{11}^n - a\sigma_{22}^n - 2\mu\frac{\partial u_1^n}{\partial x_1} + \alpha_1 p^n\right)\left(b\tilde{\sigma}_{11}^n - a\tilde{\sigma}_{22}^n - 2\mu\frac{\partial \tilde{u}_1^n}{\partial x_1} + \alpha_1 \tilde{p}^n\right),$$

$$B_2 = \left(\sigma_{21}^n - \mu(\frac{\partial u_1^n}{\partial x_2} + \frac{\partial u_2^n}{\partial x_1})\right)\left(\tilde{\sigma}_{21}^n - \mu(\frac{\partial \tilde{u}_1^n}{\partial x_2} + \frac{\partial \tilde{u}_2^n}{\partial x_1})\right),$$

$$B_3 = \left(\sigma_{12}^n - \mu(\frac{\partial u_2^n}{\partial x_1} + \frac{\partial u_1^n}{\partial x_2})\right)\left(\tilde{\sigma}_{12}^n - \mu(\frac{\partial \tilde{u}_2^n}{\partial x_1} + \frac{\partial \tilde{u}_1^n}{\partial x_2})\right),$$

$$B_4 = \left(b\sigma_{22}^n - a\sigma_{11}^n - 2\mu\frac{\partial u_2^n}{\partial x_2} + \alpha_1 p^n\right)\left(b\tilde{\sigma}_{22}^n - a\tilde{\sigma}_{11}^n - 2\mu\frac{\partial \tilde{u}_2^n}{\partial x_2} + \alpha_1 \tilde{p}^n\right),$$

$$B_5 = \left(-\frac{\partial \sigma_{11}^n}{\partial x_1} - \frac{\partial \sigma_{21}^n}{\partial x_2} + \beta u_1^n\right)\left(-\frac{\partial \tilde{\sigma}_{11}^n}{\partial x_1} - \frac{\partial \tilde{\sigma}_{21}^n}{\partial x_2} + \beta \tilde{u}_1^n\right),$$

$$B_6 = \left(-\frac{\partial \sigma_{12}^n}{\partial x_1} - \frac{\partial \sigma_{22}^n}{\partial x_2} + \beta u_2^n\right)\left(-\frac{\partial \tilde{\sigma}_{12}^n}{\partial x_1} - \frac{\partial \tilde{\sigma}_{22}^n}{\partial x_2} + \beta \tilde{u}_2^n\right),$$

$$B_7 = \left(\frac{\alpha}{2}(\sigma_{11}^n + \sigma_{22}^n) + \frac{\alpha p^n}{B} - k^*\frac{\partial p_1^n}{\partial x_1} - k^*\frac{\partial p_2^n}{\partial x_2}\right)\left(\frac{\alpha}{2}(\tilde{\sigma}_{11}^n + \tilde{\sigma}_{22}^n) + \frac{\alpha \tilde{p}^n}{B} - k^*\frac{\partial \tilde{p}_1^n}{\partial x_1} - k^*\frac{\partial \tilde{p}_2^n}{\partial x_2}\right),$$

$$B_8 = \left(-\frac{\partial p^n}{x_1} + p_1^n\right)\left(-\frac{\partial \tilde{p}^n}{x_1} + \tilde{p}_1^n\right),$$

$$B_9 = \left(-\frac{\partial p^n}{x_2} + p_2^n\right)\left(-\frac{\partial \tilde{p}^n}{x_2} + \tilde{p}_2^n\right),$$

with $k^* = kK\Delta t$. Also,

$$(f,\, Lv)_{\mathbf{L}} = \int_\Omega (D_5 + D_6 + D_7)\, d\Omega,$$

where

$$D_5 = f_1^n\left(-\frac{\partial \tilde{\sigma}_{11}^n}{\partial x_1} - \frac{\partial \tilde{\sigma}_{21}^n}{\partial x_2} + \beta \tilde{u}_1^n\right),$$

$$D_6 = f_2^n\left(-\frac{\partial \tilde{\sigma}_{12}^n}{\partial x_1} - \frac{\partial \tilde{\sigma}_{22}^n}{\partial x_2} + \beta \tilde{u}_2^n\right),$$

$$D_7 = \left(h^n K\Delta t + \frac{\alpha}{2}(\sigma_{11}^{n-1} + \sigma_{22}^{n-1}) + \frac{\alpha p^{n-1}}{B}\right)\left(\frac{\alpha}{2}(\tilde{\sigma}_{11}^n + \tilde{\sigma}_{22}^n) + \frac{\alpha\tilde{p}^n}{B} - k^*\frac{\partial\tilde{p}_1^n}{\partial x_1} - k^*\frac{\partial\tilde{p}_2^n}{\partial x_2}\right).$$

4.6 Error analysis

Theorem 4.6. *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with sufficiently smooth boundary $\partial\Omega$. The boundary value problem $Lu = f$ in Ω with $u_1^n = u_2^n = p^n = 0$ on $\partial\Omega$ has a uniques solution $u \in [L^2(\Omega)]^9$ for every $f \in [L^2(\Omega)]^9$.*

Proof. The linear operator L is symmetric positive operator. Also, the boundary condition $u_1^n = u_2^n = p^n = 0$ on $\partial\Omega$ is admissible. so the claim follows. □

Theorem 4.7. *There is a normed subspace $V \subset \mathbf{L} = [L^2(\Omega)]^9$ and two positive constants c_1 and c_2 such that $c_1\|u\|_V \leq \|Lu\|_{\mathbf{L}} \leq c_2\|u\|_V$ for every $u \in V$.*

Proof. Define $V = \{u \in \mathbf{L};\ Lu \in \mathbf{L},\ u_1^n = u_2^n = p^n = 0 \text{ on } \partial\Omega\}$ with inner product $(u,\, v)_V = (u,\, v)_{\mathbf{L}} + (Lu, Lv)_{\mathbf{L}}$ for every $u, v \in V$. Then V is a closed subspace of $\mathbf{L}$ and hence a Hilbert space. Now consider $L : V \to \mathbf{L}$. L is clearly continuous with respect to the induced norm $\|u\|_V^2 = \|u\|_{\mathbf{L}}^2 + \|Au\|_{\mathbf{L}}^2$. Also, L is symmetric positive operator. So, L is an isomorphism between V and $\mathbf{L}$. The conclusions follow from the fact L is an isomorphism. □

Theorem 4.8. *Consider the problem, find $u \in V$ such that $(Lu,\, Lv)_{\mathbf{L}} = (f,\, Lv)_{\mathbf{L}}$ for every $v \in V$. Then, the problem has a unique solution.*

Proof. Sketch of proof: Define a new bilinear form, $\tilde{a} : V \times V \to \mathbb{R}$ as $\tilde{a}(u,\, v) = (Lu,\, Lv)_{\mathbf{L}}$ for every $u, v \in V$. As L is an isomorphism between V and $\mathbf{L}$ as proved in the last theorem, $\tilde{a}$ is clearly coercive and continuous. Since, $f \in \mathbf{L}$, then $(f,\, L\cdot)_{\mathbf{L}}$ is a continuous form on V. The conclusion follows from the Lax-Milgram Lemma. Note that, the solution minimizes the quadratic function $E(v) = \|Lv - f\|_{\mathbf{L}}$ for $v \in V$. □

Theorem 4.9. *Let $V_h \subset V$ be a finite dimensional space and consider the problem, find $u_h \in V_h$ such that $(Lu_h,\, Lv_h)_{\mathbf{L}} = (f,\, Lv_h)_{\mathbf{L}}$ for every $v_h \in V_h$. Then, the problem has a unique solution.*

Proof. Define the bilinear form, $\tilde{a_h} : V_h \times V_h \to \mathbb{R}$ as $\tilde{a_h}(u_h,\, v_h) = (Lu_h,\, Lv_h)_{\mathbf{L}}$ for every $u_h, v_h \in V_h$. As $V_h \subset V$, $\tilde{a_h}$ is clearly coercive and continuous. Since, $f \in \mathbf{L}$, then $(f,\, L\cdot)_{\mathbf{L}}$ is a continuous form on V_h. The conclusion follows from the Lax-Milgram Lemma. □

Theorem 4.10. *Let $V_h \subset V$ be a finite dimensional space and consider two problems, find $u_h \in V_h$ such that $(Au_h,\, Av_h)_{\mathbf{L}} = (f,\, Av_h)_{\mathbf{L}}$ for every $v_h \in V_h$ and find $u \in V$ such that $(Au,\, Av)_{\mathbf{L}} = (f,\, Av)_{\mathbf{L}}$ for every $v \in V$. Moreover, if $u \in H^{m+1}(\Omega)$ for some integer $m \geq 1$. Then, there is a $c > 0$ such that for every $h > 0$,*

$$I \leq J$$

where

$$I = \|\sigma - \sigma_h\|_{0,\Omega} + \|\nabla.(\sigma - \sigma_h)\|_{0,\Omega} + \|u - u_h\|_{1,\Omega} + \|p - p_h\|_{1,\Omega} + \|\nabla.(p^* - p_h^*)\|_{0,\Omega}$$

with

$$p^* = (p_1 = \frac{\partial p}{\partial x}, p_2 = \frac{\partial p}{\partial y})^T \quad \textit{and} \quad J = ch^m(\|\sigma\|_{m+1,\Omega} + \|u\|_{m+1,\Omega} + \|p\|_{m+1,\Omega})$$

Proof. By Lax-Milgram Lemma, we have the existence and uniqueness of u and u_h. Consider the bilinear form, $\tilde{a} : V \times V \to \mathbb{R}$ as $\tilde{a}(u, v) = (Lu, Lv)_{\mathbf{L}}$ for every $u,\, v \in V_h$. By the symmetry and coercivity of $\tilde{a}$ together with the Galerkin orthogonality, it is easy to show

$$\|u - u_h\|_V \leq C \inf_{w_h \in V_h} \|u - w_h\|_V$$

The conclusion follows by choosing an appropriate interpolation of u with desired properties.

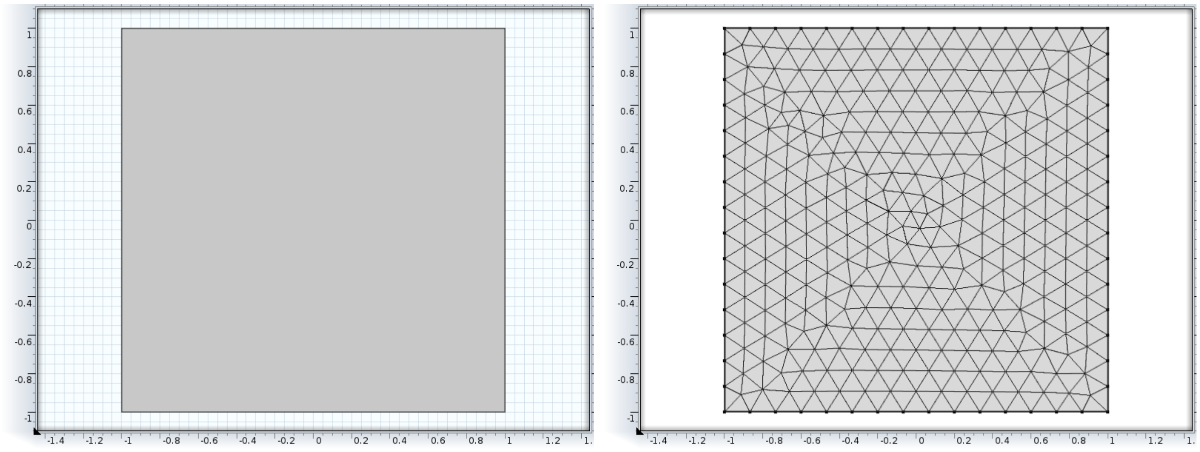
□

4.7 Numerical solution

The PDE domain is a square of side 2 with center at (0, 0). The PDEs are supplemented by the boundary condition ($p = u_1 = u_2 = 0$). We also set $\beta = 10^{-200}(1,\, 1)^T$. We choose $\Delta t = 0.001,\, 0.01,\, 0.05,\, 0.08,\, 0.1,\, 0.3,\, 0.6,\, 1.0$. COMSOL 4.3 weak form PDE console is used to implement the corresponding weak formulation. In this finite element implementation, 578 elements, 8519 degrees of freedom, Lagrange shape functions with quadratic element order are used. Also, following parameters have been used.

Parameters	Value	Parameters	Value
λ, μ	1	f_1^n	1
k, K	1	f_1^n	1
α, B	0.5	h^n	1

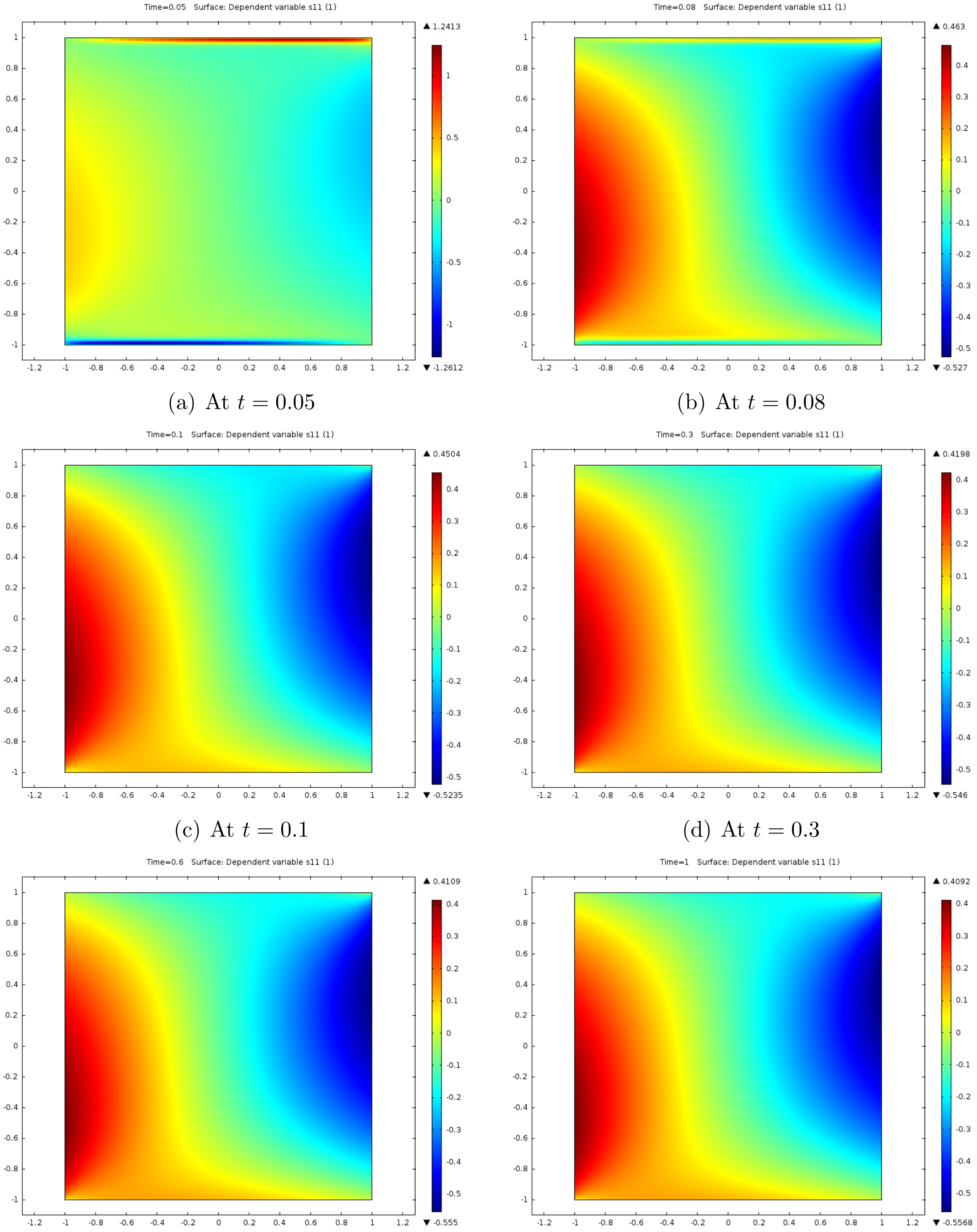
Table 4.4: Different Parameters for COMSOL for the second formulation



(a) Physical domain

(b) Meshed domain

Figure 4.1: Domain and its meshing for the second formulation



(a) At $t = 0.05$

(b) At $t = 0.08$

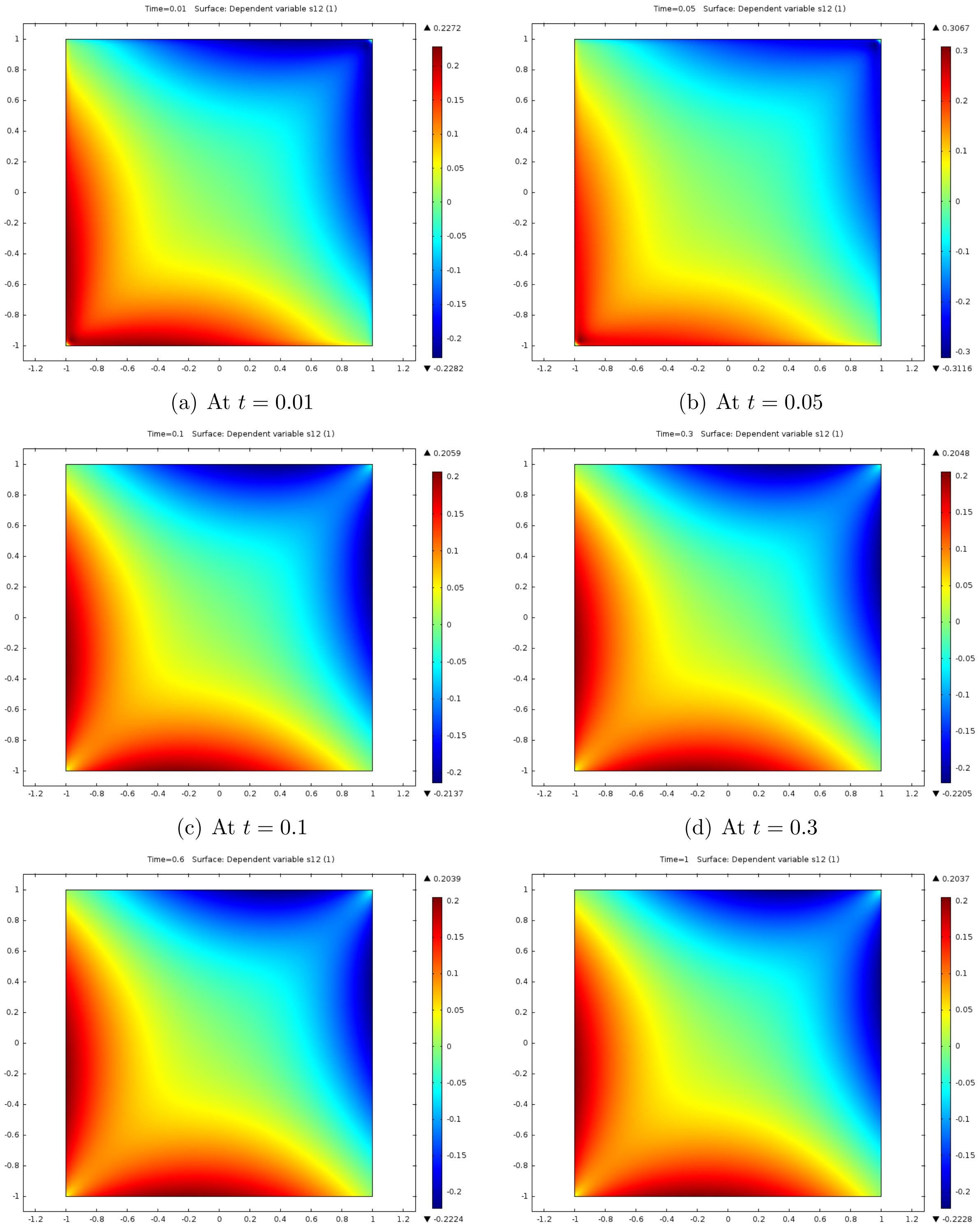
(c) At $t = 0.1$

(d) At $t = 0.3$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.2: σ_{11} at different times



(a) At $t = 0.01$

(b) At $t = 0.05$

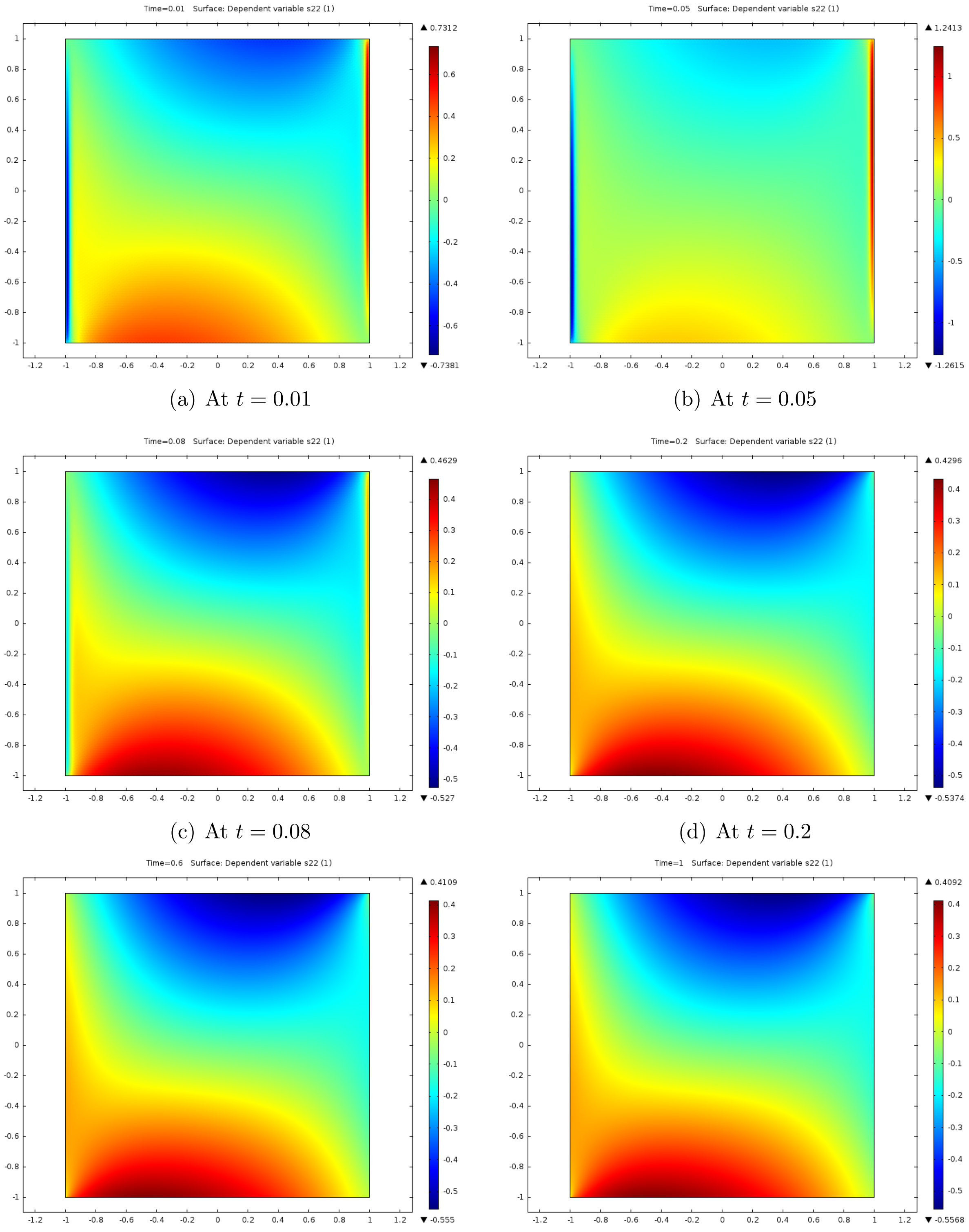
(c) At $t = 0.1$

(d) At $t = 0.3$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.3: $\sigma_{12} = \sigma_{21}$ at different times



(a) At $t = 0.01$

(b) At $t = 0.05$

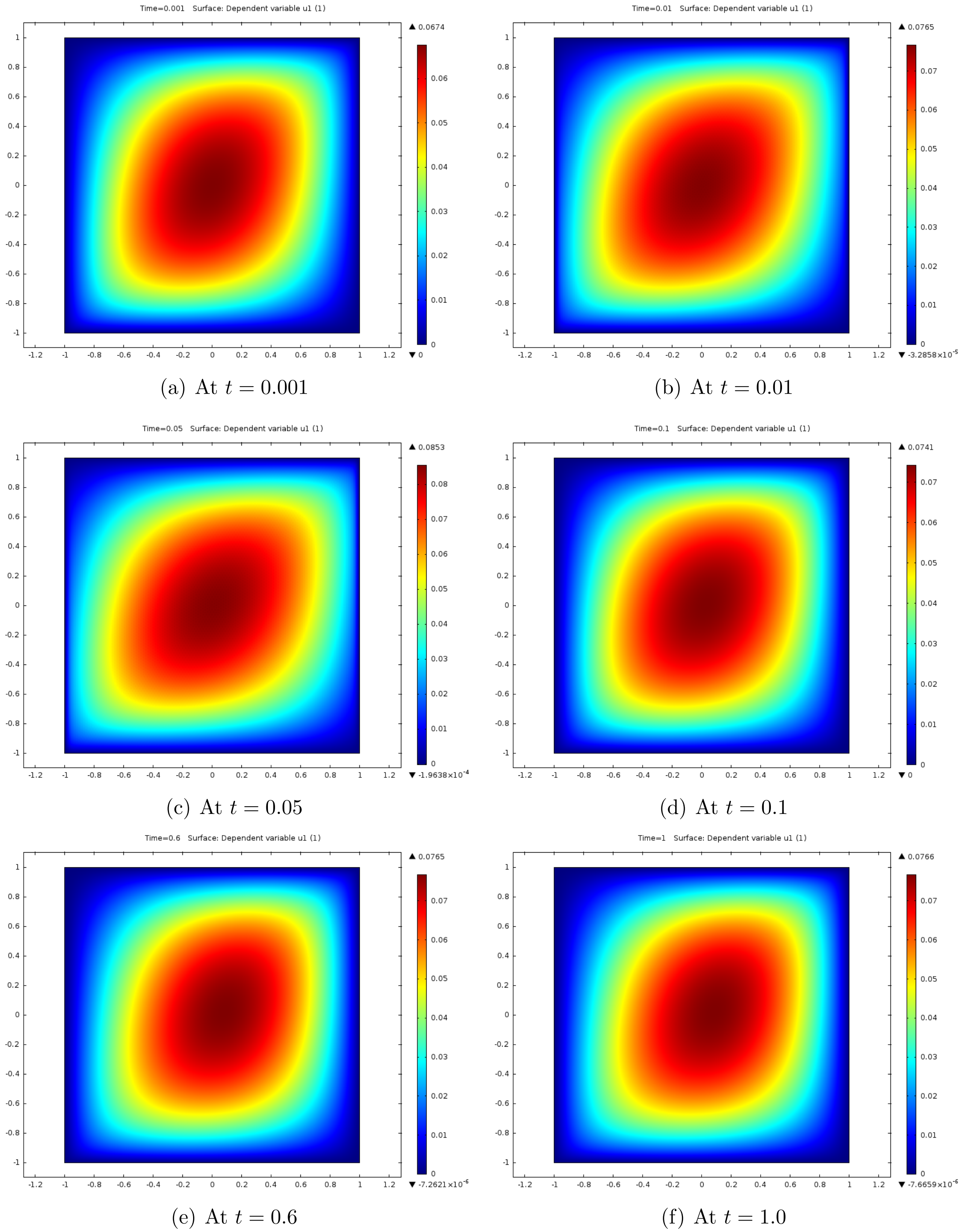
(c) At $t = 0.08$

(d) At $t = 0.2$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.4: σ_{22} at different times



(a) At $t = 0.001$

(b) At $t = 0.01$

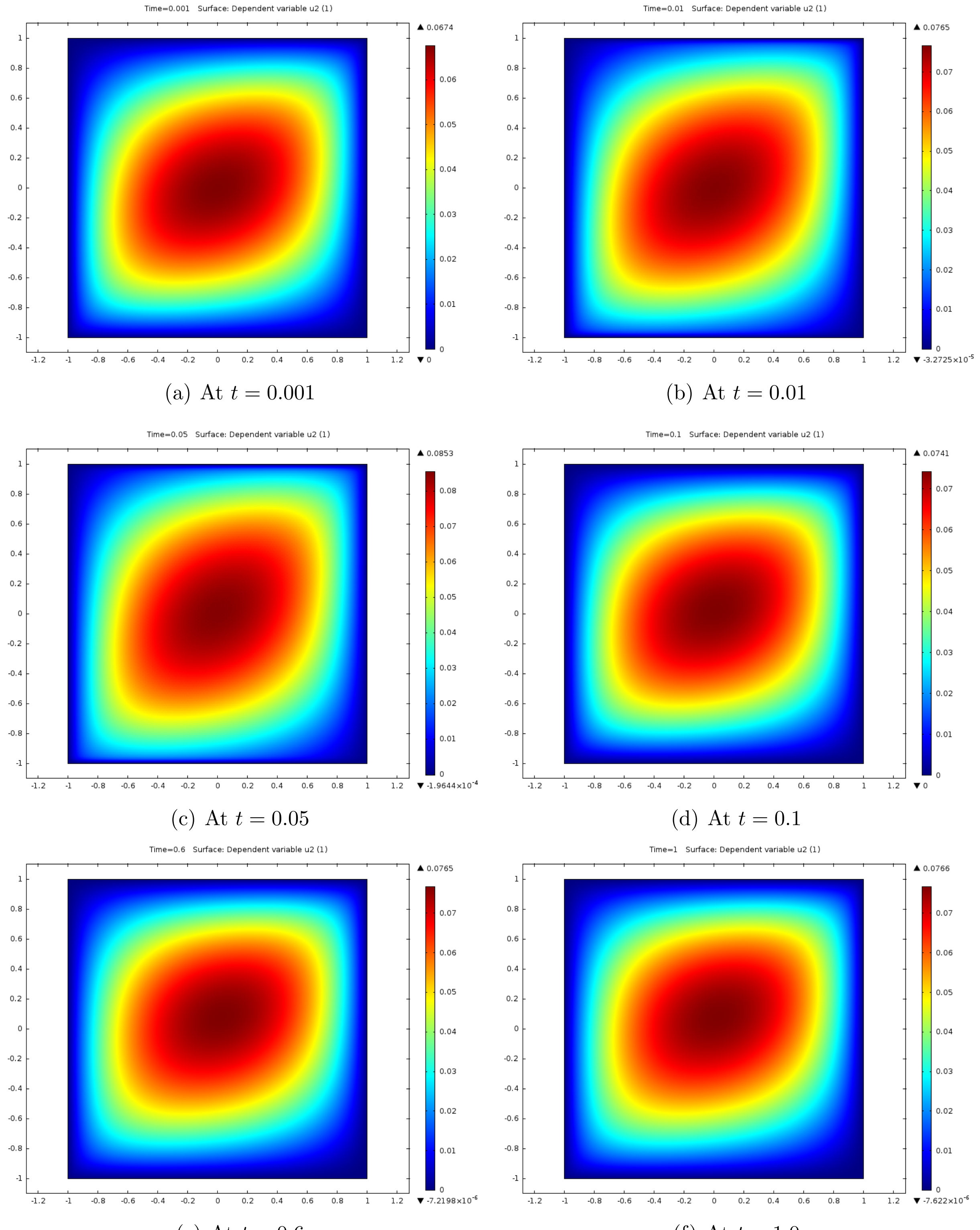
(c) At $t = 0.05$

(d) At $t = 0.1$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.5: u_1 at different times



(a) At $t = 0.001$

(b) At $t = 0.01$

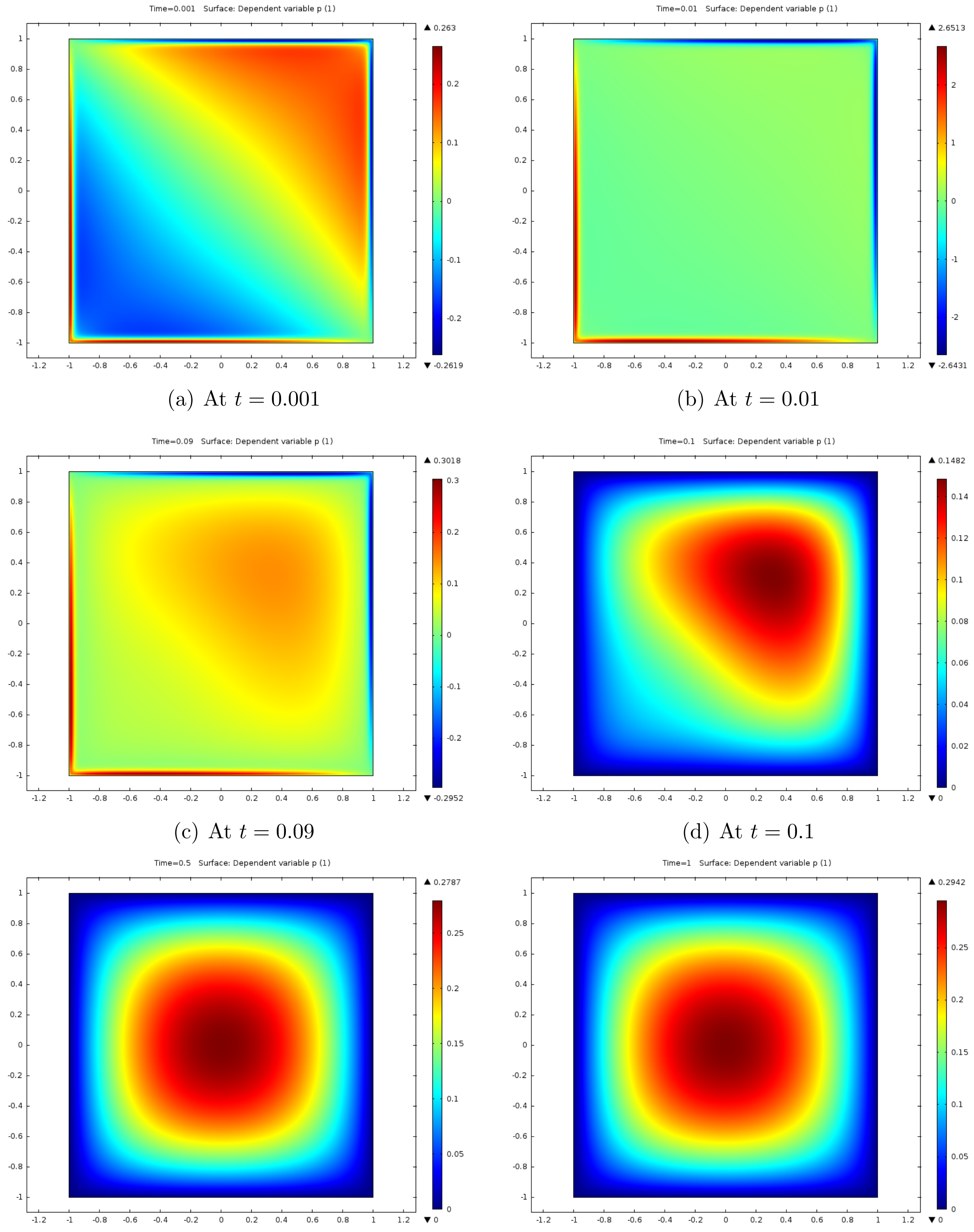
(c) At $t = 0.05$

(d) At $t = 0.1$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.6: u_2 at different times



(a) At $t = 0.001$

(b) At $t = 0.01$

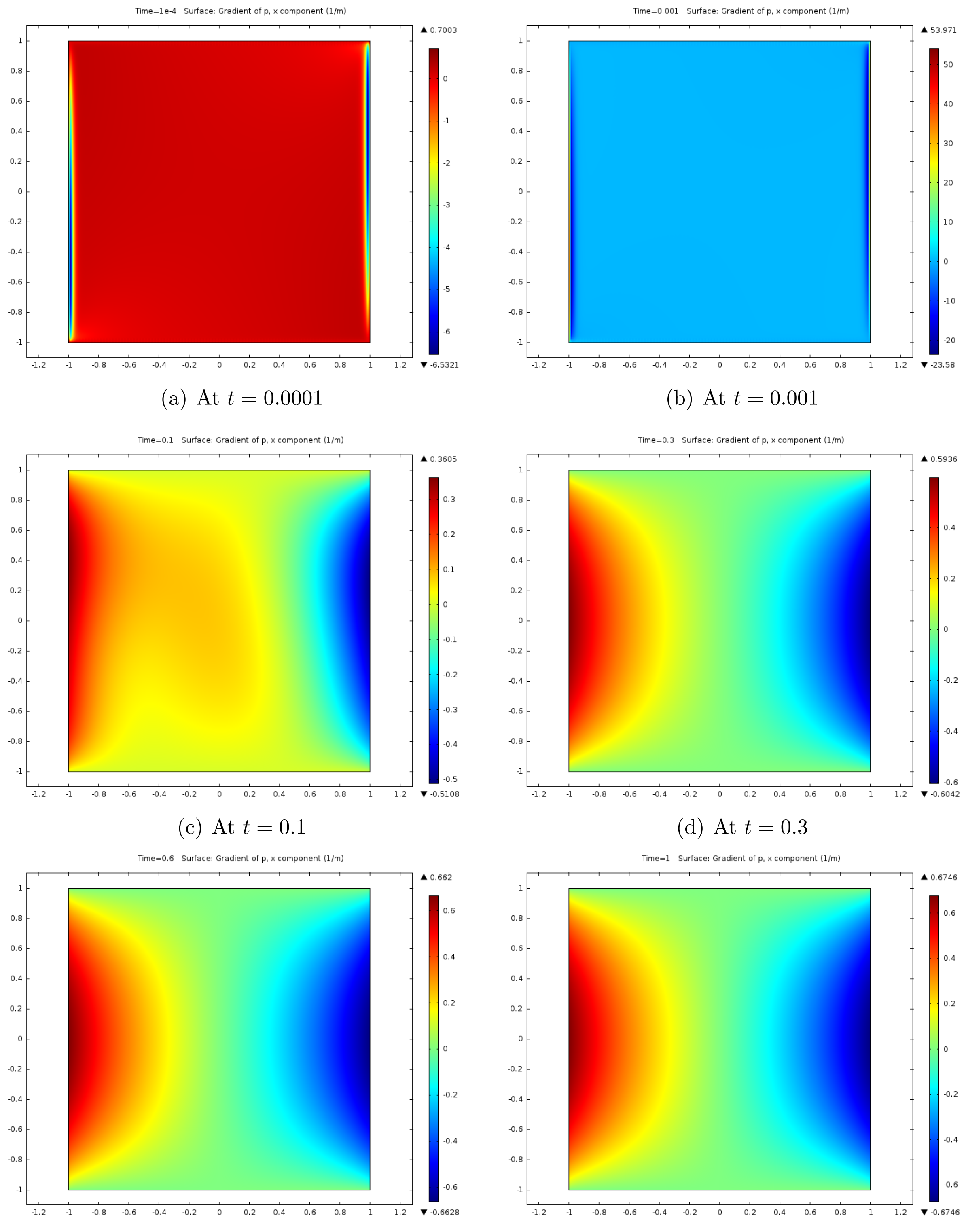
(c) At $t = 0.09$

(d) At $t = 0.1$

(e) At $t = 0.5$

(f) At $t = 1.0$

Figure 4.7: p at different times



(a) At $t = 0.0001$

(b) At $t = 0.001$

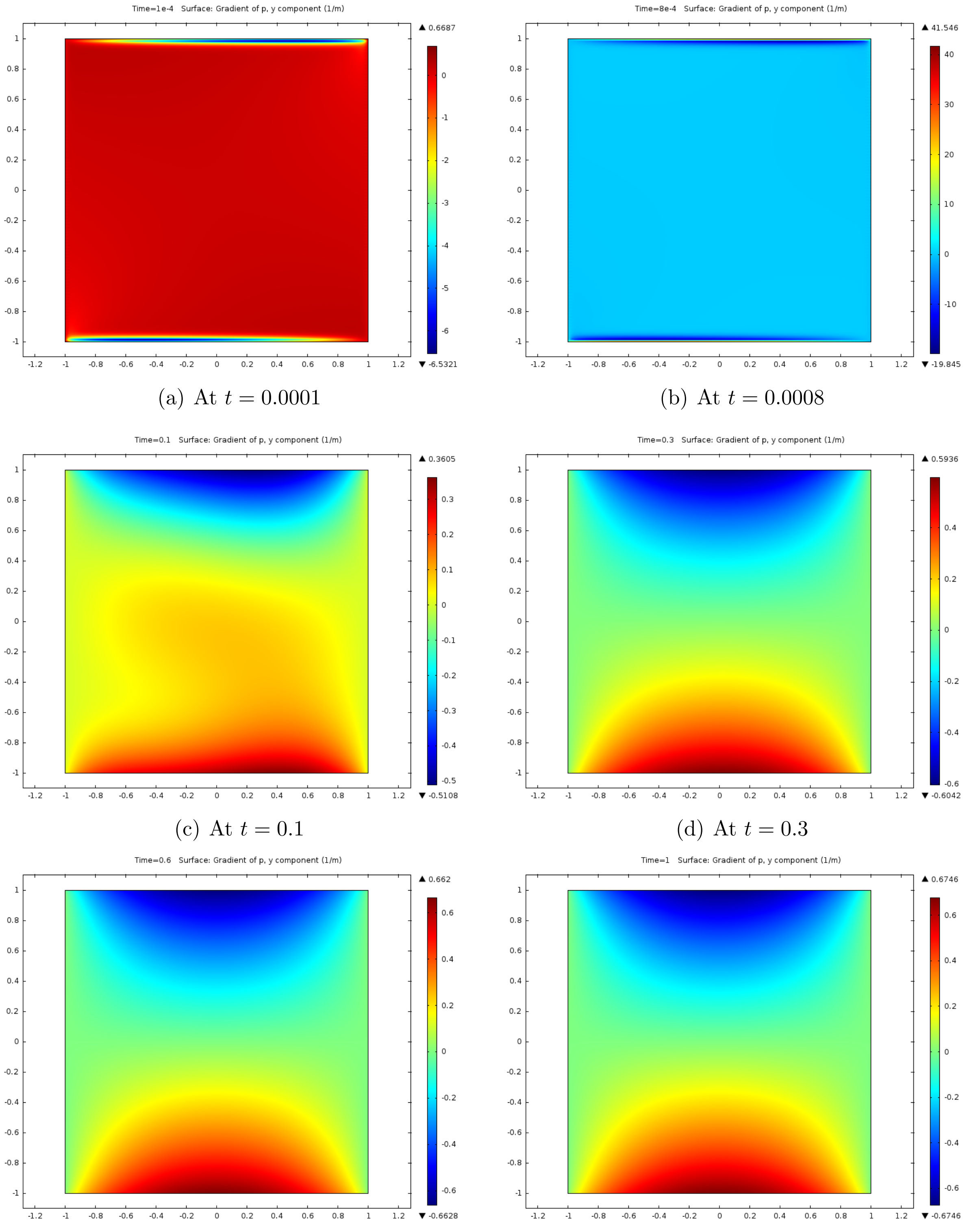
(c) At $t = 0.1$

(d) At $t = 0.3$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.8: p_x at different times



(a) At $t = 0.0001$

(b) At $t = 0.0008$

(c) At $t = 0.1$

(d) At $t = 0.3$

(e) At $t = 0.6$

(f) At $t = 1.0$

Figure 4.9: p_y at different times

4.8 Convergence study

Let $\Omega = (-1,1) \times (-1,1)$ be the domain. We run an experiment with following parameters $\Delta t = 1$, $\lambda = 1$, $\mu = 1$, $k = 1$, $K = 1$, $\alpha = 0.5$, $B = 0.5$. We also set the following variables such that the actual unique solution is known.

$$f_1^n = I_1 + I_2, \quad \text{with}$$

$$I_1 = 2\pi^2 \mu \sin \pi x \sin \pi y - \mu \left[4xy - \pi^2 \sin \pi x \sin \pi y\right],$$

$$I_2 = -\lambda \left[\left(4xy - \pi^2 \sin \pi x \sin \pi y\right) - \pi\alpha(y^2-1) \cos \pi x\right],$$

$$f_2^n = I_3 + I_4, \quad \text{with}$$

$$I_3 = -4\mu(x^2-1) - \lambda \left[2x^2 + \pi^2 \cos \pi x \cos \pi y - 2\right],$$

$$I_4 = -\mu \left[2y^2 + \pi^2 \cos \pi x \cos \pi y - 2\right] - 2\alpha y \sin \pi x,$$

$$h^n = J_1 + J_2 + J_3 + J_4,$$

$$J_1 = 2k \sin \pi x - \pi^2 k(y^2-1) \sin \pi x,$$

$$J_2 = \frac{1}{\Delta t} \left[2\alpha\lambda \left(2y(x^2-1) + \pi \cos \pi x \sin \pi y\right)\right],$$

$$J_3 = \frac{1}{\Delta t} \left[4\alpha\mu y(x^2-1) + 2\alpha^2(y^2-1) \sin \pi x\right],$$

$$J_4 = \frac{1}{\Delta t} \left[\frac{\pi\mu\alpha \cos \pi x \sin \pi y}{K^2} - \frac{\alpha(y^2-1) \sin \pi x}{BK^2}\right].$$

For this data, the solution is

$$\sigma_{11} = K_1 + K_2, \quad \text{with}$$

$$K_1 = \lambda \left[2y(x^2-1) + \pi \cos \pi x \sin \pi y\right],$$

$$K_2 = \alpha(y^2 - 1)\sin \pi x + 2\pi\mu \cos \pi x \sin \pi y,$$

$$\sigma_{21} = \sigma_{12} = \mu \left[2x(y^2 - 1) + \pi \cos \pi y \sin \pi x\right],$$

$$\sigma_{22} = K_3 + K_4, \quad \text{with}$$

$$K_3 = \lambda \left[2y(x^2 - 1) + \pi \cos \pi x \sin \pi y\right],$$

$$K_4 = 4\mu y(x^2 - 1) + \alpha(y^2 - 1)\sin \pi x,$$

$$u_1 = \sin \pi x \sin \pi y,$$

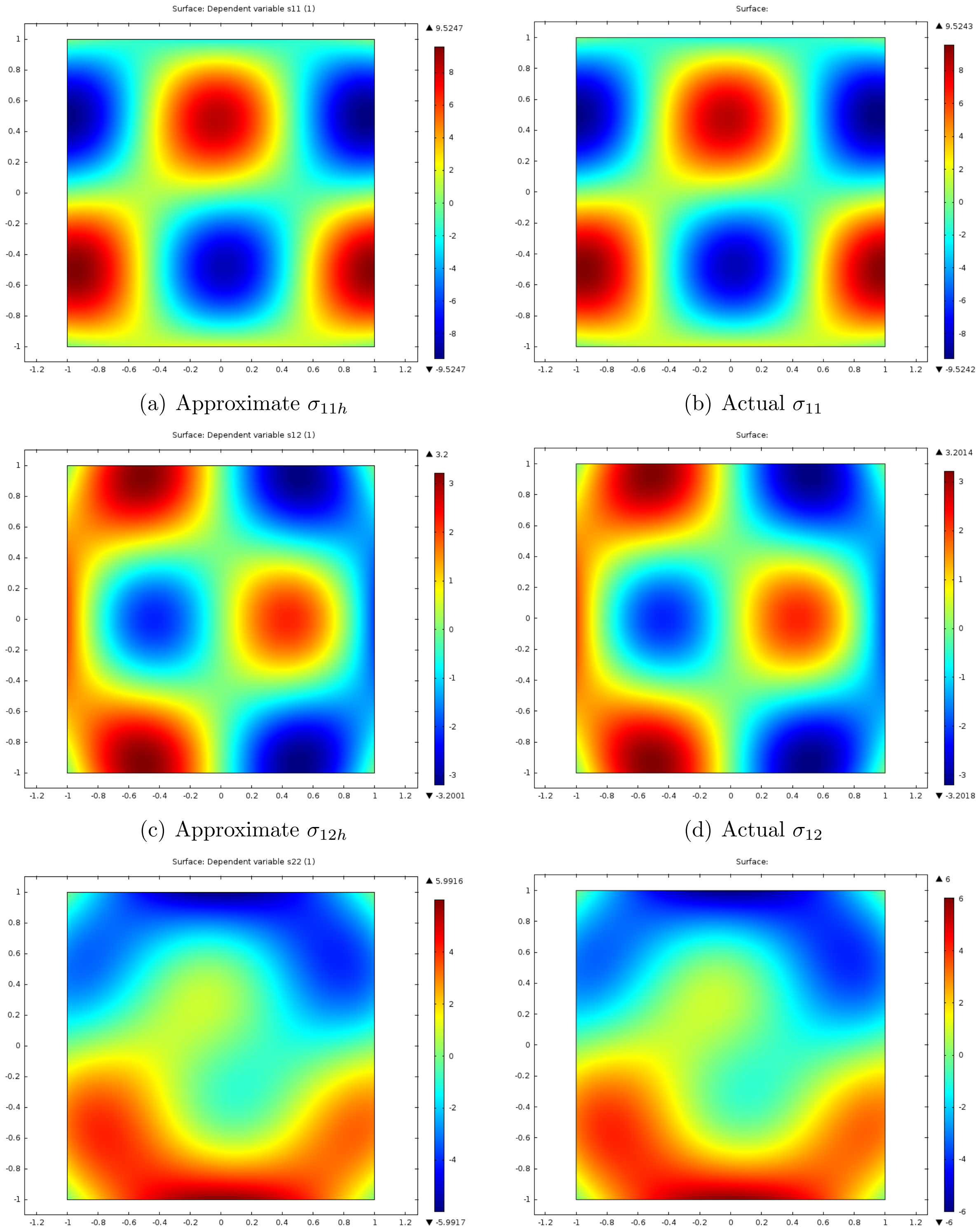
$$u_2 = (1 - x^2)(1 - y^2),$$

$$p = (1 - y^2)\sin \pi x,$$

$$p_1 = -\pi(y^2 - 1)\cos \pi x,$$

$$p_2 = -2y \sin \pi x.$$

In this experiment, 268 elements, 11403 degrees of freedom, Lagrange shape functions with cubic element order are used. The following figures show finite element solution and actual solution for comparison.



(a) Approximate σ_{11h}

(b) Actual σ_{11}

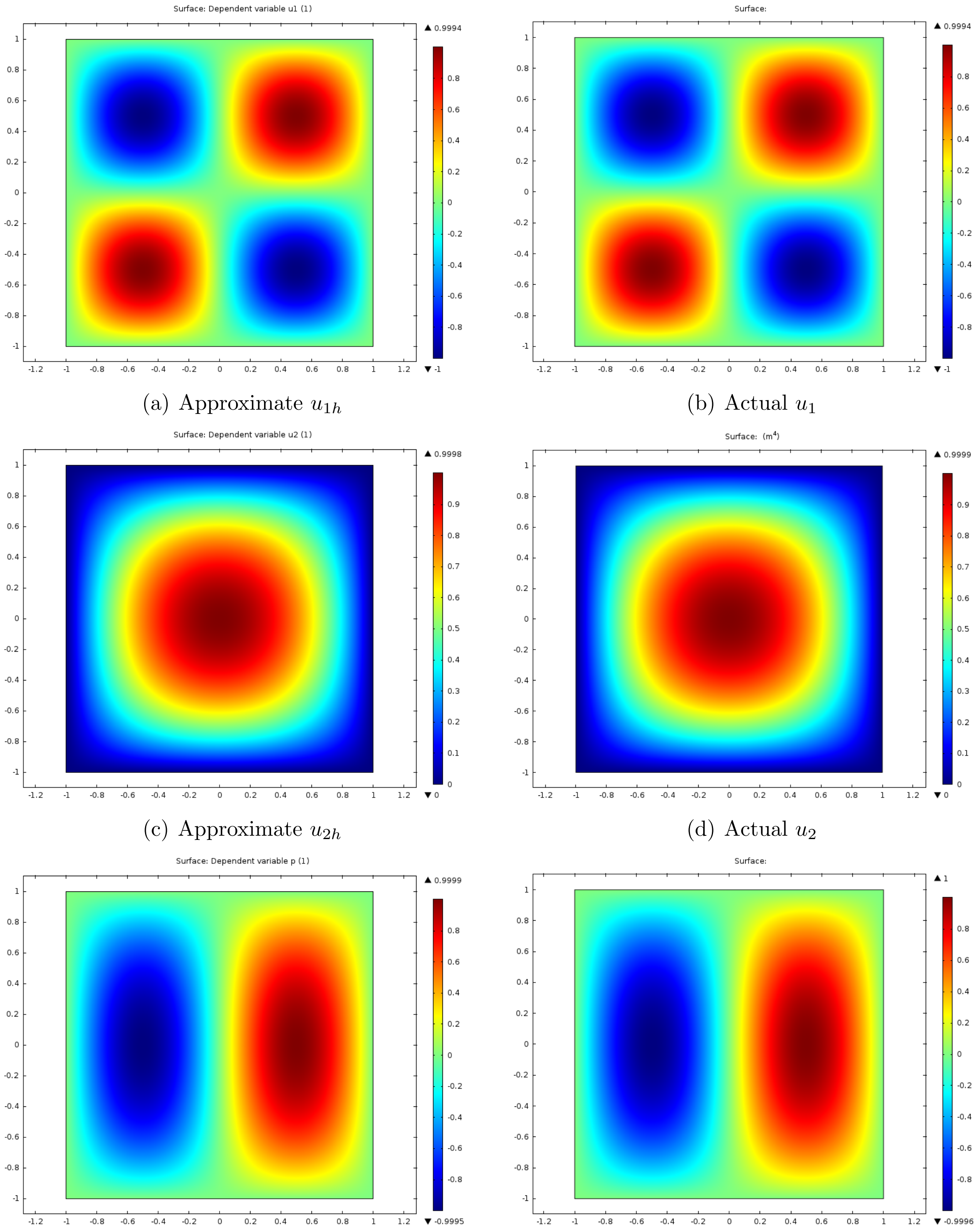
(c) Approximate σ_{12h}

(d) Actual σ_{12}

(e) Approximate σ_{22h}

(f) Actual σ_{22}

Figure 4.10: Approximate and actual solution for σ



(a) Approximate u_{1h}

(b) Actual u_1

(c) Approximate u_{2h}

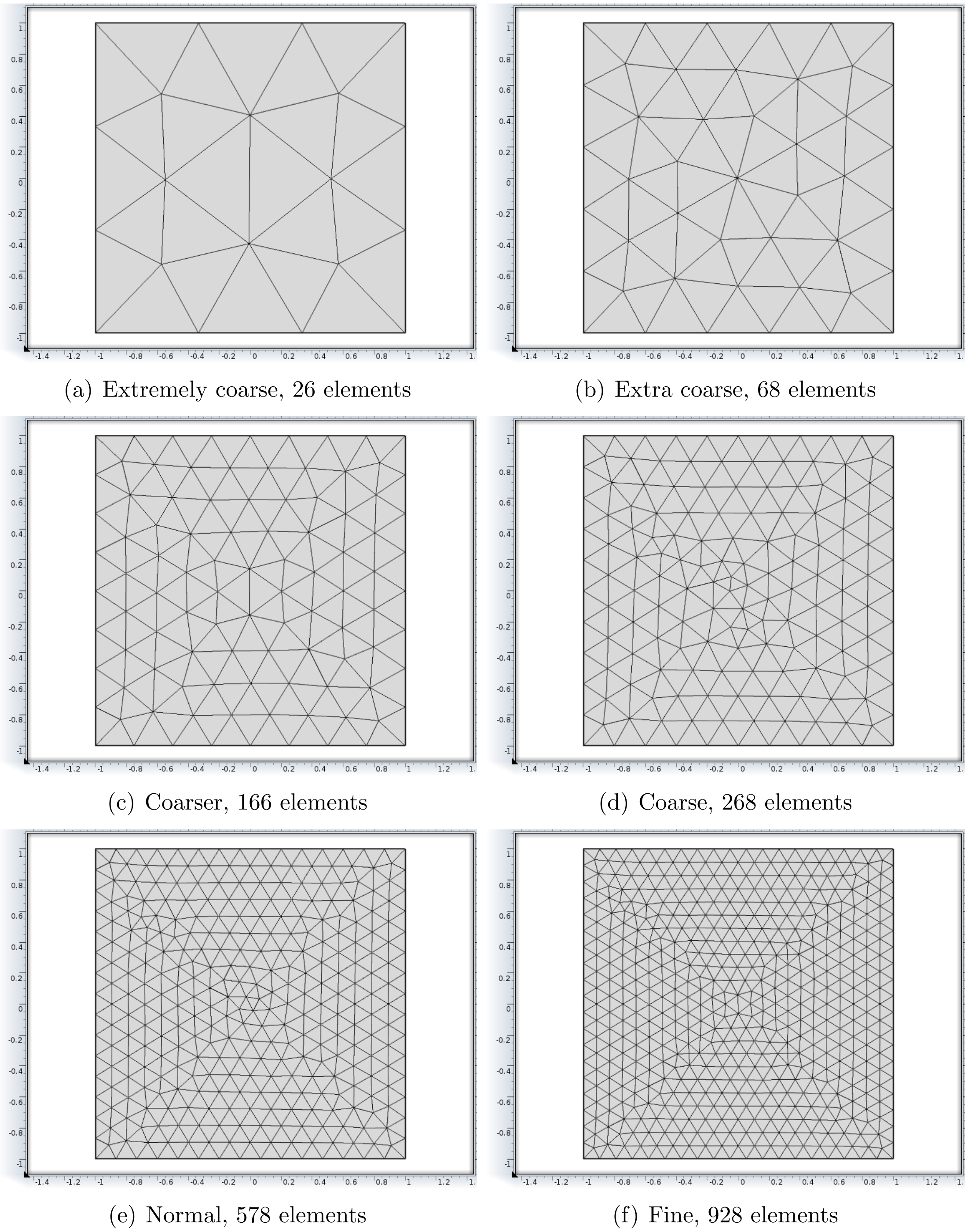
(d) Actual u_2

(e) Approximate p_h

(f) Actual p

Figure 4.11: Approximate and actual solution for u and p

We try to find different error norm with varying mesh structures. We have the following meshes for comparison.



(a) Extremely coarse, 26 elements

(b) Extra coarse, 68 elements

(c) Coarser, 166 elements

(d) Coarse, 268 elements

(e) Normal, 578 elements

(f) Fine, 928 elements

Figure 4.12: Different mesh structures

Mesh	No of Elements	DOF	$\|\sigma - \sigma_h\|_2$	$\|u - u_h\|_2$	$\|p - p_h\|_2$
Extremely coarse	26	1224	0.175	0.0185	0.00483
Extra coarse	68	3033	0.0382	0.00202	7.03E-04
Coarser	166	7164	0.00893	3.43E-04	1.08E-04
Coarse	268	11403	0.00431	1.33E-04	4.47E-05
Normal	578	24228	0.00127	2.71E-05	8.79E-06
Fine	928	38619	6.12E-04	1.05E-05	3.38E-06

Table 4.5: Different norms with varying mesh using Lagrange shape function with cubic element at $t = 1$

Element order	No of Elements	DOF	$\|\sigma - \sigma_h\|_2$	$\|u - u_h\|_2$	$\|p - p_h\|_2$
Linear	68	405	4.036	0.409	0.124
Quadratic	68	1413	0.343	0.0195	0.00777
Cubic	68	3033	0.0382	0.00202	7.03E-04
Quartic	68	5265	0.00388	1.09E-04	2.82E-05
Quintic	68	8109	2.85E-04	7.87E-06	1.61E-06

Table 4.6: Different norms with varying element order using Lagrange shape function with extra coarse mesh at $t = 1$

h	$\|\sigma - \sigma_h\|_2$	Conv. Rate	$\|u - u_h\|_2$	Conv. Rate	$\|p - p_h\|_2$	Conv. Rate
0.66	4.684		0.751		0.279	
0.40	3.333	0.679	0.409	1.213	0.139	1.395
0.26	2.304	0.857	0.209	1.550	0.0711	1.551
0.20	1.764	1.019	0.129	1.837	0.0445	1.785
0.134	1.186	0.991	0.0569	2.052	0.0204	1.943
0.106	0.948	0.955	0.0361	1.949	0.0128	1.992
0.074	0.658	1.014	0.0177	1.978	0.00628	1.984
0.04	0.357	0.993	0.00511	2.020	0.00181	2.0162

Table 4.7: Convergence rate at fixed time step $\Delta T = k = 0.01$ in $L^2(0, 1, L^2(\Omega))$ using Lagrange shape function with linear element

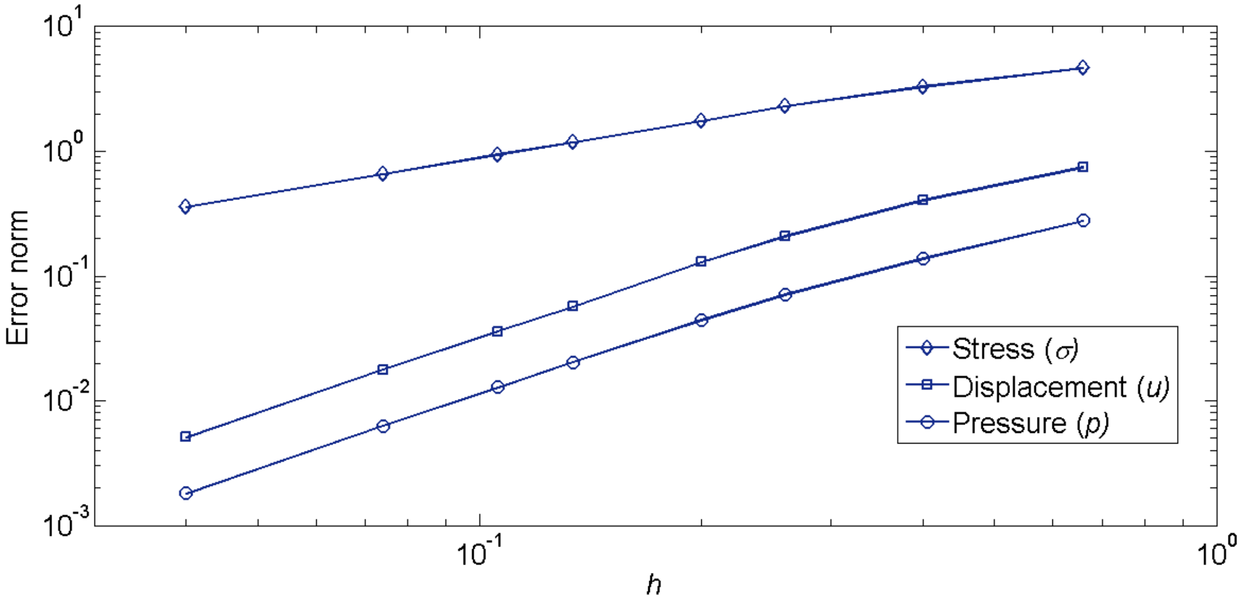


Figure 4.13: Convergence rate for linear element

h	$\|\sigma - \sigma_h\|_2$	Conv. Rate	$\|u - u_h\|_2$	Conv. Rate	$\|p - p_h\|_2$	Conv. Rate
0.66	1.013		0.0838		0.0293	
0.40	0.532	1.286	0.0256	2.364	0.00867	2.435
0.26	0.271	1.565	0.00843	2.580	0.00277	2.648
0.20	0.173	1.707	0.00397	2.869	0.00129	2.904
0.134	0.0758	2.063	0.00120	2.972	3.93E-04	2.975
0.106	0.0481	1.945	5.97E-04	2.994	1.91E-04	3.074
0.074	0.0236	1.982	1.99E-04	3.066	6.48E-05	3.010
0.04	0.00678	2.024	3.10E-05	3.021	1.02E-05	3.007

Table 4.8: Convergence rate at fixed time step $\Delta T = k = 0.01$ in $L^2(0, 1, L^2(\Omega))$ using Lagrange shape function with quadratic element

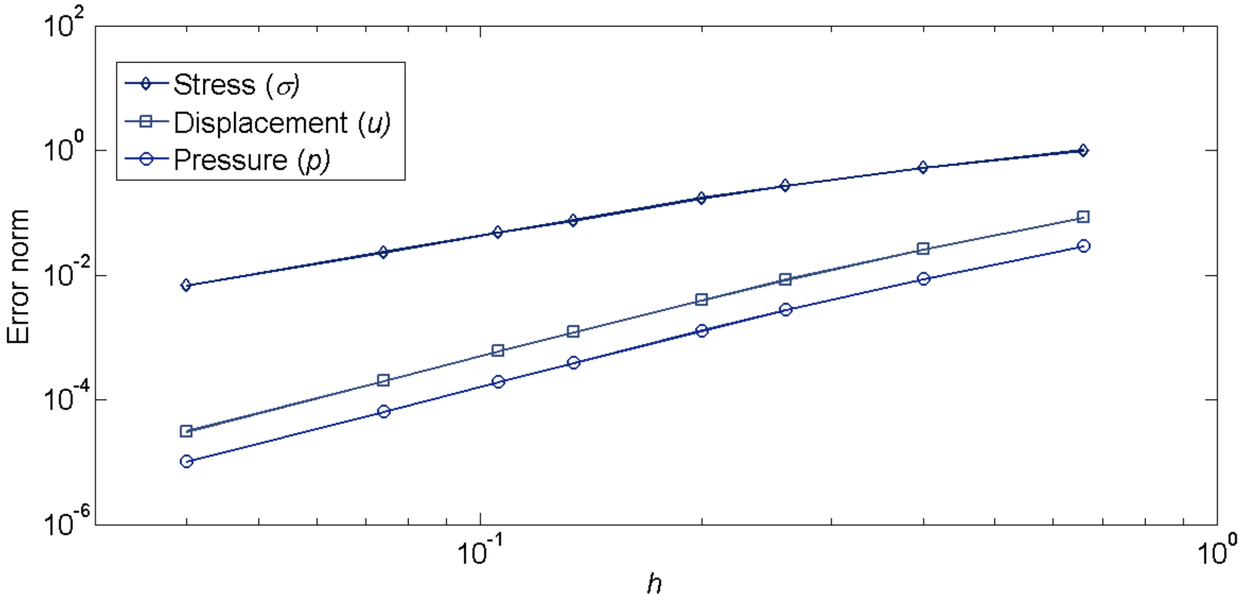


Figure 4.14: Convergence rate for quadratic element

Chapter 5

Conclusions and future works

In this work, we have studied a system of PDE modeling poroelasticity. Starting from momentum and mass conservation equations at macroscopic level, along with the linear constitutive equations, we formulated the equations describing the coupled processes of elastic deformation and the pore fluid pressure in a porous medium. Upon reasonable assumptions on fully developed poroelasticity equations, we have the quasi-static form and analysis including well-posedness and approximation of this form is the main purpose of our study. We have proved the existence and uniqueness of the quasi-static form with admissible boundary conditions using the concept of symmetric positive system, as introduced by Friedrich [30] in 1958. We expressed the quasi-static form as symmetric positive system for two different formulations, namely fluid content-rotation-pressure gradient and stress-displacement-pressure formulation. The main advantage of having different formulations is to provide varied supplemental boundary data, necessary for its well-posedness. For both formulation, the unknown variables are often physically important. For the stress-displacement-pressure formulation, we have the existence and uniqueness results depending on the physical parameters of the poroelastic system, which is kind of restrictive. Using the scaling technique, we have found the requirement is not very strong, as it allows some poroelastic systems, which are non-positive without scaling, to be symmetric positive. So, the scaling technique adapts a lot of geophysical system into the symmetric positive framework. Having well-posedness of the system, we have conducted numerical experiments approximating weak solutions of both formulations using the least square finite element method. Convergence and numerical results are presented for both cases to show that the numerical technique is working.

A number of extension to the present work are possible. First, we have ignored the time derivative term of the following equation in the mathematical modeling section

$$\rho\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla\cdot\sigma(\mathbf{x},t) = f(\mathbf{x},t).$$

Our numerical approximation is based on time discretization, and thus we may allow this term to represent better modeling. A certain difficulty will arise to make the system symmetric positive because of the additional term. For real life poroelastic system, there are usually complicated boundary conditions, not just Dirichlet or Neumann boundary conditions. Although, in symmetric positive framework, boundary condition are dictated by the algebraic form of the first order representation, still we can try accommodating more complex boundary conditions. In poroelasticity, it is known that as $\lambda \to \infty$, the error estimate in poroelasticity might be unreliable. This locking phenomena can be studied in our numerical setting in order to understand why it happens and how to get rid of this. In our numerical technique, we have found solution vector for different times as our approximation is based on time discretization formulation. An attempt to construct complete solution vector from the discrete solution can be made. Then, an analysis on the continuity, differentiability i.e. the regularity properties of the complete solution vector can be conducted. There are other efficient numerical scheme as approximation technique, such as discontinuous Galerkin method, finite volume method etc. It is worth mentioning that general symmetric positive system can be approximated by using discontinuous Galerkin method [32]. So, these numerical techniques can be considered for better approximation.

Bibliography

[1] K. Terzaghi. Die berechnung der durchlassigkeitsziffer des tones aus dem verlauf der hydrodynamischen spannungserscheinungen, Sitz. Akad. Wissen. Wien Math. Naturwiss. Kl., Abt. IIa, 132, 105-124, 1923.

[2] L. Rendulic. Porenziffer und Porenwasserdrunk in Tonen. Der Bauingenieur, 17, 559-564, 1936.

[3] M. A. Biot. Le problme de la consolidation des matires argileuses sous une charge. Ann. Soc. Sci. Bruxelles, B55, 110-113, 1935.

[4] M. A. Biot. General theory of three-dimensional consolidation. J. Appl. Phys., 12, 155-164, 1941.

[5] M. A. Biot. Theory of elasticity and consolidation for a porous anisotropic solid. J. Appl. Phys., 26, 182-185, 1955.

[6] M. A. Biot. General solutions of the equations of elasticity and consolidation for a porous material. J. Appl. Mech., Trans. ASME, 78, 91-96, 1956.

[7] M. A. Biot. Thermoelasticity and irreversible thermodynamics. J. Appl. Phys. 27, 240-253, 1956.

[8] M. A. Biot. Mechanics of deformation and acoustic propagation in porous media. J. Appl. Phys., 33, 14821498, 1962.

[9] J.R. Rice and M.P. Cleary. Some basic stress-diffusion solutions for fluid saturated elastic porous media with compressible constituents. Rev. Geophys. Space Phys., 14, 227-241, 1976.

[10] J. L. Nowinski and C. F. Davis . A model of the human skull as a poroelastic spherical shell subjected to a quasistatic load. Mathematical Biosciences, 8, 397-416, 1970.

[11] J. L. Nowinski and C. F. Davis . The flexure and torsion of bones viewed as anisotropic poroelastic bodies, International Journal of Engineering Science, 10, 1063-1079, 1972.

[12] J. L. Nowinski. Bone articulations as systems of poroelastic bodies in contact. AIAA Journal, 9, 62-69, 1971.

[13] J. L. Nowinski. Stress concentrations around a cylindrical cavity in a bone treated as a poroelastic body. AIAA Journal,Acta Mechanica, 13, 281-292, 1972.

[14] M. W. Johnson, D.A. Chakkalakal, R.A. Harper, J.L. Katz and S.W. Rouhana. Fluid flow in bone. Journal of Biomechanics, 11, 881-885, 1982.

[15] S.I. Barry and G.N. Mercer. Flow and deformation in poroelasticity - I unusual exact solutions. Mathematical and Computer Modeling, 30:2329, 1999.

[16] O. Coussy. Poromechanics. Wiley, 2004.

[17] Herbert F. Wang. Theory of Linear Poroelasticity with Applications to Geomechanics and Hydrology. Princeton Series in Geophysics, Princeton University Press, Princeton, 2000.

[18] J. F. Poland and G. H. Davis. Land Ssubsidence due to withdrawal of fluid. Reviews in Engineering Geology, 2, 187-270, 1969.

[19] N. Lubick. Modeling complex, Multiphase porous media systems. Siam News, 5(3), 2002.

[20] Arthur F. T. Mak, Lidu Huang and Qinque Wang. A biphasic poroelastic analysis of the flow dependent subcutaneous tissue pressure and compaction due to epidermal loadings: issues in pressure sore. J Biomech Eng, 116(4), 421-429, 1994.

[21] B. R. Simon, M. V. Kaufmann, M. A. McAfee and A. L. Baldwin. Finite Element Models for Arterial Wall Mechanics. J Biomech Eng, 115(4B), 489-496, 1993.

[22] Ming Yang, Larry A. Taber. The possible role of poroelasticity in the apparent viscoelastic behavior of passive cardiac muscle. J Biomech, 24(7), 587-597, 1991.

[23] A. Pena, M.D. Bolton and J.D. Pickard. Cellular poroelasticity: A theoretical model for soft tissue mechanics. First Biot Conference on Poromechanics, Louvain la Neuve, Belgium, 1998.

[24] P. A. Netti, L.T. Baxter, Y. Coucher, R.K. Skalak and R.K. Jain. A poroelastic model for interstitial pressure in tumors. Biorheology, 32(2), 346, 1995.

[25] T. Roose, P. A. Netti, L. L. Munn, Y. Boucher and R.K. Jain. Solid stress generated by spheroid growth estimated using a linear poroelasticity model. Microvascular Research, 66(3), 204212, 2003.

[26] X. G. Li, H. von Holst, J. Ho and S. Kleiven. Three Dimensional Poroelastic Simulation of Brain Edema: Initial Studies on Intracranial Pressure. IFMBE Proceeding, 25(4), 1478-1481, 2010.

[27] A. Eisentrager. Finite Element Simulation of a Poroelastic Model of the CSF System in the Human Brain during an Infusion Test. DPhil thesis, University of Oxford, 2012.

[28] J. M. Skotheim , L. Mahadevan. Dynamics of poroelastic filaments. Proc. R. Soc. Lond. A, 460, 19952020, 2004.

[29] R. E. Showalter. Diffusion in poro-elastic media. Jour. Math. Anal. Appl., 251:310340, 2000.

[30] K. O. Friedrichs. Symmetric positive linear differential equations. Comm. Pure Appl. Math., 11:333-418, 1958.

[31] W. Min-Hua. On applications of symmetric positive systems to elasticity. Chinise Science Bulletin, 35(21), 1769-1773, 1990.

[32] D. A. Di Pietro, A. Ern. Mathematical Aspects of Discontinuous Galerkin Methods. Springer, 2012.

[33] A. Ern Guermond. Theory and practice of finite elements. Springer, 2004.

[34] N. Antonic, K. Burazin, and M. Vrdoljak. Second-order equations as Friedrichs systems. Nonlinear Analysis: Real World Applications, 15, 290-305, 2014.

[35] N. Antonic, K. Burazin, and M. Vrdoljak. Heat equation as a Friedrichs system. Journal of Mathematical Analysis and Applications, 404(2), 537-553, 2013.

[36] B. N. Jiang. The least-square finite element method. Springer, 1998.

[37] T. Bui-Thanh, L. Demkowicz and O. Ghattas. A Unified Discontinuous Petrov–Galerkin Method and Its Analysis for Friedrichs' Systems. SIAM Journal on Numerical Analysis, 2013.

[38] A. Aziz, R. Kellogg, and A. Stephens. Least-squares methods for elliptic systems. Math. of Comp., 44:53-70, 1985.

[39] P. B. Bochev. Analysis of least-squares finite element methods for the Navier-Stokes equations. SIAM J. Num. Anal., 1997.

[40] P. B. Bochev, M.D. Gunzburger. Finite element methods of least-squares type. SIAM Rev., 40:789-837, 1998.

[41] P. B. Bochev, M.D. Gunzburger. Least-squares finite element methods. Proceedings of the international congress of Mathematicians, 2006

[42] S. C. Brenner and L. R. Scott. The Mathematical Theory of Finite Element Methods. Springer-Verlag, volume 15 of Texts in Applied Mathematics, second edition, 2002.

[43] L. C. Evans. Partial differential equations. American Mathematical Society, volume 19 of Graduate Studies in Mathematics, second edition, 2010.

[44] K. Atkinson and W. Han. Theoretical numerical analysis, a functional analysis framework. Springer, 2001.

[45] B.-N. Jiang and L. Povinelli. Least-squares finite element method for fluid dynamics. Comput. Meth. Appl. Mech. Engr., 81:13-37, 1990.

[46] B.-N. Jiang. The least-squares finite element method. Springer, 1998.

[47] J. H. Bramble and A. H. Schatz. Least Squares Methods for 2mth Order Elliptic Boundary-Value Problems. Mathematics of Computation, 25(113):1-32, 1971.

[48] G. A. Baker. Simplified proofs of error estimates for the least squares method for Dirichlet's problem. Mathematics of computation, 27(122), 1973.