Implementation of linear and non-linear elastic biphasic porous media problems into FEATFLOW and comparison with PANDAS

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik der
Technischen Universität Dortmund
vorgelegt von

Abdulrahman Sadeq Obaid

im März 2017

Dissertation

Implementation of linear and non-linear elastic biphasic porous media problems into FEATFLOW and comparison with PANDAS

Fakultät für Mathematik
Technische universität Dortmund

Erstgutachter: Prof. Dr. Stefan Turek

Zweitgutachter: Prof. Dr. Bernd Markert

Tag der mündlichen Prüfung: 07. 07. 2017

Abstract

This dissertation presents a fully implicit, monolithic finite element solution scheme to effectively solve the governing set of differential algebraic equations of incompressible poroelastodynamics. Thereby, a two-dimensional, biphasic, saturated porous medium model with intrinsically coupled and incompressible solid and fluid constituents is considered.

Our schemes are based on some well-accepted CFD techniques, originally developed for the efficient simulation of incompressible flow problems, and characterized by the following aspects: (1) a special treatment of the algebraically coupled volume balance equation leading to a reduced form of the boundary conditions; (2) usage of a higher-order accurate mixed LBB-stable finite element pair with piecewise discontinuous pressure for the spatial discretization; (3) application of the fully implicit 2nd-order Crank-Nicolson scheme for the time discretization; (4) use of a special fast multigrid solver of Vanka-type smoother available in FEATFLOW to solve the resulting discrete linear equation system. Furthermore, a new adaptive time stepping scheme combined with Picard iteration method is introduced to solve a non-linear elastic problem with special hyper-elastic model.

For the purpose of validation and to expose the merits and benefits of our new solution strategies in comparison to other established approaches, canonical one- and two-dimensional wave propagation problems are solved and a large-scale, dynamic soil-structure interaction problem serves to reveal the efficiency of the special multigrid solver and to evaluate its performance for different formulations.

To my parents, wife,
my sons: Muhammad (2008), Abdullah (2013) and Ahmad (2014) and my daughter Khadija (2016)

Acknowledgments

I am sincerely grateful to my supervisor Prof. Dr. Stefan Turek, for his kindness, support, advice, and motivation during my research work which led me quickly to have full understanding of my research task. As my teacher, I attribute to him my knowledge in numerical techniques throughout the valuable math lectures regularly delivered by him.

Prof. Turek's innovative ideas opened for me the avenue to the world of scientific research and publication. I am very thankful for his time, devotion, encouragement, positive and constructive critique and optimism while providing good working environment.

Many thanks go to Dr. Hilmar Wobker for introducing orally and through a written report my initial PhD research task and providing me with his implementation of the elasticity problem into FEATFLOW2 which was my initial step toward mastering FEAFLOW. Thanks to Dr. Mohamed El-Borhamy who shortly introduced devisor grid (the domain and mesh generator of FEATFLOW) in the initial phase of my PhD studies.

I would like to express my heartiest regards and blessings to FEATFLOW team for constant support in the stage of implementing the presented model into FEATFLOW. In particular, Prof. Turek and Hilmar Wobker for suggesting a test procedure that helped me to check the correctness of the implementation of my schemes, the main developers Michael Köster and Peter Zajac for answering general questions and helping to remove very hard types of bugs to track that tortured me for weeks, Jens Acker and Sven Buijssen for answering some questions about FEATFLOW postprocessing tool and creation of simulation movies and Dominik Göddeke and Christian Becker for their nice manual about preprocessing tool of FEATFLOW, the so-called devisor grid.

I would also like to thank my supervisor Prof. Stefan Turek and our colleagues from Aachen university, Prof. Bernd Markert and Dr. Yousef Heider, for taking the time and energy to revise our joined papers and for the insightful comments and professional, editorial advice that help improve them.

Special thanks go to people like Sven Buijssen and Christian Becker for constant administrative support and to Jens Acker for sharing and bearing me in same room and helping me resolve some miscellaneous technical problems. In particular, those related to Linux.

I am unable to find the appropriate words to properly express the deep feeling of gratitude to my parents for always encouraging me and for the prayers and well wishes to continue my studies.

Finally, I would like to express my gratitude to the German Research Foundation (DFG) for partial financial support under grant TU102/34-1 and grant TU102/34-2 and for TU Dortmund for helping me to fund myself by giving me an opportunity to work.

Dortmund, December 11, 2016

Abdulrahman Sadeq Obaid

Contents

1 Introduction

The theoretical and numerical investigation of fluid-saturated porous solid behavior under dynamic or quasi-static loading is a challenging and important application in different fields of engineering, such as in Geomechanics and Biomechanics, see, e. g., [60, 61, 50, 49]. The behavior of such materials is mainly governed by the mutual interaction between the solid skeleton and the pore fluid (see [23, 37]) which leads to algebraic equations representing some essential side condition or Lagrangian constraint, for instance, forcing incompressibility or continuity. This typically results in ill-conditioned saddle-point problems.

From a computational perspective, the most difficult situation (see [63, 62]) is given in dynamic problems if the solid and fluid constituents are materially incompressible, the hydraulic conductivity is very low implying a strong coupling and no restriction to the considered frequency range exists, such that reduced displacement-pressure formulations are not feasible. In addition, the pore-fluid pressure as algebraic variable takes over the role of a Lagrange multiplier associated with the continuity-like volume balance yielding a system of differential-algebraic equations (DAEs) of higher differentiation index, readily complicating the numerical solution [63].

At the time when I started my PhD study, there had been a recent publication, [63], by Markert and others, who introduced two finite element solution strategies which were implemented into the finite element package PANDAS to treat some dynamical problems of porous media. Many challenges were resolved and some others have remained issues. For instance, (1) the problem of finding LBB stable finite element pair producing accurate solutions in small layers below loaded permeable boundaries, in particular, in case of strong coupling, (2) the problem of pressure instability associated with the application of Crank-Nicholson method, (3) the divergence of the proposed methods when proceeding from momentum balance of the whole mixture to avoid the need for separation of the boundary conditions and have the more convenient problem, (4) the problem of the arising volume efflux as third boundary condition which prevents the boundary conditions from being chosen independently.

The above problems together with the high demands for sophisticated fast solver to handle systems with large condition numbers and unknowns as well as the need for an adaptive time stepping scheme in combination with special non-linear solver to avoid the computation of the expensive elastic material tangent matrix is the motivation and the main objective of this work.

My supervisor, Prof. Turek, therefore asked me to solve the porous media problems discussed in [63, 50] using the free finite element package FEATFLOW in order to see if FEATFLOW is capable of resolving the above issues and to compare with the results of PANDAS. The goal is to see if some special CFD techniques available in FEATFLOW are able to overcome the problems discussed in [63]. These techniques include the use of Q2/P1 element, the fully implicit Crank-Nicolson time integrator and multigrid solver with special Vanka smoother and for the nonlinear elastic problem, we additionally introduce an adaptive time stepping time integration scheme in combination with pure Picard iteration method to solve a problem with special hyper-elastic model.

On this basis, I started up my PhD study in which two papers ([106, 71]) were published in collaboration with our colleagues from RWTH Aachen University, Prof. Markert and Dr. Heider, where two formulations were solved. In this dissertation, I show the important results of these papers as well as new results for comparing and evaluating the multigrid performance for our linear algorithms.

This dissertation is structured as follows. Chapter 2 presents the basics of the TPM approach and provides the governing model equations of poro-elastodynamics. Chapter 3 introduces the concerned initial boundary value problems and describes the numerical treatment of the coupled problem including the weak formulation, the spatial and temporal discretization as well as the final matrix system. Chapter 4 is concerned with the numerical validation and evaluation of the proposed solution strategies and comparison with published data. Chapter 5 gives a brief summary and conclusions of the presented research work and future insights. Finally, I provide a short Appendix in which I discuss the existence and uniqueness as well as the stability of our problem and I further provide detailed description for the calculation of the internal stress power.

2
Theoretical Background

A porous medium, such as sponge, ceramic, foam, soil, biological tissue, etc., is a solid material (often referred to as solid matrix or solid constituent) with uniform or random interconnected networks of pores (voids) filled with fluids. Poromechanics is specialized in studying the response of deformable fluid-saturated porous media under external loads and poroelasticity is a branch of poromechanics concerned with porous media of elastic solid constituent. This chapter aims to give an easy and simple introduction to some fundamentals of poromechanics with emphasis on poroelasticity under dynamic loading. This includes the porous media modeling approach, the corresponding kinematics and kinetics as well as the equilibrium and constitutive relationships.

The author tried his best to make this introduction, almost self-contained, simple and more accessible to people new to this field in particular people from mathematics departments, who may not be specialized in this branch of engineering science. The presentation is limited to binary aggregates, composed of a porous solid skeleton (constituent φ^S) saturated by a single pore fluid (constituent φ^F). The solid material is assumed to be linear elastic or hyper-elastic, non-polar and incompressible while the fluid material is assumed to be incompressible and viscous. The two-phase model is further simplified by assuming isotropic permeability and excluding the thermal effects and mass exchanges between the constituents.

As this thesis concerns the numerical treatment of the related PDEs, the detailed physical description is mostly restricted to the above assumptions and simplifications. For historical development of the science of continuum mechanics, we refer to [99] and for outstanding elementary textbooks in this field we used [79, 41, 43, 85, 84, 17, 55, 44, 58, 51, 18], where the last contains a basic chapter on continuum mixture theory. We also benefit from Prof. Rebecca Brannon's excellent draft books on tensors [12], deformations [13] and rotations [11] and Ogden textbooks such as [75] and many others will be mentioned later. Clifford Ambrose Truesdell is the pioneer of continuum mechanics and had great contributions that made him the grandfather of this science and here are some of his famous books we used [102, 103, 97, 96, 100, 94, 101]. In fact most of modern presentations of theory of mixtures (including the general form of the entropy inequality of mixtures) are attributed to work of Truesdell and Bowen, see for example, [15, 102, 7, 10]. For further references on theory of porous media, interested readers may also consult [54, 20, 21, 37, 27, 38, 24] where the last contains intensive study on the historical

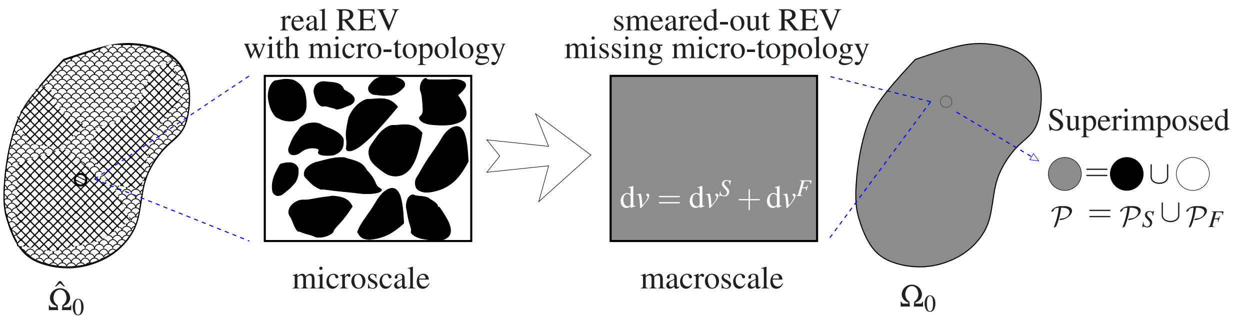


Figure 2.1: REV of saturated binary porous medium showing irregular internal micro-structure and the smeared-out model with overlapped constituents. The point $\mathcal{P}$ is picked arbitrarily from Ω_0.

development and the current state of porous media.

2.1 Theory of Porous Media (TPM)

Beside discussing the TPM, the main purpose of this section is to precisely define the reference configuration Ω_0 we are going to deal numerically with, which in most cases will be a simplified version of the actual complex structure $\hat{\Omega}_0$ of the considered porous medium.

2.1.1 Mixture Theory

The internal geometry of the solid skeleton is (in most cases) of random pattern which remains totally unknown. For example, soils consist of a mixture of particles of different sizes and arbitrary and irregular shapes, the internal structure of bones and sponges and the cells and cavities of both polymer and metal foams may be completely irregular. Hence, a detailed description of the pore shapes and sizes are impossible except for some man-made porous objects.

To handle this sort of problems, it is a standard practice to pick a sub-region (usually refereed to as RVE [1]) from the actual porous medium $\hat{\Omega}_0$. The dimensions of this sub-region must be relatively large compared to the characteristic pore sizes to allow for a valid statistical statement but extremely small compared to the size of the considered domain, so that we can obtain macroscopic state variables (pressure, velocities, densities..., etc.).

This sub-region RVE is then looked at (in imagination) as heterogeneous mixture, of two constituents (solid: φ^S and fluid: φ^F) with volume fractions n^S and n^F. After that, a real or virtual mathematical averaging is applied on this RVE to get smeared-out (blended) RVE, in which the micro-topology information (i. e., the geometrical description of the internal micro-

[1] RVE is an acronym for representative volume element

structure and the local position of the individual constituents) are lost [2]. The smeared-out RVE will represent each continuum (or macroscopic) point $\mathcal{P}$ in a new substitute domain denoted by Ω_0 and whose geometry will be defined in the subsequent subsection. The homogeneous simplified substitute domain Ω_0 will be the initial (undeformed or reference) configuration we deal numerically with.

Obviously, each point $\mathcal{P}$ in this 'theoretically well-mixed' (i. e., statistically averaged) representative domain $\mathcal{P} \in \Omega_0$ is assumed to be concurrently occupied by two materially-different particles, one for the solid constituent $\mathcal{P}_S$ and one for the fluid constituent $\mathcal{P}_F$ (see Figue 2.2). That is,

$$\mathcal{P} = \mathcal{P}_S \cup \mathcal{P}_F. \tag{2.1}$$

Moreover, each of the above two local particles will definitely interact with its partner and will also follow a separate motion path. In the course of the body deformation they will generally not continue to share the same spatial point. They will probably separate and each of them will join another partner of different phase and position. For more information about the classical Theory of Mixtures, we refer the reader to [15, 98, 24, 7, 95] and citations therein. Discussing the averaging and homogenization procedures is out of scope of this thesis and interested readers are referred to [24, 25, 47, 48, 90, 27, 78, 16, 83] and for the quantitative evaluation of the REV, we refer to [30] and [31].

2.1.2 Concept of volume fractions

There are still two important questions from the previous subsection left unanswered; (1) what is the geometry of substitute domain Ω_0? (2) Can we recover some kinds of micro-topology information, lost in the smeared-out RVE? For the first question, we assume that the control space Ω_0 is modeled and spanned by the porous solid and that only the pore fluid can escape the control space.

For the second question, the missing micro-topology information is somehow recaptured using the 'concept of volume fractions', in which we introduce (space- and time-dependent) volume functions n^S and n^F to recover the actual volumes of the solid phase and fluid phase:

$$\mathrm{d}v^S = n^S \,\mathrm{d}v \qquad \text{and} \qquad \mathrm{d}v^F = n^F \,\mathrm{d}v \tag{2.2}$$

Here, dv indicates the differential volume element associated with point $\mathcal{P}$ (see figure 2.2), which consists of two partial volume elements (dv^S and dv^F) corresponding to the solid particle $\mathcal{P}^S$ and fluid particle $\mathcal{P}^F$. By the saturation condition, all the pores are assumed to be filled with fluid. Thus,

$$dv = dv^S + dv^F \quad \rightarrow \quad n^S + n^F = 1. \tag{2.3}$$

Obviously the initial values $n_0^S = n^S(t_0)$ and $n_0^F = n^F(t_0)$ are homogeneously distributed, because they are connected to the (homogenized) initial configuration Ω_0.

[2] This forms a basis for the continuum (macroscopic) approach to describe the porous medium, because in continuum approaches, we ignore the description on the micro-level.

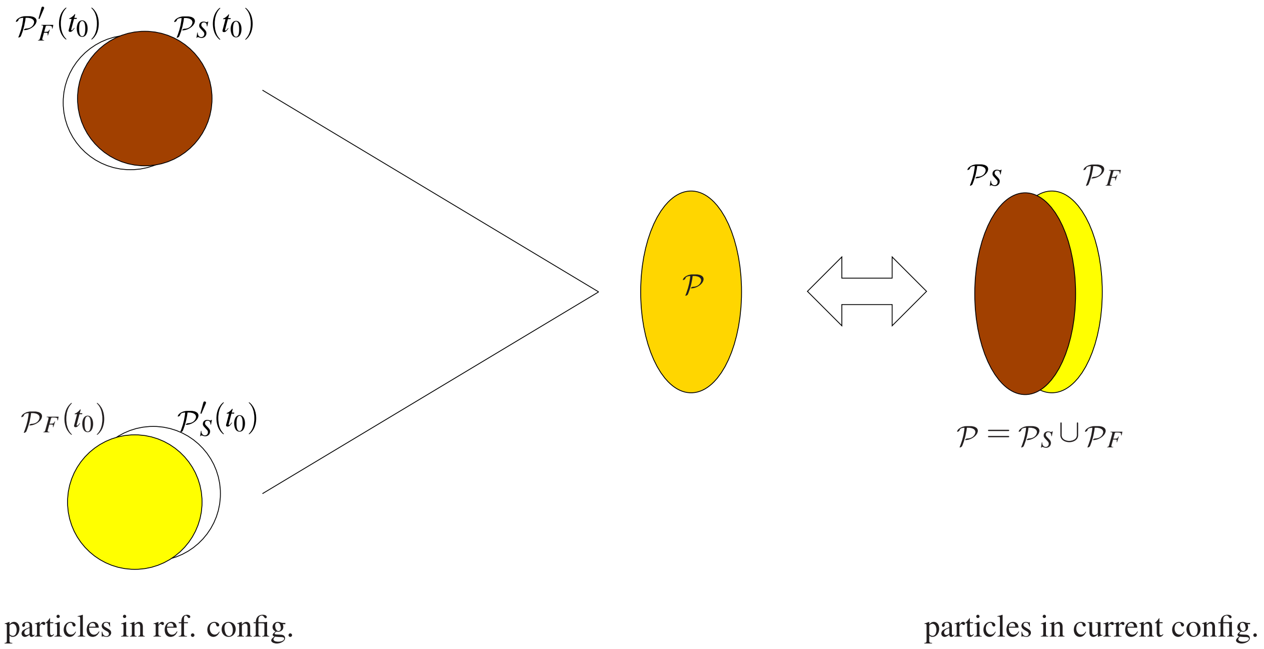


Figure 2.2: Two materially different particles ($\mathcal{P}_S(t_0)$ and $\mathcal{P}_F(t_0)$) with distinct positions in the reference configuration deformed (to $\mathcal{P}_S = \mathcal{P}_S(t)$ and $\mathcal{P}_F = \mathcal{P}_F(t)$) and met at one spatial point $\mathcal{P}$ in the current configuration. Particle $\mathcal{P}_S$ and $\mathcal{P}_F$ occupy together a differential volume dv. Particle $\mathcal{P}_S(t_0)$ is associated with the differential volume dV_S and differential area dA_S. Similarly, particle $\mathcal{P}_F(t_0)$ is associated with the differential volume dV_F and differential area dA_F

The introduction of the concept of volume fractions gives rise to two types of density functions:

- material density:

$$\rho^{SR} = \frac{\mathrm{d}m^S}{\mathrm{d}v^S} \qquad \text{and} \qquad \rho^{FR} = \frac{\mathrm{d}m^F}{\mathrm{d}v^F}\,, \tag{2.4}$$

- partial density:

$$\rho^{S} = \frac{\mathrm{d}m^S}{\mathrm{d}v} \qquad \text{and} \qquad \rho^{F} = \frac{\mathrm{d}m^F}{\mathrm{d}v}\,, \tag{2.5}$$

where

$$\rho^{\alpha R} = \text{const.}, \qquad \alpha \in \{S, F\}\,, \tag{2.6}$$

due to the material incompressiblity assumption. Now, by virtue of (2.2), (2.4) and (2.5), the partial density functions ρ^S and ρ^F can be expressed by

$$\rho^S = \rho^{SR}\, n^S \qquad \text{and} \qquad \rho^F = \rho^{FR}\, n^F\,, \tag{2.7}$$

which reveals the deformation dependency of the partial densities through the volume fractions n^S and n^F. Mills [64] and Morland [66] were the first scientists who used the concept of volume fractions together with the mixture theory. Mills used the volume fraction concept for mixtures of fluids only, whereas Morland was the first to incorporate this concept in porous media.

2.2 Continuum Kinematics

2.2.1 Particle derivatives

In poromechanics, the biphasic porous medium (see Figure 3.2.1) is seen as a collection of points $\mathcal{P}$, each of which is simultaneously occupied by a solid particle ($\mathcal{P}_S$) and a fluid particle ($\mathcal{P}_F$) as stated in (2.1) and shown in Figure 2.2. Following the standard notations, we shall use the vector [3] $\mathbf{x} = [x_1\; x_2\; x_3]^T$ to refer to the position of any spatial point in the current configuration Ω. Two phase-different (or materially different) particles $\mathcal{P}_S$ and $\mathcal{P}_F$, which accidentally found to meet at position $\mathbf{x}$ in the current configuration are generally coming from different locations in the reference configurations because according to TPM, each constituent moves differently and follows a separate motion path as already depicted in Figure 2.2. Thus, each point $\mathbf{x}$ in the

[3] All vector fields in this thesis are written with respect to the standard orthonormal Schauder basis $\{e_i\}$ of the d-dimensional Euclidean space R^d. For instance,

$$\mathbf{x} = \sum_{i=1}^{d} x_i\, e_i \quad \text{with} \quad x_i = \mathbf{x} \cdot e_i\,, \tag{2.8}$$

where '$\cdot$ 'denotes the scalar (dot) product.

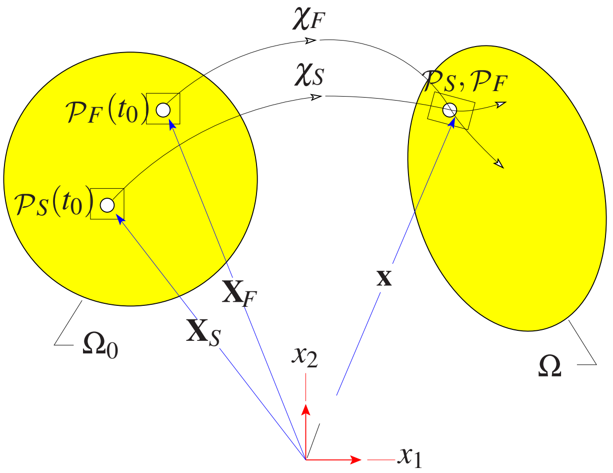


Figure 2.3: Biphasic solid-fluid mixture with individual motion paths. Reference configuration (left) and current configuration (right). See Figure 2.2 for definition of some symbols.

current configuration is generally related to two position vectors in the reference configuration Ω_0; $\mathbf{X}_S$ to characterize particle $\mathcal{P}_S(t_0)$ and $\mathbf{X}_F$ to characterize particle $\mathcal{P}_F(t_0)$ (see Figure 2.2). Now when computing the time derivative of any physical quantity $\eta(\mathbf{x},t)$ defined on the domain, we usually apply the standard Eulerian time derivative,

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial \mathbf{x}} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}, \quad \text{where} \quad \eta = \eta(\mathbf{x},t). \tag{2.9}$$

However, the above relation needs to be modified because, as stated before, at position $\mathbf{x}$, there are particle $\mathcal{P}_S$ and particle $\mathcal{P}_F$ each moving differently and probably going to separate in the next time step. Accordingly, when computing the above $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$, we have to decide which particle in $\mathbf{x}$ we are going to follow. Therefore, we normally replace $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$ with $\frac{\mathrm{d}_\alpha \mathbf{x}}{\mathrm{d}t}$ to indicate the particle we follow. Hence, $\frac{\mathrm{d}_\alpha}{\mathrm{d}t}$ is defined as the time derivative following particle $\mathcal{P}^\alpha$. Based on that, (2.9) is modified to

$$(\eta)'_\alpha = \frac{\mathrm{d}_\alpha \eta}{\mathrm{d}t} = \frac{\partial \eta}{\partial t} + \frac{\partial \eta}{\partial \mathbf{x}} \underbrace{\frac{\mathrm{d}_\alpha \mathbf{x}}{\mathrm{d}t}}_{=\frac{\partial \mathbf{x}_\alpha}{\partial t}}, \quad \text{where} \quad \alpha \in \{S,F\} \quad \text{and} \quad \eta = \eta(\mathbf{x},t). \tag{2.10}$$

The above modified setting is known as the material (or particle) time derivative of $\eta(\mathbf{x},t)$ following the motion of φ^α. The displacement $\mathbf{u}_\alpha$ of particle $\mathcal{P}_\alpha$ is simply a measure of how far that particle has moved from its initial position in the reference configuration and hence, defined as

$$\mathbf{u}_\alpha = \mathbf{x} - \mathbf{X}_\alpha, \quad \text{where} \quad \alpha \in \{S,F\}. \tag{2.11}$$

The particle derivative of the above displacement leads to the definition of the so-called particle velocity and particle acceleration fields

$$\mathbf{v}_\alpha = \overset{\prime}{\mathbf{x}}_\alpha = \frac{\mathrm{d}_\alpha \mathbf{x}}{\mathrm{d}t} = \frac{\partial \mathbf{x}_\alpha}{\partial t} \quad \text{and} \quad (\mathbf{v}_\alpha)'_\alpha = \overset{\prime\prime}{\mathbf{x}}_\alpha \,. \tag{2.12}$$

Equation (2.10) can be extended from scalar $\eta(\mathbf{x},t)$ to any tensor quantity, so that the particle derivative works as follows:

$$(\,\cdot\,)'_\alpha := \frac{\mathrm{d}_\alpha(\,\cdot\,)}{\mathrm{d}t} = \frac{\partial(\,\cdot\,)}{\partial t} + \operatorname{grad}(\,\cdot\,)\cdot \mathbf{v}_\alpha \,. \tag{2.13}$$

2.2.2 Deformation gradient

The deformation gradient is known as the best measure for deformation, from which we can extract many information about length change, area change, volume change, rotation and some more things which will be partially discussed in this subsection. Recall that the deformation gradient $\mathbf{F}_\alpha$ maps any differential length vector $\mathrm{d}\mathbf{X}_\alpha$ in the reference configuration Ω_0 to its new length vector $\mathrm{d}\mathbf{x}$ in the current configuration via the relation

$$\mathrm{d}\mathbf{x} = \underbrace{\frac{\partial \mathbf{x}}{\partial \mathbf{X}_\alpha}}_{=\mathbf{F}_\alpha} \mathrm{d}\mathbf{X}_\alpha \,. \tag{2.14}$$

Then, the volume $\mathrm{d}v$ (read the caption of Figure 2.2) of any differential parallelepiped in the current configuration spanned by the three independent vectors $(\mathrm{d}\mathbf{x}_1, \mathrm{d}\mathbf{x}_2, \mathrm{d}\mathbf{x}_3)$ is connected to the differential size $\mathrm{d}V_\alpha$ in the reference configuration using the triple product and equation (2.14)

$$\begin{aligned} \mathrm{d}v &= \mathrm{d}\mathbf{x}_1 \cdot (\mathrm{d}\mathbf{x}_2 \times \mathrm{d}\mathbf{x}_3) \\ &= |\mathrm{d}\mathbf{x}_1, \mathrm{d}\mathbf{x}_2, \mathrm{d}\mathbf{x}_3| \\ &= |\mathbf{F}_\alpha\, \mathrm{d}\mathbf{X}_{\alpha_1}, \mathbf{F}_\alpha\, \mathrm{d}\mathbf{X}_{\alpha_2}, \mathbf{F}_\alpha\, \mathrm{d}\mathbf{X}_{\alpha_3}| \\ &= \underbrace{|\mathbf{F}_\alpha|}_{J_\alpha}\; \underbrace{|\mathrm{d}\mathbf{X}_{\alpha_1}, \mathrm{d}\mathbf{X}_{\alpha_2}, \mathrm{d}\mathbf{X}_{\alpha_3}|}_{\mathrm{d}V_\alpha} = J_\alpha\, \mathrm{d}V_\alpha \end{aligned} \tag{2.15}$$

or shortly,

$$\mathrm{d}v = J_\alpha\, \mathrm{d}V_\alpha \,. \tag{2.16}$$

Herein, $J_\alpha = |\mathbf{F}_\alpha|$ is the determinant of the deformation gradient $\mathbf{F}_\alpha$ and called in the literature the Jacobian [4]. Since the mass cannot be destroyed according to the law of conservation of

[4] The name 'Jacobian' is used in mutlivariable calculus when making change of variables for multiple integrals as we do, for instance, in finite elements if we assume that $\mathbf{X}$ is the position vector in the element local coordinates and $\mathbf{x}$ is the position vector in the global coordinates.

mass, it is essential that J_α does not equal to zero (and hence $\mathbf{F}_\alpha$ must be invertible or "globally admissible") otherwise a vanishing volume is not physically realistic. Therefore, material inter-penetration is not only physically unacceptable but also mathematically prohibited since this will allow for two different solid particles in the reference configuration to be mapped into one particle (as a result of being merged together) or in other words, the mapping is not one-to-one and hence it is not invertible. The so-called contact algorithms in computational mechanics are used to ensure the global admissibility. Another important condition is called the "compatibility condition", which ensures the existence of the mapping $\mathbf{F}_\alpha$. The occurrence of the cracks is an example that violates this condition because at the crack one particle is broken into two separate ones or in other words, one solid particle in the reference configuration is mapped into two solid particles in the current configuration. Mathematically, this means that $\mathbf{F}_S$ is no longer a mapping and the cracks are automatically avoided by adopting the so-called Lipschitz domain. Furthermore, $J_\alpha = |\mathbf{F}_\alpha|$ must be positive because realistic volumes ($\mathrm{d}V_\alpha$ and $\mathrm{d}v$) cannot be negative quantities and because the differential volume in (2.16) was generated through triple product, one must take care that the three vectors are oriented such that this product remains positive. Therefore, no single material should be permitted to invert as this will invert the direction of the cross-products in (2.15) turning $\mathrm{d}v$ into negative quantity. The last mentioned requirement is referred to as "local admissibility" of the deformation and in finite element simulations it is ensured through the so-called "hourglass control". In addition, the choice of finite element test functions vanishing at the element boundaries is necessary to obtain admissible $\mathbf{F}_\alpha$. Based on what mentioned, **a deformation gradient that fulfills the global admissibility, local admissibility and compatibility condition must be a positive definite tensor**.

Morland in [66] pointed out that the deformation gradient of solid matrix can be multiplicatively decomposed into spherical part and and partial density preserving (or partial volume preserving in case of our material incompressibility assumption) part as follows:

$$\mathbf{F}_S = \left(J_S^{1/3}\,\mathbf{I}\right)\bar{\mathbf{F}}_S \quad \Leftrightarrow \quad \bar{\mathbf{F}}_S = J_S^{-1/3}\mathbf{F}_S\,. \tag{2.17}$$

Herein, $\bar{\mathbf{F}}_S$ is obviously the partial volume preserving part because

$$\det\bar{\mathbf{F}}_S = J_S = 1 \quad \Rightarrow \quad n^S = n^S_{0S} \tag{2.18}$$

and $J_S^{1/3}\,\mathbf{I}$ is the spherical part with

$$\det\left(J_S^{1/3}\,\mathbf{I}\right) = J_S\,. \tag{2.19}$$

The deformation gradient can be used to link any differential mixture area vector $\mathrm{d}a$ in the current configuration to its constituents differential areas $\mathrm{d}A_\alpha$ in the reference configuration. To illustrate this, we first use (2.14) and the triple product to link $\mathrm{d}v$ to $\mathrm{d}a$:

$$\begin{aligned} \mathrm{d}v &= \mathrm{d}\mathbf{x}_1 \cdot (\mathrm{d}\mathbf{x}_2 \times \mathrm{d}\mathbf{x}_3) \\ &= \mathrm{d}\mathbf{x}_1 \cdot \mathrm{d}a = \mathrm{d}\mathbf{x}_1^T \mathrm{d}a \\ &= (\mathbf{F}_\alpha\, \mathrm{d}\mathbf{X}_{\alpha_1})^T \mathrm{d}a = (\mathrm{d}\mathbf{X}_{\alpha_1}^T \mathbf{F}_\alpha^T)\mathrm{d}a\,, \end{aligned} \tag{2.20}$$

and similarly (by means of triple product), $\mathrm{d}V^{\alpha}$ is linked to $\mathrm{d}A_{\alpha}$;

$$\begin{aligned}\mathrm{d}V^{\alpha} &= \mathrm{d}\mathbf{X}_{\alpha_1} \cdot (\mathrm{d}\mathbf{X}_{\alpha_2} \times \mathrm{d}\mathbf{X}_{\alpha_3}) \\ &= \mathrm{d}\mathbf{X}_{\alpha_1} \cdot \mathrm{d}A_{\alpha} = \mathrm{d}\mathbf{X}_{\alpha_1}^T \mathrm{d}A_{\alpha} \,.\end{aligned} \tag{2.21}$$

The sought-after relation is then obtained after substituting (2.20) and (2.21) in (2.16), which yields the so-called Nanson‘s formula:

$$\begin{aligned}\underbrace{(\mathrm{d}\mathbf{X}_{\alpha_1}^T \mathbf{F}_{\alpha}^T)\mathrm{d}a}_{\mathrm{d}v} &= J_{\alpha} \underbrace{\mathrm{d}\mathbf{X}_{\alpha_1}^T \mathrm{d}A_{\alpha}}_{\mathrm{d}V^{\alpha}} \,, \\ \mathbf{F}_{\alpha}^T \mathrm{d}a &= J_{\alpha} \mathrm{d}A_{\alpha} \,, \\ \mathrm{d}a &= J_{\alpha} \mathbf{F}_{\alpha}^{-T} \mathrm{d}A_{\alpha} \,.\end{aligned} \tag{2.22}$$

In addition to providing information about volume and area changes, the deformation gradient can also be used to compute the stretches as well as the rotations. This is based on the fact that any differential length vector $\mathrm{d}\mathbf{X}_S$ in the reference configuration can be written as

$$\mathrm{d}\mathbf{X}_S = \|\mathrm{d}\mathbf{X}_S\|_2 \,\tilde{\mathbf{N}} \qquad \text{with} \qquad \tilde{\mathbf{N}} = \frac{\mathrm{d}\mathbf{X}_S}{\|\mathrm{d}\mathbf{X}_S\|_2} \,. \tag{2.23}$$

Herein, $\|\mathrm{d}\mathbf{X}_S\|_2$ is the Euclidean norm of the vector $\mathrm{d}\mathbf{X}_S$ and $\tilde{\mathbf{N}}$ is a unit vector in the direction of $\mathrm{d}\mathbf{X}_S$. Next, using (2.14) and then (2.23), we obtain

$$\begin{aligned}\|\mathrm{d}\mathbf{x}\|_2^2 &= \mathrm{d}\mathbf{x}^T \;\mathrm{d}\mathbf{x} \\ &= (\mathbf{F}_S \;\mathrm{d}\mathbf{X}_S)^T \,(\mathbf{F}_S \;\mathrm{d}\mathbf{X}_S) \\ &= \left(\mathbf{F}_S \;\|\mathrm{d}\mathbf{X}_S\|_2 \tilde{\mathbf{N}}\right)^T \left(\mathbf{F}_S \;\|\mathrm{d}\mathbf{X}_S\|_2 \tilde{\mathbf{N}}\right) \\ &= \|\mathrm{d}\mathbf{X}_S\|_2^2 \;\tilde{\mathbf{N}}^T \left(\mathbf{F}_S^T \;\mathbf{F}_S\right) \tilde{\mathbf{N}} \,.\end{aligned} \tag{2.24}$$

Dividing both sides of (2.24) by $\|\mathrm{d}\mathbf{X}_S\|_2^2$, we conclude that the stretch λ is given such that

$$\lambda^2 = \frac{\|\mathrm{d}\mathbf{x}\|_2^2}{\|\mathrm{d}\mathbf{X}_S\|_2^2} = \tilde{\mathbf{N}}^T \left(\mathbf{F}_S^T \;\mathbf{F}_S\right) \tilde{\mathbf{N}} \,. \tag{2.25}$$

Through any particle, locating at $\mathbf{X}_S$, in a general three-dimensional reference configuration, we can draw an infinite number of unit vectors $\tilde{\mathbf{N}}$, each of which will in general experience different stretch values. Here, we are interested in the directions $\tilde{\mathbf{N}}$ that give the extreme values of stretches. For this purpose, we need to find the extreme values of the RHS of (2.25) by solving the following optimization problem,

$$\begin{aligned}&\text{extremize} \quad \mathrm{obj} = \tilde{\mathbf{N}}^T \left(\mathbf{F}_S^T \;\mathbf{F}_S\right) \tilde{\mathbf{N}} \\ &\text{subject to} \quad \|\tilde{\mathbf{N}}\| = \tilde{\mathbf{N}}^T \;\tilde{\mathbf{N}} = 1 \,.\end{aligned} \tag{2.26}$$

The resulting Lagrange function reads

$$L = \tilde{\mathbf{N}}^T \left(\mathbf{F}_S^T \, \mathbf{F}_S\right) \tilde{\mathbf{N}} + \lambda^2 \left(1 - \tilde{\mathbf{N}}^T \, \tilde{\mathbf{N}}\right) , \tag{2.27}$$

where λ^2 is the Lagrange multiplier. To find the extreme values of the function obj, we need to solve the following Kuhn-Tucker condition:

$$\frac{\partial L}{\partial \tilde{\mathbf{N}}} = \mathbf{0} \quad \Rightarrow \quad \left(\mathbf{F}_S^T \, \mathbf{F}_S\right) \, \tilde{\mathbf{N}} = \lambda^2 \tilde{\mathbf{N}} . \tag{2.28}$$

The solution is simply the three eigenvalues of $\mathbf{F}_S^T \, \mathbf{F}_S$. In continuum mechanics literature, the square roots of these eigenvalues (i. e., λ_1, λ_2, λ_3) are called "principal stretches" and the corresponding eigenvectors ($\tilde{\mathbf{N}}_1$, $\tilde{\mathbf{N}}_2$ and $\tilde{\mathbf{N}}_3$) are called the principal axes. Using Einstein's summation convention [5], we can write

$$\mathbf{F}_S^T \, \mathbf{F}_S = \lambda_i^2 \, \tilde{\mathbf{N}}_\mathbf{i} \otimes \tilde{\mathbf{N}}_\mathbf{i} . \tag{2.29}$$

Remember from (2.23) that the orthonormal eigenvectors $\mathbf{N}_\mathbf{i}$ are associated with $\mathrm{d}\mathbf{X}_S$'s in the reference configuration. To compute the stretches with respect to the current configuration, we follow a similar procedure. Namely, we first define a unit vector in the current configuration so that

$$\mathrm{d}\mathbf{x} = \|\mathrm{d}\mathbf{x}\|_2 \, \tilde{\mathbf{n}} \qquad \text{with} \qquad \tilde{\mathbf{n}} = \frac{\mathrm{d}\mathbf{x}}{\|\mathrm{d}\mathbf{x}\|_2} . \tag{2.30}$$

Then, making use of (2.14), we obtain

$$\begin{aligned}
\|\mathrm{d}\mathbf{X}_S\|_2^2 &= \mathrm{d}\mathbf{X}_S^T \, \mathrm{d}\mathbf{X}_S \\
&= \left(\mathbf{F}_S^{-1} \, \mathrm{d}\mathbf{x}\right)^T \left(\mathbf{F}_S^{-1} \, \mathrm{d}\mathbf{x}\right) \\
&= \left(\mathbf{F}_S^{-1} \, \|\mathrm{d}\mathbf{x}\|_2 \, \tilde{\mathbf{n}}\right)^T \left(\mathbf{F}_S^{-1} \, \|\mathrm{d}\mathbf{x}\|_2 \, \tilde{\mathbf{n}}\right) \\
&= \|\mathrm{d}\mathbf{x}\|_2^2 \, \tilde{\mathbf{n}}^T \left(\mathbf{F}_S \, \mathbf{F}_S^T\right)^{-1} \tilde{\mathbf{n}} .
\end{aligned} \tag{2.31}$$

Hence,

$$\lambda^{-2} = \frac{\|\mathrm{d}\mathbf{X}_S\|_2^2}{\|\mathrm{d}\mathbf{x}_S\|_2^2} = \tilde{\mathbf{n}}^T \left(\mathbf{F}_S \, \mathbf{F}_S^T\right)^{-1} \tilde{\mathbf{n}} . \tag{2.32}$$

With the extreme values λ_1^{-2}, λ_2^{-2} and λ_3^{-2} being the eigenvalues of the problem below

$$\left(\mathbf{F}_S \, \mathbf{F}_S^T\right)^{-1} \, \tilde{\mathbf{n}}_\mathbf{i} = \lambda_i^{-2} \, \tilde{\mathbf{n}}_\mathbf{i} . \tag{2.33}$$

[5] This convention will always be used for indices j and i.

Thus,

$$\left(\mathbf{F}_S\,\mathbf{F}_S^T\right)^{-1} = \lambda_i^{-2}\,\tilde{\mathbf{n}}_\mathbf{i}\otimes\tilde{\mathbf{n}}_\mathbf{i} \quad\Leftrightarrow\quad \mathbf{F}_S\,\mathbf{F}_S^T = \lambda_i^2\,\tilde{\mathbf{n}}_\mathbf{i}\otimes\tilde{\mathbf{n}}_\mathbf{i}\,. \tag{2.34}$$

From spectral analysis, it is well known that the real-valued symmetric positive definite tensors $\mathbf{F}_S^T\,\mathbf{F}_S$ and $\mathbf{F}_S\,\mathbf{F}_S^T$ possess the same eigenvalues and are the squares of those of the real-valued positive definite $\mathbf{F}_S$. However, the principal axes ($\tilde{\mathbf{n}}_1$, $\tilde{\mathbf{n}}_2$ and $\tilde{\mathbf{n}}_3$) for $\mathbf{F}_S\,\mathbf{F}_S^T$ belong to the current configuration as defined in (2.30) and different from ($\tilde{\mathbf{N}}_1$, $\tilde{\mathbf{N}}_2$ and $\tilde{\mathbf{N}}_3$), which belong to the reference configuration as defined in (2.23). The right Cauchy-Green deformation tensor is defined as

$$\mathbf{C}_S = \mathbf{F}_S^T\,\mathbf{F}_S = \underbrace{\lambda_i^2\,\tilde{\mathbf{N}}_\mathbf{i}\otimes\tilde{\mathbf{N}}_\mathbf{i}}_{\text{cf. } (2.29)} \tag{2.35}$$

and the left Cauchy-Green deformation tensor is defined as

$$\mathbf{B}_S = \mathbf{F}_S\,\mathbf{F}_S^T = \underbrace{\lambda_i^2\,\tilde{\mathbf{n}}_\mathbf{i}\otimes\tilde{\mathbf{n}}_\mathbf{i}}_{\text{cf. } (2.34)}\,, \tag{2.36}$$

where the first refers to the reference configuration and the last to the current configuration. Beside the stretches, the deformation gradient provides information about the rotations. To extract such information, we recall that the singular value decomposition allows for multiplicative splitting of $\mathbf{F}_S$ into

$$\mathbf{F}_S = \mathbf{n}^T\,\Lambda\,\mathbf{N} \quad\Leftrightarrow\quad \mathbf{F}_S = \lambda_i\,\tilde{\mathbf{n}}_\mathbf{i}\otimes\tilde{\mathbf{N}}_\mathbf{i}\,, \tag{2.37}$$

where Λ is a diagonal tensor that contains the eigenvalues (λ_1, λ_2, λ_3) of $\mathbf{F}_S$ and $\mathbf{n}$ and $\mathbf{N}$ are orthogonal tensors such that $\mathbf{n} = [\tilde{\mathbf{n}}_1, \tilde{\mathbf{n}}_2, \tilde{\mathbf{n}}_3]$ and $\mathbf{N} = \left[\tilde{\mathbf{N}}_1, \tilde{\mathbf{N}}_2, \tilde{\mathbf{N}}_3\right]$. The above decomposition is unique and it follows that

$$\mathbf{F}_S = \mathbf{n}^T\,\Lambda\,\mathbf{N} = \underbrace{\mathbf{n}^T\,\Lambda\,\mathbf{n}}_{\mathbf{V}_S}\,\underbrace{\mathbf{n}^T\,\mathbf{N}}_{\mathbf{R}} = \mathbf{V}_S\,\mathbf{R}\,. \tag{2.38}$$

Herein, the symmetric positive definite $\mathbf{V}_S$ is, for obvious reason, called the left stretch tensor and the orthonormal $\mathbf{R}$ is the sought-after rotation tensor. Such a decomposition is usually referred to as left polar decomposition. Another possibility is to bring up the rotation first then the stretch. That is

$$\mathbf{F}_S = \mathbf{n}^T\,\mathbf{N}\,\mathbf{N}^T\,\Lambda\,\mathbf{N} = \underbrace{\mathbf{n}^T\,\mathbf{N}}_{\mathbf{R}}\,\underbrace{\mathbf{N}^T\,\Lambda\,\mathbf{N}}_{\mathbf{U}_S} = \mathbf{R}\,\mathbf{U}_S\,, \tag{2.39}$$

where the positive definite tensor $\mathbf{U}_S$ is named the right Cauchy stretch tensor. It is obvious that

$$\mathbf{C}_S = \mathbf{F}_S^T\,\mathbf{F}_S = \mathbf{U}_S^2 = \underbrace{\lambda_i^2\,\tilde{\mathbf{N}}_\mathbf{i}\otimes\tilde{\mathbf{N}}_\mathbf{i}}_{\text{cf. } (2.35)} \qquad\text{and}\qquad \mathbf{B}_S = \mathbf{F}_S\,\mathbf{F}_S^T = \mathbf{V}_S^2 = \underbrace{\lambda_i^2\,\tilde{\mathbf{n}}_\mathbf{i}\otimes\tilde{\mathbf{n}}_\mathbf{i}}_{\text{cf. } (2.36)}\,. \tag{2.40}$$

The above relation is essential for restricting the deformation energy function of the isotropic hyper-elastic materials as will be shown later. The deformation gradient $\mathbf{F}_S$ is also useful for computing the strain measures. To illustrate this, let us start with a simple one-dimensional problem, a bar subjected to uni-axial extension or compression. Let L_0 be the initial length and l be the current length of the bar, then we can define a strain measure as follows:

$$\varepsilon = \frac{l - L_0}{L_0} = \frac{l}{L_0} - \frac{L_0}{L_0} = \lambda - 1\,. \tag{2.41}$$

The above strain relation belongs to material strain measures because the above change in length is compared against the initial length L_0. If the comparison is against the current length, then we obtain the so-called spatial strain measure,

$$\varepsilon = \frac{l - L_0}{l} = \frac{l}{l} - \frac{L_0}{l} = -\left(\lambda^{-1} - 1\right)\,. \tag{2.42}$$

Notice from the power sign, we can easily differentiate between the material and spatial measures. In fact, the above two strain measures belong to Seth-Hill family of generalized strain tensors (cf. [86]), which possesses the general form

$$\varepsilon = \frac{1}{k}\left(\lambda^k - 1\right)\,, \quad \text{where} \quad k \in \mathbb{R}\,. \tag{2.43}$$

Observ that a negative real-valued k is obviously associated with strains measured with respect to the current configuration, while positive k indicates that the strain is measured with respect to the reference configuration as seen for special two cases right above. For example, for $k = 2$ and $k = 1/2$, we get the Green-Lagrangian strain and Biot strain, respectively, for $k = -2$, we obtain the so-called Almansi strain measure and for $k = 0$, we obtain the so-called (logarithmic) Hencky strain [6]

$$\varepsilon = \ln\lambda\,. \tag{2.46}$$

[6] Given that $\lambda \in \mathbb{R}^+$, (2.46) can be proven using the fact that

$$e^h = \sum_{n=0}^{\infty} \frac{h^n}{n!} = 1 + h + \frac{h^2}{2!} + \frac{h^3}{3!} + \cdots \quad \Rightarrow \quad \frac{e^h - 1}{h} = 1 + o(h)\,. \tag{2.44}$$

Hence,

$$\begin{aligned} \frac{\lambda^k - 1}{k} &= \ln(\lambda)\frac{e^{(k\;\ln(\lambda))} - 1}{(k\;\ln(\lambda))} \\ &= \ln(\lambda)\frac{e^h - 1}{h}\,, \quad \text{with} \quad h = k\;\ln(\lambda) \\ &= \ln(\lambda)\,(1 + o(h)) \\ &= \ln(\lambda) \quad \text{as} \quad h \to 0\,. \end{aligned} \tag{2.45}$$

Where λ is a fixed positive real as stated above.

However, if the deformation (stretch) is infinitesimal ($l \approx L_0 \Rightarrow \lambda \approx 1$), then all the previous measures have almost equal values. In the one-dimensional case, the Seth-Hill strain measures are based on λ, a ratio between lengths. For the three-dimensional case, one would naturally assume that an analogous strain measure would be based on the ratio between volumes, J_S (cf. , (2.16)) as follows

$$\varepsilon_{\text{vol}} = \begin{cases} \ln J_S & \text{if } k = 0\,, \\ \frac{1}{k}\left(J_S^k - 1\right) & \text{if } k \neq 0\,. \end{cases} \tag{2.47}$$

Note that the strains in (2.47), as being pure function of partial volume change J_S, are indeed volumetric measures that do not account for distortion (change in shape). A problem arises if we speak about the case of large stretching, in which the associated deformation energy W^S_{vol} is to be considered as linear function of ε_{vol}, for simplicity, say for example

$$W^S_{\text{vol}} = \mu_S\, \varepsilon_{\text{vol}} \qquad \text{where} \qquad \mu_S = \text{const.}. \tag{2.48}$$

Physically, a finite load cannot compress a compressible cube down to nothing (i. e., to $J_S = 0$) or expand it to infinity (i. e., to $\lambda = \infty$) and therefore, we theoretically assume that these two extreme cases would require infinite load or energy. Following this principle, if we considered k to be a fixed positive real number, then the non-logarithmic choices of (2.47) together with (2.48) would imply for $J_S \to 0$ we need a finite energy whereas negative choice of k, tells us that expanding the cube to infinity can be achieved by a finite energy in complete disagreement with this physical requirement. The logarithmic strain measure,

$$\varepsilon_{\text{vol}} = \ln J_S\,, \tag{2.49}$$

is the only choice that saves us from this dilemma. The above strain measure is recommended at large ratios of stretching (as in rubbers) and simply results from adding up the three principal logarithmic strains $\ln\lambda_1$, $\ln\lambda_2$ and $\ln\lambda_3$ corresponding to the three principal stretch ratios λ_1, λ_2 and λ_3:

$$\varepsilon_v = \ln\lambda_1 + \ln\lambda_2 + \ln\lambda_3 = \ln(\lambda_1\lambda_2\lambda_3) = \ln(\det\mathbf{F}_S) = \ln J_S\,. \tag{2.50}$$

Similarly, we can get more general strain if we add up the three principal Seth-Hill strains $\frac{1}{k}\left(\lambda_1^k - 1\right)$, $\frac{1}{k}\left(\lambda_2^k - 1\right)$ and $\frac{1}{k}\left(\lambda_3^k - 1\right)$ corresponding to the three principle stretch ratios λ_1, λ_2 and λ_3

$$\begin{aligned} \hat{\varepsilon}^k &= \frac{1}{k}\left(\lambda_1^k - 1\right) + \frac{1}{k}\left(\lambda_2^k - 1\right) + \frac{1}{k}\left(\lambda_3^k - 1\right) \\ &= \frac{1}{k}\left(\lambda_1^k + \lambda_2^k + \lambda_3^k - 3\right), \end{aligned} \tag{2.51}$$

where 2.50 is recovered by setting $k = 0$ in 2.51 and for $k = 2$, we obtain a special strain measure $\hat{\varepsilon}^2$, used in the following incompressible Neo-Hooke hyper-elastic material:

$$W^S_{\text{iso}} = \mu_S \, \hat{\varepsilon}^2 = \frac{1}{2}\mu_S \, (I - 3) \, . \tag{2.52}$$

Here, $I = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$ is referred to as the first invariant of $\mathbf{C}_S$ or $\mathbf{B}_S$ and will be discussed later and μ_S is the very well known standard notation for Lamé's second parameter. In addition to the above practical applications of $\hat{\varepsilon}^k$, a weighted combination of several $\hat{\varepsilon}^k$ from 2.51 was used by Ogden for modeling rubber at large strains:

$$\begin{aligned} W^S &= \sum_{i=1}^{N} \mu_{k_i} \, \hat{\varepsilon}^{k_i} \\ &= \sum_{k=i}^{N} \frac{\mu_{k_i}}{k_i} \left(\lambda_1^{k_i} + \lambda_2^{k_i} + \lambda_3^{k_i} - 3 \right) \quad \text{where} \quad k_i \in \mathbb{R} \, . \end{aligned} \tag{2.53}$$

Herein, W^S is the so-called Ogden's strain energy density and the weighting coefficients k_i and μ_{k_i} are material constants that satisfy certain mathematical conditions and are specified experimentally. For profound knowledge, interested readers are referred to [74, 73].

For a special choice ($k_m = 0$ with $m \in \{1, \dots N\}$) the Hencky strain will be included in W^S and for $N = 1$ and $k_1 = 2$ the strain energy function W^S boils down to the Hookean model. By choosing a special Cartesian coordinate system, so that the three orthogonal axes are aligned with $\tilde{\mathbf{N}}_\mathbf{1}$, $\tilde{\mathbf{N}}_\mathbf{2}$ and $\tilde{\mathbf{N}}_\mathbf{3}$, then the three axial strain components, $\frac{1}{k}\left(\lambda_1^k - 1\right)$, $\frac{1}{k}\left(\lambda_2^k - 1\right)$ and $\frac{1}{k}\left(\lambda_3^k - 1\right)$, can be combined using 2.40 and Einstein's summation convention as follows:

$$\tilde{\varepsilon} = \frac{1}{k}\left(\lambda_i^k - 1\right) \, \tilde{\mathbf{N}}_\mathbf{i} \otimes \tilde{\mathbf{N}}_\mathbf{i} = \frac{1}{k}\left((\mathbf{U}_S)^k - \mathbf{I} \right) \tag{2.54}$$

for material strain (Lagrangian) measures. For spatial (Eulerian) strain measure, a compact expression can be obtained using the left of 2.34 and the fact that $\mathbf{V}_S^{-2} = \left(\mathbf{F}_S \, \mathbf{F}_S^T\right)^{-1}$ (cf. 2.40):

$$\tilde{\varepsilon} = -\frac{1}{k}\left(\lambda_i^{-k} - 1\right) \, \tilde{\mathbf{n}}_\mathbf{i} \otimes \tilde{\mathbf{n}}_\mathbf{i} = -\frac{1}{k}\left((\mathbf{V}_S)^{-k} - \mathbf{I} \right) \tag{2.55}$$

for $\tilde{\mathbf{n}}_\mathbf{1}$, $\tilde{\mathbf{n}}_\mathbf{2}$ and $\tilde{\mathbf{n}}_\mathbf{3}$ being the directions of the Cartesian coordinate axes. Observe for 2.55 or 2.54, only if we set $k = 0$, we obtain a volumetric Hencky strain. The material Green-Lagrangian strain is obtained from 2.54 by choosing $k = 2$ and then making use of 2.40 to get

$$\mathbf{E}_S = \frac{1}{2}\left((\mathbf{U}_S)^2 - \mathbf{I} \right) = \frac{1}{2}\left(\mathbf{F}_S^T \, \mathbf{F}_S - \mathbf{I} \right) = \frac{1}{2}\left(\mathbf{C}_S - \mathbf{I} \right) . \tag{2.56}$$

Similarly, we obtain the spatial Almansi strain tensor by setting $k = 2$ in 2.55 and using 2.40

$$\varepsilon_S = -\frac{1}{2}\left((\mathbf{V}_S)^{-2} - \mathbf{I}\right) = -\frac{1}{2}\left(\mathbf{F}_S^{-T}\,\mathbf{F}_S^{-1} - \mathbf{I}\right) = -\frac{1}{2}\left(\mathbf{B}_S^{-1} - \mathbf{I}\right). \tag{2.57}$$

Notice that by means of deformation gradient, the above two quantities can be associated through the pull-back and push-forward operation

$$\mathbf{E}_S \underset{\mathbf{F}_S^{\mathrm{T}}\,(\,\cdot\,)\,\mathbf{F}_S}{\overset{\mathbf{F}_S^{-\mathrm{T}}\,(\,\cdot\,)\,\mathbf{F}_S^{-1}}{\rightleftarrows}} \varepsilon_S\,. \tag{2.58}$$

Admitting the Hencky strain measure $\ln J_S$ works fine for compressible non-porous bodies, it does not account (after certain limit of deformation) for special physical aspects related to porous media of materially incompressible constituents; the point of compaction and permeability effect.

Despite that the solid skeleton is materially incompressible, the solid matrix is compressible; large volume variations are possible due to pore expansion or shrinking. The point of compaction is an extreme deformation state, reached when all pores are closed and further volume reductions are impossible. This limit is reached if J_S hits its constant physical greatest lower bound [7], n^S_{0S}. Namely,

$$J_S \in \left(n^S_{0S}\ ,\ \infty\right) \quad \text{where} \quad n^S_{0S} \text{ is constant positive fraction}\,. \tag{2.59}$$

Hence, an admissible strain energy function (see 2.48) must take into account that the two physical impossibilities ($J_S = n^S_{0S}$ and $J_S = \infty$) require infinite energy W^S_{vol} (cf. 2.48) . Careless look may suggest the following strain measure:

$$\varepsilon_{\text{vol}} = \ln\left(J_S - n^S_{0S}\right), \tag{2.60}$$

which accounts for these two physical impossibilities. However, such a strain measure is not acceptable because it does not vanish in the undeformed [8] (reference) configuration. Therefore, 2.60 should be modified to

$$\varepsilon_{\text{vol}} = \ln\left(\frac{J_S - n^S_{0S}}{1 - n^S_{0S}}\right) \tag{2.61}$$

as suggested in [39]. The second physical phenomenon, not respected by the Hencky strain measure, is the effect of tortuosity, which can be better explained in the following equivalent

[7] Since the concentration n^S of the solid constituent is bounded above by 1, we get the sought-after lower bound for J_S using 2.121.

[8] in the reference (undeformed) configuration we have $J_S = 1$. Thus $\tilde{\varepsilon}_v = \ln\left(1 - n^S_{0S}\right) \neq 0$ since $n^S_{0S} < 1$.

expression

$$\varepsilon_{\text{vol}} = \ln\left(\frac{J_S - n_{0S}^S}{1 - n_{0S}^S}\right) = \ln\left(\frac{n^F}{n_0^F}\, J_S\right), \qquad \text{where} \qquad W_{\text{vol}}^S = \text{const.}\ \varepsilon_{\text{vol}}\,. \tag{2.62}$$

Smaller values of $\frac{n^F}{n_0^F}$ somehow indicate how much the pore channels are twisted or being tortuous. According to experimental observations, if $\frac{n^F}{n_0^F}$ decreases (which makes the pore channels thinner with more turns, it becomes harder for an incompressible pore fluid to travel inside the pore channels and escape to allow for pores to shrink further. As a result, higher compression load or higher W_{vol}^S will be required to cause further pore shrinking. The extreme physical impossibility (corresponding to $n^F \to 0$) therefore requires infinite compression load or infinite energy. For the extension case, where the pore channels become wider and hence more permeable, the extreme physical impossibility (corresponding to $n^F \to 1$) must be be associated with infinite extension load or energy W_{vol}^S, which is ensured by the factor J_S in (2.62) which goes infinite as n^F goes to 1.

The previous discussion also tells us a permeability function of the form $k^F = \left(\frac{n^F}{n_0^F}\right)^\kappa$ (suggested in [39]) as being free from the extra term J_S is suitable if we are only interested in lower limiting case ($n^F \to 0$) while increasing permeability will need the use of J_S so that $k^F = \left(\frac{n^F}{n_0^F}\, J_S\right)^\kappa$ is able to describe the upper limiting case ($n^F \to 1$) together with the lower one. The last (k^F formula) seems to be more consistent with the definition of strain and energy in (2.62) and was adopted by Markert. For intensive study on this issue, interested readers are referred to [59].

The deformation gradient is not only useful for computing volumes, areas, lengths changes, rotations and strain measures, but is also used in many other push-forward and pull-back operations in continuum mechanics. In what follows, we shall discuss this matter. Based on (2.11) and (2.14), the deformation gradient can be expressed by

$$\mathbf{F}_\alpha = \frac{\partial \mathbf{x}}{\partial \mathbf{X}_\alpha} = \frac{\partial\left(\mathbf{X}_\alpha + \mathbf{u}_\alpha\right)}{\partial \mathbf{X}_\alpha} = \mathbf{I} + \text{Grad}_\alpha\left(\mathbf{u}_\alpha\right), \tag{2.63}$$

for the Lagrangian (material) specification, where

$$\text{Grad}_\alpha\left(\,\cdot\,\right) = \frac{\partial\left(\,\cdot\,\right)}{\partial \mathbf{X}_\alpha}\,. \tag{2.64}$$

And for the Eulerian (spatial) specification, we first use (2.11) and (2.63)

$$\mathbf{F}_\alpha^{-1} = \frac{\partial \mathbf{X}_\alpha}{\partial \mathbf{x}} = \frac{\partial\left(\mathbf{x} - \mathbf{u}_\alpha\right)}{\partial \mathbf{x}} = \mathbf{I} - \text{grad}\left(\mathbf{u}_\alpha\right), \tag{2.65}$$

where

$$\operatorname{grad}(\,\boldsymbol{\cdot}\,) = \frac{\partial(\,\boldsymbol{\cdot}\,)}{\partial \mathbf{x}}\,, \tag{2.66}$$

and then by taking the inverse of (2.65), $\mathbf{F}_\alpha$ can be expressed by

$$\mathbf{F}_\alpha = (\mathbf{I} - \operatorname{grad}(\mathbf{u}_\alpha))^{-1}\,. \tag{2.67}$$

For any arbitrary vector $\mathbf{v} \in \{\mathbf{v}_\alpha, \mathbf{u}_\alpha\}$, the deformation gradient $\mathbf{F}_\alpha$ can be used to link $\operatorname{grad}(\mathbf{v})$ to $\operatorname{Grad}_\alpha(\mathbf{v})$ using the chain rule

$$\frac{\partial \mathbf{v}}{\partial \mathbf{X}_\alpha} = \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{X}_\alpha} \Leftrightarrow \operatorname{Grad}_\alpha(\mathbf{v}) = \operatorname{grad}(\mathbf{v})\,\mathbf{F}_\alpha\,. \tag{2.68}$$

Equation (2.68) is written in the following matrix form:

$$\operatorname{grad}(\mathbf{v}) = \operatorname{Grad}_\alpha(\mathbf{v})\,\mathbf{F}_\alpha^{-1} \qquad \forall \mathbf{v} \in \{\mathbf{v}_\alpha, \mathbf{u}_\alpha\}\,. \tag{2.69}$$

A similar procedure of using the chain rule is applied to link the spatial and material gradients of scalar quantities. To pull back $\operatorname{div}(\mathbf{v}_\alpha)$, we do the following

$$\operatorname{div}(\mathbf{v}_\alpha) = \operatorname{tr}(\operatorname{grad}\mathbf{v}_\alpha)\,, \tag{2.70}$$

and by means of (2.69), we obtain that

$$\operatorname{div}(\mathbf{v}_\alpha) = \operatorname{tr}(\operatorname{Grad}_\alpha(\mathbf{v}_\alpha)\,\mathbf{F}_\alpha^{-1})\,, \tag{2.71}$$

which can also be written in terms of double contraction using the relation $\operatorname{tr}(\mathbf{A}\mathbf{B}^T) = \mathbf{A} : \mathbf{B}$ as

$$\operatorname{div}(\mathbf{v}_\alpha) = \operatorname{Grad}_\alpha \mathbf{v}_\alpha : \mathbf{F}_\alpha^{-\mathsf{T}}\,. \tag{2.72}$$

We also use deformation gradient to derive fundamental relationships, required for computing porosity, for linearization of the weak form and for deriving a modified transport theorem as shown in the following section. For this, we use the chain rule to derive the Jacobian with respect to time [9]

$$(J_\alpha)'_\alpha = \frac{\mathrm{d}_\alpha J_\alpha}{\mathrm{d}t} = \frac{\partial J_\alpha}{\partial F_{\alpha_{ij}}}\frac{\mathrm{d}_\alpha F_{\alpha_{ij}}}{\mathrm{d}t} \quad \Leftrightarrow \quad (J_\alpha)'_\alpha = \frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha} : \frac{\mathrm{d}_\alpha \mathbf{F}_\alpha}{\mathrm{d}t}\,, \tag{2.73}$$

where the Einstein's summation convention (for indices i and j) was adopted and the operator ': ' denotes the double contraction. Because $J_\alpha = \det \mathbf{F}_\alpha$, then $\frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha}$ is equal to $J_\alpha \mathbf{F}_\alpha^{-\mathsf{T}}$. Thus,

$$(J_\alpha)'_\alpha = J_\alpha \mathbf{F}_\alpha^{-\mathsf{T}} : \frac{\mathrm{d}_\alpha \mathbf{F}_\alpha}{\mathrm{d}t}\,. \tag{2.74}$$

[9] Observe that $J_\alpha = \det(F_\alpha) \quad \rightarrow \quad J_\alpha = J_\alpha(F_{\alpha_{11}}, \ldots, F_{\alpha_{33}})$ and recall that $F_{\alpha_{ij}} = F_{\alpha_{ij}}(\mathbf{x}_\alpha, t)$.

According to (2.12) and (2.14), we conclude that [10]

$$\frac{\mathrm{d}_\alpha \mathbf{F}_\alpha}{\mathrm{d}t} = \frac{\mathrm{d}_\alpha}{\mathrm{d}t}\Big(\frac{\partial \mathbf{x}}{\partial \mathbf{X}_\alpha}\Big) = \frac{\partial}{\partial \mathbf{X}_\alpha}\Big(\frac{\mathrm{d}_\alpha \mathbf{x}}{\mathrm{d}t}\Big) = \frac{\partial}{\partial \mathbf{X}_\alpha}(\mathbf{v}_\alpha) = \mathrm{Grad}_\alpha\, \mathbf{v}_\alpha\,. \tag{2.75}$$

Substituting (2.75) in (2.74) and using (2.72) results in the following relation

$$(J_\alpha)'_\alpha = J_\alpha\, \mathrm{div}\,(\mathbf{v}_\alpha) \Rightarrow \mathrm{div}\,(\mathbf{v}_\alpha) = \frac{(J_\alpha)'_\alpha}{J_\alpha}\,. \tag{2.76}$$

The above equation is useful for deriving an extremely important modified transport theorem. It will also be mixed with the mass balances to derive a relation for the porosity as will be shown in the subsequent section. Furthermore, by means of (2.69) and (2.75) we obtain another and frequently used relationship,

$$\mathrm{grad}\,(\mathbf{v}_S) = \frac{\mathrm{d}_S \mathbf{F}_S}{\mathrm{d}t}\, \mathbf{F}_S^{-1}\,. \tag{2.77}$$

The deformation gradient helps to check the objectivity of physical quantities. For example, one can prove the non objectivity of the spatial velocity gradient $\mathrm{grad}\,(\mathbf{v}_S)$ as follows [11]

$$\begin{aligned}
\frac{\partial \mathbf{v}_S^*}{\partial \mathbf{x}^*} &= \underbrace{\frac{\mathrm{d}_S \mathbf{F}_S^*}{\mathrm{d}t}\, (\mathbf{F}_S^*)^{-1}}_{\text{cf. } (2.77)} \\
&= \frac{\mathrm{d}_S\ \overbrace{(\mathbf{Q}\, \mathbf{F}_S)}^{\text{cf. } (2.197)}}{\mathrm{d}t}\, (\mathbf{Q}\, \mathbf{F}_S)^{-1} \\
&= \big(\dot{\mathbf{Q}}\, \mathbf{F}_S + \mathbf{Q}\, \dot{\mathbf{F}}_S\big)\, \big(\mathbf{F}_S^{-1}\, \mathbf{Q}^T\big) \\
&= (\mathbf{Q})'_S\, \mathbf{Q}^T + \mathbf{Q}\, \underbrace{\dot{\mathbf{F}}_S\, \mathbf{F}_S^{-1}}_{\text{cf. } (2.77)}\, \mathbf{Q}^T \\
&= (\mathbf{Q})'_S\, \mathbf{Q}^T + \mathbf{Q}\, \frac{\partial \mathbf{v}_S}{\partial \mathbf{x}}\, \mathbf{Q}^T\,.
\end{aligned} \tag{2.78}$$

Thus,

$$\frac{\partial \mathbf{v}_S^*}{\partial \mathbf{x}^*} \neq \mathbf{Q}\, \frac{\partial \mathbf{v}_S}{\partial \mathbf{x}}\, \mathbf{Q}^T\,. \tag{2.79}$$

[10] The interchange of the derivatives in (2.75) is valid because $\mathbf{X}_\alpha$ is independent of t. If we have $\mathbf{x}$ instead of $\mathbf{X}_S$ then the interchange does not commute.

[11] The reader may first read from the second paragraph of section 2.5.3 till (2.197).

Due to (2.79), we say that $\mathrm{grad}\,(\mathbf{v}_S)$ does not follow the objective rule of transformation. In contrast, we can show the objectivity of the symmetric part of $\mathrm{grad}\,(\mathbf{v}_S)$,

$$\mathbf{d}_S = \frac{1}{2}\left(\mathrm{grad}\,\mathbf{v}_S + \mathrm{grad}^T\,\mathbf{v}_S\right) \ , \tag{2.80}$$

by virtue of (2.78), it follows

$$\begin{aligned}
\mathbf{d}_s^* &= \frac{1}{2}\left(\frac{\partial\mathbf{v}_S^*}{\partial\mathbf{x}^*} + \left(\frac{\partial\mathbf{v}_S^*}{\partial\mathbf{x}^*}\right)^T\right) \\
&= \frac{1}{2}\left((\mathbf{Q})'_S\ \mathbf{Q}^T + \mathbf{Q}\ (\mathbf{Q}^T)'_S\right) + \frac{1}{2}\left(\mathbf{Q}\ \frac{\partial\mathbf{v}_S}{\partial\mathbf{x}}\ \mathbf{Q}^T + \mathbf{Q}^T\left(\frac{\partial\mathbf{v}_S}{\partial\mathbf{x}}\right)^T\mathbf{Q}\right) \\
&= \frac{1}{2}\frac{\mathrm{d}_S\,\overbrace{(\mathbf{Q}\,\mathbf{Q}^T)}^{=\mathbf{I}}}{\mathrm{d}t} + \mathbf{Q}\overbrace{\frac{1}{2}\left(\frac{\partial\mathbf{v}_S}{\partial\mathbf{x}} + \left(\frac{\partial\mathbf{v}_S}{\partial\mathbf{x}}\right)^T\right)}^{=\mathbf{d}_S}\mathbf{Q}^T \\
&= \mathbf{0} + \mathbf{Q}\ \mathbf{d}_S\ \mathbf{Q}^T
\end{aligned} \tag{2.81}$$

or shortly

$$\mathbf{d}_s^* = \mathbf{Q}\ \mathbf{d}_S\ \mathbf{Q}^T\,, \tag{2.82}$$

which proves the objectivity of $\mathbf{d}_S$. In addition, it turned out that the rate of the Green-Lagrangian is not only objective but also transforms like a scalar. That is

$$\begin{aligned}
(\mathbf{E}_S^*)'_S &= \tfrac{1}{2}\left(\mathbf{F}_S^{*T}\mathbf{F}_S^* - \mathbf{I}\right)'_S \\
&= \tfrac{1}{2}\left(\mathbf{F}_S^T\,\overbrace{\mathbf{Q}^T\ \mathbf{Q}}^{=\mathbf{I}}\mathbf{F}_S - \mathbf{I}\right)'_S \\
&= \tfrac{1}{2}\left(\mathbf{F}_S^T\ \mathbf{F}_S - \mathbf{I}\right)'_S \\
&= (\mathbf{E}_S)'_S\,,
\end{aligned} \tag{2.83}$$

where

$$\begin{aligned}
(\mathbf{E}_S)'_S &= \tfrac{1}{2}\left(\mathbf{F}_S^T\ \mathbf{F}_S - \mathbf{I}\right)'_S \\
&= \tfrac{1}{2}\left(\mathbf{F}_S^T\ (\mathbf{F}_S)'_S + (\mathbf{F}_S^T)'_S\ \mathbf{F}_S\right) \\
&= \mathrm{sym}\left(\left(\mathbf{F}_S^T\right)'_S\ \mathbf{F}_S\right)
\end{aligned} \tag{2.84}$$

is required for deriving the energetic conjugates.

In the subsequent subsection and section 2.5, the deformation gradient plays an important role in transforming between stress measures and energy conjugates, deriving objective stress rates and objective constitutive relationships.

Now, we will show how the deformation gradient helps to compute some important Gateaux-derivatives necessary for for linearization of the weak forms.

Let $D_{\delta \mathbf{u}_S} f(\mathbf{u}_S)$ be the Gateaux-derivative of $f(\mathbf{u}_S)$ in the direction of $\delta \mathbf{u}_S$ and defined by

$$D_{\delta \mathbf{u}_S} f(\mathbf{u}_S) = \frac{\partial}{\partial \eta}\Big(f(\mathbf{u}_S + \eta\, \delta \mathbf{u}_S)\Big)_{\eta=0}. \tag{2.85}$$

Then, the directional derivative of the deformation gradient $\mathbf{F}_\alpha$ in the direction of displacement $\delta \mathbf{u}_\alpha$ is calculated using (2.85)

$$\begin{aligned} D_{\delta \mathbf{u}_\alpha} \mathbf{F}_\alpha &= \frac{\partial}{\partial \eta}\Big(\mathbf{F}_\alpha\left(\mathbf{u}_\alpha + \eta\, \delta \mathbf{u}_\alpha\right)\Big)_{\eta=0} \\ &= \frac{\partial}{\partial \eta}\Big(\mathbf{I} + \mathrm{Grad}_\alpha\left(\mathbf{u}_\alpha + \eta\, \delta \mathbf{u}_\alpha\right)\Big)_{\eta=0} \\ &= \mathrm{Grad}_\alpha\left(\delta \mathbf{u}_\alpha\right). \end{aligned} \tag{2.86}$$

And then by using the chain rule and Einstein's summation convention (for indices $i, j \in \{1 \ldots ndim\}$), we obtain

$$\frac{\partial J_\alpha}{\partial u_{\alpha_k}} = \frac{\partial J_\alpha}{\partial F_{\alpha_{ij}}} \frac{\partial F_{\alpha_{ij}}}{\partial u_\alpha} \quad \Leftrightarrow \quad \frac{\partial J_\alpha}{\partial \mathbf{u}_{\alpha_k}} = \frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha} : \frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{u}_{\alpha_k}} \tag{2.87}$$

with *ndim* denoting the dimension of the considered problem and the colon operator is used to indicate the double inner product between the second-order tensor $\frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha}$ and the third-order tensor $\frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{u}_S}$ to produce the first-order tensor $\frac{\partial J_\alpha}{\partial \mathbf{u}_S}$. Since $\frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha}$ is equal to $J_\alpha \mathbf{F}_\alpha^{-\mathsf{T}}$, we see that equation (2.87) can be written as

$$\frac{\partial J_\alpha}{\partial \mathbf{u}_S} = J_\alpha \mathbf{F}_\alpha^{-\mathsf{T}} : \frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{u}_S}. \tag{2.88}$$

The directional derivative of J_α in the direction of $\delta \mathbf{u}_\alpha$ is evaluated with the aid of the above relation

$$
\begin{aligned}
D_{\delta \mathbf{u}_\alpha}(J_\alpha) &= \frac{\partial J_\alpha}{\partial \mathbf{u}_\alpha} \cdot \delta \mathbf{u}_\alpha \\
&= \frac{\partial J_\alpha}{\partial \mathbf{F}_\alpha} : \frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{u}_\alpha} \cdot \delta \mathbf{u}_\alpha \\
&= J_\alpha \mathbf{F}_\alpha^{-\mathsf{T}} : \underbrace{\left(\frac{\partial \mathbf{F}_\alpha}{\partial \mathbf{u}_\alpha} \cdot \delta \mathbf{u}_\alpha \right)}_{D_{\delta \mathbf{u}_\alpha} \mathbf{F}_\alpha} \\
&= J_\alpha \underbrace{\mathbf{F}_\alpha^{-\mathsf{T}} : \operatorname{Grad}_\alpha \delta \mathbf{u}_\alpha}_{\text{cf. } \boxed{2.72}} \\
&= J_\alpha \operatorname{div}(\delta \mathbf{u}_\alpha) \, .
\end{aligned}
\tag{2.89}
$$

Now, we are ready to derive the balance relations but, we will show first how the deformation gradient is used to derive the energetic conjugates and stress push-forward and pull-back operations in the subsequent section.

2.3 Aspects on stresses and energetic conjugates

There are several classical stress measures in continuum mechanics, which found their way in porous media applications. The total Cauchy stress tensor $\mathbf{T}$ relates to the differential mixture [12] force vector $\mathrm{d}\mathbf{f}$ acting on a current differential mixture area $\mathrm{d}a$ through the following expression:

$$
\mathbf{T}\, \mathbf{n} = \frac{\mathrm{d}\mathbf{f}}{\mathrm{d}a} \, , \tag{2.90}
$$

where $\mathbf{n}$ is a unit vector normal to $\mathrm{d}a$. The mixture boundary traction $\bar{\mathbf{t}}$, which is the boundary force $\mathrm{d}\bar{\mathbf{f}}$ per unit boundary area $\mathrm{d}\bar{a}$, simultaneously acts on the solid and fluid constituent and can be directly connected to $\mathbf{T}$ using the so-called Cauchy Theorem, which yields

$$
\mathbf{T}\, \bar{\mathbf{n}} = \mathbf{t} = \frac{\mathrm{d}\bar{\mathbf{f}}}{\mathrm{d}\bar{a}} \tag{2.91}
$$

with $\bar{\mathbf{n}}$ being the normal to the boundary $\mathrm{d}\bar{a}$. In his empirical study (cf. [91]) on one-dimensional water saturated clay, Terzaghi found that there are two components which mainly contribute to the total stress in an element of soil: the effective stress [13] $\mathbf{T}_E^S$ and the pore water pressure

[12] This force is decomposed into force $\mathrm{d}\mathbf{f}_S$ deforming the solid constituent of the mixture and force $\mathrm{d}\mathbf{f}_F$ related to the pressure of the fluid constituent.

[13] In poromechanics this is also referred to as intergranular stress.

p. He also observed that only the effective stress is responsible for deforming the solid constituent. Following this empirical observation, he stated (without mathematical proof), for his one-dimensional case, that

$$\mathbf{T} = \mathbf{T}_E^S \ - p\mathbf{I} \qquad \text{for one-dimesional problem.} \tag{2.92}$$

For more details, interested readers are referred to [91], [88], [26] and [4]. Later we will show that the above constitutive relation was proven to be thermodynamically admissible. The effective symmetric Cauchy stress tensor $\mathbf{T}_E^S$ associates a current differential force $\mathrm{d}\mathbf{f}_S$ with a current differential area $\mathrm{d}a$ of normal $\mathbf{n}$. More precisely,

$$\mathbf{T}_E^S \, \mathbf{n} \, \mathrm{d}a = \mathrm{d}\mathbf{f}_S \, . \tag{2.93}$$

Similarly, one may think of a stress measure $\mathbf{P}_E^S$ that relates the current differential force $\mathrm{d}\mathbf{f}_S$ to the initial differential area $\mathrm{d}A_S$ of normal $\mathbf{N}$. Such a stress is usually referred to as first Piola-Kirchhoff stress and is defined such that

$$\mathbf{P}_E^S \, \mathbf{N} \, \mathrm{d}A_S = \mathrm{d}\mathbf{f}_S \, . \tag{2.94}$$

By virtue of $\boxed{2.93}$ and $\boxed{2.22}$, $\mathbf{P}_E^S$ can be directly linked to $\mathbf{T}_E^S$:

$$\mathbf{P}_E^S = J_S \, \mathbf{T}_E^S \, \mathbf{F}_S^{-T} \, . \tag{2.95}$$

Obviously, the unsymmetric $\mathbf{P}_E^S$ is a two-point tensor because it relates to the current and reference configuration, whereas the symmetric $\mathbf{T}_E^S$ relates only to the current configuration. However, there is another symmetric stress tensor, that relates only to the reference configuration; the so-called second Piola-Kirchhoff stress $\mathbf{S}_E^S$. To get that stress, we first observe that $\mathbf{N} \, \mathrm{d}A_S$ in $\boxed{2.94}$ already relates to the reference configuration, while $\mathrm{d}\mathbf{f}_S$ still needs to be premultiplied [14] by $\mathbf{F}_S^{-1}$ to pull it back to the reference configuration. Doing so, we obtain

$$\underbrace{\mathbf{F}_S^{-1} \, \mathbf{P}_E^S}_{=\mathbf{S}_E^S} \, \mathbf{N} \, \mathrm{d}A_S = \underbrace{\mathbf{F}_S^{-1} \, \mathrm{d}\mathbf{f}_S}_{\text{fictitious force}} \quad \Rightarrow \quad \mathbf{S}_E^S = \mathbf{F}_S^{-1} \mathbf{P}_E^S \quad \Rightarrow \quad \mathbf{S}_E^S = \mathbf{F}_S^{-1} \, \underbrace{\overbrace{J_S \mathbf{T}_E^S \, \mathbf{F}_S^{-T}}^{=\boldsymbol{\tau}_E^S}}_{cf. \, \boxed{2.95}} \, . \tag{2.96}$$

Since $\mathbf{S}_E^S$ relates only a reference area vector $\mathrm{d}A_S \, \mathbf{N}$ to a reference fictitious force ($\mathbf{F}_S^{-1} \, \mathrm{d}\mathbf{f}_S$) or in other words, $\mathbf{S}_E^S$ is a function of those two reference quantities, it must be a quantity in

[14] Remark that

$$\mathrm{d}\mathbf{f}_S = \|\mathrm{d}\mathbf{f}_S\| \, \mathbf{n} \qquad \text{where} \qquad \mathbf{n} = \frac{\mathrm{d}\mathbf{f}_S}{\|\mathrm{d}\mathbf{f}_S\|}$$

and using $\boxed{2.14}$, we can pull back the unit vector $\mathbf{n}$ to the reference configuration.

the reference configuration too. Following 2.96, we deduce the following push-forward and pull-back stress transformations, which are different from strain transformations in 2.58:

$$\mathbf{S}_E^S \underset{\mathbf{F}_S^{-1}\,(\,\bullet\,)\,\mathbf{F}_S^{-\mathrm{T}}}{\overset{\mathbf{F}_S\,(\,\bullet\,)\,\mathbf{F}_S^{\mathrm{T}}}{\rightleftarrows}} \boldsymbol{\tau}_E^S \tag{2.97}$$

Although the symmetric $\mathbf{S}_E^S$ generally does not admit a physical interpretation, it is useful for theoretical treatment to derive an objective stress rates. An objective second-order tensor is the one that transforms objectively. That is, if

$$\mathbf{T} = \mathbf{T}_{ij}\; \mathbf{e}_i \otimes \mathbf{e}_j \tag{2.98}$$

and if we rotate our Cartesian coordinate system by an orthogonal rotation tensor $\mathbf{Q}^T$, which leads to the orthonormal basis being transformed as follows

$$\mathbf{e}_i^* = \mathbf{Q}^T\; \mathbf{e}_i\,. \tag{2.99}$$

Then, $\mathbf{T}$, in term of the new basis, reads

$$\begin{aligned}\mathbf{T}_{ij}\; \mathbf{e}_i \otimes \mathbf{e}_j &= \mathbf{T}_{ij}\;\; (\mathbf{Q}\mathbf{e}_i^*) \otimes \left(\mathbf{Q}\mathbf{e}_j^*\right) \\ &= \mathbf{T}_{ij}\;\; (\mathbf{Q}\mathbf{e}_i^*) \left(\mathbf{Q}\mathbf{e}_j^*\right)^T \\ &= \mathbf{Q}\,\underbrace{\left(\mathbf{T}_{ij}\; \mathbf{e}_i^* \otimes \mathbf{e}_j^*\right)}_{\mathbf{T}_{ij}^*}\,\mathbf{Q}^T\end{aligned} \tag{2.100}$$

or shortly, we have

$$\mathbf{T}^* = \mathbf{Q}^T\; \mathbf{T}\,\mathbf{Q}\,. \tag{2.101}$$

And similarly,

$$\mathbf{T}_E^{S*} = \mathbf{Q}^T\; \mathbf{T}_E^S\; \mathbf{Q}, \qquad \boldsymbol{\tau}_E^{S*} = \mathbf{Q}^T\; \boldsymbol{\tau}_E^S\; \mathbf{Q}\,. \tag{2.102}$$

Although the first two quantities ($\mathbf{T}_E^S$ and $\boldsymbol{\tau}_E^S$) obey the objective rule of transformation, their rate of change are not so because [15]

$$\begin{aligned}\left(\boldsymbol{\tau}_E^{S*}\right)_S' &= \left(\mathbf{Q}^T\; \boldsymbol{\tau}_E^S\; \mathbf{Q}\right)_S' \\ &= \left(\mathbf{Q}^T\right)_S'\; \boldsymbol{\tau}_E^S\; \mathbf{Q} \;+\; \mathbf{Q}^T\; \boldsymbol{\tau}_E^S\; (\mathbf{Q})_S' \;+\; \mathbf{Q}^T\; \left(\boldsymbol{\tau}_E^S\right)_S'\; \mathbf{Q}\,.\end{aligned} \tag{2.103}$$

[15] The reader may first read from the second paragraph of section 2.5.3 till 2.197 .

Or shortly,

$$\left(\boldsymbol{\tau}_E^{S*}\right)_S' \neq \mathbf{Q}^T \left(\boldsymbol{\tau}_E^{S}\right)_S' \mathbf{Q}. \tag{2.104}$$

For that reason, it is not clear how to directly establish a useful constitutive relation starting from the following form

$$\left(\boldsymbol{\tau}_E^{S}\right)_S' \propto \mathbf{d}_S, \tag{2.105}$$

since $\mathbf{d}_S$ transforms objectively as shown in 2.82, while $\left(\boldsymbol{\tau}_E^{S}\right)_S'$ does not as shown in 2.104. The significance of $\mathbf{S}_E^S$ lies in the fact that it relates to a fixed domain (reference configuration) and it transforms objectively like a scalar. Namely,

$$\begin{aligned}\left(\mathbf{S}_E^{S*}\right)_S' &= \underbrace{\left((\mathbf{F}_S^*)^{-1}\,\boldsymbol{\tau}_E^{S*}\,(\mathbf{F}_S^*)^{-T}\right)_S'}_{\text{cf. } 2.97}\\ &= \underbrace{\left(\mathbf{F}_S^{-1}\overbrace{\mathbf{Q}^T\,\mathbf{Q}}^{=\mathbf{I}}\boldsymbol{\tau}_E^S\overbrace{\mathbf{Q}^T\,\mathbf{Q}}^{=\mathbf{I}}\mathbf{F}_S^{-T}\right)_S'}_{\text{cf. } 2.102}\\ &= \underbrace{\left(\mathbf{F}_S^{-1}\,\boldsymbol{\tau}_E^S\,\mathbf{F}_S^{-T}\right)_S'}_{\text{cf. } 2.97}\\ &= \left(\mathbf{S}_E^S\right)_S'. \end{aligned} \tag{2.106}$$

Therefore, one can exploit this nice property of $\mathbf{S}_E^S$ and the transformation 2.97 to derive objective stress rates (for instance, Truesdell rate, Green-Naghdi rate and Jaumann rate of the Cauchy stress) for other stress measures and also directly generate a meaningful constitutive relation of the form

$$\left(\mathbf{S}_E^{S}\right)_S' \propto \left(\mathbf{E}_S\right)_S', \tag{2.107}$$

where $(\mathbf{E}_S)_S'$ also relates to the reference configuration and transforms objectively like a scalar as shown in 2.83. Notice in 2.107 and 2.105, we associated $\mathbf{S}_E^S$ with $\mathbf{E}_S$ and $\boldsymbol{\tau}_E^S$ with $\mathbf{d}_S$. Indeed, these quantities are usually referred to as energetic conjugates and they are based on the definition of the internal stress power $\mathfrak{p}_{\text{int}}$

$$\mathfrak{p}_{\text{int}} = \int_\Omega \mathbf{T}_E^S : \operatorname{grad} \mathbf{v}_S, \tag{2.108}$$

which arises from the balance of energy and will be discussed in section 2.4.4. Due to symmetry of $\mathbf{T}_E^S$, the above expression is equivalent to

$$\mathfrak{p}_{\text{int}} = \int_\Omega \mathbf{T}_E^S : \mathbf{d}_S \,, \qquad \text{where} \qquad \mathbf{d}_S = \frac{1}{2}\left(\operatorname{grad}\mathbf{v}_S + \operatorname{grad}^T \mathbf{v}_S\right) . \tag{2.109}$$

The last is preferred as $\mathbf{d}_S$ is objective. The above relation can be used to derive the famous four energetic conjugates:

$$\begin{aligned}
\mathfrak{p}_{\text{int}} &= \int_\Omega \mathbf{T}_E^S : \operatorname{grad}\mathbf{v}_S \,\mathrm{d}v \\
&= \int_\Omega \mathbf{T}_E^S : \operatorname{grad}\mathbf{v}_S \,\underbrace{(J_S \,\mathrm{d}V_S)}_{\text{see } \boxed{2.16}} \\
&= \int_\Omega \underbrace{J_S\,\mathbf{T}_E^S}_{=\boldsymbol{\tau}_E^S} : \operatorname{grad}\mathbf{v}_S \,\mathrm{d}V_S \,, \quad \underline{\text{1st and 2nd conjugates}} \\
&= \int_\Omega J_S\,\mathbf{T}_E^S : \underbrace{\left((\mathbf{F}_S)'_S\;\mathbf{F}_S^{-1}\right)}_{\text{cf. } \boxed{2.77}} \,\mathrm{d}V_S \\
&= \int_\Omega \underbrace{\left(J_S\,\mathbf{T}_E^S\,\mathbf{F}_S^{-T}\right)}_{=\mathbf{P}_E^S,\ \text{cf. } \boxed{2.95}} : \; (\mathbf{F}_S)'_S \,\mathrm{d}V_S \\
&= \int_\Omega \mathbf{P}_E^S : \; (\mathbf{F}_S)'_S \,\mathrm{d}V_S \,, \quad \underline{\text{3rd conjugate}} \\
&= \int_\Omega \underbrace{\left(\mathbf{F}_S\,\mathbf{S}_E^S\right)}_{\text{cf. } \boxed{2.96}} : \; (\mathbf{F}_S)'_S \,\mathrm{d}V_S \\
&= \int_\Omega \mathbf{S}_E^S : \; \left(\mathbf{F}_S^T\;(\mathbf{F}_S)'_S\right) \mathrm{d}V_S \\
&= \int_\Omega \mathbf{S}_E^S : \operatorname{sym}\left(\mathbf{F}_S^T\;(\mathbf{F}_S)'_S\right) \mathrm{d}V_S \\
&= \int_\Omega \mathbf{S}_E^S : \; \underbrace{(\mathbf{E}_S)'_S}_{\boxed{2.84}} \,\mathrm{d}V_S \,, \quad \underline{\text{4th conjugate}} \,.
\end{aligned} \tag{2.110}$$

Hence, the energetic conjugates read

$$\boldsymbol{\tau}_E^S : \operatorname{grad}\mathbf{v}_S = J_S\,\mathbf{T}_E^S : \operatorname{grad}\mathbf{v}_S = \mathbf{P}_E^S : \; (\mathbf{F}_S)'_S = \mathbf{S}_E^S : \; (\mathbf{E}_S)'_S \,, \tag{2.111}$$

where all the above stress measures (except $\mathbf{P}_E^S$) are symmetric. Therefore, $\boxed{2.111}$ can alternatively be written as

$$\boldsymbol{\tau}_E^S : \mathbf{d}_S = J_S\,\mathbf{T}_E^S : \mathbf{d}_S = \mathbf{P}_E^S : \; (\mathbf{F}_S)'_S = \mathbf{S}_E^S : \; (\mathbf{E}_S)'_S \,. \tag{2.112}$$

2.4 Balance equations

Following Truesdell's metaphysical principles for mixtures [16], the balance equations (i. e., the balance of mass, the balance of linear momentum, the balance of angular momentum as well as the balance of energy) can be described for each ingredient φ^α individually provided that the interaction effects between ingredients are taken into account.

2.4.1 Balance of masses and porosity

Before turning to the derivations of the mass balances, we state that if the integration to be carried out is taken over the current configuration Ω, then the differentiation and integration do not commute. That is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega f \mathrm{d}v \neq \int_\Omega \frac{\mathrm{d}f}{\mathrm{d}t}\mathrm{d}v\,, \tag{2.113}$$

does not hold because Ω is deforming with time t. For instance, consider integrating $f = f(\mathbf{x},t)$ over the deformation-dependent region Ω then taking the derivative with respect to time:

$$\begin{aligned}
\frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_\Omega f \mathrm{d}v &= \frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_{\Omega_0} f\,\underbrace{(J_\alpha\,\mathrm{d}V^\alpha)}_{\text{see } \boxed{2.16}} \\
&= \int_{\Omega_0} \frac{\mathrm{d}_\alpha(f\,J_\alpha)}{\mathrm{d}t}\,\mathrm{d}V^\alpha \\
&= \int_{\Omega_0} J_\alpha\,\frac{\mathrm{d}_\alpha(f)}{\mathrm{d}t} + f\,\underbrace{\frac{\mathrm{d}_\alpha(J_\alpha)}{\mathrm{d}t}}_{\text{see } \boxed{2.76}}\,\mathrm{d}V^\alpha \\
&= \int_{\Omega_0} \left(\frac{\mathrm{d}_\alpha(f)}{\mathrm{d}t} + f\,\operatorname{div}(\mathbf{v}_\alpha)\right)\,(J_\alpha\,\mathrm{d}V^\alpha) \\
&= \int_{\Omega} \left(\frac{\mathrm{d}_\alpha(f)}{\mathrm{d}t} + f\,\operatorname{div}(\mathbf{v}_\alpha)\right)\,\mathrm{d}v \\
&= \int_{\Omega} \big((f)'_\alpha + f\,\operatorname{div}(\mathbf{v}_\alpha)\big)\,\mathrm{d}v\,.
\end{aligned} \tag{2.114}$$

[16] Truesdell's metaphysical principles stated in [93, p. 83] or [95, p. 221]

- All properties of the mixture must be mathematical consequences of properties of the constituents.
- So as to describe the motion of a constituent, we may in imagination isolate it from the rest of the mixture, provided we allow properly for the actions of the other constituents upon it.
- The motion of a mixture is governed by the same equations as is a single body.

Hence [17],

$$\frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_\Omega f(\mathbf{x},t)\ \mathrm{d}v = \int_\Omega \big((f(\mathbf{x},t))'_\alpha + f(\mathbf{x},t)\ \mathrm{div}\,(\mathbf{v}_\alpha)\big)\ \mathrm{d}v\,. \tag{2.115}$$

Therefore, when deriving a spatial integration (integration over Ω) with respect to time or when computing the Gateauxderivative, it is essential to switch the spatial domain of integration to the fixed domain Ω_0 (which is time and deformation independent) and express the spatial quantities in terms of the reference configuration using a transformation mechanism such as (2.16), (2.86), (2.58), (2.72), (2.89) and (2.97).

As we stated before, our binary aggregate is treated here as a heterogeneous mixture which indicates no mass exchanges [18] ($\hat{\rho}^\alpha = 0$) between the solid phase and the fluid phase. Accordingly, the balances of mass must be satisfied individually. The total mass (M^α) of each individual constituent φ^α is computed by virtue of (2.4) as follows

$$M^\alpha = \int_\Omega \mathrm{d}m^\alpha = \int_\Omega \rho^\alpha \mathrm{d}v \quad \forall \alpha \in \{S,F\}\,. \tag{2.116}$$

The individual mass balance is given by

$$\frac{\mathrm{d}_\alpha M^\alpha}{\mathrm{d}t} = \frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_\Omega \rho^\alpha \mathrm{d}v = 0\,. \tag{2.117}$$

The above equation is ready for the direct application of the transport theorem (2.115)

$$\frac{\mathrm{d}_\alpha M^\alpha}{\mathrm{d}t} = \int_\Omega \big((\rho^\alpha)'_\alpha + \rho^\alpha\ \mathrm{div}\,(\mathbf{v}_\alpha)\big)\ \mathrm{d}v = 0\,. \tag{2.118}$$

Since the above equation must be satisfied for any arbitrary domain, the localization of (2.118) yields the so-called partial mass balance

$$(\rho^\alpha)'_\alpha + \rho^\alpha\ \mathrm{div}\,(\mathbf{v}_\alpha) = 0\,. \tag{2.119}$$

Due to the material incompressibility assumption (i. e., $\rho^{\alpha R}$ is assumed to be constant real) and by virtue of (2.7), the above equation can be reduced to the so-called partial volume balance:

$$(n^\alpha)'_\alpha + n^\alpha\ \mathrm{div}\,(\mathbf{v}_\alpha) = 0\,. \tag{2.120}$$

[17] The relation (2.115) is nothing but a slightly modified Reynolds' transport

[18] An example of mass exchanges: A cup contains piece of ice and water or fluid-solid chemical interactions.

To compute n^S, we combine (2.120) and (2.76) (for $\alpha = S$) as follows

$$\begin{aligned} -\frac{(n^S)'_s}{n^S} &= \frac{(J_S)'_s}{J_S} \Leftrightarrow \\ \frac{1}{n^S}\frac{\mathrm{d}_S(n^S)}{\mathrm{d}t} &= -\frac{1}{J_S}\frac{\mathrm{d}_S(J_S)}{\mathrm{d}t} \Leftrightarrow \\ \frac{1}{n^S}\mathrm{d}_S n^S &= -\frac{1}{J_S}\mathrm{d}_S J_S \Leftrightarrow \\ \int_{n_{0S}}^{n_S} \frac{1}{n^S}\mathrm{d}_S(n^S) &= -\int_{J_{0S}=1}^{J_S} \frac{1}{J_S}\mathrm{d}_S J_S \Leftrightarrow \\ n^S &= \frac{n^S_{0S}}{J_S}\,. \end{aligned} \tag{2.121}$$

The above relation is attributed to Goodman and Cowin [46]. Next, From saturation condition (2.3) and (2.121), we obtain

$$n^F = 1 - \frac{n^S_{0S}}{J_S}\,. \tag{2.122}$$

To derive the continuity equation or the total (or bulk) volume balance, the left hand side of (2.120) must be first reformulated using (2.12) and (2.10):

$$\begin{aligned} &(n^\alpha)'_\alpha + n^\alpha \operatorname{div}(\mathbf{v}_\alpha) = 0 \\ =&\frac{\partial n^\alpha}{\partial t} + \operatorname{grad}(n^\alpha)\cdot\mathbf{v}_\alpha + n^\alpha \operatorname{div}(\mathbf{v}_\alpha) \\ =&\frac{\partial n^\alpha}{\partial t} + \operatorname{div}(n^\alpha \mathbf{v}_\alpha)\,, \end{aligned} \tag{2.123}$$

and then, we add up the above partial volume balances for $\alpha \in \{F,S\}$, which leads to

$$\frac{\partial\left(n^F + n^S\right)}{\partial t} + \operatorname{div}(n^F\mathbf{v}_F) + \operatorname{div}(n^S\mathbf{v}_S) = 0\,. \tag{2.124}$$

Next, by virtue of the saturation condition (2.3), we obtain the so-called continuity equation

$$\operatorname{div}(n^F\mathbf{v}_F) + \operatorname{div}(n^S\mathbf{v}_S) = 0\,, \tag{2.125}$$

which can also be written as

$$\begin{aligned} \operatorname{div}(n^S\mathbf{v}_S) + \operatorname{div}(n^F\mathbf{v}_F) &= n^S\operatorname{div}(\mathbf{v}_S) + n^F\operatorname{div}(\mathbf{v}_F) + \operatorname{grad}(n^S)\cdot\mathbf{v}_S + \operatorname{grad}(n^F)\cdot\mathbf{v}_F \\ &= n^S\operatorname{div}(\mathbf{v}_S) + n^F\operatorname{div}(\mathbf{v}_F) + \operatorname{grad}(n^F)\cdot(\mathbf{v}_F - \mathbf{v}_S) \\ &= n^S\mathbf{I}\colon \operatorname{grad}(\mathbf{v}_S) + n^F\mathbf{I}\colon \operatorname{grad}(\mathbf{v}_F) + \operatorname{grad}(n^F)\cdot(\mathbf{v}_F - \mathbf{v}_S) = 0\,. \end{aligned} \tag{2.126}$$

Where by (2.3), we have $\operatorname{grad}(n^S) = -\operatorname{grad}(n^F)$. Finally, we use (2.119) and (2.115) to derive the following important relation

$$\begin{aligned}\frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_\Omega \rho^\alpha f(\mathbf{x},t)\ \mathrm{d}v &= \int_\Omega (\rho^\alpha f(\mathbf{x},t))'_\alpha + \rho^\alpha f(\mathbf{x},t)\ \operatorname{div}(\mathbf{v}_\alpha)\ \mathrm{d}v \\ &= \int_\Omega \rho^\alpha\,(f(\mathbf{x},t))'_\alpha + f(\mathbf{x},t)\underbrace{(\rho^\alpha)'_\alpha + \rho^\alpha\ \operatorname{div}(\mathbf{v}_\alpha)}_{=0,\ \text{cf. } \boxed{2.119}}\ \mathrm{d}v\,, \qquad \boxed{2.127}\end{aligned}$$

which gives

$$\frac{\mathrm{d}_\alpha}{\mathrm{d}t}\int_\Omega \rho^\alpha f(\mathbf{x},t)\ \mathrm{d}v = \int_\Omega \rho^\alpha\,(f(\mathbf{x},t))'_\alpha\ \mathrm{d}v\,. \qquad \boxed{2.128}$$

By componentwise application, (2.128) and (2.115) can be generalized for f as vector or tensor quantity.

2.4.2 Balance of linear momentum

For any particle $\mathcal{P}^\alpha \in \Omega$, the momentum ι^α is defined to be the product of two quantities, the particle mass $\mathrm{d}m^\alpha$ and the particle velocity $\mathbf{v}_\alpha$. That is,

$$\iota^\alpha = \mathbf{v}_\alpha\ \mathrm{d}m^\alpha\,. \qquad \boxed{2.129}$$

Since each constituent φ^α is considered to be a collection of an infinite number of particles, its total momentum $\mathbf{I}^\alpha$ is obtained by summing all the particle momenta, which is attained via

$$\mathbf{I}^\alpha = \int_\Omega \mathbf{v}_\alpha\ \underbrace{\mathrm{d}m^\alpha}_{\text{see } \boxed{2.5}} = \int_\Omega \rho^\alpha \mathbf{v}_\alpha\ \mathrm{d}v\,. \qquad \boxed{2.130}$$

The material time derivative of (2.130) is computed by applying (2.128), which gives

$$(\mathbf{I}^\alpha)'_\alpha = \int_\Omega \rho^\alpha\,(\mathbf{v}_\alpha)'_\alpha\ \mathrm{d}v\,, \qquad \boxed{2.131}$$

where $(\mathbf{v}_\alpha)'_\alpha$ is the particle acceleration. According to the balance of linear momentum, $(\mathbf{I}^\alpha_i)'_\alpha$ must be equal to the sum of external forces. That is,

$$\int_\Omega \rho^\alpha\,(\mathbf{v}_\alpha)'_\alpha\ \mathrm{d}v = \underbrace{\int_\Omega \rho^\alpha \mathbf{b}^\alpha\ \mathrm{d}v}_{\text{Volume force}} + \underbrace{\int_\Omega \hat{\mathbf{p}}^\alpha \mathrm{d}v}_{\text{interaction force}} + \underbrace{\int_\Gamma \mathbf{t}^\alpha \mathrm{d}a}_{\text{surface force}}\,, \qquad \boxed{2.132}$$

with

$$\hat{\mathbf{p}}^S + \hat{\mathbf{p}}^F = \mathbf{0} \qquad \boxed{2.133}$$

Applying Cauchy's Theorem [19] and then the divergence theorem, (2.132) can be written as

$$\int_{\Omega} \left(\rho^{\alpha} \left(\mathbf{v}_{\alpha}\right)'_{\alpha} - \rho^{\alpha}\mathbf{b}^{\alpha} - \hat{\mathbf{p}}^{\alpha} - \operatorname{div}\left(\mathbf{T}^{\alpha}\right)\right) \mathrm{d}v = 0 \,. \tag{2.134}$$

Hence, the local form of the balance of momentum reads

$$\rho^{\alpha} \left(\mathbf{v}_{\alpha}\right)'_{\alpha} = \rho^{\alpha}\mathbf{b}^{\alpha} + \hat{\mathbf{p}}^{\alpha} + \operatorname{div}\left(\mathbf{T}^{\alpha}\right). \tag{2.135}$$

2.4.3 Balance of angular momentum

For any particle $\mathcal{P}^{\alpha} \in \Omega$, the angular momentum $\mathbf{h}^{\alpha}$ (rarely, rotational momentum) about the spatial reference point $\mathcal{O}$ is defined as

$$\text{angular momentum} = \mathbf{x} \times \text{linear momentum}\,.$$

Therefore, it is sometimes called the 'moment' of momentum and given by

$$\mathbf{h}^{\alpha} = \mathbf{x} \times \mathbf{v}^{\alpha} \mathrm{d}m^{\alpha}\,. \tag{2.136}$$

Since each constituent φ^{α} is considered to be a collection of an infinite number of particles, its total angular momentum $\mathbf{H}^{\alpha}$ is obtained by summing all the particle angular momenta, which is attained via

$$\mathbf{H}^{\alpha} = \int_{\Omega} \mathbf{x} \times \mathbf{v}_{\alpha} \underbrace{\mathrm{d}m^{\alpha}}_{\text{see } (2.5)} = \int_{\Omega} \rho^{\alpha}\, \mathbf{x} \times \mathbf{v}_{\alpha}\, \mathrm{d}v\,. \tag{2.137}$$

The material time derivative of (2.137) is computed by applying (2.128) and making use of the relation ($\frac{\mathrm{d}_{\alpha}\mathbf{x}}{\mathrm{d}t} \times \mathbf{v}_{\alpha} = \mathbf{v}_{\alpha} \times \mathbf{v}_{\alpha} = \mathbf{0}$), which gives

$$\left(\mathbf{H}^{\alpha}\right)'_{\alpha} = \int_{\Omega} \rho^{\alpha}\, \mathbf{x} \times \left(\mathbf{v}_{\alpha}\right)'_{\alpha}\, \mathrm{d}v\,, \qquad \forall \alpha \in \{S, F\}\,. \tag{2.138}$$

According to the balance of angular momentum, $\left(\mathbf{H}_{i}^{\alpha}\right)'_{\alpha}$ must be equal to the total external moments. That is,

$$\left(\mathbf{H}^{\alpha}\right)'_{\alpha} = \int_{\Omega} \rho^{\alpha}\mathbf{x} \times \mathbf{b}\, \mathrm{d}v + \int_{\Omega} \mathbf{x} \times \hat{\mathbf{p}}^{\alpha}\, \mathrm{d}v + \int_{\Omega} \hat{\mathbf{m}}^{\alpha}\, \mathrm{d}v + \int_{\Gamma} \mathbf{x} \times \mathbf{t}^{\alpha}\, \mathrm{d}a\,, \tag{2.139}$$

where the interaction vector, $\hat{\mathbf{m}}^{\alpha}$, is known as the moment of momentum production and satisfies the relation,

$$\hat{\mathbf{m}}^{S} + \hat{\mathbf{m}}^{F} = \mathbf{0}\,. \tag{2.140}$$

[19] Based on Cauchy's theorem, for any surface load $\mathbf{t}^{\alpha}$ and the unit-length vector $\mathbf{n}$ normal to the surface, there exists a unique second-order tensor $\mathbf{T}^{\alpha}$ with nine components $\mathbf{T}_{ij}$ such that $\mathbf{t} = \mathbf{T}\mathbf{n}$.

The axial vector $\hat{\mathbf{m}}^\alpha$ is usually associated [20] with its skew symmetric couple shear tensor $\hat{\mathbf{M}}^\alpha$ as shown below

$$\hat{\mathbf{m}}^\alpha = \begin{pmatrix} \hat{m}_1^\alpha \\ \hat{m}_2^\alpha \\ \hat{m}_3^\alpha \end{pmatrix} \quad \rightarrow \quad \hat{\mathbf{M}}^\alpha = \begin{pmatrix} 0 & \hat{m}_3^\alpha & -\hat{m}_2^\alpha \\ -\hat{m}_3^\alpha & 0 & \hat{m}_1^\alpha \\ \hat{m}_2^\alpha & -\hat{m}_3^\alpha & 0 \end{pmatrix} . \tag{2.141}$$

Applying Cauchy's theorem followed by divergence theorem to the most right term in 2.139, we then get

$$(\mathbf{H}^\alpha)'_\alpha = \int_\Omega \rho^\alpha \mathbf{x} \times \mathbf{b} \,\mathrm{d}v + \int_\Omega \mathbf{x} \times \hat{\mathbf{p}}^\alpha \,\mathrm{d}v + \int_\Omega \hat{\mathbf{m}}^\alpha \,\mathrm{d}v + \int_\Gamma \operatorname{div}(\mathbf{x} \times \mathbf{T}^\alpha) \,\mathrm{d}v . \tag{2.142}$$

Observe that $\operatorname{div}(\mathbf{x} \times \mathbf{T}^\alpha)$ can be expressed by [21]

$$\operatorname{div}(\mathbf{x} \times \mathbf{T}^\alpha) = \mathbf{x} \times \operatorname{div}(\mathbf{T}^\alpha) - \omega^\alpha_{2\mathrm{skw}(\mathbf{T}^\alpha)} , \tag{2.143}$$

where $\omega^\alpha_{2\mathrm{skw}(\mathbf{T}^\alpha)} = \mathbf{I} \times \mathbf{T}^\alpha$ is the axial vector associated with its skew symmetric tensor $\left(\mathbf{T}^\alpha - \mathbf{T}^{\alpha T}\right)$ exactly as $\hat{\mathbf{m}}^\alpha$ was associated with $\hat{\mathbf{M}}^\alpha$ in 2.141. Finally, after substitution of 2.143 and 2.138 in 2.142, the balance of angular momentum reads

$$\int_\Omega \mathbf{x} \times \underbrace{\left(\rho^\alpha (\mathbf{v}_\alpha)'_\alpha - \rho^\alpha \mathbf{b} - \hat{\mathbf{p}}^\alpha - \operatorname{div}(\mathbf{T}^\alpha)\right)}_{= 0 \text{ by balance of momentum } 2.135} - \hat{\mathbf{m}}^\alpha + \omega^\alpha_{2\mathrm{skw}(\mathbf{T}^\alpha)} \,\mathrm{d}v = \int_\Omega -\hat{\mathbf{m}}^\alpha + \omega^\alpha_{2\mathrm{skw}(\mathbf{T}^\alpha)} \,\mathrm{d}v . \tag{2.144}$$

Hence, the localized form results in

$$\omega^\alpha_{2\mathrm{skw}(\mathbf{T}^\alpha)} = \hat{\mathbf{m}}^\alpha \quad \Leftrightarrow \quad \mathbf{T}^\alpha - (\mathbf{T}^\alpha)^T = \hat{\mathbf{M}}^\alpha . \tag{2.145}$$

Since we deal with non-polar material (i. e., $\mathbf{T}^\alpha$ is symmetric), the balance of angular momentum 2.145 is boiled down to

$$\hat{\mathbf{M}}^\alpha = \mathbf{0} \qquad \text{and} \qquad \mathbf{T}^\alpha = (\mathbf{T}^\alpha)^T . \tag{2.146}$$

Accordingly, for non-polar materials, there is no moment of momentum production and the balance of angular momentum is automatically satisfied provided that the mass and momentum balances are already met.

[20] Let $\mathbf{w}$ be any vector and let $\mathbf{W}$ be the corresponding skew-symmetric tensor as described in 2.141. Then for any arbitrary vector $\mathbf{a}$, we have

$$\mathbf{W}\mathbf{a} = \mathbf{w} \times \mathbf{a} .$$

Hence, $\mathbf{w}$ is a real eigenvector of $\mathbf{W}$ because

$$\mathbf{W}\mathbf{w} = \mathbf{w} \times \mathbf{w} = \mathbf{0} = 0\mathbf{w} .$$

In fact, $\mathbf{w}$ is the only real eigenvector. This can be deduced by noticing that the characteristic equation $|\mathbf{W} - \lambda \mathbf{I}| = 0$ has only one real solution (i. e., $\lambda = 0$).

[21] The proof is simple. See , for example, page 117 in [1]

2.4.4 First law of thermodynamics (Balance of energy)

The total specific energy[22] stored in a particle $\mathcal{P}^\alpha \in \Omega$, is defined to be the summation of two quantities, the particle specific internal energy ε^α and the particle specific kinetic energy $\frac{1}{2}\|\mathbf{v}_\alpha\|^2$. Since each constituent φ^α is considered to be a collection of an infinite number of particles, the total energy of φ^α is the summation of all the particle energies and is attained via

$$E^\alpha = \int_\Omega \left(\varepsilon^\alpha + \frac{1}{2}\|\mathbf{v}_\alpha\|^2\right) \mathrm{d}m^\alpha = \int_\Omega \rho^\alpha \left(\varepsilon^\alpha + \frac{1}{2}\|\mathbf{v}_\alpha\|^2\right) \mathrm{d}v \quad \forall \alpha \in \{S, F\}. \tag{2.147}$$

To compute the rate of change of the energy $\frac{\mathrm{d}_\alpha E^\alpha}{\mathrm{d}t}$, we observe that (2.147) is in a form right for the application of (2.128). Hence,

$$(E^\alpha)'_\alpha = \int_\Omega \rho^\alpha \left((\varepsilon^\alpha)'_\alpha + \mathbf{v}_\alpha \cdot (\mathbf{v}_\alpha)'_\alpha\right) \mathrm{d}v. \tag{2.148}$$

According to the first law of thermodynamics, the above rate of change of the total energy must equal to the sum of external mechanical power [23] and the rate of change of heat[24] as well as an additional energy production $\hat{e}^\alpha$ coming from the other constituent as local interaction of energy due to Truesdell's metaphysical principles for mixtures

$$\begin{aligned} \underbrace{\int_\Omega \rho^\alpha (\varepsilon^\alpha)'_\alpha + \rho^\alpha \mathbf{v}_\alpha \cdot (\mathbf{v}_\alpha)'_\alpha \ \mathrm{d}v}_{\text{rate of change of energy } (E^\alpha)'_\alpha} &= \underbrace{\int_\Omega \mathbf{v}_\alpha \cdot \mathbf{b}^\alpha \, \mathrm{d}v + \int_\Gamma \mathbf{v}_\alpha \cdot \mathbf{t}^\alpha \mathrm{d}a}_{\text{mechanical power}} \\ &+ \underbrace{\int_\Omega \rho^\alpha r^\alpha \, \mathrm{d}v - \int_\Gamma \mathbf{q}^\alpha \mathbf{n} \mathrm{d}a}_{\text{rate of change of heat}} + \underbrace{\int_\Omega \hat{e}^\alpha \, \mathrm{d}v}_{\text{energy production}} \end{aligned} \tag{2.149}$$

Since energy $\hat{e}^\alpha$ given to one constituent is taken from another one, their total sum must vanish. Namely,

$$\sum_\alpha \hat{e}^\alpha = \hat{e}^S + \hat{e}^F = 0. \tag{2.150}$$

Then, (2.149) is modified by the following changes, (1) by Cauchy's theorem, set ($\mathbf{t}^\alpha = \mathbf{T}^\alpha \mathbf{n}$), (2) apply the divergence theorem on ($\int_\Gamma \mathbf{v}_\alpha \cdot \mathbf{T}^\alpha \mathbf{n} \, \mathrm{d}a$) and use a suitable identity[25], (3) apply

[22] The prefix 'specific' is commonly used in thermodynamics to mean division by mass. Thus, specific energy of particle $\mathcal{P}^\alpha$ is energy per particle mass, where the particle mass is $\mathrm{d}m^\alpha$.

[23] The power is defined as force times velocity. The external mechanical power is caused by three external forces: the body load $\mathbf{b}^\alpha$, the traction load $\mathbf{t}^\alpha$, and the the external interaction load $\hat{\mathbf{p}}^\alpha$. However the last is implicitly included in the energy production $\hat{e}^\alpha$.

[24] r denotes the heat production per unit mass per unit time and $\mathbf{q}$ is the heat flux vector and defined as heat per unit area per unit time

[25] The identity reads $\operatorname{div}(\mathbf{v}_\alpha \cdot \mathbf{T}^\alpha) = \mathbf{v}_\alpha \cdot \operatorname{div}(\mathbf{T}^\alpha) + \operatorname{grad}(\mathbf{v}_\alpha) : \mathbf{T}^\alpha$

the divergence theorem on ($\int_\Gamma \mathbf{q}^\alpha \mathbf{n}\,\mathrm{d}a$), (4) add and subtract ($\int_\Omega \mathbf{v}_\alpha \cdot \hat{\mathbf{p}}^\alpha\,\mathrm{d}v$), and finally, (5) substitute 2.148 in 2.149 and after some re-arrangements, we obtain

$$\begin{aligned}\int_\Omega \rho^\alpha(\varepsilon^\alpha)'_\alpha - \hat{e}^\alpha - \rho^\alpha r^\alpha + \mathbf{v}_\alpha\cdot\hat{\mathbf{p}}^\alpha - \operatorname{grad}(\mathbf{v}_\alpha):\,\mathbf{T}^\alpha + \operatorname{div}(\mathbf{q}^\alpha)\,\mathrm{d}v \\ = \int_\Omega \mathbf{v}_\alpha\cdot\underbrace{\big(-(\mathbf{v}_\alpha)'_\alpha + \operatorname{div}(\mathbf{T}^\alpha) + \mathbf{b}^\alpha + \hat{\mathbf{p}}^\alpha\big)}_{=\,0,\text{ cf. }\boxed{2.135}}\mathrm{d}v = 0\,.\end{aligned} \tag{2.151}$$

Accordingly, the local form of the first law of thermodynamics for constituent α reads

$$\rho^\alpha(\varepsilon^\alpha)'_\alpha = \hat{e}^\alpha + \rho^\alpha r^\alpha + \operatorname{grad}(\mathbf{v}_\alpha):\,\mathbf{T}^\alpha - \mathbf{v}_\alpha\cdot\hat{\mathbf{p}}^\alpha - \operatorname{div}(\mathbf{q}^\alpha)\,. \tag{2.152}$$

By virtue of 2.150 and 2.133, the local form of the first law of thermodynamics for the mixture is given by:

$$\sum_\alpha \underbrace{\rho^\alpha(\varepsilon^\alpha)'_\alpha - \operatorname{grad}(\mathbf{v}_\alpha):\,\mathbf{T}^\alpha}_{\text{internal energy terms}} + \underbrace{\operatorname{div}(\mathbf{q}^\alpha) - \rho^\alpha r^\alpha}_{\text{thermal terms}} \quad \underbrace{+\mathbf{v}_\alpha\cdot\hat{\mathbf{p}}^\alpha}_{\text{interaction terms}} \quad = 0\,, \tag{2.153}$$

Because in this work we only deal with purely mechanical problems, in which thermal effects are insignificant, the above equation boils down to so-called balance of mechanical energy

$$\sum_\alpha \rho^\alpha(\varepsilon^\alpha)'_\alpha = \sum_\alpha \mathbf{T}^\alpha:\,\operatorname{grad}(\mathbf{v}_\alpha) - \sum_\alpha \mathbf{v}_\alpha\cdot\hat{\mathbf{p}}^\alpha\,. \tag{2.154}$$

It turned out that 2.154 is not an additional independent equation[26], and it is automatically satisfied if the previous balances are met. Therefore, for our purely mechanical problems, the first law of thermodynamics will be removed from the list of equations, we seek to solve.

[26] To see this, do the following:

- multiply 2.135 by $\mathbf{v}^\alpha$, integrat over Ω and build the sum:

$$\sum_\alpha \int_\Omega \mathbf{v}_\alpha\cdot\big(\rho^\alpha(\mathbf{v}_\alpha)'_\alpha - \rho^\alpha\mathbf{b} - \hat{\mathbf{p}}^\alpha - \operatorname{div}(\mathbf{T}^\alpha)\big)\,\mathrm{d}v = 0\,.$$

- use the identity $\Big(\operatorname{div}(\mathbf{v}_\alpha\cdot\mathbf{T}^\alpha) = \mathbf{v}_\alpha\cdot\operatorname{div}(\mathbf{T}^\alpha) + \operatorname{grad}(\mathbf{v}_\alpha):\,\mathbf{T}^\alpha\Big)$ and then use Cauchy's theorem. We finally obtain:

$$\underbrace{\sum_\alpha \int_\Omega \mathbf{T}^\alpha:\,\operatorname{grad}(\mathbf{v}_\alpha) - \mathbf{v}_\alpha\cdot\hat{\mathbf{p}}^\alpha\mathrm{d}v + \int_\Omega \rho^\alpha\mathbf{v}_\alpha\cdot(\mathbf{v}_\alpha)'_\alpha\,\mathrm{d}v}_{\text{rate of change of energy, see }\boxed{2.149}\text{ where }(\varepsilon^\alpha)'_\alpha\text{ is given in }\boxed{2.154}} = \sum_\alpha \underbrace{\int_\Omega \mathbf{v}_\alpha\cdot\mathbf{b}^\alpha\,\mathrm{d}v + \int_\Gamma \mathbf{v}_\alpha\cdot\mathbf{t}^\alpha\mathrm{d}a}_{\text{Mechanical power}}$$

From the above equation, we can see that the balance of mechanical energy of the mixture is satisfied (i. e., the sum of 2.149 for all constituents after using 2.150 and removing all thermal terms).

2.4.5 Legendre transform and Helmholtz free energy

The specific internal energy ε^{α} is known to be a function of two independent arguments, for example, the deformation tensor $\mathbf{F}_{\alpha}$ and the entropy η^{α}. That is,

$$\varepsilon^{\alpha} = \varepsilon^{\alpha}\left(\mathbf{F}_{\alpha}, \eta^{\alpha}\right) . \tag{2.155}$$

Since there is no laboratory equipment that allows us to control or even measure the entropy η^{α} of our system, we usually replace it with another quantity one can control and characterize such as the temperature which can be measured and controlled by (for instance) thermometers and thermostats. To do so, we define a function χ^{α} such that

$$\chi^{\alpha} = \varepsilon^{\alpha} - \frac{\partial \varepsilon^{\alpha}}{\partial \eta^{\alpha}} \eta^{\alpha} = \varepsilon^{\alpha} - \Theta^{\alpha} \eta^{\alpha} , \tag{2.156}$$

where $\Theta^{\alpha} = \frac{\partial \varepsilon^{\alpha}}{\partial \eta^{\alpha}}$ is the temperature in Kelvin and χ^{α} is an example of a Legendre transform function[27] and called the free energy function[28]. Next, we compute the differential $\mathrm{d}\chi^{\alpha}$ as

$$\begin{aligned}
\mathrm{d}\chi^{\alpha} &= \mathrm{d}\varepsilon^{\alpha} - \mathrm{d}\Theta^{\alpha}\eta^{\alpha} - \Theta^{\alpha}\mathrm{d}\eta^{\alpha} \\
&= \frac{\partial \varepsilon^{\alpha}}{\partial \mathbf{F}_{\alpha}} : \mathrm{d}\mathbf{F}_{\alpha} + \frac{\partial \varepsilon^{\alpha}}{\partial \eta^{\alpha}} \mathrm{d}\eta^{\alpha} - \mathrm{d}\Theta^{\alpha}\eta^{\alpha} - \Theta^{\alpha}\mathrm{d}\eta^{\alpha} \\
&= \frac{\partial \varepsilon^{\alpha}}{\partial \mathbf{F}_{\alpha}} : \mathrm{d}\mathbf{F}_{\alpha} + \Theta^{\alpha}\mathrm{d}\eta^{\alpha} - \mathrm{d}\Theta^{\alpha}\eta^{\alpha} - \Theta^{\alpha}\mathrm{d}\eta^{\alpha} \\
&= \frac{\partial \varepsilon^{\alpha}}{\partial \mathbf{F}_{\alpha}} : \mathrm{d}\mathbf{F}_{\alpha} - \eta^{\alpha}\mathrm{d}\Theta^{\alpha} .
\end{aligned} \tag{2.157}$$

The above equation tells us that any change in χ^{α} is actually a result of change in Θ^{α} and $\mathbf{F}_{\alpha}$. Hence, χ^{α} must be a function of two measurable arguments, the deformation gradient $\mathbf{F}_{\alpha}$ and absolute the temperature Θ^{α}. That is,

$$\chi^{\alpha} = \chi^{\alpha}\left(\mathbf{F}_{\alpha}, \Theta^{\alpha}\right) . \tag{2.158}$$

[27] Let $f = f(x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n)$ be a function of n variables. Assume that $\{x_1, x_2, \ldots, x_k\}$ is the set of variables we intend to keep and assume that $\{x_{k+1}, x_{k+2}, \ldots, x_n\}$ is the set of variables to be switched. Next, define $\Theta_i = \frac{\partial f}{\partial x_i}$ $\forall i > k$. Then, the Legendre transform of f is given by

$$g = f - \sum_{i=k+1}^{n} x_i \, \Theta_i .$$

Following the same argument in 2.157, we observe that $g = g\left(x_1, x_2, \ldots, x_k, \Theta_{k+1}, \ldots, \Theta_n\right)$ is now independent of x_i $\forall i > k$.

[28] The free energy (or Helmholtz free energy due to Hermann von Helmholtz) is the internal energy after subtracting the wasted energy (the heat $Q^{\alpha} = \Theta^{\alpha}\eta^{\alpha}$). Hence, it is the energy which we can get a maximum useful work from in an isothermal thermodynamic process for free. Why for free? The reason is that the heat is by itself useful unless we turn it into useful energy using a heat engine such as refrigerators or car engines which consume energy (for example electricity, fuel) and hence not free.

$\boxed{2.157}$ also tells us that the entropy η^α is now left free to change, which makes the experiments more convenient as we no longer need a device to put a control on η^α. It is interesting that the above equation will allow us to eliminate the undesirable entropy term from the second law of thermodynamics as will be shown in section 2.4.7. If we divide $\boxed{2.157}$ by dt and further assume an isothermal process $\left(\frac{\mathrm{d}_\alpha \Theta^\alpha}{\mathrm{d}t} = 0\right)$, we find that $(\chi^\alpha)'_\alpha$ is given by

$$\begin{aligned}
\frac{\mathrm{d}_\alpha \chi^\alpha}{\mathrm{d}t} &= \frac{\partial \chi^\alpha}{\partial \mathbf{F}_\alpha} : \frac{\mathrm{d}_\alpha \mathbf{F}_\alpha}{\mathrm{d}t} \\
&= \frac{\partial \chi^\alpha}{\partial \mathbf{F}_\alpha} : \underbrace{(\operatorname{grad}(\mathbf{v}_\alpha)\ \mathbf{F}_\alpha)}_{\text{cf. } \boxed{2.69} \,\&\, \boxed{2.75}} \\
&= \frac{\partial \chi^\alpha}{\partial \mathbf{F}_\alpha} : (\operatorname{grad}(\mathbf{v}_\alpha)\mathbf{F}_\alpha) \\
&= \frac{\partial \chi^\alpha}{\partial \mathbf{F}_\alpha} \mathbf{F}_\alpha^{\mathrm{T}} : \operatorname{grad}(\mathbf{v}_\alpha)\,. \qquad \boxed{2.159}
\end{aligned}$$

Under incompressibility assumption for the solid and fluid constituent, the Helmholtz free energy functions and their rate of changes (for isothermal process) are modeled by [29]

$$\begin{aligned}
\chi^S = \chi^S(\mathbf{F}_S) \quad &\rightarrow \quad \frac{\mathrm{d}_S \chi^S}{\mathrm{d}t} = \frac{\partial \chi^S}{\partial \mathbf{F}_S} \mathbf{F}_S^{\mathrm{T}} : \operatorname{grad}(\mathbf{v}_S) \\
\chi^F = \chi^F(-) \quad &\rightarrow \quad \frac{\mathrm{d}_F \chi^F}{\mathrm{d}t} = 0\,. \qquad \boxed{2.160}
\end{aligned}$$

The above relation is required for deriving thermodynamically consistent stresses. The graphical method is another way to introduce the Legendre transform. For more details about this transform and its graphical representation, interested readers are referred to [52, 113, 2, 87]

2.4.6 Second law of thermodynamics (entropy principal)

We use the second law of thermodynamics to derive the Clausius-Duhem inequality, which we shall use to gain restrictions for constitutive equations. The second law of thermodynamics relates the rate of change of the total entropy to the net heat and the absolute temperature. For mixture theory, this has to further account for Truesdell's metaphysical principles for mixtures when considering constituent φ^α alone. Hence, the mathematical statement of the second law reads

$$\frac{\mathrm{d}_\alpha}{\mathrm{d}t} \int_\Omega \rho^\alpha \eta^\alpha \mathrm{d}v \geq \int_\Omega \hat{\eta}^\alpha \mathrm{d}v + \int_\Omega \rho^\alpha \frac{r^\alpha}{\Theta^\alpha} \mathrm{d}v - \int_\Gamma \frac{\mathbf{q}^\alpha \mathbf{n}}{\Theta^\alpha} \mathrm{d}a\,, \qquad \boxed{2.161}$$

where η^α is the specific entropy of constituent φ^α and $\hat{\eta}^\alpha$ is the entropy production per unit volume such that

$$\hat{\eta}^S + \hat{\eta}^F \geq 0\,. \qquad \boxed{2.162}$$

[29] for more details, see [38]

The left hand side of 2.161 is in a form right for application of 2.128, while the most right term is ready for the application of the divergence theorem, which give

$$\int_\Omega \rho^\alpha (\eta^\alpha)'_\alpha - \rho^\alpha \frac{r^\alpha}{\Theta^\alpha} + \operatorname{div}\left(\frac{\mathbf{q}^\alpha}{\Theta^\alpha}\right) - \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta^\alpha)}{(\Theta^\alpha)^2} - \hat{\eta}^\alpha \,\mathrm{d}v \geq 0 . \tag{2.163}$$

Hence, the localized form of the second law of thermodynamics for individual constituent φ^α reads

$$\rho^\alpha (\eta^\alpha)'_\alpha \geq \rho^\alpha \frac{r^\alpha}{\Theta^\alpha} - \operatorname{div}\left(\frac{\mathbf{q}^\alpha}{\Theta^\alpha}\right) + \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta^\alpha)}{(\Theta^\alpha)^2} + \hat{\eta}^\alpha \tag{2.164}$$

and for the whole mixture

$$\sum_\alpha \rho^\alpha (\eta^\alpha)'_\alpha \geq \sum_\alpha \left(\rho^\alpha \frac{r^\alpha}{\Theta^\alpha} - \operatorname{div}\left(\frac{\mathbf{q}^\alpha}{\Theta^\alpha}\right) + \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta^\alpha)}{(\Theta^\alpha)^2}\right) \tag{2.165}$$

and for the constrained mixture we set $\Theta^\alpha = \Theta$.

2.4.7 Clausius-Duhem inequality

The Clausius-Duhem inequality is a combination of the first and second law of thermodynamics by means of the Helmholtz free energy. The derivations are simple and start by applying the material time derivative $\frac{\mathrm{d}_\alpha}{\mathrm{d}t}$ on 2.156 then pre-multiplying the result by ρ^α and substituting 2.152 and 2.164 as follows:

$$\begin{aligned} \rho^\alpha (\chi^\alpha)'_\alpha + \rho^\alpha (\Theta^\alpha)'_\alpha = & \underbrace{\rho^\alpha \left((\varepsilon^\alpha)'_\alpha - \Theta^\alpha (\eta^\alpha)'_\alpha\right)}_{= \text{1st law } \boxed{2.152} \text{ - } \Theta^\alpha \times \text{ 2nd law } \boxed{2.164}} \\ & \leq \operatorname{grad}(\mathbf{v}_\alpha) : \mathbf{T}^\alpha - \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta^\alpha)}{\Theta^\alpha} + \hat{e}^\alpha - \Theta^\alpha \, \hat{\eta}^\alpha - \mathbf{v}_\alpha \cdot \hat{\mathbf{p}}^\alpha . \end{aligned} \tag{2.166}$$

For the constrained mixture ($\Theta^\alpha = \Theta$), this reads

$$\rho^\alpha (\chi^\alpha)'_\alpha + \rho^\alpha (\Theta)'_\alpha \leq \operatorname{grad}(\mathbf{v}_\alpha) : \mathbf{T}^\alpha - \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta)}{\Theta} + \hat{e}^\alpha - \Theta \, \hat{\eta}^\alpha - \mathbf{v}_\alpha \cdot \hat{\mathbf{p}}^\alpha . \tag{2.167}$$

Assuming constrained mixture and using 2.162 and 2.150, the Clausius-Duhem inequality for the whole mixture is given by

$$\sum_\alpha \rho^\alpha (\chi^\alpha)'_\alpha + \rho^\alpha (\Theta)'_\alpha \leq \sum_\alpha \operatorname{grad}(\mathbf{v}_\alpha) : \mathbf{T}^\alpha - \frac{\mathbf{q}^\alpha \cdot \operatorname{grad}(\Theta)}{\Theta} - \mathbf{v}_\alpha \cdot \hat{\mathbf{p}}^\alpha . \tag{2.168}$$

As we shall deal with constrained type of mixtures that undergo isothermal ($\Theta = \text{const.}$) process, the above inequality reduces to the so-called Clausius-Planck inequality:

$$\sum_\alpha \rho^\alpha (\chi^\alpha)'_\alpha \leq \sum_\alpha \operatorname{grad}(\mathbf{v}_\alpha) : \mathbf{T}^\alpha - \mathbf{v}_\alpha \cdot \hat{\mathbf{p}}^\alpha . \tag{2.169}$$

The above two relations are used to gain restrictions for constitutive equations (will be discussed in the next section). It tells us that for any thermodynamically consistent (or admissible) process, the free energy (which is a positive scalar quantity) should never increase. In other words, the isolated systems reach at state of equilibrium when the free energy is minimum.

2.5 Constitutive equations

The fundamental principles of constitutive modeling for single continua are well established and dated back to the early works by Truesdell, Noll and Coleman (cf. , [92, 68, 69, 19]) and are also discussed in many modern text books such as [18]. Passman, Nunziato and Walsh in [77] confirmed the applicability of these principles on porous media and further introduced the so-called principle of phase separation for porous media. Most of these principles will be mentioned in different places inside this section for isothermal processes.

2.5.1 Principle of dissipation

Following the principle of dissipation for constitutive modeling, the second law of thermodynamics must be fulfilled. To derive thermodynamically admissible (or consistent) constitutive relations, we first recall that our solutions must satisfy all the following equations:

- The linear momentum balances

$$\underbrace{\rho^S (\mathbf{v}_S)'_S = \rho^S \mathbf{b} + \hat{\mathbf{p}}^S + \operatorname{div}(\mathbf{T}^S)}_{=\mathbf{g}_S} = \mathbf{0} \qquad \text{and} \qquad \underbrace{\rho^F (\mathbf{v}_F)'_F = \rho^F \mathbf{b} + \hat{\mathbf{p}}^F + \operatorname{div}(\mathbf{T}^F)}_{=\mathbf{g}_F} = \mathbf{0}, \tag{2.170}$$

- The angular momentum balances:

$$\underbrace{\mathbf{T}^S - \left(\mathbf{T}^S\right)^T}_{=\mathbf{G}_S} = \mathbf{0} \qquad \text{and} \qquad \underbrace{\mathbf{T}^F - \left(\mathbf{T}^F\right)^T}_{=\mathbf{G}_F} = \mathbf{0}, \tag{2.171}$$

- The continuity equation (see (2.126)):

$$\underbrace{n^S \mathbf{I} : \operatorname{grad}(\mathbf{v}_S) + n^F \mathbf{I} : \operatorname{grad}(\mathbf{v}_F) + \operatorname{grad}(n^F)(\mathbf{v}_F - \mathbf{v}_S)}_{=g_p} = 0, \tag{2.172}$$

- The Clausius-Planck inequality:

$$\underbrace{-\rho^S \left(\chi^S\right)'_S + \operatorname{grad}(\mathbf{v}_S) : \mathbf{T}^S - \rho^F \left(\chi^F\right)'_F + \operatorname{grad}(\mathbf{v}_F) : \mathbf{T}^F - (\mathbf{v}_F - \mathbf{v}_S) \cdot \hat{\mathbf{p}}^F}_{=\mathcal{D}_{\text{int}}} \geq 0. \tag{2.173}$$

Besides the above relations, it is also necessary to maintain the saturation condition (2.3) for any moment of time. Mathematically, this condition is fulfilled if the following two requirements are met:

- The saturation condition (2.3) is satisfied at the initial time $t = t_0$. That is,

$$\left(n^S + n^F\right)_{t=0} = n_0^S + n_0^F = 1\,, \tag{2.174}$$

- no change in (2.3) with time. Namely,

$$-\left(n^F + n^S\right)'_\alpha = 0\,, \qquad \text{where} \qquad \alpha \in \{S, F\} \tag{2.175}$$

The first requirement is satisfied by assumption and the second is equal to (2.172) [30] . Since $\mathcal{D}_{\text{int}} \geq \min \mathcal{D}_{\text{int}}$, the inequality (2.173) will be unconditionally satisfied if we find that $\min \mathcal{D}_{\text{int}} \geq 0$. This motivates us to examine the following minimization problem:

$$\begin{aligned} &\text{minimize} \quad \mathcal{D}_{\text{int}} \\ &\text{subject to} \quad (2.170),\ (2.171) \text{ and } (2.172). \end{aligned}$$

In order to save some spaces, assume that

$$\mathbf{L}_S = \operatorname{grad}(\mathbf{v}_S) \quad \text{and} \quad \mathbf{L}_F = \operatorname{grad}(\mathbf{v}_F)\,. \tag{2.176}$$

Then Lagrangian function, corresponding to our minimization problem, reads

$$\mathcal{L} = \mathcal{D}_{\text{int}} + p\, g_p + \hat{\mathbf{v}}_S \cdot \mathbf{g}_S + \hat{\mathbf{v}}_F \cdot \mathbf{g}_F + \hat{\mathbf{D}}^S : \mathbf{G}_S + \hat{\mathbf{D}}^F : \mathbf{G}_F \tag{2.177}$$

[30] This can be shown as follows:

$$\begin{aligned}
-\left(n^F + n^S\right)'_S &= -\left(n^S\right)'_S - \left(n^F\right)'_S \\
&= \underbrace{n^S \operatorname{div} \mathbf{v}_S}_{\text{cf. } (2.120)} - \underbrace{\left(\frac{\partial n^F}{\partial t} + \operatorname{grad}(n^F) \cdot \mathbf{v}_S\right)}_{\text{cf. } (2.9)} + \underbrace{\left(\operatorname{grad}(n^F) \cdot \mathbf{v}_F - \operatorname{grad}(n^F) \cdot \mathbf{v}_F\right)}_{= 0. \text{ (add and subtract)}} \\
&= \underbrace{n^S \operatorname{div} \mathbf{v}_S}_{\text{cf. } (2.120)} - \underbrace{\left(\frac{\partial n^F}{\partial t} + \operatorname{grad}(n^F) \cdot \mathbf{v}_F\right)}_{= (n^F)'_F \text{ cf. } (2.9)} + \operatorname{grad}(n^F) \cdot (\mathbf{v}_F - \mathbf{v}_S) \\
&= n^S \operatorname{div} \mathbf{v}_S + \underbrace{n^F \operatorname{div} \mathbf{v}_F}_{\text{cf. } (2.120)} + \operatorname{grad}(n^F) \cdot (\mathbf{v}_F - \mathbf{v}_S) \\
&= n^S \mathbf{I} : \operatorname{grad}(\mathbf{v}_S) + n^F \mathbf{I} : \operatorname{grad}(\mathbf{v}_F) + \operatorname{grad}(n^F) \cdot (\mathbf{v}_F - \mathbf{v}_S)\,,
\end{aligned}$$

The same result is obtained if we compute $\left(n^F + n^S\right)'_F$.

Where $\mathcal{D}_{\text{int}}$, g_p, $\mathbf{g}_S$, $\mathbf{g}_F$, $\mathbf{G}_S$ and $\mathbf{G}_F$ are defined in 2.170 - 2.173 and p is a Lagrange multiplier of a scalar quantity which posseses the unit of pressure, $\hat{\mathbf{v}}_S$ and $\hat{\mathbf{v}}_F$ are Lagrange multipliers of vector quantities with velocity unit and $\hat{\mathbf{D}}^S$ and $\hat{\mathbf{D}}^F$ are Lagrange multipliers of second-order tensors that have the unit of a velocity gradient. Next, expanding the first two terms in 2.177 and then making use of 2.160, we can express 2.177 as

$$\begin{aligned}\mathcal{L} = &\left(\mathbf{T}^S - \rho^S \frac{\partial \chi^S}{\partial \mathbf{F}_S}\mathbf{F}_S^{\mathrm{T}} + p\, n^S \mathbf{I}\right) : \mathbf{L}_S + \left(\mathbf{T}^F + p\, n^F \mathbf{I}\right) : \mathbf{L}_F - \left(\hat{\mathbf{p}}^F - p\,\operatorname{grad}(n^F)\right)\cdot(\mathbf{v}_F - \mathbf{v}_S)\\ &+ \underbrace{\hat{\mathbf{v}}_S \cdot \mathbf{g}_S + \hat{\mathbf{v}}_F \cdot \mathbf{g}_F + \hat{\mathbf{D}}^S : \mathbf{G}_S + \hat{\mathbf{D}}^F : \mathbf{G}_F}_{\text{no need to expand these terms}}\end{aligned} \tag{2.178}$$

A region β in which $\min \mathcal{D}_{\text{int}}$ may reside must satisfy the Kuhn-Tucker conditions altogether. Namely,

$$\beta = \{\text{solutions} : \begin{cases} \dfrac{\partial \mathcal{L}}{\partial p} = 0 & (\text{this gives } 2.172)\\[2ex] \dfrac{\partial \mathcal{L}}{\partial \mathbf{L}_S} = \mathbf{0} & \left(\text{this gives}\quad \mathbf{T}^S = \underbrace{\rho^S \dfrac{\partial \chi^S}{\partial \mathbf{F}_S}\mathbf{F}_S^{\mathrm{T}}}_{=\mathbf{T}_E^S} - p\, n^S \mathbf{I}\right)\\[2ex] \dfrac{\partial \mathcal{L}}{\partial \mathbf{L}_F} = \mathbf{0} & \left(\text{this gives}\quad \mathbf{T}^F = -p\, n^F \mathbf{I}\right)\\ \dfrac{\partial \mathcal{L}}{\partial \hat{\mathbf{v}}_\alpha} = \mathbf{0} & \left(\text{this gives } 2.170 \text{ where } \alpha \in \{S,F\}\right)\\ \dfrac{\partial \mathcal{L}}{\partial \hat{\mathbf{D}}^\alpha} = \mathbf{0} & \left(\text{this gives } 2.171 \text{ where } \alpha \in \{S,F\}\right)\\ & \vdots\\ & \vdots \end{cases} \tag{2.179}$$

We have not written the rest of these conditions (i. e., $\frac{\partial \mathcal{L}}{\partial \mathbf{v}_S} = \mathbf{0}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{v}_F} = \mathbf{0}$, etc.), because what mentioned above is indeed sufficient to show that the $\min \mathcal{D}_{\text{int}}$ is non-negative; Plugging the second and the third results of 2.179 into 2.173 gives

$$\begin{aligned}\mathcal{D}_{\text{int}} &= -pn^S \mathbf{I} \colon \operatorname{grad}(\mathbf{v}_S) - pn^F \mathbf{I} \colon \operatorname{grad}(\mathbf{v}_F) - (\mathbf{v}_F - \mathbf{v}_S)\cdot \hat{\mathbf{p}}^F\\ &= -p \underbrace{\left(n^S \mathbf{I} \colon \operatorname{grad}(\mathbf{v}_S) + n^F \mathbf{I} \colon \operatorname{grad}(\mathbf{v}_F) + \operatorname{grad}(n^F)(\mathbf{v}_F - \mathbf{v}_S)\right)}_{=0,\ \text{cf. } \frac{\partial \mathcal{L}}{\partial p} = 0 \text{ in } 2.179}\\ &\quad - \underbrace{\left(\hat{\mathbf{p}}^F - p\,\operatorname{grad}(n^F)\right)}_{\hat{\mathbf{p}}_E^F}\cdot(\mathbf{v}_F - \mathbf{v}_S) \quad \text{for all solutions in } \beta\end{aligned} \tag{2.180}$$

or shortly,

$$\min \mathcal{D}_{\text{int}} = -\hat{\mathbf{p}}_E^F \cdot (\mathbf{v}_F - \mathbf{v}_S) \qquad \text{with} \qquad \hat{\mathbf{p}}_E^F = \hat{\mathbf{p}}^F - p\,\operatorname{grad}(n^F) \tag{2.181}$$

for some solutions in β. Using the fact that the norm of any vector can never be negative, the above equation will be non-negative for arbitrary $\mathbf{v}_F - \mathbf{v}_S$ if $\hat{\mathbf{p}}_E^F$ is expressed by [31]

$$\hat{\mathbf{p}}_E^F = -\alpha^+(\mathbf{v}_F - \mathbf{v}_S) \qquad \text{with} \qquad \alpha^+ > 0\,, \tag{2.183}$$

where α^+ is a scalar positive function (which is based on experimental data and related to the original work of Darcy) and is approximated by

$$\alpha^+ = \frac{\left(n^F\right)^2 \gamma^{FR}}{k^F}\,. \tag{2.184}$$

The relation (2.183) is consistent with the principle of phase separation since the production (growth) terms depend on the kinematics of both constituents. Remark that $\hat{\mathbf{p}}_E^F$ parallel to the direction of the motion of the fluid relative to the solid constituent, therefore it must be associated with the drag force as the lift force acts perpendicular to the fluid motion. In this model, the effect of pore fluid viscosity (fluid friction) is considered only at the fluid/ solid interface, while the inertial effects and the capillary forces are assumed to be negligible (see [76, 70]).

In summary, from the calculations in (2.180), we have found that the second law of thermodynamics (2.173) will be automatically satisfied if (2.172) is met and $\mathbf{T}^S$, $\mathbf{T}^F$ and $\hat{\mathbf{p}}^F$ are chosen such that

$$\begin{aligned}
&\hat{\mathbf{p}}_E^F = -\frac{\left(n^F\right)^2 \gamma^{FR}}{k^F}(\mathbf{v}_F - \mathbf{v}_S) \qquad \text{and} \qquad \hat{\mathbf{p}}^F = \hat{\mathbf{p}}_E^F + p\,\operatorname{grad}(n^F)\\
&\mathbf{T}^S = \mathbf{T}_E^S - n^S\,p\mathbf{I} \qquad \text{where} \qquad \mathbf{T}_E^S = \rho^S \frac{\partial \chi^S}{\partial \mathbf{F}_S}\mathbf{F}_S^{\mathrm{T}}\\
&\mathbf{T}^F = -n^F p\mathbf{I}
\end{aligned} \tag{2.185}$$

An attempt to just fulfill the inequality mathematically without considering the real physical behavior of the examined material can lead to thermodynamically consistent results that are unable to predict this physical behavior. An important physical phenomenon needs to be checked here is related to the total stress,

$$\mathbf{T} = \mathbf{T}^S + \mathbf{T}^F \quad \rightarrow \quad \mathbf{T} = \mathbf{T}_E^S - p\mathbf{I}\,,$$

which appeared to be in full agreement with Terzaghi's classical concept of effective stress mentioned in section 2.3. The above thermodynamically admissible quantities ($\hat{\mathbf{p}}^F$, $\mathbf{T}^S$ and $\mathbf{T}^F$) are to be substituted in (2.170)-(2.171) [32], which yields the following PDEs:

[31] This holds only for the isotropic case. For general non-isotropic behavior, we replace the scalar function with the so-called permeability tensor $\mathbf{S}_\upsilon^+$ which is a positive definite tensor:

$$\hat{\mathbf{p}}_E^F = -(\mathbf{v}_F - \mathbf{v}_S)^{\mathrm{T}}\mathbf{S}_\upsilon^+\,. \tag{2.182}$$

However, in this work we will only deal with the isotropic permeability.

[32] This is equivalent to finding solutions which concurrently satisfy the last two statements in (2.179) and the other three statements right above them.

- Balance of momentum of the solid phase:

$$\rho^S (\mathbf{v}_S)'_S = \mathrm{div}\,(\mathbf{T}^S_E) + \rho^S \mathbf{b} + \frac{(n^F)^2\,\gamma^{FR}}{k^F}\,(\mathbf{v}_F - \mathbf{v}_S) - n^S \mathrm{grad}\,(p)\,, \tag{2.186}$$

- Balance of momentum of the fluid phase:

$$\rho^F\,(\mathbf{v}_F)'_S + \rho^F\ \mathrm{grad}\,(\mathbf{v}_F)\,(\mathbf{v}_F - \mathbf{v}_S) = \rho^F\,\mathbf{b} - \frac{(n^F)^2\,\gamma^{FR}}{k^F}\,(\mathbf{v}_F - \mathbf{v}_S) - n^F \mathrm{grad}\,(p)\,, \tag{2.187}$$

- Volume balance of the overall aggregate:

$$\mathrm{div}\,(n^F \mathbf{v}_F) + \mathrm{div}\,(n^S \mathbf{v}_S) = 0\,, \tag{2.188}$$

we seek to solve. Where

$$(\mathbf{u}_S)'_S = \mathbf{v}_S \tag{2.189}$$

is the velocity displacement relationship and

$$\mathbf{T}^S_E = \left(\mathbf{T}^S_E\right)^{\mathrm{T}}\,. \tag{2.190}$$

By using 2.185, the vector $\hat{\mathbf{p}}^F$ has been eliminated from 2.186-2.187 and a new scalar p has arisen instead. Despite of this reduction in the unknowns in the above underdetermined system, we are still far beyond from having unique solutions. This issue will be discussed and resolved in the subsequent subsection.

The idea of exploiting the Clausius-Duhem inequality for deriving restrictions on the constitutive relation is dated back to Coleman and Noll (cf. [19]) and was later modified by Müller and Liu (cf. [56] and [57]), who are the first to employ the Lagrange multipliers for obtaining such restrictions. In fact, Müller in [67] was the first to utilize a general entropy principle to gain restrictions on the constitutive relations for mixtures of fluids. These ideas inspired others (like Goodman and Cowin [46], Bowen [9], Bedford and Drumheller [3] and Nunziato and Passman [70]) for constructing the thermodynamically consistent theory for porous media. Actually, Goodman and Cowin in [46] are known to be the first scientists to use a Lagrangen multiplier to enter the constraint 2.121 into the Clausius-Duhem inequality for granular materials, while Nunziato and Passman [70] used a Lagrange multiplier to include the saturation condition 2.3 as constraint in the entropy inequality. Readers who are interested in basic techniques for generating thermodynamically admissible constitutive porous media models may consult [7, 8, 9, 32, 34, 35, 36, 29, 23, 24, 5].

2.5.2 Closure problem

Two identical structures, subjected to identical conditions, must produce identical responses. From a mathematical perspective, the fundamental problem here is to ensure the existence of unique solutions for our porous media problems. However, by comparing the number of unknowns in (2.186)-(2.190) with the number of given independent equations, we quickly recognize that the system is indeed underdetermined. This can be shown by counting the number of variables (where ndim is 1, 2 and 3 for 1D, 2D and 3D problems, respectively):

- 1 variable associated with p,
- ndim variables associated with $\mathbf{v}_F$,
- ndim variables associated with $\mathbf{v}_S$,
- ndim variables associated with $\mathbf{u}_S$,
- $\frac{\text{ndim}^2+\text{ndim}}{2}$ variables associated with $\mathbf{T}_E^S$ due to (2.190).

This results in a total of $\left(1 + 3\ \text{ndim} + \frac{\text{ndim}^2+\text{ndim}}{2}\right)$ unknowns versus $\left(1 + 3\ \text{ndim}\right)$ given independent equations. The set of equations is completed (closed) by linking $\mathbf{T}_E^S$ directly to the kinematics (e. g., $\mathbf{u}_S$) which results in a constitutive relation that depends on the constituent (rubber, steel, liquid, etc) itself. The use of this kind of linking is justified by two principles in continuum mechanics; the first is the so-called principle of determinism (cf. [97, 96]) which, for our isothermal process, states that the present state of $\mathbf{T}_E^S$ can be completely determined from its kinematic history and the second is referred to as the principle of phase separation (cf. [111, 33] which states that the constitutive quantities of constituent φ^α depend only on variables that belong to φ^α. Hence, the material modeler will need to link the kinetics (for example $\mathbf{P}_E^S$, $\mathbf{T}_E^S$, etc.) of solid constituent to kinematics ($\mathbf{F}_S$, $\mathbf{u}_S$, etc.) of the solid constituent only. To do so, the empirical observations on hyper-elastic constituents (materials) in labs revealed that

- for every macroscopic kinetic measure ($\mathbf{P}_E^S$, etc.), there is only one kinematic measure ($\mathbf{F}_S$, etc.) regardless of the chosen deformation path [33]. Mathematically, this implies that there must be a unique one-to-one mapping between kinematics and kinetics. Namely,

$$\mathbf{P}_E^S = \mathbf{P}_E^S(\mathbf{F}_S) \quad \Rightarrow \quad W^S = W^S(\mathbf{F}_S) \quad \text{because} \quad W^S = \mathbf{P}_E^S : \mathbf{F}_S\,, \qquad (2.191)$$

  where W^S is the deformation energy density function.

[33] A deformation path is frequently-used terminology in continuum mechanics which we shall use to indicate a deformation that starts from one configuration and (through a sequence of applied loading/ unloading) move through other configurations. A deformation path is **closed** if the starting and end configuration are identical.

- for every arbitrary closed deformation path, the constituent returns to its original kinetic and kinematic states retracing the exact same path. This observation tells us that the system is indeed conservative and that the kinetic measures and kinematic measures are state variables (path-independent variables). Hence, since $W^S = W^S(\mathbf{F}_S)$,

$$\text{then for every closed path } \int_{t_1}^{t_2} \frac{\mathrm{d}_S \mathbf{F}_S}{\mathrm{d}t}\,\mathrm{d}t = \mathbf{0}\text{, we must have } \int_{t_1}^{t_2} \frac{\mathrm{d}_S W^S}{\mathrm{d}t}\,\mathrm{d}t = 0 \ , \tag{2.192}$$

and using one of the well-known energetic conjugates in (2.112), we obtain

$$\int_{t_1}^{t_2} \frac{\mathrm{d}_S W^S}{\mathrm{d}t} = \int_{t_1}^{t_2} \mathbf{P}_E^S : \frac{\mathrm{d}_S \mathbf{F}_S}{\mathrm{d}t} \, . \tag{2.193}$$

Because of (2.192), the right hand side of the above equation must vanish. This is satisfied by setting

$$\mathbf{P}_E^S = \frac{\partial W^S}{\partial \mathbf{F}_S} \, . \tag{2.194}$$

But we know that the first Piola-Kirchhoff stress $\mathbf{P}_E^S$ can be directly linked to the Cauchy stress $\mathbf{T}_E^S$ using (2.95), which completes our system. Substituting the thermodynamically admissible $\mathbf{T}_E^S$ in the balance equation (2.186), then (2.186)-(2.189) become field equations and their solutions ($\mathbf{u}_S$, $\mathbf{v}_S$. $\mathbf{v}_F$ and p) are called a thermodynamic process. In addition, by using (2.95) together with (2.194), (2.185), (2.121) and (2.7), we obtain

$$W^S = \rho_0^S\, \chi^S \quad \text{where} \quad \rho_0^S = n_0^S\, \rho^{SR} \, . \tag{2.195}$$

It is a task of the material modeler to provide us with a suitable function W^S based on his empirical observations. However, there are some mathematical restrictions on W^S, which need to be maintained always. Interesting for us here are the material frame-indifference (frame-invariance or material objectivity) and the material isotropy (material symmetry), which will be discussed in the following subsection.

2.5.3 Material objectivity

Many scalar quantities such as temperature (Θ), mass ($\mathrm{d}m^S$), density (ρ^S), volume ($\mathrm{d}v^S$), porosity (n^F), etc. should have their magnitudes not effected by a chosen coordinate system (or observer). The same applies to our deformation energy density function W^S. In the literature, this principle is usually referred to as objectivity (also frame-indifference or frame-invariance) of W^S.

Suppose you (as observer) are sitting in front of tv or pc monitor. If you rotate the screen 15 degrees counter clockwise or if you (instead of rotating the screen) rotate your seat 15 degrees but in the opposite direction (that is 15 degrees clockwise), you will have the same view. The same applies to the translation. That is, if you shift the screen table 2 meters forward or if you move the seat 2 meters backward, you will also get the same view. Mathematically, this means that rotating our coordinate system by a rotational matrix $\mathbf{Q}^{\mathrm{T}}$ is equivalent to rotating our object by $\mathbf{Q}$. Hence, we can avoid switching the coordinate system (frame) of lab devices by simply applying rigid body motions (pure translation and/or pure rotation) on the object. In general, the rigid body motion for any instant of time [34] t is expressed by

$$\mathbf{x}^{*} = \mathbf{c}(t) + \mathbf{Q}(t)\,\mathbf{x}\,, \tag{2.196}$$

Where the vector $\mathbf{c}(t)$ and the second-order orthogonal[35] rotational tensor $\mathbf{Q}(t)$ represent a certain choice of pure translation and pure rotation, respectively. The main objective is to guarantee the frame-invariance of W^S for each configuration separately. However, since we assume that the reference configuration is undeformed (or stress free), the deformation energy W^S is zero or in other words, we in fact do not have W^S in the reference configuration and consequently, we do not need to check the frame invariance of W^S there. For this reason, it is standard practice to fix the reference configuration (and hence, any associated referential position $\mathbf{X}$) and only the current configuration will undergo rigid-body motions.

Because W^S is a function of $\mathbf{F}_S$, we need first to examine the behavior of $\mathbf{F}_S$ with respect to different rigid body motions. This is attained by applying $\frac{\partial}{\partial \mathbf{X}_S}$ on 2.196, which yields

$$\overset{*}{\mathbf{F}}_S = \mathbf{Q}(t)\,\mathbf{F}_S\,. \tag{2.197}$$

Observe that on the contrary to $\mathbf{F}_S$, the determinant $J_S = \det \mathbf{F}_S$ is not effected by rigid body motions because

$$\det \overset{*}{\mathbf{F}}_S = \underbrace{\det \mathbf{Q}(t)}_{=+1}\,\det \mathbf{F}_S = \det \mathbf{F}_S\,. \tag{2.198}$$

This guarantees the frame-invariance of the previously mentioned scalar quantities as they are functions of J_S as shown in 2.16, 2.121 and 2.16. Analogously, W^S should not be effected by rigid body motions. Namely, we must have

$$W^S(\mathbf{F}_S) = W^S(\underbrace{\mathbf{Q}(t)\,\mathbf{F}_S}_{=\overset{*}{\mathbf{F}}_S})\,. \tag{2.199}$$

[34] The time here is a pseudo time, which indicates only a certain choice of rigid-body motion or location of observer.

[35] Since $\mathbf{Q}\,\mathbf{Q}^T = \mathbf{I}$. A simple proof can be found, for example, in page 55 in [18].

Since (2.199) has to be satisfied for any arbitrary $\mathbf{Q}$, it must also work for the spacial case,

$$\mathbf{Q} = \mathbf{R}^T\,, \tag{2.200}$$

where $\mathbf{R}$ is a special orthogonal rotation tensor (see (2.38)-(2.39)), which has already been discussed in section 2.2.2 and defined such that

$$\mathbf{F}_S = \mathbf{R}\,\mathbf{U}_S \quad \text{where} \quad \mathbf{U}_S \text{ is the right stretch tensor}\,. \tag{2.201}$$

Substitution of (2.201) and (2.200) in the right hand side of (2.199), we obtain

$$W^S\left(\mathbf{F}_S\right) = W^S\left(\mathbf{U}_S\right)\,. \tag{2.202}$$

The above equation tells us that W^S depends on $\mathbf{F}_S$ only through the stretch tensor $\mathbf{U}_S$. This fact is consistent with what we observe in the real life. For example, a spring stores deformation energy if we deflect it but if we walk while carrying the deflected spring this will not cause any change in the stored energy. Moreover, from (2.40) and (2.197), we can easily conclude that

$$\mathbf{C}_S = \mathbf{U}_S^2 = \mathbf{F}_S^T\mathbf{F}_S = \overset{*}{\mathbf{F}}{}_S^{T}\overset{*}{\mathbf{F}}_S = \overset{*}{\mathbf{C}}_S\,, \tag{2.203}$$

which proves the frame-invariance of $\mathbf{C}_S$. Accordingly, by expressing W^S in term of $\mathbf{C}_S$, the objectivity of W^S is ensured. Thus, a frame-invariant W^S can be given by

$$W^S = W^S\left(\mathbf{U}_S\right) = W^S_{\mathbf{C}}\left(\mathbf{C}_S\right)\,. \tag{2.204}$$

Furthermore, we know that the three invariants[36] of $\mathbf{C}_S$, as a result of being functions of $\mathbf{C}_S$, must be frame-indifferent and consequently, we can express W^S as

$$W^S = W^S_I\left(I,\ I\!I,\ I\!I\!I\right)\,. \tag{2.205}$$

[36] From mathematical analysis, it is well-known that for any symmetric real second-order tensor $\mathbf{C} \in R^3 \times R^3$, the three eigenvalues λ_1^2, λ_2^2, λ_3^2 are real numbers. These eigenvalues are the solutions of the characteristic equation that is given below:

$$-\left|\mathbf{C}_S - \lambda^2\mathbf{I}\right| = \left(\lambda^2\right)^3 - I\ \left(\lambda^2\right)^2 + I\!I\ \left(\lambda^2\right) - I\!I\!I = 0\,,$$

where

$$\begin{aligned} I &= \operatorname{tr}\mathbf{C}_S = \lambda_1^2 + \lambda_2^2 + \lambda_3^2\,, \\ I\!I &= \frac{1}{2}\left(\left(\operatorname{tr}\mathbf{C}_S\right)^2 - \operatorname{tr}\mathbf{C}_S^2\right) = \lambda_1^2\,\lambda_2^2 + \lambda_1^2\,\lambda_3^2 + \lambda_2^2\,\lambda_3^2\,, \\ I\!I\!I &= \det\mathbf{C}_S = \left(\det\mathbf{F}_S\right)^2 = \left(J_S\right)^2 = \lambda_1^2\,\lambda_2^2\,\lambda_3^2\,, \end{aligned}$$

and the square roots of these eigenvalues give λ_1, λ_2, λ_3, which are the principal stretches and, as discussed in detail in section 2.2.2, these eigenvalues are also the eigenvalues of $\mathbf{B}_S$ and therefore, the principle invariants of $\mathbf{B}_S$ and $\mathbf{C}_S$ are identical.

The frame-invariance of the principal stretches λ_1, λ_2, λ_3 follows from being functions of the frame-invaraint *I*, *II* and *III* as stated in footnote [36]. Therefore, we can also write W^S in term of the pricipal stretches

$$W^S = W^S_\lambda\,(\lambda_1, \lambda_2, \lambda_3)\,. \tag{2.206}$$

Remember from section 2.2.2, the three principal stretches are nothing but the eigenvalues of $\mathbf{F}_S$, $\mathbf{U}_S$, $\mathbf{C}_S$ raised to power 1, 1 and $1/2$, respectively. Hence, they form a bridge that links (2.202) and (2.204) and so are the principal invariants as a result of being functions of these eigenvalues [36].

2.5.4 Material symmetry

Here, the orientation of the material fibers (or micro-structures) plays no role. For easy mathematical interpretation, we select an arbitrary infinitesimal ball[37] $B(\mathbf{X}_S, \varepsilon)$ from the reference configuration and then, theoretically (in imagination) rotate it with arbitrary rotation $\mathbf{Q}$ to change the direction of ball fibers or the orientation of the ball micro-structure. If the material is isotropic, then the response W^S in the current configuration should not be effected by this rotation. This must hold true for any arbitrary values of $\mathbf{X}_S$ and $\mathbf{Q}$. Symbolically, this is expressed as

$$\mathbf{X}_S^{\oslash} = \underbrace{\mathbf{Q}\,\mathbf{X}_S}_{\text{ball rotation}} \qquad \Rightarrow \qquad \mathbf{F}_S^{\oslash} = \underbrace{\frac{\partial \mathbf{x}_S}{\partial \mathbf{X}_S^{\oslash}} = \frac{\partial \mathbf{x}_S}{\partial \mathbf{X}_S}\frac{\partial \mathbf{X}_S}{\partial \mathbf{X}_S^{\oslash}}}_{\text{effect of rotation on } \mathbf{F}_S} = \mathbf{F}_S\,\mathbf{Q}^{\mathrm{T}}\,, \tag{2.207}$$

and we must have

$$W^S\,(\mathbf{F}_S) = W^S\,\underbrace{\left(\mathbf{F}_S\,\mathbf{Q}^{\mathrm{T}}\right)}_{=\mathbf{F}_S^{\oslash}}\,. \tag{2.208}$$

If the above relation is satisfied, then we say that the porous solid constituent is such that the orientation of the micro-structures is insignificant, the solid material is invariant to the rotation or the solid material possesses full-symmetry around the three coordinate axes. For simplicity, we may even (for mechanical purpose) assume that there exist no fibers or micro-structures. Since the above relation must hold true for any arbitrary rotation, it must also hold in particular for

$$\mathbf{Q} = \mathbf{R}\,, \qquad \text{where} \qquad \mathbf{F}_S = \mathbf{V}_S\,\mathbf{R} \tag{2.209}$$

[37] In mathematics, for example functional analysis books, the ball is expressed by $B(\mathbf{X}_S, \varepsilon)$ where $\mathbf{X}_S$ is the ball center and the very small ε is the ball radius. In continuum mechanics, this ball definition is usually referred to as neighborhood of $\mathbf{X}_S$.

with the left stretch tensor $\mathbf{V}_S$ and the special orthogonal rotation tensor $\mathbf{R}$ (see 2.38-2.39) already discussed in detail in section 2.2.2. Substituting 2.209 in 2.208, we obtain a mathematical restriction (for W^S) for the isotropic solid constituent,

$$W^S(\mathbf{F}_S) = W^S(\mathbf{V}_S)\,. \tag{2.210}$$

According to 2.40, the left Cauchy stress tensor $\mathbf{B}_S$ [38] is a function of $\mathbf{V}_S$. Namely,

$$\mathbf{B}_S = \mathbf{F}_S\,\mathbf{F}_S^{\mathrm{T}} = \mathbf{V}_S^2. \tag{2.213}$$

Therefore, we can also express W^S by

$$W^S(\mathbf{F}_S) = W^S(\mathbf{V}_S) = W^S(\mathbf{B}_S)\,. \tag{2.214}$$

Furthermore, we know that the three invariants[36] of $\mathbf{B}_S$, as a result of being functions of $\mathbf{B}_S$, must be isotropic. Hence, we can also write

$$W^S = W_I^S(I,\, II,\, III)\,. \tag{2.215}$$

We also know that the eigenvalues [36] of $\mathbf{B}_S$ are functions of those three principal invariants. Thus, we also obtain

$$W^S = W_\lambda^S(\lambda_1, \lambda_2, \lambda_3)\,. \tag{2.216}$$

2.5.5 Isotropy and objectivity

There are four available choices, 2.204-2.206, we can pick one of them to meet the objectivity restriction as well as another four available choices, 2.214-2.216, one of them is sufficient to meet the isotropy restriction. To satisfy isotropy and objectivity together, we have to select one of the mutual choices. Namely,

$$W^S = W_I^S(I,\, II,\, III) \qquad \text{or} \qquad W^S = W_\lambda^S(\lambda_1, \lambda_2, \lambda_3)\,. \tag{2.217}$$

Because $III = (J_S)^2$ [36], the above equation can be expressed by

$$W^S = W_I^S(I,\, II,\, J_S)\,. \tag{2.218}$$

[38] It is worth to mention that the right Cauchy stress tensor $\mathbf{C}_S$, mentioned in the previous subsection, is isotropic because

$$\mathbf{C}_S^{\oslash} = \mathbf{F}_S^{\oslash\mathrm{T}}\mathbf{F}_S^{\oslash} = \mathbf{F}_S^{\mathrm{T}}\underbrace{(\mathbf{Q}^{\mathrm{T}}\mathbf{Q})}_{=\mathbf{I}}\mathbf{F}_S = \mathbf{F}_S^{\mathrm{T}}\mathbf{F}_S = \mathbf{C}_S\,. \tag{2.211}$$

However, $\mathbf{B}_S$ does change when rotating the coordinate system and this can be seen as follows:

$$\mathbf{B}_S^{\oslash} = \mathbf{F}_S^{\oslash}\mathbf{F}_S^{\oslash\mathrm{T}} = \mathbf{Q}\underbrace{(\mathbf{F}_S\mathbf{F}_S^{\mathrm{T}})}_{\mathbf{B}_S}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}\mathbf{B}_S\mathbf{Q}^{\mathrm{T}}\,. \tag{2.212}$$

2.5.6 Hyper-elastic material model

Based on (2.17) and following Flory in [42], W^S can be uniquely decoupled into isochoric (partial volume-preserving, distortional) part W^S_{iso} and spherical (volumetric, dilatational) part W^S_{vol}:

$$W^S = W^S_{\text{iso}} + W^S_{\text{vol}} \,. \tag{2.219}$$

We can capture the partial volume preserving deformation (as in pure shearing and incompressible deformation) by setting $J_S = 1$ (see (2.16)). Thus the strain energy function W^S in (2.218) becomes

$$W^S_{\text{iso}} = W^S_I \,(I,\; II) \,. \tag{2.220}$$

Mooney and Rivlin (in [65, 80, 81]) indicated that W^S as being in c^∞ can be written as power series. Hence, the power series expansion of the above W^S_{iso} around the reference configuration reads

$$W^S_{\text{iso}} = \sum_{i,j=0}^{\infty} c_{ij}\; (I - I_0)^i\; (II - II_0)^j \,. \tag{2.221}$$

Remark on the reference configuration, we have $\mathbf{x} = \mathbf{X}_S$ and according to (2.14), $\mathbf{F}_S = \mathbf{I}$. Therefore, all the three eigenvalues (principal stretches λ_1 , λ_3 and λ_3) of $\mathbf{F}_S$ are equal to 1. Following footnote ^{36}, we obtain $I_0 = II_0 = 3$ and $III_0 = 1$. Consequently, (2.221) is equivalent to

$$W^S_{\text{iso}} = \sum_{i,j=0}^{\infty} c_{ij}\; (I - 3)^i\; (II - 3)^j \,. \tag{2.222}$$

Obviously, a vanishing W^S_{iso} in the undeformed configuration requires that $c_{00} = 0$. Based on (2.222), the most general form of linear (in I and II) W^S_{iso} for incompressible hyper-elastic material takes the form

$$W^S_{\text{iso}} = c_{10}\; (I - 3)\; + c_{01}\; (II - 3) \,. \tag{2.223}$$

A further simplification, results in the so-called Neo-Hookean model, referred to in (2.52),

$$W^S_{\text{iso}} = c_{10}\; (I - 3) \qquad \text{with} \qquad c_{10} = \frac{\mu_S}{2} \,, \tag{2.224}$$

where μ_S is called shear modulus. The single material parameter c_{10} is determined empirically. For example, one can use a lab device (such as tensile testing machine) that performs a uniaxial test for a bar of material and (at the same time) plots the stress-strain curve, which will be used to find c_{10}. The above Neo-Hooken model is in fact linear in strain as can be seen from (2.52).

For W^S_{vol} that is linear (in volumetric strain), the Hencky strain measure will be chosen for the reasons found in the paragraph directly below equation 2.48. Hence,

$$W^S_{\text{vol}} = \mu_S \; \ln J_S \tag{2.225}$$

is the good choice for large volumetric changes. Accordingly, a compressible Neo-Hooken model (cf. [75]), such that

$$W^S = \underbrace{\tfrac{1}{2}\mu_S\,(I-3)}_{W^S_{\text{iso}}} + \underbrace{\mu_S \; \ln J_S}_{W^S_{\text{vol}}}\,, \tag{2.226}$$

may seem to work fine to certain extent. However, the above model does not account for two physical observations; the point of compaction and permeability effect. Ehlers and Eipper in [39] therefore proposed the following constitutive relation (inspired by Ogden model[39]):

$$W^S = \tfrac{1}{2}\mu_S\,(I-3) + \mu_S \; \ln J_S + \underbrace{\lambda^S \; n^F_{0S}\,(J_S - 1) + \lambda^S \; \left(n^F_{0S}\right)^2 \; \ln\left(\frac{n^F}{n^F_0}\;J_S\right)}_{\text{additional volumetric expansion term } \tilde{U_S}}\,. \tag{2.227}$$

Here, the strain measure $\ln\left(\frac{n^F}{n^F_0}\;J_S\right)$ accounts for theses two physical observations as discussed in subsection 2.2.2, while the term $(J_S - 1)$ is another volumetric strain measure that belongs to the Seth family of volumetric strains as shown in 2.47. It is obvious that W^S is also linear in the additional strains $\ln\left(\frac{n^F}{n^F_0}\;J_S\right)$ and (J_S-1). Using 2.185, the effective Cauchy stress tensor $\mathbf{T}^S_E$ reads

$$\mathbf{T}^{\text{S}}_{\text{E}} = \frac{\mu^S}{J_{\text{S}}}(\mathbf{F}_S\mathbf{F}^{\text{T}}_S - \mathbf{I}) + \lambda^S(1-n^{\text{S}}_{0S})^2\Big(\frac{1}{1-n^{\text{S}}_{0S}} - \frac{1}{J_{\text{S}} - n^{\text{S}}_{0S}}\Big)\mathbf{I}. \tag{2.228}$$

[39] Just for comparison, Ogden in [72] proposed the following volumetric expansion of the form

$$\tilde{U_S} = \frac{\lambda^S}{\gamma^2}\left((J_S)^\gamma - 1 - \gamma \ln J_S\right)$$

and for $\gamma = 1$, we obtain

$$\tilde{U_S} = \lambda^S \; (J_S - 1) - \lambda^S \; \ln J_S\,.$$

To know why $\ln J_S$ was replaced by $\ln\left(\frac{n^F}{n^F_0}\;J_S\right)$, please look at subsection 2.2.2. Notice, $(J_S - 1)$ in the above equation is premultiplied by the relative pore size n^F_{0S} in 2.227, because a local volume change in a spatial point (differential volume element) is indeed attributed to the change in the size of the pores in that point.

Just for comparison, observe that the term $\frac{n^F}{n_0^F} J_S$ which is used to model the $\tilde{U_S}$ in 2.227 was also used by Markert (see section 5.2.2 in [59]) to model the permeability as follows:

$$k^{\mathrm{F}}(n^{\mathrm{F}}) = k_0^{\mathrm{F}} \Big(\frac{n^{\mathrm{F}}}{n_0^{\mathrm{F}}} \, J_S \Big)^{\kappa} . \tag{2.229}$$

This makes the above relation seems more consistent than Eipper's expression (see [39]),

$$k^{\mathrm{F}}(n^{\mathrm{F}}) = k_0^{\mathrm{F}} \Big(\frac{n^{\mathrm{F}}}{n_0^{\mathrm{F}}} \Big)^{\kappa} . \tag{2.230}$$

In fact, Markert expression shows better fit to the experimental data from the lab (see section 5.2.2 in [59]). The two expressions were discussed in the paragraphs directly below and before equation 2.62.

Material modeling is a complete and large topic in its own right and is out of scope of this work, which mainly concerns the numerical treatment of the governing equations. There are many things and conditions (polyconvexity, etc.) which are to be taken into account when developing a hyper-elastic constitutive relations and what has been mentioned so far is just like a drop in the ocean. The above constitutive relation for us is just a function, chosen for numerical testing. For profound knowledge about this material model and others, interested readers are referred to the PhD work of Eipper [40] and citations therein.

2.5.7 Linear elastic model

To compare with results from [63], we shall adopt the Hooke's elasticity law for infinitesimal deformation (small strain regime), for which the solid extra stress is determined by

$$\mathbf{T}_E^S = 2\mu^S \boldsymbol{\varepsilon}_S + \lambda^S (\boldsymbol{\varepsilon}_S \cdot \mathbf{I})\,\mathbf{I} \qquad \text{with} \qquad \boldsymbol{\varepsilon}_S = \tfrac{1}{2}\,(\operatorname{grad}\mathbf{u}_S + \operatorname{grad}^T\mathbf{u}_S) \tag{2.231}$$

as the geometrically linear solid strain tensor and μ^S, λ^S being the macroscopic Lamé constants of the porous solid matrix. The porosity, permeability are assumed to be constants and consequently the hydraulic conductivity will be so. Namely,

$$n^S = n_{0S}^S \qquad \text{and} \qquad n^F = n_{0S}^F \qquad \Rightarrow k^F = k_0^F \, . \tag{2.232}$$

3

Numerical Treatment

3.1 Initial boundary value problems (IBVP's)

3.1.1 IBVP 1: linear $\mathbf{uv}p$-formulation

Here, we deal only with infinitesimal linear elastic problem, in which the leading coefficients are assumed to be constant (see (2.231)-(2.232)) and the convective term is negligible.

We consider a region Ω in R^2 with Lipschitz boundary. The solid displacement , the solid velocity and the fluid velocity respectively, are the vector-valued functions $\mathbf{u}_S(\mathbf{x},t)$, $\mathbf{v}_S(\mathbf{x},t)$ and $\mathbf{v}_F(\mathbf{x},t)$ for $\mathbf{x} \in \bar{\Omega}$ and $0 \leq t \leq T$, while the pressure is the scalar-valued function $p(\mathbf{x},t)$ for $\mathbf{x} \in \Omega$ and $0 \leq t \leq T$. These functions, we seek to find, must satisfy (2.186)-(2.189), which we repeat here (considering the above assumptions) for the convenience of the reader:

- Balance of momentum of the solid phase (SMB):

$$\rho_0^S \, (\mathbf{v}_S)'_S = \operatorname{div}(\mathbf{T}_E^S) + \rho_0^S \mathbf{b} + \frac{(n_{0S}^F)^2 \, \gamma^{FR}}{k_0^F} \, (\mathbf{v}_F - \mathbf{v}_S) - n_{0S}^S \operatorname{grad}(p) \,, \tag{3.1}$$

- Balance of momentum of the fluid phase (FMB):

$$\rho_0^F \, (\mathbf{v}_F)'_S = \rho_0^F \, \mathbf{b} - \frac{(n_{0S}^F)^2 \, \gamma^{FR}}{k_0^F} \, (\mathbf{v}_F - \mathbf{v}_S) - n_{0S}^F \operatorname{grad}(p) \,, \tag{3.2}$$

- Volume balance of the overall aggregate:

$$\operatorname{div}(n_{0S}^F \mathbf{v}_F) + \operatorname{div}(n_{0S}^S \mathbf{v}_S) = 0 \,, \tag{3.3}$$

- velocity displacement relationship:

$$(\mathbf{u}_S)'_S = \mathbf{v}_S \,, \tag{3.4}$$

with the initial conditions

$$\mathbf{u}_S(\mathbf{x},0)=\mathbf{0}, \quad \mathbf{v}_S(\mathbf{x},0)=\mathbf{0} \quad \text{and} \quad \mathbf{v}_F(\mathbf{x},0)=\mathbf{0} \quad \forall \mathbf{x}\in\Omega\,. \tag{3.5}$$

The boundary $\Gamma=\partial\,\Omega$ is divided into Dirichlet ($\Gamma_{\mathbf{u}_S}$ and $\Gamma_{\mathbf{v}_F}$) and Neumann ($\Gamma_{\mathbf{t}^S}$ and $\Gamma_{\mathbf{t}^F}$) regions so that

$$\begin{aligned} \Gamma &= \Gamma_{\mathbf{u}_S}\cup\Gamma_{\mathbf{t}^S} \quad \text{with} \quad \Gamma_{\mathbf{u}_S}\cap\Gamma_{\mathbf{t}^S}=\phi \quad \text{for SMB}\,,\\ \Gamma &= \Gamma_{\mathbf{v}_F}\cup\Gamma_{\mathbf{t}^F} \quad \text{with} \quad \Gamma_{\mathbf{v}_F}\cup\Gamma_{\mathbf{t}^F}=\phi \quad \text{for FMB}\,. \end{aligned} \tag{3.6}$$

For Dirichlet conditions, we deal particularly with inviscid (frictionless) rigid wall boundary conditions

$$\begin{aligned} \mathbf{u}_S\cdot\mathbf{n}&=0 \quad \text{on } \Gamma_{\mathbf{u}_S}\,,\\ \mathbf{v}_F\cdot\mathbf{n}&=0 \quad \text{on } \Gamma_{\mathbf{v}_F}\,, \end{aligned} \tag{3.7}$$

where $\mathbf{n}$ denotes the normal to domain boundary Γ and for Neumann conditions, we have

$$\bar{\mathbf{t}}^S=\left(\mathbf{T}^S_E-n^S p\mathbf{I}\right)\cdot\mathbf{n}\,, \qquad\qquad \bar{\mathbf{t}}^F=-n^F p\mathbf{n}\,. \tag{3.8}$$

The rest of the BCs which are not defined as Dirichlet or Neumann are automatically considered 'Do-nothing' (zero Neumann). Another important requirement (will be discussed in the paragraph directly below equation (3.42)) on p necessary for the uniqueness of the solution is the vanishing mean. That is,

$$\int_\Omega p\,\mathrm{d}v=0\,. \tag{3.9}$$

3.1.2 IBVP 2: linear uw$p$-formulation

As in previous section, we only deal with convective-less linear elastic problem but now with the Darcy velocity,

$$\mathbf{w}=n^F_{0S}\,(\mathbf{v}_F-\mathbf{v}_S)\,, \tag{3.10}$$

being the second primary variable instead of the fluid velocity $\mathbf{v}_F$ and with the balance of momentum of the mixture in place of the balance of momentum of the solid constituent. Consequently, the IBVP goes as follows.

Consider a region Ω in R^2 with Lipschitz boundary. The solid displacement, the solid velocity and the Darcy velocity respectively, are the vector-valued functions $\mathbf{u}_S(\mathbf{x},t)$, $\mathbf{v}_S(\mathbf{x},t)$ and $\mathbf{w}(\mathbf{x},t)$ for $\mathbf{x}\in\bar{\Omega}$ and $0\le t\le T$, while the pressure is the scalar-valued function $p(\mathbf{x},t)$ for $\mathbf{x}\in\Omega$ and $0\le t\le T$. These functions, we seek to find, must satisfy the set of PDEs:

- Balance of momentum of the binary saturated mixture ((3.1)+(3.2)):

$$\rho_0\,(\mathbf{v}_S)'_S + \rho_0^{FR}\ (\mathbf{w})'_S - \operatorname{div}\mathbf{T}^S_E - \rho_0\mathbf{b} + \operatorname{grad} p = \mathbf{0}\,, \tag{3.11}$$

- Balance of momentum of the fluid phase (divided by n^F_{0S}):

$$\rho_0^{FR}\ (\mathbf{v}_S)'_S + \frac{\rho_0^{FR}}{n^F_{0S}}\,(\mathbf{w})'_S + \frac{\gamma^{FR}}{k^F}\,\mathbf{w} - \rho_0^{FR}\,\mathbf{b} + \operatorname{grad} p = \mathbf{0}\,, \tag{3.12}$$

- Volume balance of the binary saturated mixture:

$$\operatorname{div}(\mathbf{v}_S) + \operatorname{div}(\mathbf{w}) = 0\,, \tag{3.13}$$

- Velocity-displacement relationship:

$$(\mathbf{u}_S)'_S = \mathbf{v}_S \tag{3.14}$$

with the initial conditions

$$\mathbf{u}_S(\mathbf{x},0) = \mathbf{0}, \quad \mathbf{v}_S(\mathbf{x},0) = \mathbf{0} \quad \text{and} \quad \mathbf{w}(\mathbf{x},0) = \mathbf{0} \quad \forall \mathbf{x} \in \Omega\,. \tag{3.15}$$

The boundary $\Gamma = \partial\,\Omega$ is divided into Dirichlet ($\Gamma_{\mathbf{u}_S}$ and $\Gamma_{\mathbf{w}}$) and Neumann ($\Gamma_{\mathbf{t}}$ and $\Gamma_{\mathbf{t}^F}$) parts so that

$$\begin{aligned} \Gamma &= \Gamma_{\mathbf{u}_S} \cup \Gamma_{\mathbf{t}} \quad \text{with} \quad \Gamma_{\mathbf{u}_S} \cap \Gamma_{\mathbf{t}} = \phi \quad \text{for SMB}\,, \\ \Gamma &= \Gamma_{\mathbf{w}} \cup \Gamma_{\mathbf{t}^F} \quad \text{with} \quad \Gamma_{\mathbf{w}} \cup \Gamma_{\mathbf{t}^F} = \phi \quad \text{for FMB}\,. \end{aligned} \tag{3.16}$$

For the Dirichlet conditions, we deal particularly with inviscid (frictionless) rigid wall boundary conditions

$$\begin{aligned} \mathbf{u}_S \cdot \mathbf{n} &= 0 \quad \text{on } \Gamma_{\mathbf{u}_S}\,, \\ \mathbf{w} \cdot \mathbf{n} &= 0 \quad \text{on } \Gamma_{\mathbf{w}}\,, \end{aligned} \tag{3.17}$$

where $\mathbf{n}$ denotes the normal to domain boundary Γ and for Neumann conditions, we have

$$\bar{\mathbf{t}} = \left(\mathbf{T}^S_E - p\mathbf{I}\right) \cdot \mathbf{n}\,, \qquad\qquad \bar{\mathbf{t}}^F = -p\mathbf{n}\,. \tag{3.18}$$

Observe that the above Neumann conditions are more convenient than (3.8) since the surface traction $\bar{\mathbf{t}}$ acts simultaneously on both the solid and the fluid phase such that the separation of the boundary conditions is not needed anymore. $\mathbf{t}^F$ is nothing but the vector $\mathbf{n}$ scaled by the negative of the ambient atmospheric pressure.

The rest of the boundary edges, which are not assigned Dirichlet or Neumann condition, are automatically considered 'Do-nothing' (zero Neumann). Finally, for the uniqueness of p, we further require vanishing mean pressure as stated in (3.9).

3.1.3 IBVP 3: non-linear uw$p$-formulation

The non-linear PDEs (2.187)-(2.189) need to be first transformed into a Stokes-like or Navier-Stokes-like structure in order to exploit the powerful capabilities of CFD solvers available in FEATFLOW [1]. This means that our system of equations must be transformed into the following form

$$\begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{u}} \\ \mathrm{p} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \tag{3.19}$$

with $\mathbf{B}^T$ representing the gradient operator acting on the pressure p, $\mathbf{B}$ the adjoint divergence operator, $\tilde{\mathbf{A}}$ is a differential operator acting on non-pressure variables. For example, for stationary Stokes problem, $\tilde{\mathbf{A}}$ is related to the Laplace operator. To achieve this goal, a series of derivations and reformulations should be carried out.

To allow for the direct use of the standard θ-scheme, we start by differentiating $(\rho\ \mathbf{v}_S)$ by parts, which gives the following useful relation

$$(\rho\ \mathbf{v}_S)'_S = (\rho)'_S\ \mathbf{v}_S + \rho\ (\mathbf{v}_S)'_S \quad \rightarrow \quad \rho\ (\mathbf{v}_S)'_S = (\rho\ \mathbf{v}_S)'_S - (\rho)'_S\ \mathbf{v}_S \tag{3.20}$$

with the mixture density

$$\rho = \rho^S + \rho^F\,. \tag{3.21}$$

For the same reason, $\mathbf{w}/n^F$ is also differentiated by parts to obtain another useful relation

$$\left(\frac{\mathbf{w}}{n^F}\right)'_S = \frac{n^F\,(\mathbf{w})'_S - \left(n^F\right)'_S\,\mathbf{w}}{\left(n^F\right)^2} \quad \rightarrow \quad \rho^F\left(\frac{\mathbf{w}}{n^F}\right)'_S = -\rho^{FR}\frac{\left(n^F\right)'_S}{n^F}\mathbf{w} + \rho^{FR}\ (\mathbf{w})'_S\,. \tag{3.22}$$

Here $\left(n^F\right)'_S$ is computed by virtue of (2.120) and condition (2.3) as

$$\left(n^F\right)'_S = \left(1 - n^S\right)'_S = -\left(n^S\right)'_S = n^S\ \mathrm{div}\,(\mathbf{v}_S)\,. \tag{3.23}$$

Multiplying both sides of (3.23) by ρ^{FR} and using (2.7), we obtain

$$\left(\rho^F\right)'_S = \left(\rho^{FR}\ n^F\right)'_S = n^S\ \rho^{FR}\ \mathrm{div}\,(\mathbf{v}_S)\,, \tag{3.24}$$

which can be exploited together with (3.21) as well as (2.119), (2.7) and (2.6) (with $\alpha = S$) to compute $(\rho)'_S$ as

$$(\rho)'_S = \left(\rho^S + \rho^F\right)'_S = -\left(\rho^{SR} - \rho^{FR}\right)\ n^S\ \mathrm{div}\,(\mathbf{v}_S)\,. \tag{3.25}$$

[1] http://www.featflow.de

Next, using (2.3) and (3.21) and adding up (2.186) and (2.187) results in the mixture momentum balance, which has a positive effect on the performance of our multigrid solver as will be shown later:

$$\rho^S (\mathbf{v}_S)'_S + \rho^F (\mathbf{v}_F)'_S + \rho^F \operatorname{grad}\mathbf{v}_F\,(\mathbf{v}_F - \mathbf{v}_S) - \operatorname{div}\mathbf{T}^S_E - \rho\,\mathbf{b} + \operatorname{grad} p = \mathbf{0}\,. \tag{3.26}$$

The main purpose of this step is to remove the variable n^S in front of $\operatorname{grad} p$ in (2.186) in order to get a Stokes-like form and to avoid arising of $\operatorname{grad} n^S$ (which contains second-order derivatives) in the weak form to allow for testing with lower order finite elements for the no-convection assumption. The appearance of $\operatorname{grad} n^S$ cannot be avoided if (2.186) is applied because the intention is to use discontinuous pressure finite elements that do not carry derivatives and hence the indispensable integration by parts (to remove the differential operator grad, acting on p) will generate the undesirable $\operatorname{grad} n^S$. Furthermore, this step, as will be shown latter, leads to the more convenient boundary conditions, already mentioned in (3.18).

For the same main purpose, (2.187) is divided by $n^F > 0$ yielding

$$\rho^{FR}\,(\mathbf{v}_F)'_S + \rho^{FR}\operatorname{grad}\mathbf{v}_F\,(\mathbf{v}_F - \mathbf{v}_S) - \rho^{FR}\,\mathbf{b} + \frac{n^F\,\gamma^{FR}}{k^F}\,(\mathbf{v}_F - \mathbf{v}_S) + \operatorname{grad} p = \mathbf{0}\,, \tag{3.27}$$

which removes the leading coefficient n^F before $\operatorname{grad} p$ and leads to a solution-independent external load vector ($\mathbf{t}^F$), which depends only on the ambient pressure as will be shown in the next section. To continue with this reformulation to get a Stokes-like form, (2.188) should also be modified by use of the Darcy velocity vector $\mathbf{w}$

$$\mathbf{v}_F = \mathbf{w}/n^F + \mathbf{v}_S, \quad \text{where} \quad n^F > 0 \tag{3.28}$$

and then by substitution of (3.28) in (3.26), (3.27) and (2.188) and then making use of (3.20) and (3.22), we get the modified $\mathbf{uw}p$ formulation (IBVP 3) written in the actual configuration as:

- Balance of momentum of the binary saturated mixture:

$$\begin{aligned}\Big(\rho\;\mathbf{v}_S\Big)'_S + \rho^{FR}\;(\mathbf{w})'_S + \rho^{FR}\,(\operatorname{grad}\mathbf{v}_F\,\mathbf{w}) - \operatorname{div}\mathbf{T}^S_E - \rho\,\mathbf{b}&\\ - (\rho)'_S\;\mathbf{v}_S - \frac{(\rho^F)'_S}{n^F}\mathbf{w} + \operatorname{grad} p = \mathbf{0}\,,&\end{aligned} \tag{3.29}$$

- Balance of momentum of the fluid phase:

$$\begin{aligned}\rho^{FR}\;(\mathbf{v}_S)'_S + \rho^{FR}\left(\frac{\mathbf{w}}{n^F}\right)'_S + \frac{\rho^{FR}}{n^F}\;(\operatorname{grad}\mathbf{v}_F\,\mathbf{w})&\\ + \frac{\gamma^{FR}}{k^F}\,\mathbf{w} - \rho^{FR}\,\mathbf{b} + \operatorname{grad} p = \mathbf{0}\,,&\end{aligned} \tag{3.30}$$

- Volume balance of the binary saturated mixture:

$$\operatorname{div}(\mathbf{v}_S) + \operatorname{div}(\mathbf{w}) = 0\,, \tag{3.31}$$

- Velocity-displacement relationship:

$$(\mathbf{u}_S)'_S = \mathbf{v}_S\,. \tag{3.32}$$

The convective terms and the volume fraction changes are colored with orange and green, respectively, because we will refer to them frequently when later studying their influence. Note that the chosen primary unknowns for this set of PDE are $\mathbf{u}_S$, $\mathbf{w}$ and p. Hence, $\mathbf{v}_S(\mathbf{u}_S)$ as well as $\mathbf{T}^S_E(\mathbf{u}_S)$, $n^S(\mathbf{u}_S)$, $n^F(\mathbf{u}_S)$ and $\mathbf{v}_F$ represent the secondary variables of the problem. We shall adopt a hyper-elastic porous material model, which assumes a non-strained (or unstressed) initial configuration with the initial conditions

$$\mathbf{u}_S(\mathbf{X},0) = \mathbf{0}, \quad \mathbf{v}_S(\mathbf{X},0) = \mathbf{0} \quad \text{and} \quad \mathbf{w}(\mathbf{X},0) = \mathbf{0} \quad \forall \mathbf{X} \in \Omega_0\,. \tag{3.33}$$

The boundary $\Gamma = \partial\Omega$ is divided into Dirichlet ($\Gamma_{\mathbf{u}_S}$ and $\Gamma_{\mathbf{w}}$) and Neumann ($\Gamma_{\mathbf{t}}$ and $\Gamma_{\mathbf{t}^F}$) regions so that

$$\begin{aligned} \Gamma = \Gamma_{\mathbf{u}_S} \cup \Gamma_{\mathbf{t}} \quad &\text{with} \quad \Gamma_{\mathbf{u}_S} \cap \Gamma_{\mathbf{t}} = \phi \quad \text{for SMB}\,, \\ \Gamma = \Gamma_{\mathbf{w}} \cup \Gamma_{\mathbf{t}^F} \quad &\text{with} \quad \Gamma_{\mathbf{w}} \cup \Gamma_{\mathbf{t}^F} = \phi \quad \text{for FMB}\,. \end{aligned} \tag{3.34}$$

For the Dirichlet conditions, we deal particularly with frictionless non-moving rigid wall boundary conditions with no external suction/ injection applied, so that

$$\begin{aligned} \mathbf{u}_S \cdot \mathbf{n} = 0 \quad &\text{on } \Gamma_{\mathbf{u}_S}\,, \\ \mathbf{w} \cdot \mathbf{n} = 0 \quad &\text{on } \Gamma_{\mathbf{w}}\,, \end{aligned} \tag{3.35}$$

where $\mathbf{n}$ denotes the normal to domain boundary Γ and for Neumann conditions, we have

$$\bar{\mathbf{t}} = \left(\mathbf{T}^S_E - p\mathbf{I}\right) \cdot \mathbf{n}\,, \qquad \bar{\mathbf{t}}^F = -p\mathbf{n}\,. \tag{3.36}$$

As stated in the previous subsection, the above Neumann conditions are more convenient than 3.8, since the surface traction $\bar{\mathbf{t}}$ acts simultaneously on both the solid and the fluid phase such that the separation of the boundary conditions is not needed anymore and $\mathbf{t}^F$ is the negative of the ambient atmospheric pressure, which is usually considered as reference and set to zero.

The rest of the boundary edges, which are not assigned Dirichlet or Neumann condition are automatically considered 'Do-nothing' (zero Neumann). Finally, for the uniqueness of p, we further require vanishing mean pressure as stated in 3.9.

3.2 Weak formulation and discretization in space and time

3.2.1 Weak formulation and discretization in space and time for IBVP 1

Our subsequent variational form of IBVP 1 is created by multiplying (3.1)-(3.4) with the displacement test function $\delta\mathbf{u}_S$, the fluid velocity test function $\delta\mathbf{v}_F$, the pressure test function δp, the solid velocity test function $\delta\mathbf{v}_S$, integrating over the whole domain Ω and performing partial integration to get:

$$\begin{aligned}&\int_\Omega \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}^S_E \,\mathrm{d}v - \int_\Omega \frac{(n^F_{0S})^2\,\gamma^{FR}}{k^F_0}\,\delta\mathbf{u}_S\cdot\mathbf{v}_F\,\mathrm{d}v - \int_\Omega n^S_{0S}\operatorname{div}\delta\mathbf{u}_S\; p\,\mathrm{d}v + \\ &\int_\Omega \frac{(n^F_{0S})^2\,\gamma^{FR}}{k^F_0}\,\delta\mathbf{u}_S\cdot\mathbf{v}_S\,\mathrm{d}v + \int_\Omega \rho^S_0\delta\mathbf{u}_S\cdot\{(\mathbf{v}_S)'_S-\mathbf{b}\}\,\mathrm{d}v = \int_{\Gamma_{\mathbf{t}S}} \delta\mathbf{u}_S\cdot\bar{\mathbf{t}}^S\,\mathrm{d}a\end{aligned} \tag{3.37}$$

$$\begin{aligned}&\int_\Omega \frac{(n^F_{0S})^2\,\gamma^{FR}}{k^F_0}\,\delta\mathbf{v}_F\cdot\mathbf{v}_F\,\mathrm{d}v - \int_\Omega n^F_{0S}\operatorname{div}\delta\mathbf{v}_F\;p\,\mathrm{d}v \;-\int_\Omega \frac{(n^F_{0S})^2\,\gamma^{FR}}{k^F_0}\,\delta\mathbf{v}_F\cdot\mathbf{v}_S\,\mathrm{d}v \\ &+\int_\Omega \rho^F_0\,\delta\mathbf{v}_F\cdot\{(\mathbf{v}_F)'_S-\mathbf{b}\}\,\mathrm{d}v = \int_{\Gamma_{\mathbf{t}F}} \delta\mathbf{v}_F\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a\end{aligned} \tag{3.38}$$

$$\int_\Omega n^S_{0S}\;\delta p\,\operatorname{div}\mathbf{v}_S\,\mathrm{d}v + \int_\Omega n^F_{0S}\;\delta p\,\operatorname{div}\mathbf{v}_F\,\mathrm{d}v = 0 \tag{3.39}$$

$$\int_\Omega \delta\mathbf{v}_S\cdot\{(\mathbf{u}_S)'_S-\mathbf{v}_S\}\,\mathrm{d}v = 0\,. \tag{3.40}$$

Due to the special boundary conditions and smoothness requirements in the above equations, we require that

$$\{\delta\mathbf{u}_S,\mathbf{u}_S\}\in H^1_{0,S}(\Omega)^n,\quad \{\delta\mathbf{v}_S,\mathbf{v}_S\}\in H_{0,S}(\operatorname{div},\Omega),\quad \{\delta\mathbf{v}_F,\mathbf{v}_F\}\in H_{0,F}(\operatorname{div},\Omega)\quad\text{and}\quad p\in L^2, \tag{3.41}$$

where

$$\begin{aligned} H^1_{0,S}(\Omega)^n &= \left\{\mathbf{u}_S\middle|\ \mathbf{u}_S\in H^1(\Omega)^n \ \text{ and }\ \mathbf{u}_S\cdot\mathbf{n}=0 \ \text{ on } \Gamma_{\mathbf{u}_S}\right\},\\ H_{0,S}(\operatorname{div},\Omega) &= \left\{\mathbf{v}_S\middle|\ \mathbf{v}_S\in H(\operatorname{div},\Omega) \ \text{ and }\ \mathbf{v}_S\cdot\mathbf{n}=0 \ \text{ on } \Gamma_{\mathbf{u}_S}\right\},\\ H_{0,F}(\operatorname{div},\Omega) &= \left\{\mathbf{v}_F\middle|\ \mathbf{v}_F\in H(\operatorname{div},\Omega) \ \text{ and }\ \mathbf{v}_F\cdot\mathbf{n}=0 \ \text{ on } \Gamma_{\mathbf{v}_F}\right\}.\end{aligned} \tag{3.42}$$

Observe that (for instance) $\int_\Omega \delta\mathbf{u}_S\cdot\operatorname{grad}p = -\int_\Omega \operatorname{div}\delta\mathbf{u}_S\;p\,\mathrm{d}v =$ for $\delta\mathbf{u}_S\in H^1_{0,S}$. Hence, if p is a solution, then $p+\text{const.}$ is also a solution. Therefore, to have a unique solution, one needs to remove the set of constant numbers (const.) from L^2. This is possible since the Hilbert space L^2 can be decomposed into two disjoint sets: (the set of constant numbers, const. and the set of L^2-elements orthogonal to const.). Namely,

$$L^2 = \{\text{const.}\} \cup \{\text{const.}^{\perp}\}, \quad \text{where} \quad \{\text{const.}\} \cap \{\text{const.}^{\perp}\} = \{0\}\,. \tag{3.43}$$

const.$^{\perp}$ is usually referred to in the literature as L_0^2 space and computed by setting the L^2-inner product $<$ const.$, p >$ to zero,

$$L_0^2 = \left\{ p \middle|\ p \in L^2, \quad p \perp \text{const.} \quad \Rightarrow \underbrace{\int_\Omega \text{const. } p \text{ d}v = 0}_{\text{inner product} = 0} \quad \Rightarrow \int_\Omega p \text{ d}v = 0 \right\} \tag{3.44}$$

and the uniqueness of p require that p is restricted to L_0^2 instead of L^2. That is,

$$p \in L_0^2\,. \tag{3.45}$$

Since L_0^2 is a closed subspace of L^2, it is a complete space. Keep in mind that due to the fact that the pressure (as Lagrange multiplier regarding the incompressibility constraint) provides typically less regularity than displacement and velocity, the pressure derivatives in the weak formulation have been eliminated by partial integration. For the same reason and as usual for the treatment of the incompressible Navier-Stokes equations, no integration by parts has been carried out in (3.39).

Using such a weak form, which avoids derivatives acting on the pressure functions, one can use standard FEM pairs for velocity/displacement and pressure as typical for incompressible flow problems, which are based on piecewise discontinuous pressure approximations (as shown in Figure 3.2.1), while the boundary conditions are more convenient and chosen independently because the volume effluxes are not needed anymore. As a candidate for LBB-stable Stokes elements, we apply in the following (2D) simulations the well-known (non-parametric) Q2/P1 element, that means biquadratic velocities and displacements and piecewise linear (discontinuous) pressure approximations (cf. [105]), which belongs currently to the 'best' FEM choices for incompressible flow problems with respect to efficiency, accuracy and robustness.

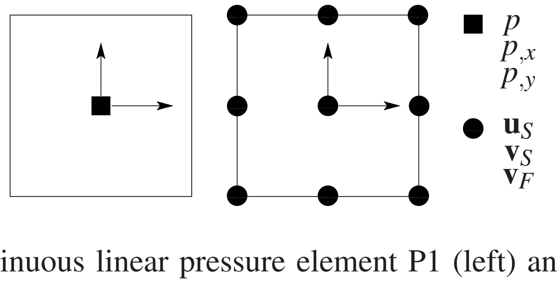
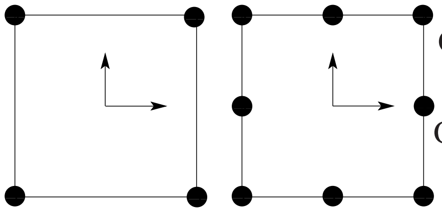


Figure 3.1: The discontinuous linear pressure element P1 (left) and the 9-node Lagrange biquadratic element Q2 (right) that we use for our **uv**p(3)-TR method.

Since we want to show explicit comparisons with a more classical (here: $\mathbf{uv}p$-TB2) approach (see [63]), we additionally introduce the Taylor-Hood-like element in Figure 3.2, with biquadratic (Q) approximations for some degrees of freedom (DOF) omitting the internal node (serendipity element), and continuous bilinear (L) approximations for other degrees of freedom.



QL: $\mathbf{u}_S$, $\mathbf{v}_S$: biquadratic; $\mathbf{v}_F$, p: bilinear

LL: $\mathbf{u}_S$, $\mathbf{v}_S$: bilinear; $\mathbf{v}_F$, p: bilinear

QQ: $\mathbf{u}_S$, $\mathbf{v}_S$: biquadratic; $\mathbf{v}_F$, p: biquadratic

Figure 3.2: The standard 4-node bilinear element L (left) and the 8-node serendipity quadrilateral element Q (right) that is used for the uvp-TB2 method.

Next, based on the discretization with the introduced FEM spaces, equations (3.37)-(3.40) can be written in the following matrix-vector block notation:

$$\left[\begin{array}{c|c} \mathbf{M} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} (\mathbf{u})'_S \\ \hline (\hat{p})'_S \end{array}\right] + \left[\begin{array}{c|c} \mathbf{K} & \mathbf{B}^{\mathrm{T}} \\ \hline \mathbf{B} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{u} \\ \hline \hat{p} \end{array}\right] \left[\begin{array}{c} \mathbf{f} \\ \hline \mathbf{0} \end{array}\right], \quad \text{where} \quad \mathbf{u} = \begin{bmatrix} \hat{\mathbf{u}}_S \\ \hat{\mathbf{v}}_S \\ \hat{\mathbf{v}}_F \end{bmatrix}. \tag{3.46}$$

In more detail with mass and stiffness matrices and right hand side vectors, one obtains

$$\left[\begin{array}{ccc|c} \mathbf{M}_{\mathbf{v}_S\mathbf{u}_S} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\mathbf{u}_S\mathbf{v}_S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_{\mathbf{v}_F\mathbf{v}_F} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} (\hat{\mathbf{u}}_S)'_S \\ (\hat{\mathbf{v}}_S)'_S \\ (\hat{\mathbf{v}}_F)'_S \\ \hline (\hat{p})'_S \end{array}\right] + \left[\begin{array}{ccc|c} \mathbf{0} & \mathbf{K}_{\mathbf{v}_S\mathbf{v}_S} & \mathbf{0} & \mathbf{0} \\ \mathbf{K}_{\mathbf{u}_S\mathbf{u}_S} & \mathbf{K}_{\mathbf{u}_S\mathbf{v}_S} & \mathbf{K}_{\mathbf{u}_S\mathbf{v}_F} & \mathbf{B}_S^{\mathrm{T}} \\ \mathbf{0} & \mathbf{K}_{\mathbf{v}_F\mathbf{v}_S} & \mathbf{K}_{\mathbf{v}_F\mathbf{v}_F} & \mathbf{B}_F^{\mathrm{T}} \\ \hline \mathbf{0} & \mathbf{B}_S & \mathbf{B}_F & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{u}}_S \\ \hat{\mathbf{v}}_S \\ \hat{\mathbf{v}}_F \\ \hline \hat{p} \end{array}\right] = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{f}_{\mathbf{u}_S} \\ \mathbf{f}_{\mathbf{v}_F} \\ \hline \mathbf{0} \end{array}\right]. \tag{3.47}$$

Here, $\mathbf{u}$ and $\hat{p}$ are the nodal unknowns. Moreover, the above matrices and right hand side vectors are given such that:

$$
\begin{aligned}
&\delta\hat{\mathbf{u}}_S^T\,\mathbf{K}_{\mathbf{u}_S\mathbf{u}_S}\,\hat{\mathbf{u}}_S = \int_\Omega \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}_E^S\,\mathrm{d}v\,, &&\delta\hat{\mathbf{u}}_S^T\,\mathbf{K}_{\mathbf{u}_S\mathbf{v}_S}\,\hat{\mathbf{v}}_S = \int_\Omega \frac{(n_{0S}^F)^2\,\gamma^{FR}}{k_{0S}^F}\,\delta\mathbf{u}_S\cdot\mathbf{v}_S\,\mathrm{d}v\\
&\delta\hat{\mathbf{u}}_S^T\,\mathbf{K}_{\mathbf{u}_S\mathbf{v}_F}\,\hat{\mathbf{v}}_F = -\int_\Omega \frac{(n_{0S}^F)^2\,\gamma^{FR}}{k_{0S}^F}\,\delta\mathbf{u}_S\cdot\mathbf{v}_F\,\mathrm{d}v\,, &&\delta\hat{\mathbf{u}}_S^T\,\mathbf{B}_S^{\mathrm{T}}\,\hat{p} = -\int_\Omega n_{0S}^S\operatorname{div}\delta\mathbf{u}_S\,p\,\mathrm{d}v\\
&\delta\hat{\mathbf{v}}_S^T\,\mathbf{K}_{\mathbf{v}_S\mathbf{v}_S}\,\delta\hat{\mathbf{v}}_S = -\int_\Omega \delta\mathbf{v}_S\cdot\mathbf{v}_S\,\mathrm{d}v\,, &&\delta\hat{\mathbf{v}}_F^T\,\mathbf{K}_{\mathbf{v}_F\mathbf{v}_S}\,\delta\hat{\mathbf{v}}_S = -\int_\Omega \frac{(n_{0S}^F)^2\,\gamma^{FR}}{k_{0S}^F}\,\delta\mathbf{v}_F\cdot\mathbf{v}_S\,\mathrm{d}v\\
&\delta\hat{\mathbf{v}}_F^T\,\mathbf{K}_{\mathbf{v}_F\mathbf{v}_F}\,\delta\hat{\mathbf{v}}_F = \int_\Omega \frac{(n_{0S}^F)^2\,\gamma^{FR}}{k_{0S}^F}\,\delta\mathbf{v}_F\cdot\mathbf{v}_F\,\mathrm{d}v\,, &&\delta\hat{\mathbf{v}}_F^T\,\mathbf{B}_F^{\mathrm{T}}\,\hat{p} = -\int_\Omega n_{0S}^F\operatorname{div}\delta\mathbf{v}_F\,p\,\mathrm{d}v\,,\\
&\delta\hat{p}^T\,\mathbf{B}_S\,\delta\hat{\mathbf{v}}_S = \int_\Omega n_{0S}^S\delta p\,\operatorname{div}\mathbf{v}_S\,\mathrm{d}v\,, &&\delta\hat{p}^T\,\mathbf{B}_F\,\delta\hat{\mathbf{v}}_F = \int_\Omega n_{0S}^F\delta p\,\operatorname{div}\mathbf{v}_F\,\mathrm{d}v\\
&\delta\hat{\mathbf{u}}_S^T\,\mathbf{M}_{\mathbf{u}_S\mathbf{v}_S}\,\delta\hat{\mathbf{v}}_S = \int_\Omega \rho_0^S\,\delta\mathbf{u}_S\cdot\mathbf{v}_S\,\mathrm{d}v\,, &&\delta\hat{\mathbf{v}}_S^T\,\mathbf{M}_{\mathbf{v}_S\mathbf{u}_S}\,\delta\hat{\mathbf{u}}_S = \int_\Omega \delta\mathbf{v}_S\cdot\mathbf{u}_S\,\mathrm{d}v\\
&\delta\hat{\mathbf{v}}_F^T\,\mathbf{M}_{\mathbf{v}_F\mathbf{v}_F}\,\delta\hat{\mathbf{v}}_F = \int_\Omega \rho_0^F\,\delta\mathbf{v}_F\cdot(\mathbf{v}_F)'_S\,\mathrm{d}v\\
&\delta\hat{\mathbf{u}}_S^T\,\mathbf{f}_{\mathbf{u}_S} = \int_{\Gamma_{\mathrm{t}S}} \delta\mathbf{u}_S\cdot\bar{\mathbf{t}}^S\,\mathrm{d}a + \int_\Omega \rho_0^S\delta\mathbf{u}_S\cdot\mathbf{b}\,\mathrm{d}v\,, &&\delta\hat{\mathbf{v}}_F^T\,\mathbf{f}_{\mathbf{v}_F} = \int_{\Gamma_{\mathrm{t}F}} \delta\mathbf{v}_F\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a + \int_\Omega \rho_0^F\,\delta\mathbf{v}_F\cdot\mathbf{b}\,\mathrm{d}v
\end{aligned}
\tag{3.48}
$$

In the next step, regarding the time integration, equations 3.46 or 3.47 are treated in a monolithic implicit way leading to a fully coupled system. In our approach, we apply the standard one-step θ-scheme to 3.46, which leads to

$$
\begin{aligned}
\mathbf{M}\frac{\mathbf{u}_{n+1}-\mathbf{u}_n}{\Delta t} + \theta\mathbf{K}\mathbf{u}_{n+1} + \mathbf{B}^{\mathrm{T}}\hat{p} &= -(1-\theta)\,\mathbf{K}\mathbf{u}_n + \theta\mathbf{f}_{n+1} + (1-\theta)\mathbf{f}_n\,.\\
\mathbf{B}\,\mathbf{u}_{n+1} &= \mathbf{0}
\end{aligned}
\tag{3.49}
$$

Notice in the above equation, the continuity equation due to the incompressibility constraint and the pressure p as corresponding Lagrange multiplier are always treated fully implicitly as usual in CFD simulations, which leads to 2nd-order accuracy, too (cf. [105]). Finally, the above equation can be written in a matrix form similar to the saddle-point system of the Stokes problem:

$$
\begin{bmatrix} \tilde{\mathbf{A}}(\theta) & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}
\begin{bmatrix} \mathbf{u} \\ \mathrm{p} \end{bmatrix}
=
\begin{bmatrix} \mathbf{g} \\ \mathbf{0} \end{bmatrix}
\tag{3.50}
$$

with

$$
\tilde{\mathbf{A}}(\theta) = \mathbf{M} + \theta\Delta t\mathbf{K}\ , \quad \mathrm{p} = \Delta t\,\hat{p} \quad \text{and} \quad \mathbf{g} = \tilde{\mathbf{A}}(\theta-1)\,\mathbf{u}_n + (\theta\mathbf{f}_{n+1} + (1-\theta)\mathbf{f}_n)\,\Delta t\,. \tag{3.51}
$$

After solving the above saddle-point system, the pressure is scaled back using the relation $p = \mathrm{p}/\Delta t$. For $\theta = 1/2$, we recover the second-order Crank-Nicholson scheme (in time), which is based on the well-known trapezoidal rule (TR) and for $\theta = 1$, we obtain the the first-order implicit Euler scheme. However, also fully L-stable 2nd-order schemes like Fractional-Step-Theta-schemes can be used in an analogous way.

3.2.2 Weak formulation and discretization in space and time for IBVP 2

The variational form of IBVP 2 is created by multiplying (3.11), (3.12), (3.13) and (3.14) with the displacement test function $\delta\mathbf{u}_S$, the Darcy velocity test function $\delta\mathbf{w}$, the pressure test function δp and the velocity test function $\delta\mathbf{v}_S$, respectively, integrating over the initial domain Ω and performing some partial integrations. Finally, we obtain the following weak form:

$$\begin{aligned}&\int_\Omega \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}^S_E\,\mathrm{d}v - \int_\Omega p\,\operatorname{div}\delta\mathbf{u}_S\;\;\mathrm{d}v + \rho_0\int_\Omega \delta\mathbf{u}_S\cdot(\mathbf{v}_S)'_S\,\mathrm{d}v\\ &+\rho^{FR}\int_\Omega \delta\mathbf{u}_S\cdot(\mathbf{w})'_S\,\mathrm{d}v = \rho_0\int_\Omega \delta\mathbf{u}_S\cdot\mathbf{b}\,\mathrm{d}v + \int_{\Gamma_\mathbf{t}}\delta\mathbf{u}_S\cdot\bar{\mathbf{t}}\,\mathrm{d}a\,,\end{aligned} \tag{3.52}$$

$$\begin{aligned}&\frac{\rho^{FR}g}{k^F_{0S}}\int_\Omega \delta\mathbf{w}\cdot\mathbf{w}\,\mathrm{d}v - \int_\Omega p\,\operatorname{div}\delta\mathbf{w}\;\,\mathrm{d}V + \frac{\rho^{FR}}{n^F_{0S}}\int_\Omega \delta\mathbf{w}\cdot(\mathbf{w})'_S\,\mathrm{d}v\\ &\quad+\rho^{FR}\int_\Omega \delta\mathbf{w}\cdot(\mathbf{v}_S)'_S\,\mathrm{d}v = \rho^{FR}\int_\Omega \delta\mathbf{w}\cdot\mathbf{b}\,\mathrm{d}v + \int_{\Gamma_{\mathbf{t}^F}}\delta\mathbf{w}\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a\,,\end{aligned} \tag{3.53}$$

$$\int_\Omega \delta p\,\operatorname{div}\mathbf{v}_S\,\mathrm{d}v + \int_\Omega \delta p\,\operatorname{div}\mathbf{w}\,\mathrm{d}v = 0\,, \tag{3.54}$$

$$\int_\Omega \delta\mathbf{v}_S\cdot(\mathbf{u}_S)'_S\,\mathrm{d}v - \int_\Omega \delta\mathbf{v}_S\cdot\mathbf{v}_S\,\mathrm{d}v = 0\,. \tag{3.55}$$

Herein,

$$\{\delta\mathbf{u}_S,\mathbf{u}_S\}\in H^1_{0,S}(\Omega)^n,\quad \{\delta\mathbf{v}_S,\mathbf{v}_S\}\in H_{0,S}(\mathrm{div}\,,\Omega),\quad \{\delta\mathbf{w},\mathbf{w}\}\in H_{0,\mathbf{w}}(\mathrm{div}\,,\Omega)\,, \tag{3.56}$$

with

$$H_{0,\mathbf{w}}(\mathrm{div}\,,\Omega) = \big\{\mathbf{w}\;\big|\;\;\mathbf{w}\in H(\mathrm{div}\,,\Omega)\;\text{ and }\;\mathbf{w}\cdot\mathbf{n} = 0\;\;\text{on }\Gamma_\mathbf{w}\big\} \tag{3.57}$$

as defined in (3.42) and p is orthogonal (in L^2) to unity as stated in (3.45). For the spatial discretization, we use the same element pair (see Figure 3.2.1) and we get (after discretization) an equation of the form of (3.46) but with

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{\mathbf{v}_S\mathbf{u}_S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\mathbf{u}_S\mathbf{v}_S} & \mathbf{M}_{\mathbf{u}_S\mathbf{w}} \\ \mathbf{0} & \mathbf{M}_{\mathbf{w}\mathbf{v}_S} & \mathbf{M}_{\mathbf{w}\mathbf{w}} \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} \mathbf{0} & \mathbf{K}_{\mathbf{v}_S\mathbf{v}_S} & \mathbf{0} \\ \mathbf{K}_{\mathbf{u}_S\mathbf{u}_S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{\mathbf{w}\mathbf{w}} \end{bmatrix}, \tag{3.58}$$

$$\mathbf{B}^{\mathrm{T}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_S^{\mathrm{T}} \\ \mathbf{B}_{\mathbf{w}}^{\mathrm{T}} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{B}_S & \mathbf{B}_{\mathbf{w}} \end{bmatrix}, \tag{3.59}$$

$$\mathbf{u} = \begin{bmatrix} \hat{\mathbf{u}}_S \\ \hat{\mathbf{v}}_S \\ \hat{\mathbf{w}} \end{bmatrix}, \qquad \text{and} \quad \mathbf{f} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_m \\ \mathbf{f}_{\mathbf{w}} \end{bmatrix}. \tag{3.60}$$

Herein,

$$\begin{aligned} &\delta\hat{\mathbf{u}}_S^T\ \mathbf{K}_{\mathbf{u}_S\mathbf{u}_S}\,\hat{\mathbf{u}}_S = \int_\Omega \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}_E^S\,\mathrm{d}v\,, && \delta\hat{\mathbf{w}}^T\ \mathbf{K}_{\mathbf{w}\mathbf{w}}\,\hat{\mathbf{w}} = \frac{\gamma^{FR}}{k_{0S}^F}\int_\Omega \delta\mathbf{w}\cdot\mathbf{w}\,\mathrm{d}v\,, \\ &\delta\hat{\mathbf{v}}_S^T\ \mathbf{K}_{\mathbf{v}_S\mathbf{v}_S}\,\hat{\mathbf{v}}_S = -\int_\Omega \delta\mathbf{v}_S\cdot\mathbf{v}_S\,\mathrm{d}v\,, && \delta\hat{\mathbf{u}}_S^T\ \mathbf{M}_{\mathbf{u}_S\mathbf{v}_S}\,\hat{\mathbf{v}}_S = \rho_0\int_\Omega \delta\mathbf{u}_S\cdot\mathbf{v}_S\,\mathrm{d}v\,, \\ &\delta\hat{\mathbf{u}}_S^T\ \mathbf{M}_{\mathbf{u}_S\mathbf{w}}\,\hat{\mathbf{w}} = \rho^{FR}\int_\Omega \delta\mathbf{u}_S\cdot\mathbf{w}\,\mathrm{d}v\,, && \delta\hat{\mathbf{w}}^T\ \mathbf{M}_{\mathbf{w}\mathbf{w}}\,\hat{\mathbf{w}} = \frac{\rho^{FR}}{n_{0S}^F}\int_\Omega \delta\mathbf{w}\cdot\mathbf{w}\,\mathrm{d}v\,, \\ &\delta\hat{\mathbf{w}}^T\ \mathbf{M}_{\mathbf{w}\mathbf{v}_S}\,\hat{\mathbf{v}}_S = \rho^{FR}\int_\Omega \delta\mathbf{w}\cdot\mathbf{v}_S\,\mathrm{d}v\,, \end{aligned} \tag{3.61}$$

$$\begin{aligned} &\delta\hat{\mathbf{u}}_S^T\ \mathbf{B}_S^{\mathrm{T}}\,\hat{p} = -\int_\Omega p\operatorname{div}\delta\mathbf{u}_S\ \ \mathrm{d}v\,, && \delta\hat{p}^T\,\mathbf{B}_S\,\hat{\mathbf{u}}_S = \int_\Omega \delta p\operatorname{div}\mathbf{v}_S\,\mathrm{d}v\,, \\ &\delta\hat{\mathbf{w}}^T\ \mathbf{B}_{\mathbf{w}}^{\mathrm{T}}\,\hat{p} = -\int_\Omega p\operatorname{div}\delta\mathbf{w}\ \ \mathrm{d}v\,, && \delta\hat{p}^T\,\mathbf{B}_{\mathbf{w}}\,\hat{\mathbf{w}} = \int_\Omega \delta p\operatorname{div}\mathbf{w}\,\mathrm{d}v\,, \end{aligned} \tag{3.62}$$

$$
\begin{aligned}
\delta\hat{\mathbf{w}}^T\,\mathbf{f}_{\mathbf{w}} &= \rho^{FR}\int_{\Omega}\delta\mathbf{w}\cdot\mathbf{b}\,\mathrm{d}v+\int_{\Gamma_{\mathbf{t}^F}}\delta\mathbf{w}\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a \quad \text{and}\\
\delta\hat{\mathbf{u}}_S^T\,\mathbf{f}_m &= \int_{\Omega}\rho_0\,\delta\mathbf{u}_S\cdot\mathbf{b}\,\mathrm{d}v+\int_{\Gamma_{\mathbf{t}}}\delta\mathbf{u}_S\cdot\bar{\mathbf{t}}\,\mathrm{d}a\,.
\end{aligned}
\tag{3.63}
$$

This we integrate over time as in 3.49-3.51.

3.2.3 Weak formulation and discretization in space and time for IBVP 3

The following variational form of IBVP 3 is created by multiplying 3.29, 3.30, 3.31 and 3.32 with the displacement test function $\delta\mathbf{u}_S$, the Darcy velocity test function $\delta\mathbf{w}$, the pressure test function δp and the velocity test function $\delta\mathbf{v}_S$, respectively, integrating over the current domain $\Omega(t)$ and performing partial integrations. Finally, we obtain the following weak form:

$$
\begin{aligned}
&\int_{\Omega(t)}\mathrm{grad}\,\delta\mathbf{u}_S:\mathbf{T}_E^S\,\mathrm{d}V-\int_{\Omega(t)}p\,\mathrm{div}\,\delta\mathbf{u}_S\;\mathrm{d}V+\int_{\Omega(t)}\delta\mathbf{u}_S\cdot\Big(\rho\;\mathbf{v}_S\Big)_S'\,\mathrm{d}V\\
&+\rho^{FR}\int_{\Omega(t)}\delta\mathbf{u}_S\cdot(\mathbf{w})_S'\,\mathrm{d}V+\rho^{FR}\int_{\Omega(t)}\delta\mathbf{u}_S\cdot\left(\mathrm{grad}\,(\mathbf{v}_F)-\frac{(n^F)_S'}{n^F}\mathbf{I}\right)\mathbf{w}\,\mathrm{d}V\\
&-\int_{\Omega(t)}\delta\mathbf{u}_S\cdot(\rho)_S'\;\mathbf{v}_S\,\mathrm{d}V=\int_{\Omega(t)}\rho\;\delta\mathbf{u}_S\cdot\mathbf{b}\,\mathrm{d}v+\int_{\Gamma_{\mathbf{t}}}\delta\mathbf{u}_S\cdot\bar{\mathbf{t}}\,\mathrm{d}a\,,
\end{aligned}
\tag{3.64}
$$

$$
\begin{aligned}
&\rho^{FR}\int_{\Omega(t)}\delta\mathbf{w}\cdot\left(\frac{1}{n^F}\mathrm{grad}\,(\mathbf{v}_F)+\frac{g}{k^F}\mathbf{I}\right)\mathbf{w}\,\mathrm{d}V-\int_{\Omega(t)}p\,\mathrm{div}\,\delta\mathbf{w}\;\mathrm{d}V\\
&+\rho^{FR}\int_{\Omega(t)}\delta\mathbf{w}\cdot\left(\frac{\mathbf{w}}{n^F}\right)_S'\mathrm{d}V+\rho^{FR}\int_{\Omega(t)}\delta\mathbf{w}\cdot(\mathbf{v}_S)_S'\,\mathrm{d}V\\
&=\rho^{FR}\int_{\Omega(t)}\delta\mathbf{w}\cdot\mathbf{b}\,\mathrm{d}V+\int_{\Gamma_{\mathbf{t}^F}}\delta\mathbf{w}\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a\,,
\end{aligned}
\tag{3.65}
$$

$$
\int_{\Omega(t)}\delta p\;\mathrm{div}\,\mathbf{v}_S\,\mathrm{d}V+\int_{\Omega(t)}\delta p\;\mathrm{div}\,\mathbf{w}\,\mathrm{d}v=0\,,
\tag{3.66}
$$

$$
\int_{\Omega(t)}\delta\mathbf{v}_S\cdot(\mathbf{u}_S)_S'\,\mathrm{d}v^t-\int_{\Omega(t)}\delta\mathbf{v}_S\cdot\mathbf{v}_S\,\mathrm{d}v^t=0\,.
\tag{3.67}
$$

Herein, the boundary conditions have been already discussed in subsection 3.1.3 and function p must have a vanishing mean value as justified in the paragraph directly below equation 3.42. Furthermore, the spaces of solutions are

$$
\begin{aligned}
&\{\delta\mathbf{u}_S,\mathbf{u}_S\}\in H^1_{0,S}(\Omega)^n,\quad\{\delta\mathbf{v}_F,\mathbf{v}_F\}\in H^1_{0,F}(\Omega)^n\,,\qquad\text{where}\\
&H^1_{0,F}(\Omega)^n=\{\mathbf{v}|\;\;\mathbf{v}\in H^1(\Omega)^n\;\;\text{and}\;\;\mathbf{v}\cdot\mathbf{n}=0\;\;\text{on }\Gamma_{\mathbf{v}_F}\}
\end{aligned}
\tag{3.68}
$$

and since

$$\operatorname{grad}(\mathbf{v}_F) = \operatorname{grad}\left(\frac{\mathbf{w}}{n^F} + \mathbf{v}_S\right) = \frac{1}{n^F}\operatorname{grad}(\mathbf{w}) - \mathbf{w} \otimes \operatorname{grad}\left(\frac{1}{n^F}\right) + \operatorname{grad}(\mathbf{v}_S), \tag{3.69}$$

we further require that

$$\{\delta\mathbf{v}_S, \mathbf{v}_S\} \in H^1_{0,S}(\Omega)^n, \quad \{\delta\mathbf{w}, \mathbf{w}\} \in H^1_{0,F}(\Omega)^n. \tag{3.70}$$

Notice that our four primary variables ($\mathbf{u}_S$, $\mathbf{v}_S$, $\mathbf{w}$ and p) are now in Hilbert spaces, which is necessary for the existence of their solutions. On the other hand, the terms of the gradient of the secondary variable $\mathbf{v}_F$ in 3.64-3.65 are considered as coefficients of the bilinear form and we assume

$$\operatorname{grad}(\mathbf{v}_F) \in L^2. \tag{3.71}$$

Because $n^F \in (0,1)$, its reciprocal (appeared together with $\operatorname{grad}(\mathbf{v}_F)$ in 3.65) is bounded and hence belongs to L^2. Next, integrating 3.64-3.67 in time using θ-scheme (with p being treated implictly as in the previous two BVPs) and then discretizing in space with the already introduced FEM spaces (see Figure 3.2.1) , we end with the following task

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_S \\ \mathbf{v}_S \\ \mathbf{w} \end{bmatrix} \quad \text{and} \quad \mathrm{p} = p\,\Delta t : \tag{3.72}$$

Given $\mathbf{u}^n$ *and the time step* $\Delta t = t_{n+1} - t_n$*, then solve for* $\mathbf{u} = \mathbf{u}^{n+1}$

$$\left[\begin{array}{c|c} \tilde{\mathbf{A}}(\mathbf{u},\theta) & \mathbf{B}^{\mathrm{T}}(\Omega) \\ \hline \mathbf{B}(\Omega) & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{u} \\ \hline \mathrm{p} \end{array}\right] = \left[\begin{array}{c} \mathbf{g}(\mathbf{u}_S) \\ \hline \mathbf{0} \end{array}\right]. \tag{3.73}$$

In more detail with mass and stiffness matrices and right hand side vectors, we have

$$\mathbf{B}^{\mathrm{T}}(\Omega) = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_S^{\mathrm{T}} \\ \mathbf{B}_{\mathbf{w}}^{\mathrm{T}} \end{bmatrix}, \quad \mathbf{B}(\Omega) = \begin{bmatrix} \mathbf{0} & \mathbf{B}_S & \mathbf{B}_{\mathbf{w}} \end{bmatrix}, \tag{3.74}$$

$$\overbrace{\tilde{\mathbf{A}}\left(\mathbf{z} = \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix}, \theta\right)}^{\text{expression is used in Algorithm 1}} = \tilde{\mathbf{A}}(\mathbf{u},\theta) =$$

$$\begin{bmatrix} \mathbf{M}_{\mathbf{v}_S\mathbf{u}_S} & \theta\Delta t\mathbf{K}_{\mathbf{v}_S\mathbf{v}_S} & \mathbf{0} \\ \theta\Delta t\mathbf{K}_{\mathbf{u}_S\mathbf{u}_S}(\mathbf{u}_S) & \mathbf{M}_{\mathbf{u}_S\mathbf{v}_S}(\mathbf{u}_S) + \theta\Delta t\mathbf{K}_{\mathbf{u}_S\mathbf{v}_S}(\mathbf{u}_S,\mathbf{v}_S) & \mathbf{M}_{\mathbf{u}_S\mathbf{w}} + \theta\Delta t\mathbf{K}_{\mathbf{u}_S\mathbf{w}}(\mathbf{u}) \\ \mathbf{0} & \mathbf{M}_{\mathbf{w}\mathbf{v}_S} & \mathbf{M}_{\mathbf{w}\mathbf{w}}(\mathbf{u}_S) + \theta\Delta t\mathbf{K}_{\mathbf{w}\mathbf{w}}(\mathbf{u}) \end{bmatrix}, \tag{3.75}$$

$$\mathbf{g}(\mathbf{z}) = \mathbf{g}(\mathbf{u}_S) = \mathbf{r}_n + \theta \Delta t\, \mathbf{f}(\mathbf{u}_S)\,, \tag{3.76}$$

$$\mathbf{r}_n = \tilde{\mathbf{A}}(\mathbf{u}^n, \theta - 1)\,\mathbf{u}^n + (\theta - 1)\,\Delta t\, \mathbf{f}(\mathbf{u}^n_S)\,, \tag{3.77}$$

$$\mathbf{f}(\mathbf{u}) = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_{\mathbf{u}}(\mathbf{u}_S) \\ \mathbf{f}_{\mathbf{w}} \end{bmatrix}\,, \tag{3.78}$$

and assuming that $(\hat{\cdot})$ indicate nodal values of $(\cdot)$, we further define

$$\underbrace{\delta\hat{\mathbf{u}}_S^T\, \mathbf{K}_{\mathbf{u}_S\mathbf{u}_S}\, \hat{\mathbf{u}}_S + \delta\hat{\mathbf{u}}_S^T\, \mathbf{h}}_{\text{cf. appendix}} = \int_\Omega \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}^S_E\, dv\,, \tag{3.79a}$$

$$\delta\hat{\mathbf{u}}_S^T\, \mathbf{K}_{\mathbf{u}_S\mathbf{v}_S}\, \hat{\mathbf{v}}_S = -\int_\Omega \delta\mathbf{u}_S \cdot \underbrace{(\rho)'_S}_{\text{cf. }\boxed{3.25}}\, \mathbf{v}_S\, dv\,, \tag{3.79b}$$

$$\delta\hat{\mathbf{u}}_S^T\, \mathbf{K}_{\mathbf{u}_S\mathbf{w}}\, \hat{\mathbf{w}} = \rho^{FR} \int_{\Omega(t)} \delta\mathbf{u}_S \cdot \left(\overbrace{\operatorname{grad}(\mathbf{v}_F)}^{\text{cf. }\boxed{3.69}} - \frac{\overbrace{(n^F)'_S}^{\text{cf. }\boxed{3.23}}}{n^F}\, \mathbf{I} \right) \mathbf{w}\, dv\,, \tag{3.79c}$$

$$\delta\mathbf{w}^T\, \mathbf{K}_{\mathbf{ww}}\, \hat{\mathbf{w}} = \rho^{FR} \int_{\Omega(t)} \delta\mathbf{w} \cdot \left(\frac{1}{n^F} \operatorname{grad}(\mathbf{v}_F) + \frac{g}{k^F}\mathbf{I} \right) \mathbf{w}\, dv\,, \tag{3.79d}$$

$$\delta\hat{\mathbf{v}}_S^T\, \mathbf{K}_{\mathbf{v}_S\mathbf{v}_S}\, \hat{\mathbf{v}}_S = -\int_{\Omega(t)} \delta\mathbf{v}_S \cdot \mathbf{v}_S\, dv\,, \tag{3.79e}$$

$$\delta\hat{\mathbf{u}}_S^T\, \mathbf{B}_S^{\mathrm{T}}\, \hat{p} = -\int_\Omega p\, \operatorname{div}\delta\mathbf{u}_S\;\; dv\,, \quad \hat{p}^T\, \mathbf{B}_S\, \delta\hat{\mathbf{v}}_S = \int_\Omega \delta p\, \operatorname{div}\mathbf{v}_S\, dv\,, \tag{3.80a}$$

$$\delta\hat{\mathbf{u}}_S^T\, \mathbf{B}_{\mathbf{w}}^{\mathrm{T}}\, \hat{p} = -\int_\Omega p\, \operatorname{div}\delta\mathbf{w}\;\; dv\,, \quad \hat{p}^T\, \mathbf{B}_{\mathbf{w}}\, \delta\hat{\mathbf{w}}_S = \int_\Omega \delta p\, \operatorname{div}\mathbf{w}\, dv\,, \tag{3.80b}$$

$$\delta\hat{\mathbf{u}}_S^T\, \mathbf{M}_{\mathbf{u}_S\mathbf{v}_S}(\mathbf{u}_S)\; \hat{\mathbf{v}}_S = \int_\Omega \rho\,(\mathbf{u}_S)\;\; \delta\mathbf{u}_S \cdot \mathbf{v}_S\, dv\,, \tag{3.81a}$$

$$\delta\hat{\mathbf{w}}^T\, \mathbf{M}_{\mathbf{ww}}(\mathbf{u}_S)\; \hat{\mathbf{w}} = \int_\Omega \frac{\rho^{FR}}{n^F\,(\mathbf{u}_S)}\; \delta\mathbf{w} \cdot \mathbf{w}\, dv\,, \tag{3.81b}$$

$$\delta\hat{\mathbf{w}}^T\, \mathbf{M}_{\mathbf{wv}_S}\, \hat{\mathbf{v}}_S = \rho^{FR} \int_\Omega \delta\mathbf{w} \cdot \mathbf{v}_S\, dv\,, \tag{3.81c}$$

$$\delta\hat{\mathbf{v}}_S{}^T\, \mathbf{M}_{\mathbf{v}_S\mathbf{u}_S}\, \hat{\mathbf{u}}_S = -\int_\Omega \delta\mathbf{v}_S \cdot \mathbf{u}_S\, dv\,, \tag{3.81d}$$

$$\delta\hat{\mathbf{w}}^T\,\mathbf{f}_{\mathbf{w}} = \rho^{FR}\int_\Omega \delta\mathbf{w}\cdot\mathbf{b}\,\mathrm{d}v + \int_{\Gamma_{\mathbf{t}^F}} \delta\mathbf{w}\cdot\bar{\mathbf{t}}^F\,\mathrm{d}a \quad \text{and} \tag{3.82a}$$

$$\delta\hat{\mathbf{u}}_S^T\,\mathbf{f}_m(\mathbf{u}_S) = \int_\Omega \rho(\mathbf{u}_S)\;\delta\mathbf{u}_S\cdot\mathbf{b}\,\mathrm{d}v + \int_{\Gamma_{\mathbf{t}}} \delta\mathbf{u}_S\cdot\bar{\mathbf{t}}\,\mathrm{d}a - \underbrace{\int_\Omega h(\mathbf{u}_S)\;\mathrm{div}(\delta\mathbf{u}_S)\,\mathrm{d}v}_{\text{cf. } (3.79a)\ \&\ (A.8)}\,. \tag{3.82b}$$

3.3 Solution methods

3.3.1 Fixed-increment solver

To solve (3.73), we shall combine the updated Lagrangian approach (used in non-linear structural mechanics) with the pure Picard iteration method (often used in CFD) and we shall refer to this special non-linear solver as **U**pdated **L**agrangian-**P**icard solver or shortly **ULP** solver. In this algorithm, we only do operator evaluation with no additional Gateaux derivatives. Thus, the full Jacobian matrix is not used. In particular, the expensive full material tangent matrix $\frac{\mathrm{d}\mathbf{T}_E^S}{\mathrm{d}\mathbf{u}_S}$ is not considered for this elastic problem. For more details, see Algorithm 1. With regard to step 11 of this algorithm, namely,

$$\mathbf{z}^i = \mathbf{z}^{i-1} + \omega^{i-1}\;\left(\mathbf{A}(\mathbf{z}^{i-1})\right)^{-1}\overbrace{\left(\mathbf{g}(\mathbf{z}^{i-1}) - \mathbf{A}(\mathbf{z}^{i-1})\,\mathbf{z}^{i-1}\right)}^{=\,\mathbf{d}^{i-1}}, \tag{3.83}$$

the standard CFD procedure described in [104] is followed, in which (3.83) is split into the following three steps:

- Calculate the non-linear residual $\mathbf{d}^{i-1}$:

$$\mathbf{d}^{i-1} = \mathbf{g}(\mathbf{z}^{i-1}) - \mathbf{A}(\mathbf{z}^{i-1})\,\mathbf{z}^{i-1}, \tag{3.84}$$

- Compute $\left(\mathbf{A}(\mathbf{z}^{i-1})\right)^{-1}\;\mathbf{d}^{i-1}$ via iteratively or directly solving for $\mathbf{q}^{i-1}$

$$\mathbf{A}(\mathbf{z}^{i-1})\,\mathbf{q}^{i-1} = \mathbf{d}^{i-1}, \tag{3.85}$$

- Perform the updating step:

$$\mathbf{z}^i = \mathbf{z}^{i-1} + \omega^{i-1}\,\mathbf{q}^{i-1}, \tag{3.86}$$

where the damping parameter ω^{i-1} in our case is set to 1. The iterations will continue unless the maximum number of iterations (*imax*) is reached or the norm of the non-linear residual goes below a given tolerance:

$$\left\|\mathbf{d}^i\right\| \le Tol\,. \tag{3.87}$$

Algorithm 1: ULP: the Updated Lagrangian-Picard Iterative Solver. This algorithm is designed for dynamic non-linear problems.

```
1  Data: $n^S_{0S}$, $\rho^{FR}$, $\rho^{SR}$, $k^F_{0S}$, $\kappa$, $\lambda^S$, $\mu^S$, $\Delta t$, $\theta$
2  start value: $\mathbf{y}_0 = \begin{bmatrix} \mathbf{u}_0 \\ p \end{bmatrix}$   where   $p$ is computed from $\mathbf{u}_0$;
3  for $n \leftarrow 0$ to $nstep$ do
4      update domain shape $\Omega$: set $\mathbf{x} = \mathbf{X}_S + \mathbf{u}^n_S$;
5      compute $\mathbf{B}$, see 3.74;
6      compute $\mathbf{r}_n$, see 3.77;
7      initial start for Picard iteration: $\mathbf{z}^0 = \mathbf{y}_n$;
8      for $i \leftarrow 1$ to $imax$ do
9          compute $\mathbf{g}(\mathbf{z}^{i-1})$, see 3.76;
10         compute $\tilde{\mathbf{A}}(\mathbf{z}^{i-1}, \theta)$ and build $\mathbf{A}(\mathbf{z}^{i-1}) = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{0} \end{pmatrix}$, see 3.75;
11         compute $\mathbf{z}^i = \mathbf{z}^{i-1} + \omega^{i-1} \left(\mathbf{A}(\mathbf{z}^{i-1})\right)^{-1} \overbrace{\left(\mathbf{g}(\mathbf{z}^{i-1}) - \mathbf{A}(\mathbf{z}^{i-1})\, \mathbf{z}^{i-1}\right)}^{\mathbf{d}^{i-1}}$, see
           3.83-3.87;
12         if $\left\|\mathbf{d}\,(\mathbf{z}^i)\right\| \leq Tol$ then go to 14;
13     end
14     set $\mathbf{y}_n = \mathbf{z}^i$;
15     set $t = t + \Delta t$;
16 end
```

We just called the highly efficient multigrid solver of Vanka-type smoother available as Fortran subroutine in the free CFD software, FEATFLOW, in order to solve 3.85 iteratively. For more information about this special solver, interested readers may look at subsection 4.1.3. It remains to mention that in the case of infinitesimal linear elastic deformations, updating the domain shape has almost no influence. Therefore, step 4 in Algorithm 1 is canceled. Hence, we end with the pure **PIC**ard solver described in Algorithm 2, which we used to solve 3.50 for IBV 1 & 2 .

Algorithm 2: PIC: the Picard Iterative Solver. This algorithm is designed for dynamic infinitesimal linear elastic problems.

```
1  Data: $n^S_{0S}$, $\rho^{FR}$, $\rho^{SR}$, $k^F_{0S}$, $\kappa$, $\lambda^S$, $\mu^S$, $\Delta t$, $\theta$
2  start value: $\mathbf{y}_0$;
3  compute $\mathbf{K}$, $\mathbf{M}$ and $\mathbf{B}$ (see 3.58-3.63 for IBV2 or 3.46-3.48 for IBV 1);
4  for $n \leftarrow 0$ to $nstep$ do
5      compute $\tilde{\mathbf{A}}(\theta)$ and $\mathbf{g}$, see 3.51;
6      build $\mathbf{A} = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{0} \end{pmatrix}$;
7      initial start for Picard iteration: $\mathbf{z}^0 = \mathbf{y}_n$;
8      for $i \leftarrow 1$ to 2 do
9          compute $\mathbf{z}^i = \mathbf{z}^{i-1} + \omega^{i-1}\ (\mathbf{A})^{-1} \overbrace{\left(\mathbf{g} - \mathbf{A}\ \mathbf{z}^{i-1}\right)}^{\mathbf{d}^{i-1}}$;
10         if $\left\|\mathbf{d}^i\right\| \leq Tol$ then go to 12;
11     end
12     set $\mathbf{y}_n = \mathbf{z}^i$;
13     set $t = t + \Delta t$;
14 end
```

3.3.2 Increment-varying solver

The choice of using a fixed time step size (i. e., fixed increment) in Algorithm 2 may sometimes lead (at certain moment of time) to sudden increase in non-linear iterations followed by stagnation of the solver in the next steps and then divergence in the later steps. To avoid such unfortunate situation, we modify Algorithm 1 as shown in Algorithm 3. The algorithm uses the non-linear convergence rate ξ for iteration i,

$$\xi^i = \sqrt[i]{\frac{\|\mathbf{d}^i\|}{\|\mathbf{d}^0\|}},$$

Algorithm 3: ATS-ULP: Adaptive Time Stepping based Updated Lagrangian-Picard Iterative Solver. This algorithm is designed for dynamic problems.

```
1  Data: $n_{0S}^S$, $\rho^{FR}$, $\rho^{SR}$, $k_{0S}^F$, $\kappa$, $\lambda^S$, $\mu^S$, $\Delta t$, $\theta$, $\xi_{max} \in [0.05\ 0.3]$, $rat\% \in \{50\%, 5\%, 0.5\%\}$
2  start value: $\mathbf{y}_0 = \begin{bmatrix} \mathbf{u}_0 \\ p \end{bmatrix}$   where   $p$ is computed from $\mathbf{u}_0$;
3  for $n \leftarrow 0$ to $nstep$ do
4      update domain shape $\Omega(t)$: set $\mathbf{x} = \mathbf{X}_S + \mathbf{u}_S^n$;
5      compute $\mathbf{B}$, see 3.74;
6      compute $\mathbf{r}_n$, see 3.77;
7      initial start for Picard iteration: $\mathbf{z}^0 = \mathbf{y}_n$;
8      for $i \leftarrow 1$ to $imax$ do
9          compute $\mathbf{g}(\mathbf{z}^{i-1})$, see 3.76;
10         compute $\tilde{\mathbf{A}}(\mathbf{z}^{i-1}, \theta)$ and build $\mathbf{A}(\mathbf{z}^{i-1}) = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{0} \end{pmatrix}$, see 3.75;
11         compute $\mathbf{z}^i = \mathbf{z}^{i-1} + \omega^{i-1}\ \left(\mathbf{A}(\mathbf{z}^{i-1})\right)^{-1} \left(\mathbf{g}(\mathbf{z}^{i-1}) - \mathbf{A}(\mathbf{z}^{i-1})\ \mathbf{z}^{i-1}\right)$, see
           3.83-3.87;
12         if $\left\|\mathbf{d}\ (\mathbf{z}^i)\right\| \leq Tol$ then go to 19;
13         if $\xi^i > \xi_{max}$ AND $i > 1$ then
               /* Cancel this time step and decrease the time step
                  size by rat%                                         */
14             $t = t - \Delta t$;
15             $\Delta t = (1 - rat\%)\ \Delta t$;
16             go to 6;
17         end
18     end
19     set $\mathbf{y}_n = \mathbf{z}^i$;
20     if $\xi^i < \xi_{max}$ then
           /* depending on how far $\xi$ is from $\xi_{max}$, increase the
              time step size by less than rat%                         */
21         $\Delta t = \left(1 + (rat\%)\ \frac{\xi_{max} - \xi}{\xi_{max}}\right)\ \Delta t$;
22     end
23     set $t = t + \Delta t$;
24 end
```

as indicator (or early warning sign) to avoid the unfortunate situation by adjusting prematurely the time step size. Here, we set an upper bound (ξ_{max}) for ξ^i and if ξ^i happens to exceed ξ_{max}, the time step gets aborted (unless the solver accidentally converged to the desired tolerance to avoid time wasting) and then reduced by $rat\%$ as described in the algorithm.

4
Numerical Results

Algorithm 1 and Algorithm 2 have been implemented into our in-house open-source software FEATFLOW[1] by expanding the cc2d code designed to solve the incompressible Navier-Stokes equation. The implementation (which is saved in the enclosed CD-ROM and also available in our server) is found in several folders starting with one of the following characters:

- IBVP1: contains the source codes that use algorithm 1 for solving IBVP 1 discussed in subsection 3.1.1. Namely, it solves the **uv**p formulation, (3.1)-(3.4)) based on the weak forms (3.37)-(3.40) using the fully implicit Crank-Nicolson[2] time integration scheme (TR) as shown in (3.49) and the mixed finite element pairs Q2/P1 shown in Figure 3.2. This solver is therefore referred to as **uv**p(3)-TR-Q2/P1 solver, where the number 3 is used to distinguish our solution algorithm from those in Table I in [63] which, we shall compare our results with. The corresponding folders are IBVP1_Results_I to solve the problem of section 4.1.1 , IBVP1_Results_II to solve the problem of section 4.1.2 and IBVP1_Results_III to solve the problem of section 4.1.3.

- IBVP2: contains the source codes that use algorithm 1 for solving IBVP 2 discussed in subsection 3.1.2. That is to say, it solves the **uw**p formulation, (3.11)-(3.14)) based on the weak forms (3.52)-(3.55) using the fully implicit Crank-Nicolson time integration scheme (TR) as shown in (3.49) and the mixed finite element pairs Q2/P1 shown in Figure 3.2. This solver is hence denoted by **uw**p(2)-TR-Q2/P1 solver where, the number 2 is used to distinguish our solution algorithm from those in Table I in [63], which we shall compare our results with. The corresponding folders are IBVP2_Results_I to solve the problem of section 4.1.1 , IBVP2_Results_II to solve the problem of section 4.1.2 and IBVP2_Results_III to solve the problem of section 4.1.3.

- IBVP3: contains the source codes that use Algorithm 2 for solving IBVP 3 discussed in subsection 3.1.3. Strictly speaking, it solves the **uw**p formulation, (3.29)-(3.32) based

[1] http://www.featflow.de

[2] Following the standard denotation in previous publications, we shall use the trapezoidal rule abreviation, TR, to refer to Crank-Nicolson (where $\theta = 1/2$).

on the weak forms 3.64-3.67 using the fully implicit Crank-Nicolson time integration scheme (TR) as shown in 3.73 and the mixed finite element pairs Q2/P1 shown in Figure 3.2. Thus, this solver will be called **uw**p(3)-TR-Q2/P1 solver, where the number 3 is used to distinguish our non-linear solution algorithm from the above linear one and from those in Table I in [63]. The corresponding folders are IBVP3_Results_I to solve the problem of section 4.2.1 , IBVP3_Results_II to solve the problem of section 4.2.2 and IBVP3_Results_III to solve the problem of section 4.2.3.

The codes are very rich with explanation remarks that help to understand the code.

4.1 Numerical results for linear problems

To validate and to evaluate our **uv**p(3)-TR-Q2/P1 and **uw**p(3)-TR-Q2/P1 solver, we shall compare them with well-established methods. To this aim, two numerical examples taken from [63] are introduced and solved by our solvers. In addition, a large-scale problem with millions of DOFs is adopted from [50] in order to study the performance of special general multigrid solver of Vanka-like smoother available in FEATFLOW.

4.1.1 Results I: Saturated poroelastic column under harmonic load

The usage of the Taylor-Hood-like QL elements is, in general, the best choice among the other choices discussed in [63]. However, this element pair is not LBB stable [3] and suffers from deficiencies at permeable loaded boundaries. In this subsection, we will show that our choice of using Q2P1, overcomes this problem and gives higher accuracy.

To this aim, we shall analyze the response of a homogeneous and isotropic, water-saturated, poroelastic column (height: 10 m, width: 2 m) under plane-strain, confined compression conditions, in which the mixture domain is surrounded by impermeable, frictionless, rigid and non-moving walls except for the loaded top side, which is perfectly drained ($\bar{\mathbf{t}}^F = \mathbf{0}$). The geometry with boundary conditions and meshes are illustrated in Figure 4.1 and Table 4.1.

The constitutive material parameters are listed in Table 4.2. The exact solutions of this problem can be calculated and are the analytical solutions of the following one-dimensional PDEs:

- Balance of momentum of the solid phase:

$$\rho^S u_{S,tt} = \underbrace{(\lambda^S + 2\mu^S) u_{S,yy}}_{\operatorname{div} \mathbf{T}_E^S} - n^S p_{,y} + \frac{(n^F)^2 \gamma^{FR}}{k^F} (u_{F,t} - u_{S,t}) \, ,$$

[3] For more information, see appendix B.0.2. In particular, equation B.36 and the texts below it.

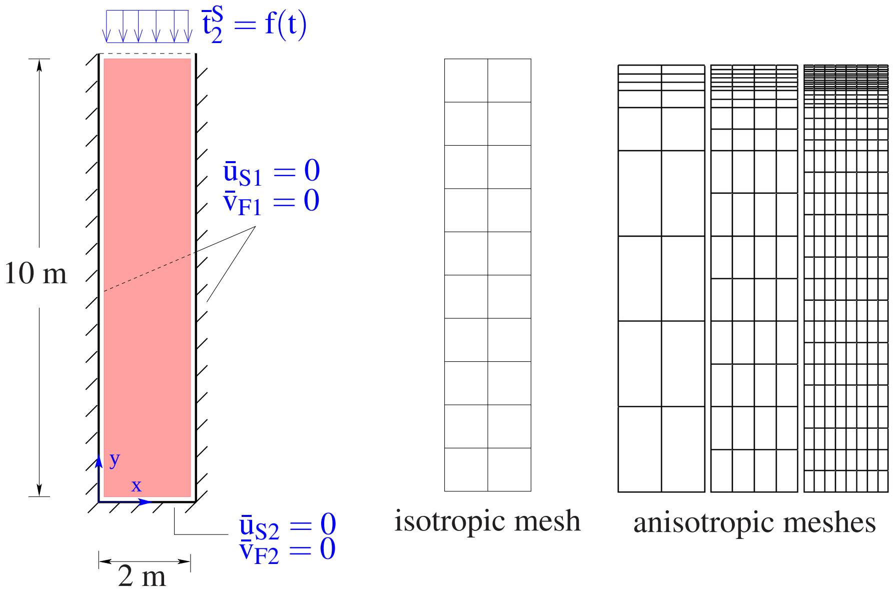


Figure 4.1: Geometry, boundary conditions (left) and isotropic FE mesh: 1 element per meter (center) and anisotropic FE mesh level 1, 2 & 3 (right)) of the dynamic confined compression of a saturated poroelastic column. Non prescribed DOFs are automatically taken as natural boundary conditions. $f(t) = 10^3[1-\cos(20\pi t)]$ $[\mathrm{N/m}^2]$. For higher mesh levels, see also Table 4.1.

Table 4.1: Total number of elements and unknowns (five primary unknowns u_{S1}, u_{S2}, v_{F1}, v_{Fy} and p plus two secondary unknowns v_{S1} and v_{S2}) for the $\mathbf{uv}p$(3)-TR-Q2/P1 approach. This table is related to Figure 4.1.

Cartesian			Rectilinear		
#Elem./m	#Elem.	#Unknowns	Level	#Elem.	#Unknowns
1	20	690	1	18	625
5	500	14226	2	72	2214
10	2000	55446	3	288	8310
15	4500	123666	4	1152	32166
20	8000	218886	5	4608	126534
25	12500	341106	-	-	-
30	18000	490326	-	-	-
40	32000	869766	-	-	-
50	50000	1357206	-	-	-

Table 4.2: Physical properties of the porous medium used for all simulations.

Parameter	Symbol	Value	SI Unit
first Lamé constant of solid	μ^S	5.583×10^6	$\mathrm{N/m^2}$
second Lamé constant of solid	λ^S	8.375×10^6	$\mathrm{N/m^2}$
Effective density of dense solid	ρ^{SR}	2000	$\mathrm{kg/m^3}$
Effective density of pore fluid	ρ^{FR}	1000	$\mathrm{kg/m^3}$
Initial volume fraction of solid	$n^S = n^S_{0S}$	0.67	–
Darcy permeability	k^F	10^{-2}, 10^{-5}	$\mathrm{m/s}$

- Balance of momentum of the fluid phase:

$$\rho^F u_{F,tt} = -n^F p_{,y} - \frac{(n^F)^2 \gamma^{FR}}{k^F} (u_{F,t} - u_{S,t}),$$

- Volume balance of the overall aggregate:

$$n^S u_{S,ty} + n^F u_{F,ty} = 0.$$

The complete details for finding the analytical solutions are provided in [28]. For the purpose of comparison, we therefore follow [63] and test two cases (1) a high permeability case with $k^F = 10^{-2}$ m/s and (2) a moderately low permeability case with $k^F = 10^{-5}$ m/s. The results of **uv**p(3)-TR-Q2/P1 and **uw**p(3)-TR-Q2/P1 solvers are almost identical and show a perfect matching with the analytical (reference) solutions as illustrated in Figures 4.2, 4.3, 4.4 and 4.5, where the convergence in time and space are clear from the figures. Observe that we used anisotropic meshes that get finer when approaching the top (perfect drainage) boundary, since at the top we have $\bar{\mathbf{t}}^F = \mathbf{0}$, which must be compensated by high pressure gradients in a small region below the top boundary. We notice that the displacement obtains full convergence at a mesh level and time step size, where the pressure is still not fully converged (see Figures 4.2 and 4.3). This indicates that a small error in the pressure does not significantly influence the full convergence in the displacement. Moreover, from Figures 4.4 and 4.5 one can notice that for smaller k^F, i. e. for a stronger coupling, more elements are required to reach full convergence to capture the large gradient in the pressure. However, in both cases the time step which yields convergence are the same.

Next, we shall compare with the results of the well-established classical methods reported in [63], which are based on the isotropic mesh in Figure 4.1. From Figure 4.6, we notice that the proposed **uw**p(2)/ **uv**p(3)-TR-Q2/P1 solvers do not show high deficiency in the solutions in the top loaded fully drained boundary and the small layer below it as does **uv**p(2)-TB2-QL and in fact they even provide the most accurate solutions at all selected heights except at the top boundary. This is attributed to the error in the interpolation of the boundary pressure (because the P1 pressure element does not contain boundary nodes) which can be overcome by using the anisotropic mesh as stated before.

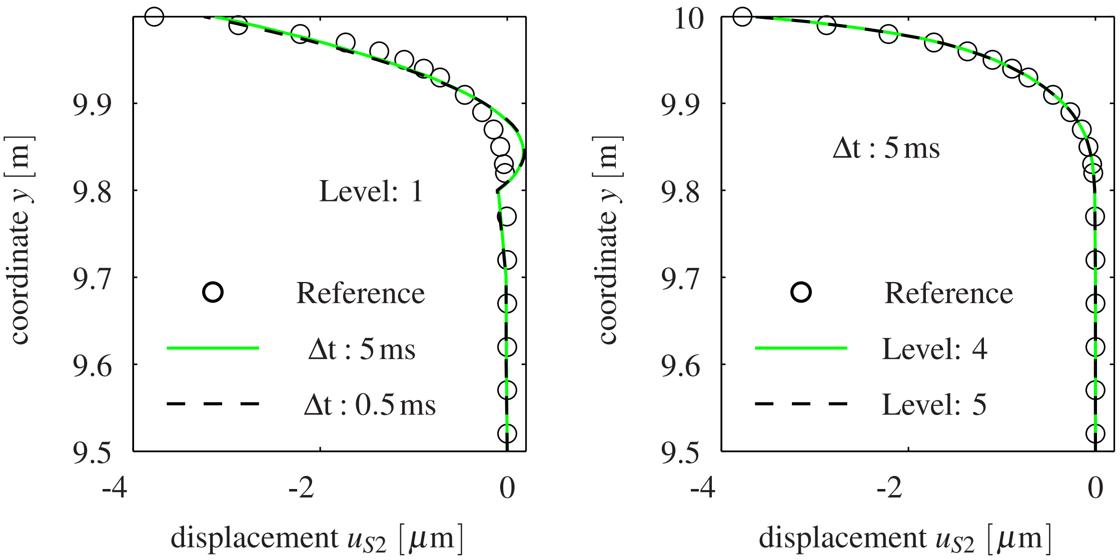


Figure 4.2: Solid displacement distribution in the first half meter under the loaded top of the column at time = 0.15 s using **uw**p(2)/ **uv**p(3)-TR-Q2/P1 for $k^F = 10^{-5}$ m/s and the rectilinear mesh (cf. Figure 4.21 right). The reference solution is taken from [63].

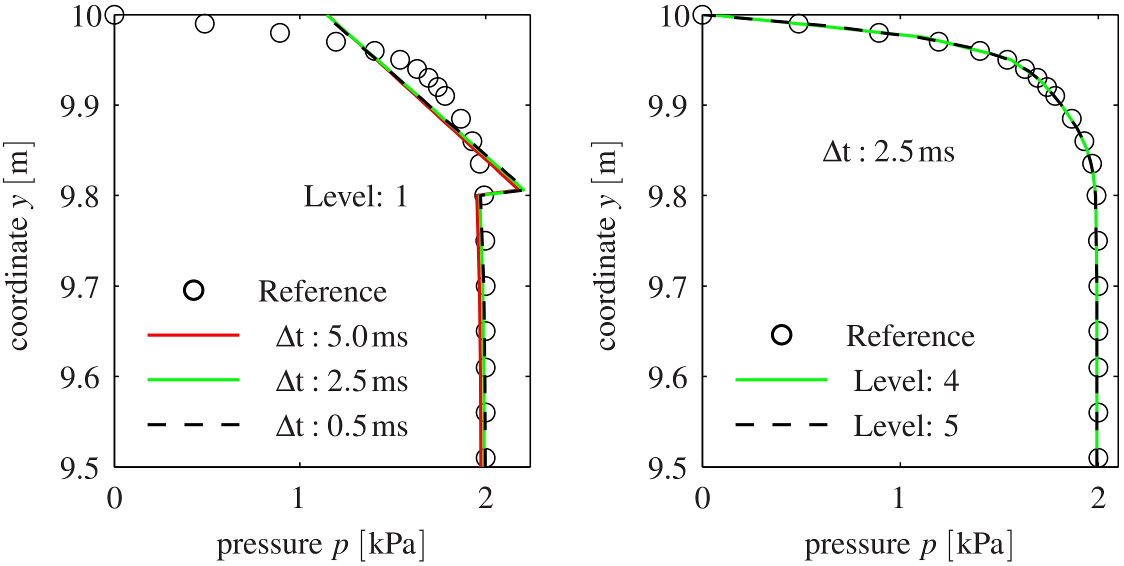


Figure 4.3: Pressure distribution in the first half meter under the loaded top of the column at time = 0.15 s using **uw**p(2)/ **uv**p(3)-TR-Q2/P1 for $k^F = 10^{-5}$m/s and the rectilinear mesh (cf. Figure 4.21 right). The reference solution is taken from [63].

Having an LBB stable and highly accurate element with low deficiency in the solutions at loaded fully drained boundaries is indeed one of the motivations for this work because the equal order approximations such as QQ and LL are not LBB stable and hence, not always trustworthy in addition to being less accurate than Q2/P1.

4.1.2 Results II: Two-dimensional wave propagation

Proceeding from the momentum balance of the whole mixture (as in **uw**p formulation) is more convenient, since the surface traction would then concurrently act on both the solid and fluid phase such that no separation of the boundary conditions is required. However, this choice was known to be problematic: pressure instability, inaccurate displacement solution or even divergence of the solution despite of using IE time integrator are one of the main issues. In addition, adopting the TR time integrator was used to lead to divergence or non-physical pressure oscillations even with **uv**p formulation and even in combination with BDF2 time integrator in strong coupling. These problems are discussed in detail in [63].

Through the numerical results of this subsection, we will show that the above issues are now resolved by adopting some FEATFLOW special CFD technique presented in the current work. In this second example, we study the 2D dynamical wave propagation in a rectangular symmetric domain under plane-strain conditions (Figure 4.17) as presented in [14]. The material parameters are the same as before (Table 4.2) and the 'earthquake event' is represented by the applied distributed impulse force

$$f(t) = 10^5 \sin(25\pi t)\,[1 - H(t-\tau)] \ \ [\mathrm{N/m^2}] \tag{4.1}$$

with $H(t-\tau)$ being the Heaviside step function and $\tau = 0.04$ s. The water saturated mixture domain is surrounded by impermeable, frictionless ($\bar{t}^F_x = 0$ for the bottom and $\bar{t}^F_y = 0$ for the left and right sides) but rigid boundaries except for the loaded top side, which is perfectly drained ($\bar{\mathbf{t}}^F = \mathbf{0}$). Since no analytical solution is available, we first compare quantitatively the accuracy of our proposed monolithic **uw**p(2)/ **uv**p(3)-TR-Q2/P1 approach with the recommended choice (**uv**p(2)-TB2-QL) in [63], which is known as well-accepted combination for solving such coupled problems. Here, we study the displacement solution at point A and the pressure history at point B in the high permeability case $k^F = 10^{-2}$ m/s and the extremely low permeability case with $k^F = 10^{-10}$ m/s.

The direct comparison of the appropriate parameters (mesh level and time step) illustrates the perfect matching as depicted in Figure 4.9. Note that **uv**p(2)-TB2-QL obtains the full convergence at level 3 as indicated in Figure 11 of [63], while our **uv**p(3)-TR-Q2/P1 converges already at mesh level 2 as shown in Figure 4.8, both leading to similar problem sizes.

Next, we switch to the extremely low permeability of $k^F = 10^{-10}$ m/s, which further demonstrates the merits of the considered Q2/P1 approach. We first conclude the optimal time step ($\Delta t \approx 10^{-3}$ s) and the optimal mesh size (level 2) from Figure 4.10. For this case both **uv**p(2)-TB2-QQ and **uv**p(2)-TB2-LL do not converge and the monolithic solution requires LBB-stable mixed FE formulations such as QL [63] and Q2/P1 element pairs.

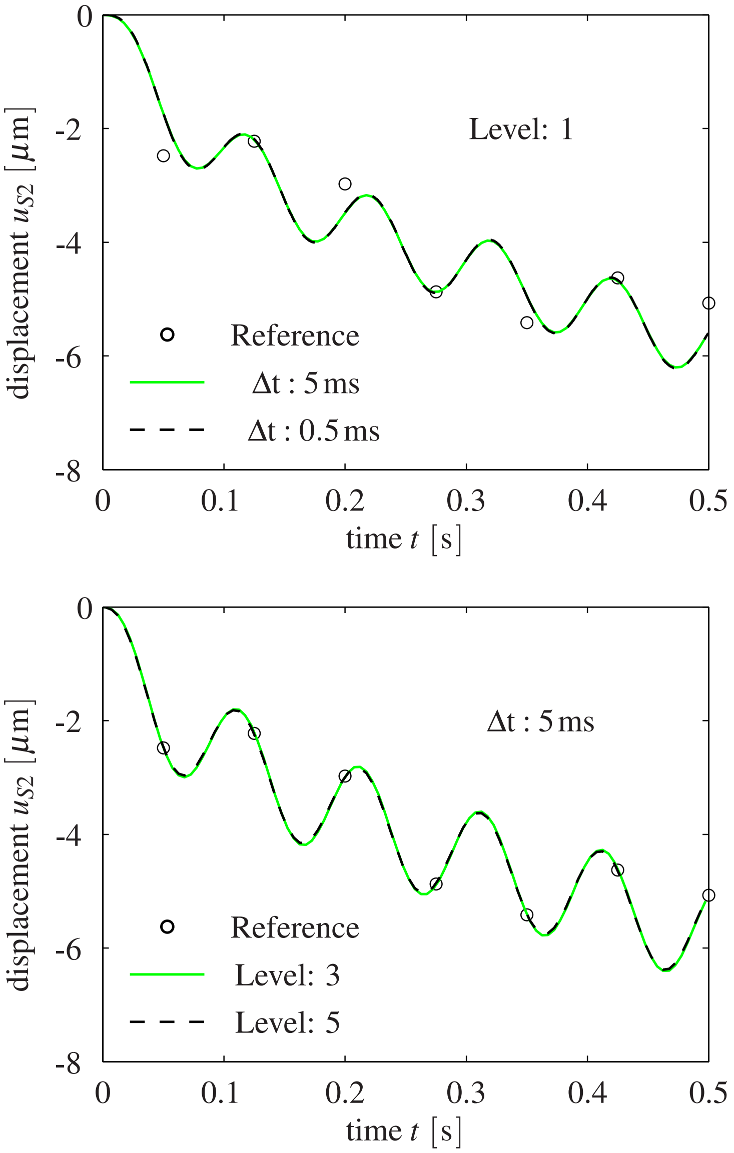


Figure 4.4: y-displacement at point (1,10) vs. time using **uw**p(2)/**uv**p(3)-TR-Q2/P1 for $k^F = 10^{-5}$m/s and rectilinear mesh (cf. Figure 4.1 right). The reference solution is taken from [63].

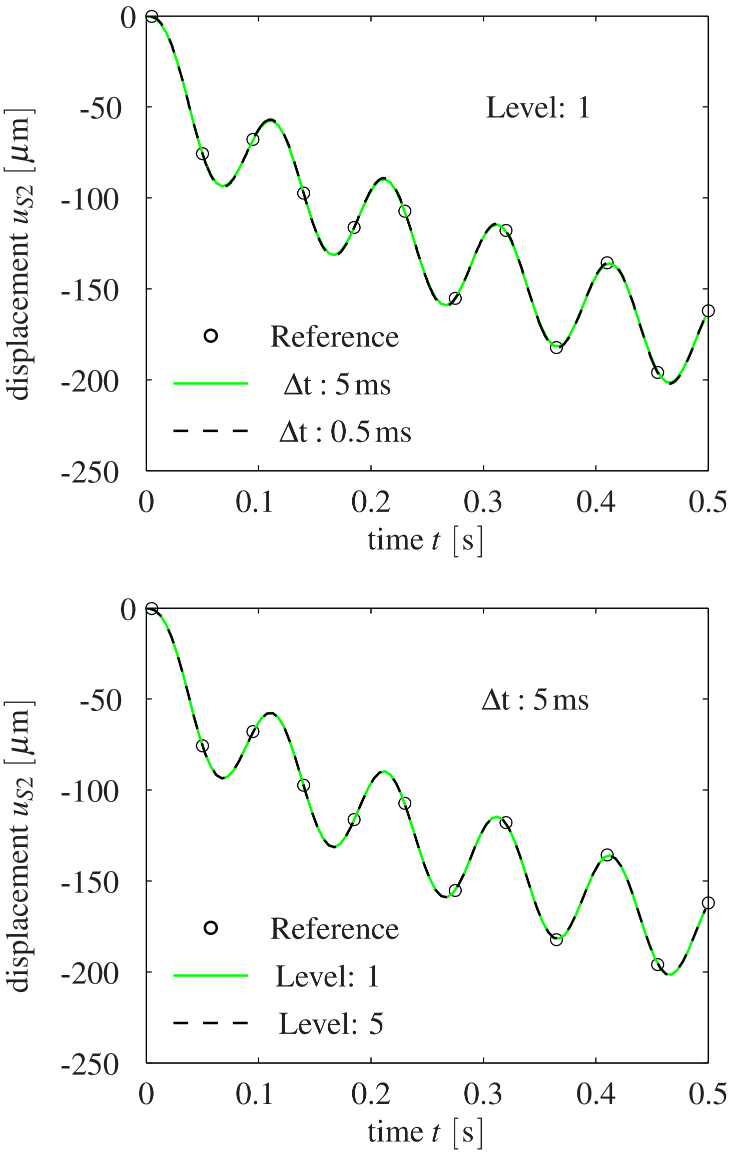
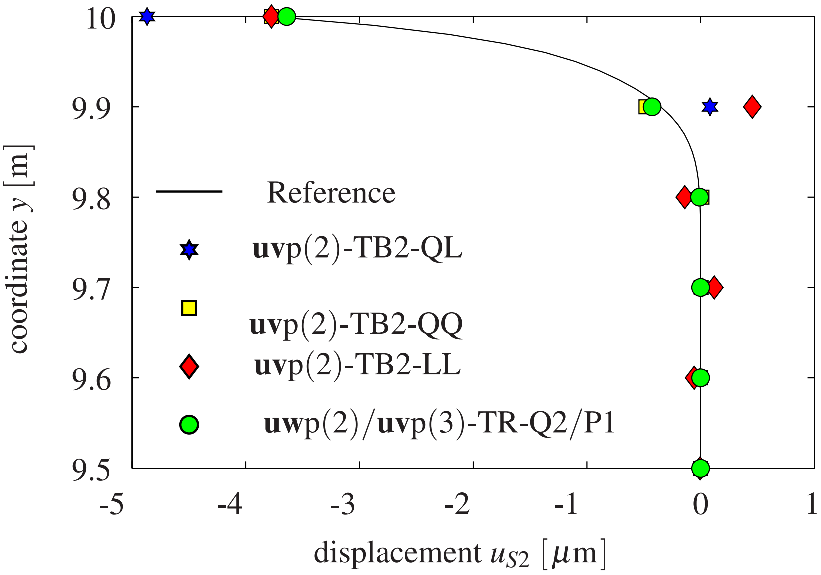
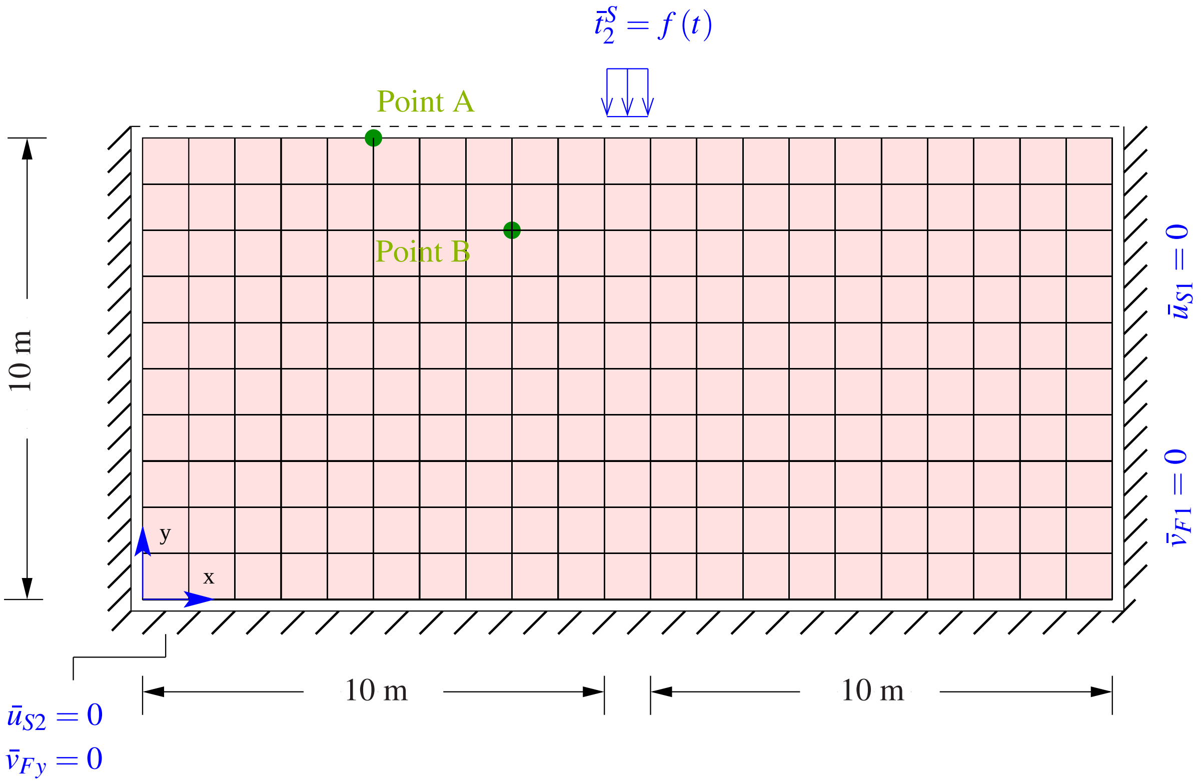


Figure 4.5: *y*-displacement at point (1,10) vs. time using **uw**p(2)/ **uv**p(3)-TR-Q2/P1 for $k^F = 10^{-2}\mathrm{m/s}$ and rectilinear mesh (cf. Figure 4.1 right). The reference solution is taken from [63].



Height	LL	QL	QQ	Q2/P1
9.7	0.12	0.019	0.0055	0.0003
9.8	0.13	0.0062	0.0193	0.0010
9.9	0.8	0.43	0.13	0.077
10	0.00	1.1	0.00	0.13

Figure 4.6: Solid displacement (top) and absolute errors in μm (bottom) for the first half meter below the top surface for the isotropic Cartesian mesh (10 elem/m) (cf. Figure 4.1, center) for $k^F = 10^{-5}\mathrm{m/s}$ at $t = 0.15$ s. All the data except Q2/P1 are taken from [63].



Mesh Level	#Elements (width-height)	#DOFs (Q2/P1)	#DOFs (QL)
1	21-10	6048	3498
2	42-20	23430	13289
3	84-40	92214	51771
4	168-80	365862	102611

Figure 4.7: Geometry, boundary conditions and mesh level 1 of the symmetric 2D wave propagation problem (top). Total number of elements and unknowns for the **uv**p(3)-TR-Q2/P1 approach (bottom). The symmetry of the problem can be exploited to reduce the problem size. However, the computation was performed on the full problem only for our Q2/P1 approach.

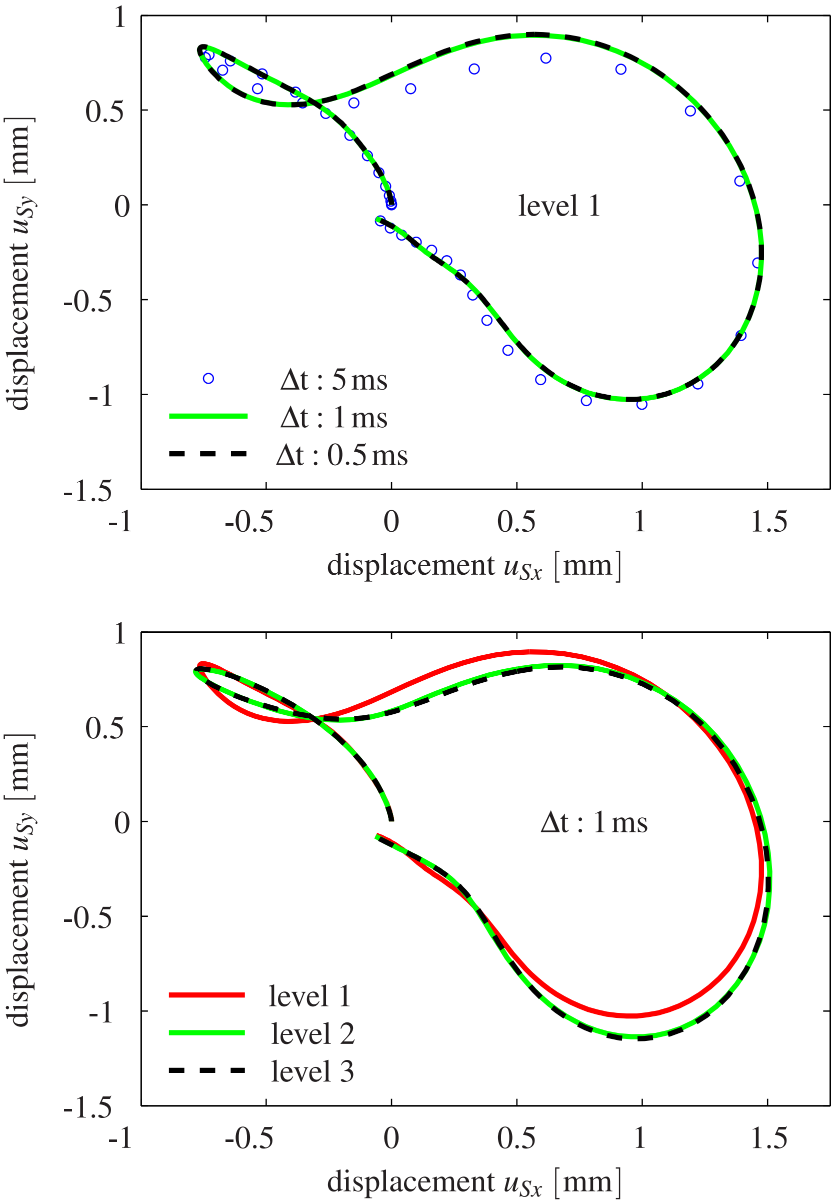


Figure 4.8: Ground rolling during the ‘earthquake event’ at point A using **uw**$p(2)$/ **uv**$p(3)$-TR-Q2/P1 with $k^F = 10^{-2}$m/s and $t \in [0\,0.2]$ s.

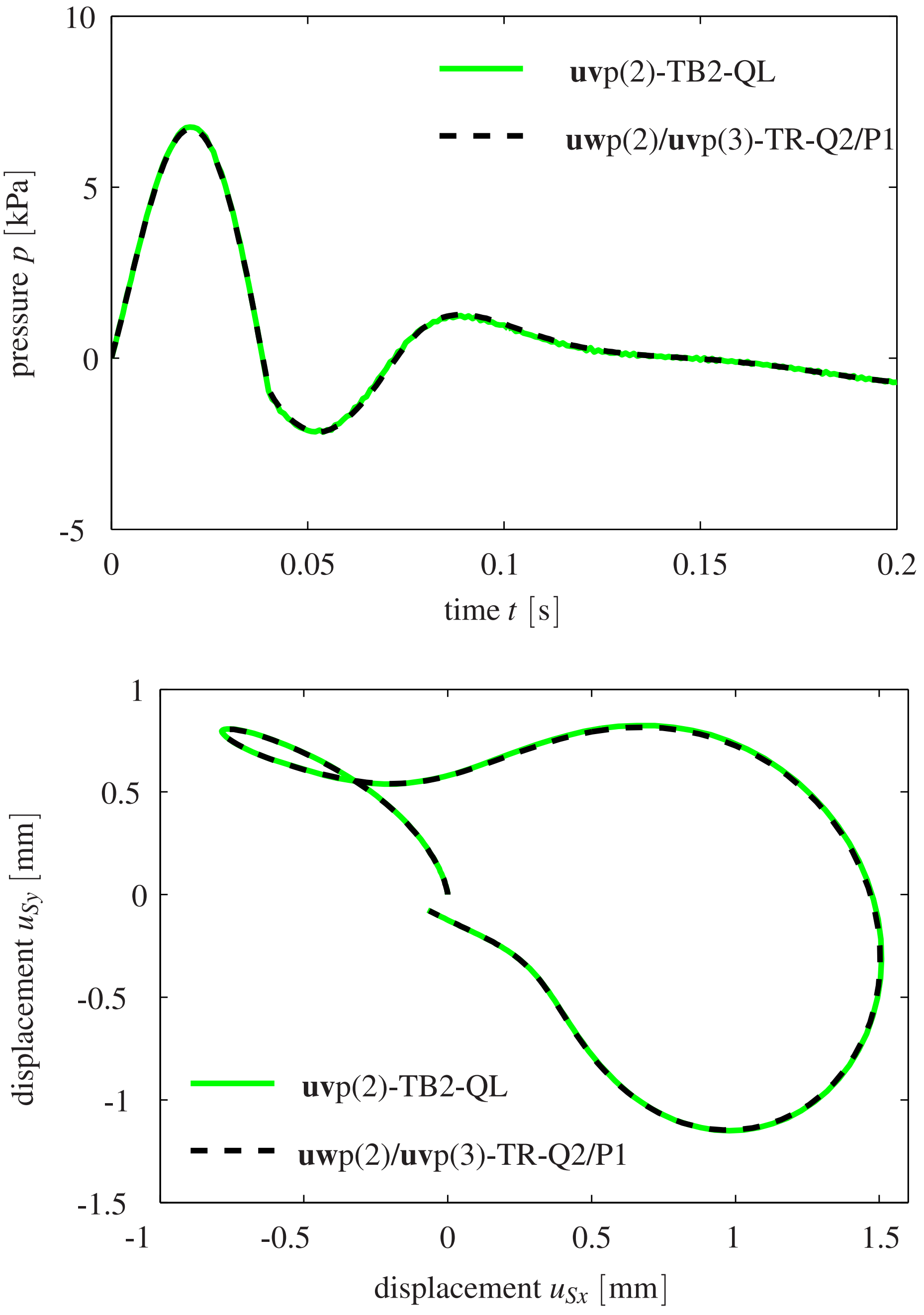


Figure 4.9: Pressure history at point B and displacement history at point A using **uw**$p(2)$/ **uv**$p(3)$-TR-Q2/P1 for $k^F = 10^{-2}\,\mathrm{m/s}$, $\Delta t = 10^{-3}\,\mathrm{s}$ and mesh level 2. The QL results are taken from [63]

.

Based on the results shown in Figures 4.6, the direct comparison between the QL solutions (on higher mesh levels) and the fully converged Q2/P1 solutions (on mesh level 2) (see Figure 4.11) reveals the less accurate displacement solution of the QL approach. In contrast to the TR-QL approach, our TR-Q2/P1 approach does not produce large pressure oscillations as seen in Figure 4.12. Such large oscillations are extremely reduced even for the trapezoidal rule (TR) by using a LBB stable element with equal-order approximations of $\mathbf{u}_S$, $\mathbf{v}_S$ and $\mathbf{v}_F$ such as the Q2/P1 element.

4.1.3 Results III: Large-scale problem

Because most of CPU time goes for solving the linear systems (3.85) and because typically, by accuracy reasons which requires small mesh widths, the arising block systems are too large to be handled by direct solvers, iterative schemes have to be preferred.

However, due to the nature of the involved partial differential equations, particularly w.r.t. the incompressibility, the condition numbers of the arising matrices typically scale with the problem size and are quite large, such that standard single-grid schemes, for instance Krylov-space methods like BICGSTAB or GMRES (cf. [82, 107]), are too slow. Therefore, an excellent alternative is to solve (3.85) via geometrical multigrid (MG) solvers (see [22] and [89, 104, 112]), which require a hierarchy of refined mesh levels and corresponding inter-grid transfer operators, which are selected w.r.t. the chosen FEM spaces. What is special for the described saddle-point problem in (3.85) (see line 10 of algorithm 1) is the choice of the so-called 'smoothing operator', which in our case can be traced back to the early work by Vanka [108]. The corresponding (basic) iterative schemes can be interpreted as block Gauß-Seidel methods applied to mixed formulations of saddle-point problems. For quantitative study on the efficiency of this special multigrid with comparison with single grid method related to this application, interested reader may consult the previous work in [106]. This special multigrid solver with all its components are available in the open-source code, FEATFLOW, as fortran subroutines, which one only needs to learn how to use them in order to perform this study. Fortunately, the FEATFLOW2 code allows for the flexible design of the multigrid structure and is rich with explanation remarks and applications that contain several implementations of these subroutines. In this subsection, we shall do quantitative comparison between **uv**p(3)-TR-Q2/P1 and **uw**p(2)-TR-Q2/P1 solver when combined with our special multigrid method. To this aim, we shall consider a large-scale problem with millions of DOFs. In particular, we shall adopt the problem of wave propagation [4] in an elastic structure-soil system illustrated in Figure 4.13. In the current problem, the structure is represented by a block, which is considered to be in a welded contact with the supporting soil. The applied shear impulse force is given by

$$f(t) = 10^4 \left[1 - \cos(20\pi t)\right]\left[1 - H(t-\tau)\right] \ \ [\mathrm{N/m^2}] \tag{4.2}$$

with $H(t-\tau)$ being the Heaviside step function and $\tau = 0.1$ s. The material parameters of

[4] Such a problem has been intensively studied in the literature, cf. [106, 50, 109, 110, 53]

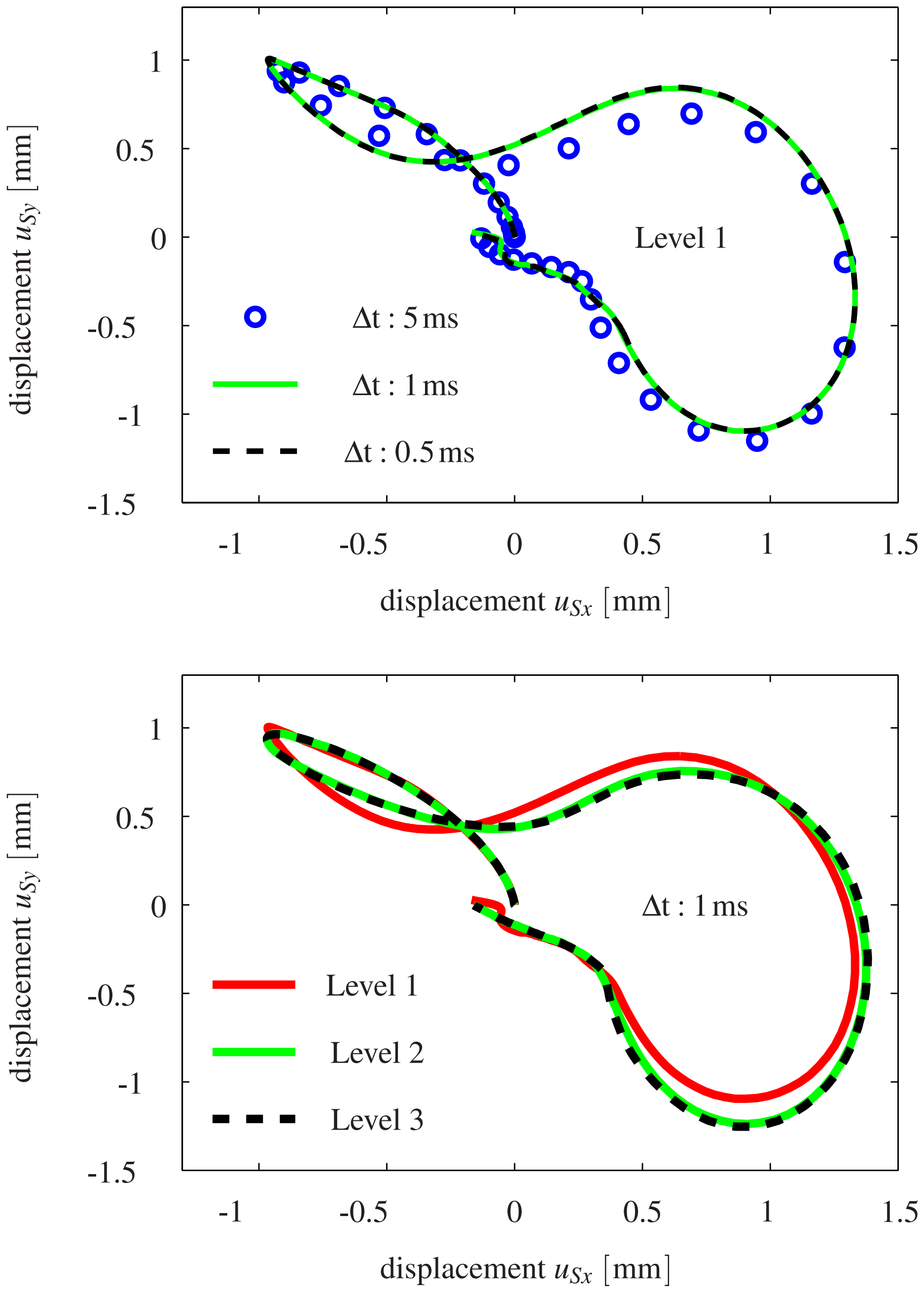


Figure 4.10: Ground rolling during the ‘earthquake event’ at point A using $\mathbf{uv}p(3)$-TR-Q2/P1 with $k^F = 10^{-10}$m/s and $t \in [0\;0.2]$ s.

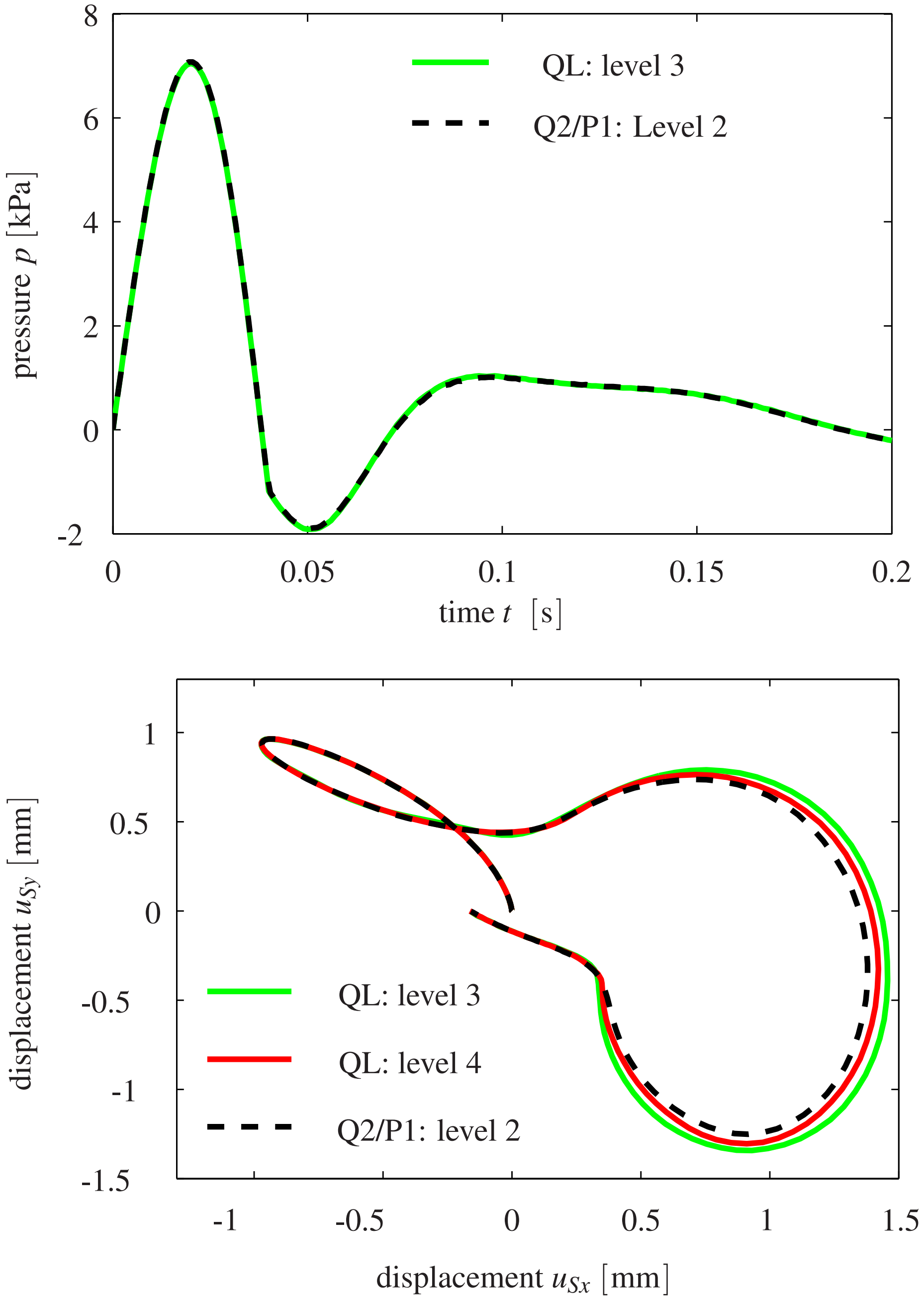


Figure 4.11: Pressure history at point B and displacement history at point A using $\mathbf{uw}p(2)$/ $\mathbf{uv}p(3)$-TR-Q2/P1 and $\mathbf{uv}p(2)$-TB2-QL for $k^F = 10^{-10}\,\mathrm{m/s}$, $\Delta t = 10^{-3}\,\mathrm{s}$, $t \in [0\;0.2]$ s. The results of the QL approach are taken from [63].

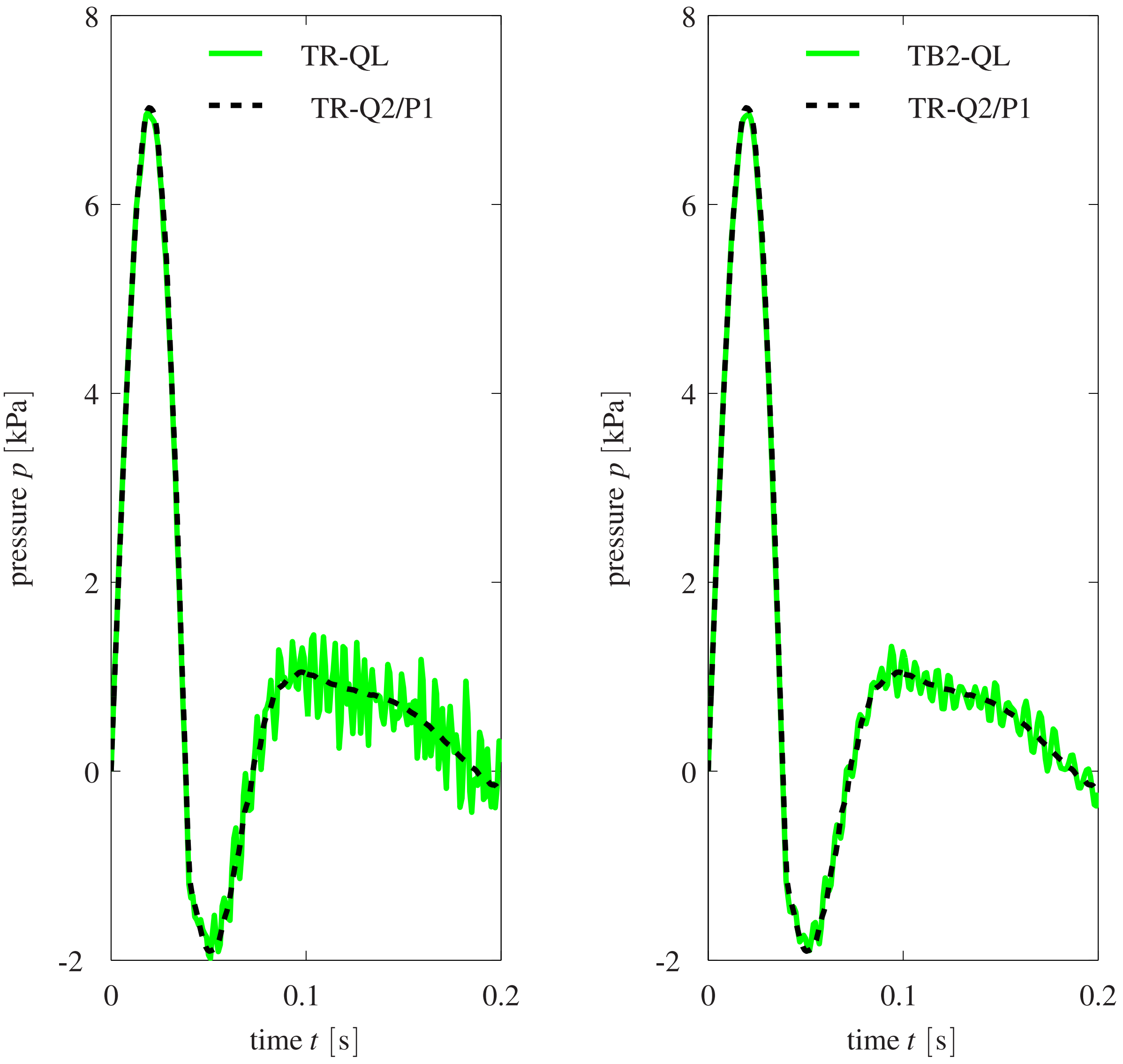


Figure 4.12: Pressure history at point B for $k^F = 10^{-10}\,\mathrm{m/s}$, $\Delta t = 10^{-3}\,\mathrm{s}$ and mesh level 2 for **uv**p(2)-TB2-QL, **uv**p(2)-TR-QL and **uw**p(2)/ **uv**p(3)-TR-Q2/P1. The results of the QL approaches are taken from [63].

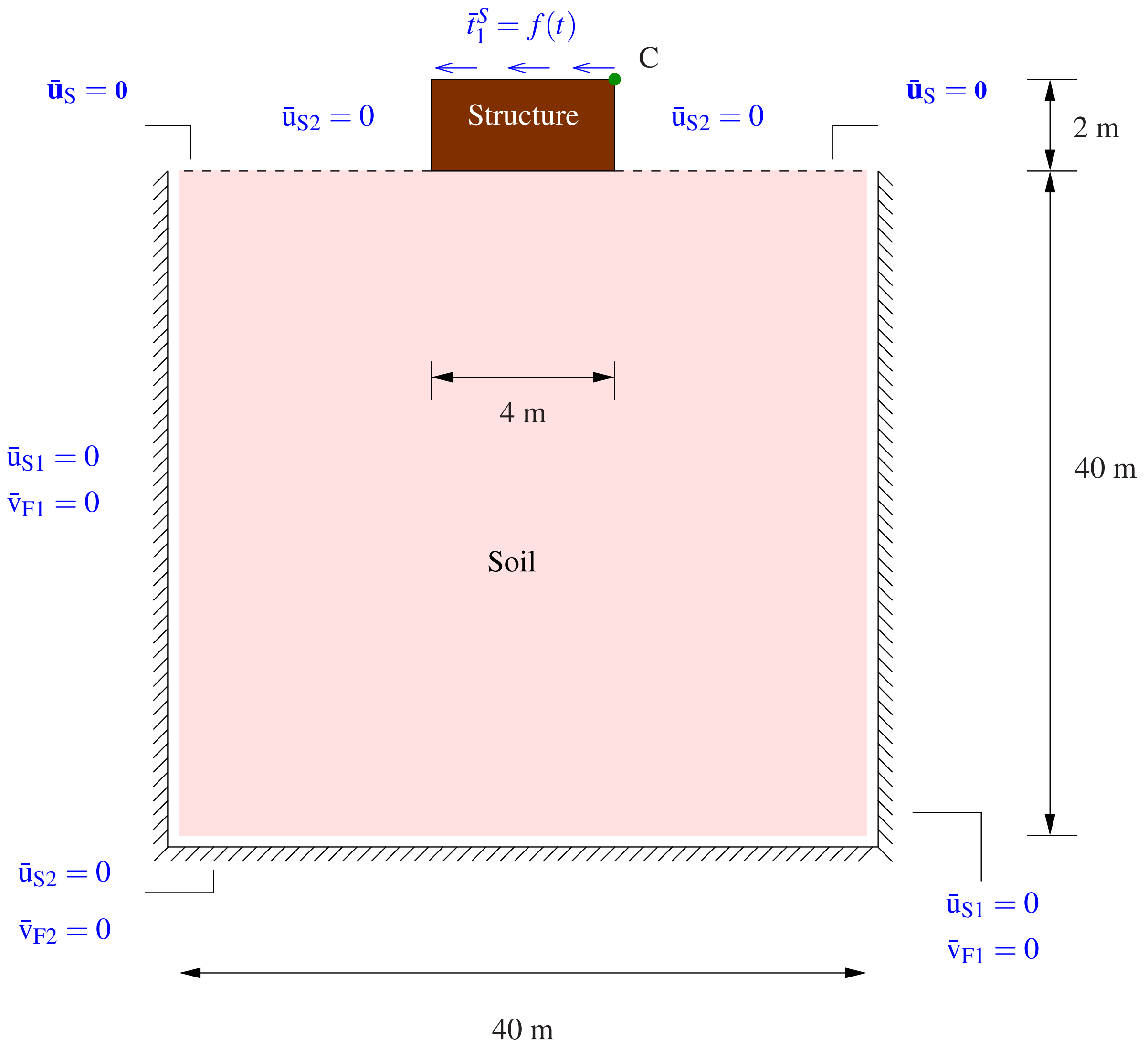


Figure 4.13: Geometry of the 2D structure-soil problem with prescribed boundary conditions. The domain is composed of a structure, represented by an elastic block (size: 4×2 m^2), founded on an infinite domain of elastic soil, replaced by a truncated domain (size: 40×40 m^2) surrounded by impermeable, non-moving and frictionless rigid walls (from left, right and bottom).

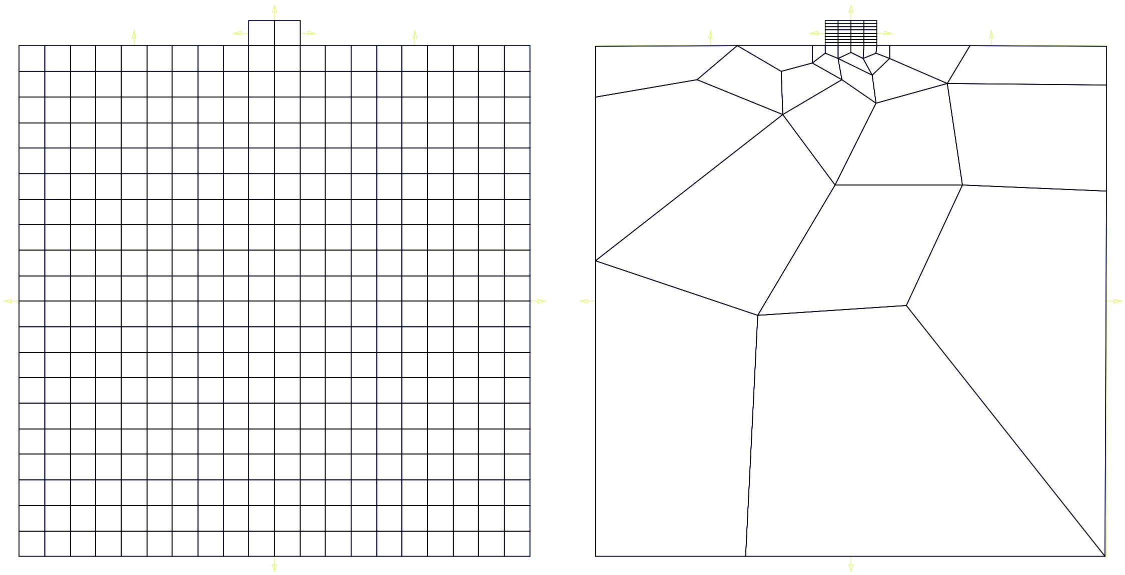


Figure 4.14: Level 1 of the Cartesian (left) and unstructured grid (right) for the problem in Figure 4.13. For higher mesh levels, see Tables 4.3 and 4.4.

the block and the soil are the same (cf. Table 4.2). This implies a weak damping of the vibrations in the loaded structure resulting in a successive wave transition into the soil [50]. The unbounded soil domain beneath the block is replaced by a finite domain with artificial, impermeable, frictionless but rigid boundaries except for the top side, which is fully drained ($\bar{\mathbf{t}}^F = \mathbf{0}$). We shall specifically consider the strong coupling case ($k^F = 10^{-10}$ m/s) together with unstructured meshes (see Figure 4.14) because the author has observed (after several test cases) that the difference in the speed of convergence between our two solvers becomes very clear in these cases. Based on the results shown in Table 4.4, we observe that for extremely low permeability, the multigrid combined with **uw**p solver seems to work remarkably better than the multigrid combined with **uv**p solver in case of irregular meshes as in our unstructured mesh. However, when going to higher mesh levels, the element shapes tend towards being more regular in shape and hence, the speed of **uv**p solver and the speed of **uw**p solver converge. The sensitivity of multigrid-**uv**p solver to mesh irregularity for strong coupling can also be seen from Table 4.5 where multigrid-**uv**p fails to converge in one smoothing step and needs at least two smoothing steps to just converge as shown in Table 4.6 and 4 smoothing steps to reach to the speed of the superior (in case of strong coupling and mesh irregularity) multigrid-**uw**p solver as shown in Table 4.7

Table 4.3: Averaged number of iterations and averaged CPU time in seconds (CPU) per time step for the multigrid solver MG (F-2-2) preconditioned by Vanka scheme in combination with **uv**p(3)-TR-Q2/P1 and **uw**p(3)-TR-Q2/P1 for $k^F = 10^{-10}$ m/s and $\Delta t = 1$ ms and $t \in [0\ 0.1]$ s for the **isotropic** mesh. Level 1 is the mesh level of the multigrid coarse grid solver (here UMFPACK).

Level	#Elem.	#DOFs	**uv**p		**uw**p	
			#Iter.	CPU	#Iter.	CPU
3	6432	175638	3.3	7.5	3.5	7.6
4	25728	698598	3.4	37	3.5	38
5	102912	2786502	3.5	161	3.5	163
6	411648	11130246	3.2	624	3.2	631

Table 4.4: Averaged number of iterations and averaged CPU time in seconds (CPU) per time step for the multigrid solver MG (F-2-2) preconditioned by Vanka scheme in combination with **uv**p(3)-TR-Q2/P1 and **uw**p(3)-TR-Q2/P1 for $k^F = 10^{-10}$ m/s and $\Delta t = 1$ ms and $t \in [0\ 0.1]$ s for the **unstructured** mesh. Level 1 is the mesh level of the multigrid coarse grid solver (here UMFPACK).

Level	#Elem.	#DOFs	**uv**p		**uw**p	
			#Iter.	CPU	#Iter.	CPU
4	3712	101862	15.1	23	5.4	8.2
5	14884	404166	14.2	92	6.1	39
6	59392	1610118	12.8	337	7.5	194
7	237568	6427398	11.6	1247	8.8	941
8	950272	25683462	11.5	5159.5	10.1	4665.4

Table 4.5: Averaged number of iterations and averaged CPU time in seconds (CPU) per time step for the multigrid solver **MG (F-1-1)** preconditioned by Vanka scheme in combination with $\mathbf{uv}p(3)$-TR-Q2/P1 and $\mathbf{uw}p(3)$-TR-Q2/P1 for $\Delta t = 1$ ms and $t \in [0\ 0.1]$ s for the isotropic and anisotropic mesh. Level 1 is the mesh level of the multigrid coarse grid solver (here UMFPACK).

Level	mesh type	k^F	$\mathbf{uv}p$		$\mathbf{uw}p$	
			#Iter.	CPU	#Iter.	CPU
4	isotropic	$k^F = 10^{-2}$	5.2	29.7	6.0	32.0
4	isotropic	$k^F = 10^{-5}$	5.2	28.2	5.7	30.8
4	isotropic	$k^F = 10^{-10}$	5.2	28.2	5.7	30.6
4	anisotropic	$k^F = 10^{-2}$	56.0	42.7	57.2	42.9
4	anisotropic	$k^F = 10^{-5}$	11.4	9.1	12.5	9.4
4	anisotropic	$k^F = 10^{-7}$	12.6	9.5	12.7	9.5
4	anisotropic	$k^F = 10^{-9}$	diverge	diverge	12.7	9.9
4	anisotropic	$k^F = 10^{-11}$	diverge	diverge	12.7	9.8

Table 4.6: Averaged number of iterations and averaged CPU time in seconds (CPU) per time step for the multigrid solver **MG (F-2-2)** preconditioned by Vanka scheme in combination with $\mathbf{uv}p(3)$-TR-Q2/P1 and $\mathbf{uw}p(3)$-TR-Q2/P1 for $\Delta t = 1$ ms and $t \in [0\ 0.1]$ s for the isotropic and anisotropic mesh. Level 1 is the mesh level of the multigrid coarse grid solver (here UMFPACK).

Level	mesh type	k^F	$\mathbf{uv}p$		$\mathbf{uw}p$	
			#Iter.	CPU	#Iter.	CPU
4	isotropic	$k^F = 10^{-2}$	3.3	35	3.5	36.8
4	isotropic	$k^F = 10^{-5}$	3.4	36	3.5	36.5
4	isotropic	$k^F = 10^{-10}$	3.4	36	3.5	37.9
4	anisotropic	$k^F = 10^{-2}$	5.8	8.7	6.0	8.9
4	anisotropic	$k^F = 10^{-5}$	4.6	6.9	4.9	7.3
4	anisotropic	$k^F = 10^{-7}$	5.4	8.0	5.4	8.0
4	anisotropic	$k^F = 10^{-9}$	10.4	15.5	5.4	7.9
4	anisotropic	$k^F = 10^{-11}$	94.3	142.2	5.4	8.3

Table 4.7: Averaged number of iterations and averaged CPU time in seconds (CPU) per time step for the multigrid solver **MG (F-4-4)** preconditioned by Vanka scheme in combination with $\mathbf{uv}p(3)$-TR-Q2/P1 and $\mathbf{uw}p(3)$-TR-Q2/P1 for $\Delta t = 1$ ms and $t \in [0\ 0.1]$ s for the isotropic and anisotropic mesh. Level 1 is the mesh level of the multigrid coarse grid solver (here UMFPACK).

Level	mesh type	k^F	$\mathbf{uv}p$		$\mathbf{uw}p$	
			#Iter.	CPU	#Iter.	CPU
4	isotropic	$k^F = 10^{-2}$	1.8	37.7	2.3	78.4
4	isotropic	$k^F = 10^{-5}$	1.9	38.3	2.3	87.3
4	isotropic	$k^F = 10^{-10}$	2.0	40.3	2.3	83.3
4	anisotropic	$k^F = 10^{-2}$	2.7	8.3	2.7	8.1
4	anisotropic	$k^F = 10^{-5}$	2.8	8.2	2.8	8.2
4	anisotropic	$k^F = 10^{-7}$	2.8	8.3	2.8	8.2
4	anisotropic	$k^F = 10^{-9}$	2.8	8.2	2.8	8.2
4	anisotropic	$k^F = 10^{-11}$	6.0	17.5	2.8	8.2

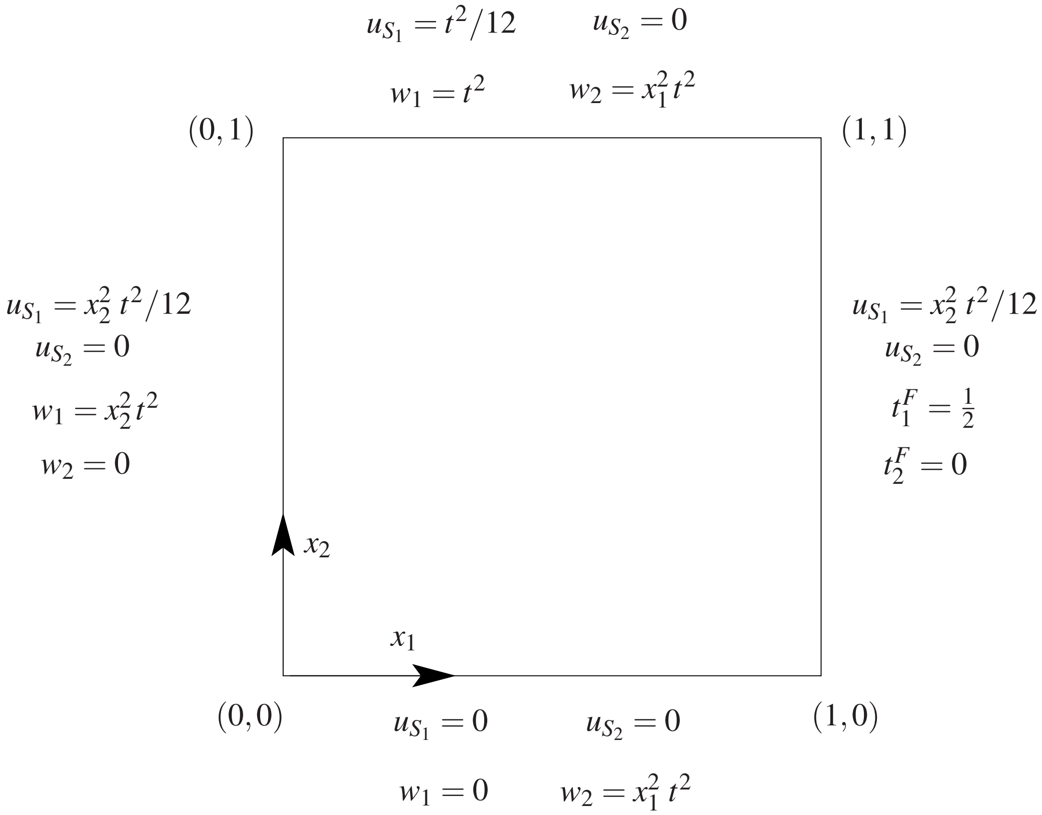


Figure 4.15: A square domain meshed with one element for mesh level 1. **uw**p-Q2/P1-TR was used to solve the problem. This problem has no real physical meaning.

4.2 Numerical results for non-linear problems

4.2.1 Results I: simulation of an analytic test problem

Since we do not have yet results for a rigorous quantitative benchmark to compare with, we first of all present results for an analytical solution. This is a pure mathematical test, which has no physical meaning. The UL formulation is deactivated (i. e., step 4 of Algorithm 1 is omitted) since $\mathbf{u}_S$ in this simulation does not indicate displacements that have a real physical meaning and are merely a mathematical function. The purpose is to debug and validate the code and to make sure that the implementation of the Picard method, the time integrators, the generation of linear and bilinear forms and the implementation of the boundary conditions were done correctly by evaluating the L2- and H1-norms of the error. The non-linear stress tensor $\mathbf{T}_E^S$ is defined in the appendix and the constant physical parameters are given in Table 4.8. Thereafter, the following

Table 4.8: Physical properties of the porous medium used only for section 4.2.1. The gravitational acceleration is set to 10 m/s^2.

Parameter	Symbol	Value	SI Unit
second Lamé constant of solid	μ^S	1	$[\mathrm{N/m^2}]$
first Lamé constant of solid	λ^S	1	$[\mathrm{N/m^2}]$
Effective density of dense solid	ρ^{SR}	2	$[\mathrm{kg/m^3}]$
Effective density of pore fluid	ρ^{FR}	1	$[\mathrm{kg/m^3}]$
Initial volume fraction of solid	n^S_{0S}	0.5	–
Initial volume fraction of fluid	n^F_{0S}	0.5	–
Initial permeability	k^F_{0S}	1	$[\mathrm{m/s}]$
Permeability exponent	κ	1	–

set of equations were solved analytically:

$$\begin{aligned}\left(\rho\ \mathbf{v}_S\right)'_S + \rho^{FR}\ (\mathbf{w})'_S + \rho^{FR}\left(\operatorname{grad}\left(\mathbf{v}_F\right)\ \mathbf{w}\right) - \operatorname{div}\mathbf{T}^S_E + \operatorname{grad} p& \\ - \rho^{FR}\left(\frac{n^S}{n^F}\right)\mathbf{w}\ \operatorname{div}\left(\mathbf{v}_S\right) + \left(\rho^{SR} - \rho^{FR}\right)\ n^S\ \operatorname{div}\left(\mathbf{v}_S\right)\ \mathbf{v}_S &= \mathbf{f}_u,\end{aligned}$$

$$\rho^{FR}\ (\mathbf{v}_S)'_S + \rho^{FR}\left(\frac{\mathbf{w}}{n^F}\right)'_S + \frac{\rho^{FR}}{n^F}\left(\operatorname{grad}\left(\mathbf{v}_F\right)\ \mathbf{w}\right) + \frac{\gamma^{FR}}{k^F}\ \mathbf{w} + \operatorname{grad} p = \mathbf{f}_w,$$

$$\operatorname{div}\left(\mathbf{v}_S\right) + \operatorname{div}\left(\mathbf{w}\right) = 0,$$

where

$$\mathbf{f}_{\mathrm{u}} = \begin{pmatrix} \left(4\ \mathrm{x}_2\ \mathrm{x}_1^2\right)\ \mathrm{t}^4\ +\ \left(\frac{1}{3}\ \mathrm{x}_2\ \mathrm{x}_1^2\right)\ \mathrm{t}^3\ -\ \left(\frac{1}{6}\right)\ \mathrm{t}^2\ +\ \left(2\ \mathrm{x}_2^2\right)\ \mathrm{t} + \left(\frac{1}{4}\mathrm{x}_2^2 - 1\right) \\ \left(4\ \mathrm{x}\ \mathrm{x}_2^2\right)\ \mathrm{t}^4\ +\ \left(2\ \mathrm{x}_1^2\right)\ \mathrm{t} \end{pmatrix},$$

$$\mathbf{f}_w = \begin{pmatrix} \left(8\ \mathrm{x}_2\ \mathrm{x}_1^2\right)\ \mathrm{t}^4\ +\ \left(\frac{2}{3}\ \mathrm{x}_2\ \mathrm{x}_1^2\right)\ \mathrm{t}^3\ +\ \left(10\ \mathrm{x}_2^2\right)\ \mathrm{t}^2\ +\ \left(4\ \mathrm{x}_2^2\right)\ \mathrm{t} + \left(\frac{1}{6}\ \mathrm{x}_2^2 - 1\right) \\ \left(8\ \mathrm{x}\ \mathrm{x}_2^2\right)\ \mathrm{t}^4\ +\ \left(10\ \mathrm{x}_1^2\right)\ \mathrm{t}^2\ +\ \left(4\ \mathrm{x}_1^2\right)\ \mathrm{t} \end{pmatrix}.$$

The domain, the boundary conditions and the mesh are depicted in Figure 4.15. The exact solutions are given by:

$$u_{S_1} = \frac{1}{12}\, x_2^2\, t^2, \qquad u_{S_2} = 0, \qquad w_1 = x_2^2\, t^2, \qquad w_2 = x_1^2\, t^2, \qquad p = \frac{1}{2} - x_1$$

It should be noted that the above solution functions are polynomials of second degree (for $\mathbf{u}_S$ and $\mathbf{w}$) and first degree (for p), which belong to the finite element space of the adopted Q2/P1 element pair (Q2 for $\mathbf{u}_S$ and $\mathbf{w}$, and P1 for p), thence using a one element mesh must be enough to eliminate the spatial errors. This allows us to focus on the time errors without being disturbed by spatial errors. The purpose of this problem is to check whether error reductions of order 1 and 2 for BE and CN, respectively, do occur as it is supposed to do.

To this end, the FE solutions of $\mathbf{uw}p$-TR-Q2/P1 are directly compared with the analytical solutions via L2 and H1 norms of the errors provided by FEATFLOW and shown in Table 4.9. Next, we focus on spatial errors by picking an extremely small time step such that the temporal errors are almost non-existent. For this purpose, the same set of equations with the same physical parameters are solved but for the following right hand side functions:

$$\mathbf{f}_{\mathrm{u}} = \begin{pmatrix} \mathrm{x}_2^3 - 1 - \frac{\mathrm{tx}_2}{4} \\ \\ 0 \end{pmatrix},$$

$$\mathbf{f}_w = \begin{pmatrix} 10\mathrm{x}_2^3\mathrm{t} + 2\ \mathrm{x}_2^3 - 1 \\ \\ 0 \end{pmatrix},$$

where the domain, the boundary conditions and the mesh are depicted in Figure 4.16. Solving the problem analytically for the full assumption gives the following exact solution functions:

$$u_{S_1} = \frac{1}{24}\, x_2^3\, t, \qquad u_{S_2} = 0, \qquad w_1 = x_2^3\, t, \qquad w_2 = 0, \qquad p = \frac{1}{2} - x_1.$$

and the results are shown in Table 4.10.

4.2.2 Results II: Two-dimensional wave propagation IBVP

In this example, we study the 2D dynamical problem, depicted in Figure 4.17 under plane-strain conditions. The solid constituent is hyper-elastic as defined in (2.228) and the Appendix A. The material parameters are given in Table 4.11.

The load $f(t)$ is given by

$$f(t) = \sin(25\,\pi\,t)\,[1 - H(t-\tau)]\ \ [\times 10^6\ \mathrm{Pa}] \tag{4.3}$$

with $H(t-\tau)$ being the Heaviside step function and $\tau = 0.04s$. The water saturated mixture domain is surrounded by impermeable, frictionless ($\bar{t}_1^F = 0$ for the bottom and $\bar{t}_2^F = 0$ for the left and right sides) but rigid boundaries except for the loaded top side, which is perfectly drained

Table 4.9: The errors in the finite element solutions at time $t = 0.01$ $[\mathrm{s}]$ and mesh level 1 of the ULP-**uw***p*-Q2/P1 solver for both Crank Nicolson (CN) and Backward Euler (BE) measured through the L2 and H1 norms.

error norm	Backward Euler (BE)		
	$\Delta t = 1 \times 10^{-5}$ $[\mathrm{s}]$	$\Delta t = 1 \times 10^{-6}$ $[\mathrm{s}]$	$\Delta t = 1 \times 10^{-7}$ $[\mathrm{s}]$
$\|\|\mathbf{u}_S - \mathrm{ref.}\|\|_{L2}$	2.0×10^{-9}	2.0×10^{-10}	2.0×10^{-11}
$\|\|\mathbf{v}_S - \mathrm{ref.}\|\|_{L2}$	2.6×10^{-8}	2.6×10^{-9}	2.6×10^{-10}
$\|\|\mathbf{w} - \mathrm{ref.}\|\|_{L2}$	3.4×10^{-8}	3.4×10^{-9}	3.4×10^{-10}
$\|\|\mathrm{p} - \mathrm{ref.}\|\|_{L2}$	2.8×10^{-6}	2.8×10^{-7}	2.8×10^{-8}
$\|\|\mathbf{u}_S - \mathrm{ref.}\|\|_{H1}$	9.0×10^{-9}	9.0×10^{-10}	9.0×10^{-11}
$\|\|\mathbf{v}_S - \mathrm{ref.}\|\|_{H1}$	1.2×10^{-7}	1.2×10^{-8}	1.2×10^{-9}
$\|\|\mathbf{w} - \mathrm{ref.}\|\|_{H1}$	1.5×10^{-7}	1.5×10^{-8}	1.5×10^{-9}
$\|\|\mathrm{p} - \mathrm{ref.}\|\|_{H1}$	6.3×10^{-6}	6.3×10^{-7}	6.3×10^{-8}
error norm	Crank Nicolson (CN)		
	$\Delta t = 1 \times 10^{-4}$ $[\mathrm{s}]$	$\Delta t = 1 \times 10^{-5}$ $[\mathrm{s}]$	$\Delta t = 1 \times 10^{-6}$ $[\mathrm{s}]$
$\|\|\mathbf{u}_S - \mathrm{ref.}\|\|_{L2}$	9.3×10^{-11}	9.3×10^{-13}	9.3×10^{-15}
$\|\|\mathbf{v}_S - \mathrm{ref.}\|\|_{L2}$	3.4×10^{-17}	3.5×10^{-17}	3.7×10^{-17}
$\|\|\mathbf{w} - \mathrm{ref.}\|\|_{L2}$	1.6×10^{-9}	1.6×10^{-11}	1.6×10^{-13}
$\|\|\mathrm{p} - \mathrm{ref.}\|\|_{L2}$	1.7×10^{-14}	1.2×10^{-14}	5.7×10^{-14}
$\|\|\mathbf{u}_S - \mathrm{ref.}\|\|_{H1}$	2.4×10^{-10}	2.4×10^{-12}	2.4×10^{-14}
$\|\|\mathbf{v}_S - \mathrm{ref.}\|\|_{H1}$	1.5×10^{-16}	1.5×10^{-16}	1.6×10^{-16}
$\|\|\mathbf{w} - \mathrm{ref.}\|\|_{H1}$	4.1×10^{-9}	4.1×10^{-11}	4.1×10^{-13}
$\|\|\mathrm{p} - \mathrm{ref.}\|\|_{H1}$	4.7×10^{-14}	3.9×10^{-14}	1.4×10^{-13}

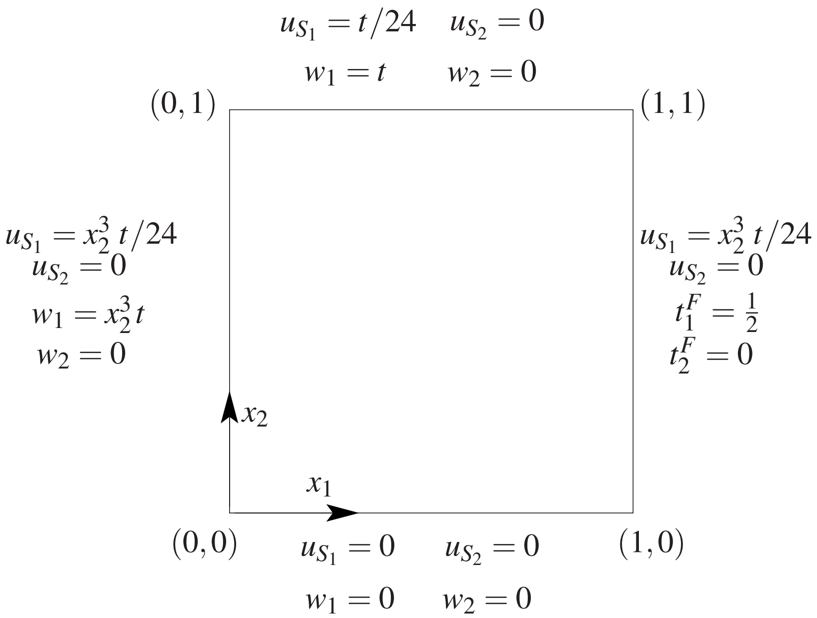
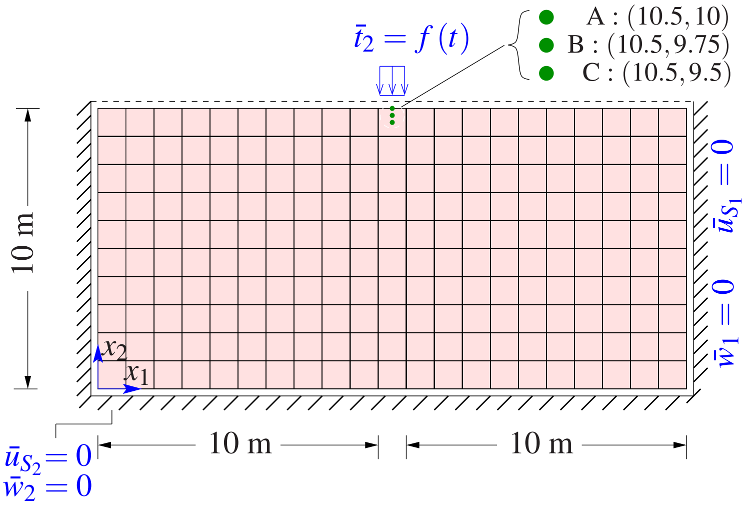


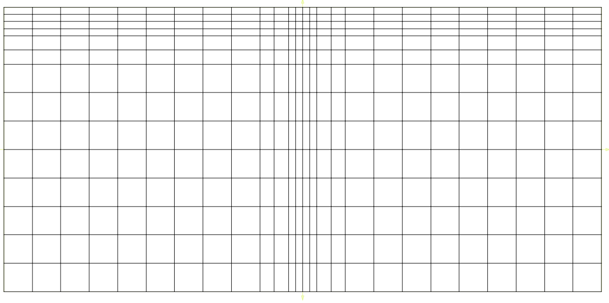
Figure 4.16: A square domain meshed with one element for mesh level 1. **uw**p-Q2/P1-TR was used to solve the problem. This problem has no real physical meaning and the UL formulation is deactivated to allow for pure testing of the Picard iterative method.

Table 4.10: The errors in the finite element solutions for different mesh levels at time $t = 0.01$ [s] of the ULP-**uw**p-Q2/P1 solver for $\Delta t = 1 \times 10^{-5}$ [s] for both Crank Nicholson and Backward Euler measured through the L2 and H1 norms.

errors	mesh level 2	mesh level 3	mesh level 4	reduction, level 3/ 4
$\|\|\mathbf{u}_S - \text{ref.}\|\|_{L2}$	1.81132×10^{-6}	2.27358×10^{-7}	2.82564×10^{-8}	8.0463
$\|\|\mathbf{v}_S - \text{ref.}\|\|_{L2}$	1.81207×10^{-4}	2.27381×10^{-5}	2.82476×10^{-6}	8.0496
$\|\|\mathbf{w} - \text{ref.}\|\|_{L2}$	4.36619×10^{-5}	5.41021×10^{-6}	6.74468×10^{-7}	8.0214
$\|\|\text{p} - \text{ref.}\|\|_{L2}$	6.11228×10^{-6}	3.35809×10^{-7}	8.77083×10^{-8}	3.8287
$\|\|\mathbf{u}_S - \text{ref.}\|\|_{H1}$	2.33429×10^{-5}	5.86860×10^{-6}	1.46165×10^{-6}	4.0150
$\|\|\mathbf{v}_S - \text{ref.}\|\|_{H1}$	2.33538×10^{-3}	5.86971×10^{-4}	1.46140×10^{-4}	4.0165
$\|\|\mathbf{w} - \text{ref.}\|\|_{H1}$	5.64424×10^{-4}	1.40021×10^{-4}	3.49498×10^{-5}	4.0063
$\|\|\text{p} - \text{ref.}\|\|_{H1}$	2.33453×10^{-5}	1.68093×10^{-6}	4.95731×10^{-7}	3.3908



Mesh Level	# Elements (width-height)	# DOFs (Q2/P1)
1	21-10	6048
2	42-20	23430
3	84-40	92214
4	168-80	365862



Mesh Level	# Elements (total)	# DOFs (Q2/P1)
1	364	10314
2	1456	40278
3	5824	159174
4	23296	632838
5	93184	2523654

Figure 4.17: Geometry, boundary conditions and isotropic mesh level 1 of the symmetric 2D wave propagation problem (top-left). Total number of elements and unknowns for the Q2/P1 approach (top-right). Anisotropic mesh level 1 (bottom-left). Total number of elements and unknowns for the Q2/P1 approach (bottom-right). Going from mesh level i to mesh level i+1, every old local element is isotropically refined into 4 new elements. The symmetry of the problem can be exploited to reduce the problem size. However, the computation was performed for the full problem.

Table 4.11: Physical properties of the porous medium used for all simulations except the ones in section 4.2.1. The gravitational acceleration is set to 10 m/s^2.

Parameter	Symbol	Value	SI Unit
first Lamé constant of solid	μ^S	5.583×10^6	[N/m^2]
second Lamé constant of solid	λ^S	8.375×10^6	[N/m^2]
Effective density of dense solid	ρ^{SR}	2000	[kg/m^3]
Effective density of pore fluid	ρ^{FR}	1000	[kg/m^3]
Initial volume fraction of solid	n^S_{0S}	0.67	–
Darcy permeability	k^F_{0S}	from 10^{-5} to 0.5	[m/s]

($\bar{\mathbf{t}}^F = \mathbf{0}$). This numerical example, including the loading, the geometry and the material parameters, was adopted from [63], in which a comparison between different equation formulations was carried out. It has already been shown within small strain settings that an accurate solution especially for higher permeabilities demands the use of uvp or uwp formulations, whereas the simplified displacement-pressure (up) formulation yields an inaccurate approximation. The objective in the following is to study quantitatively the effect of the volume fractions rate of change (the green terms in (3.64)) and the effect of convective term (the orange terms in (3.64)-(3.65)) on the solutions. Here, we differentiate between three reduced cases: (1) 'fully reduced ', in which all the green terms and all the orange terms are neglected, (2) 'no orange ' that excludes only the convection, (3) 'no green' case where only the changes in volume fractions are ignored. To do so, we prefer to do the comparison on a mesh level that leads to full convergence of the solutions (u_{S_2} and p) for the full $\mathbf{uw}p$ formulation. Therefore, three equidistant points initially located on the axis of symmetry and in the first half meter below the top loaded boundary were opted for this purpose. The results are depicted in Figure 4.18 and show the full convergence on mesh level 5.

Based on the results of this problem (for sample solutions, see Figure 4.19), we observed that the convection may noticeably influence the fluidic solution components (p and $\mathbf{v}_F$) if the considered k^F_{0S} is high enough. However, the influence on the deformation ($\mathbf{u}_S$) by convection remains negligible. On the other hand, the influence on the pressure is much weaker than the deformation in case of volume fraction changes (the green terms) and remains very small for both $\mathbf{u}_S$ and p . To gain a better picture, the deviations of the solutions p_r (subscript r refers to one of three reduced cases) from the solutions p of the 'full ' $\mathbf{uw}p$ form is quantitatively

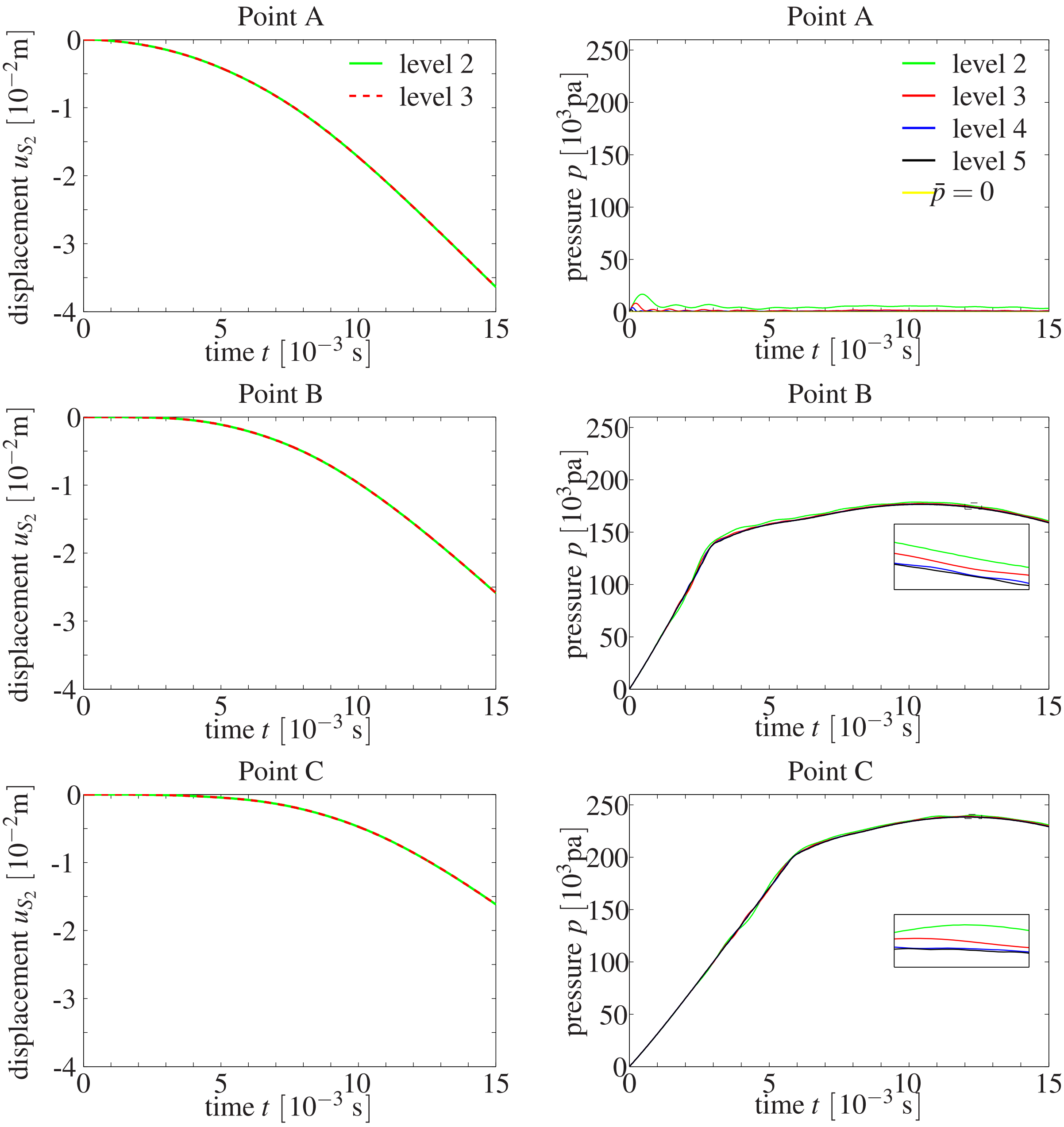


Figure 4.18: The displacement and the pressure versus time for different points located on the axis of symmetry in the first meter below the top loaded boundary for $k_{0S}^F = 10^{-2}$ [m/s], $\kappa = 1$ and for the three reduced cases using **uw**$p(3)$-TR-Q2/P1 method with $\Delta t = 2.5 \times 10^{-5}$ [s] and the anisotropic mesh level 5.

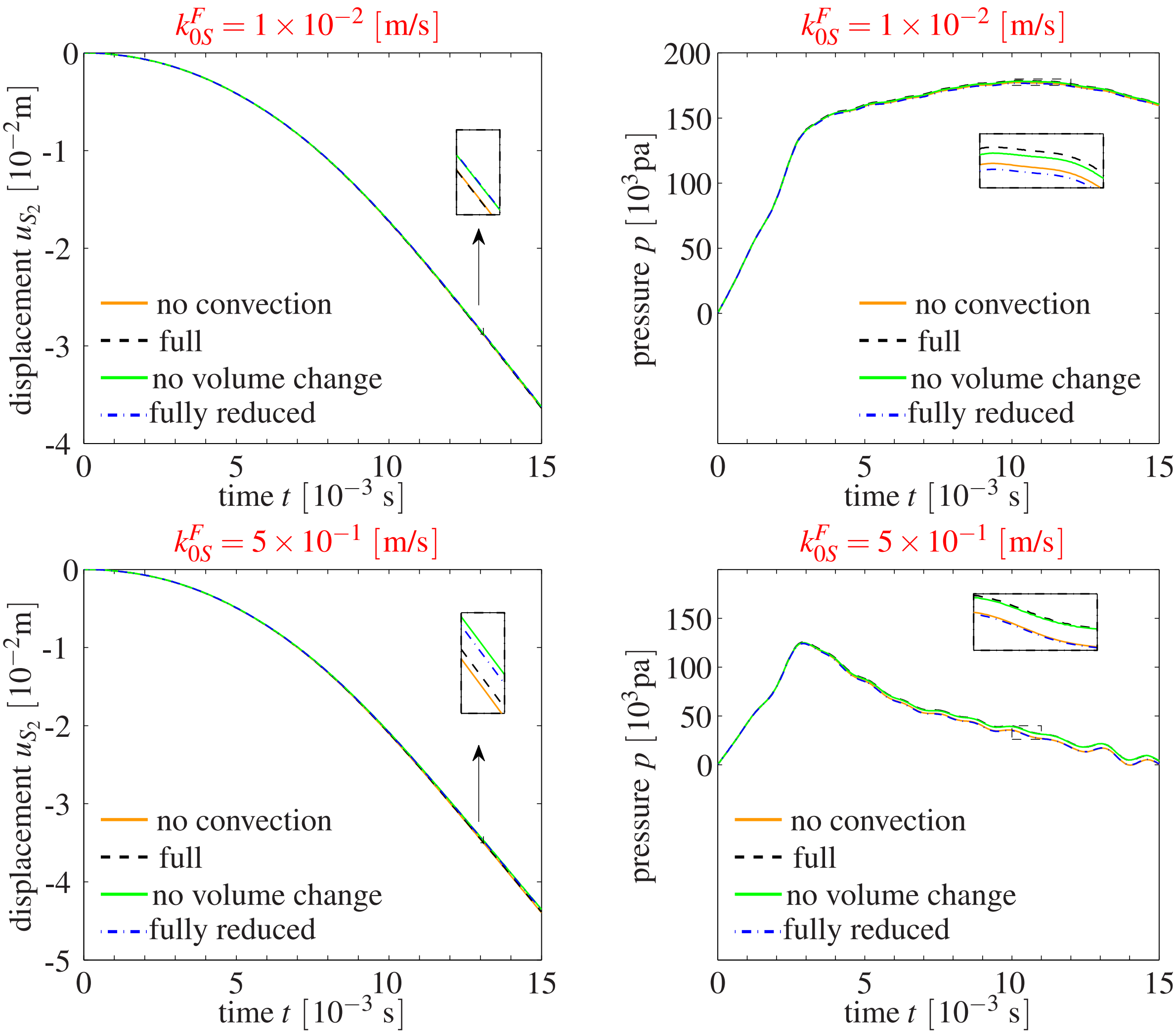


Figure 4.19: The displacement and the pressure at a point B for $k^F_{0S} = 10^{-2}$ [m/s], $\kappa = 1$ and for the three reduced cases using $\mathbf{uw}p(3)$-TR-Q2/P1 method with $\Delta t = 2.5 \times 10^{-5}$ [s] s and the anisotropic mesh level 5.

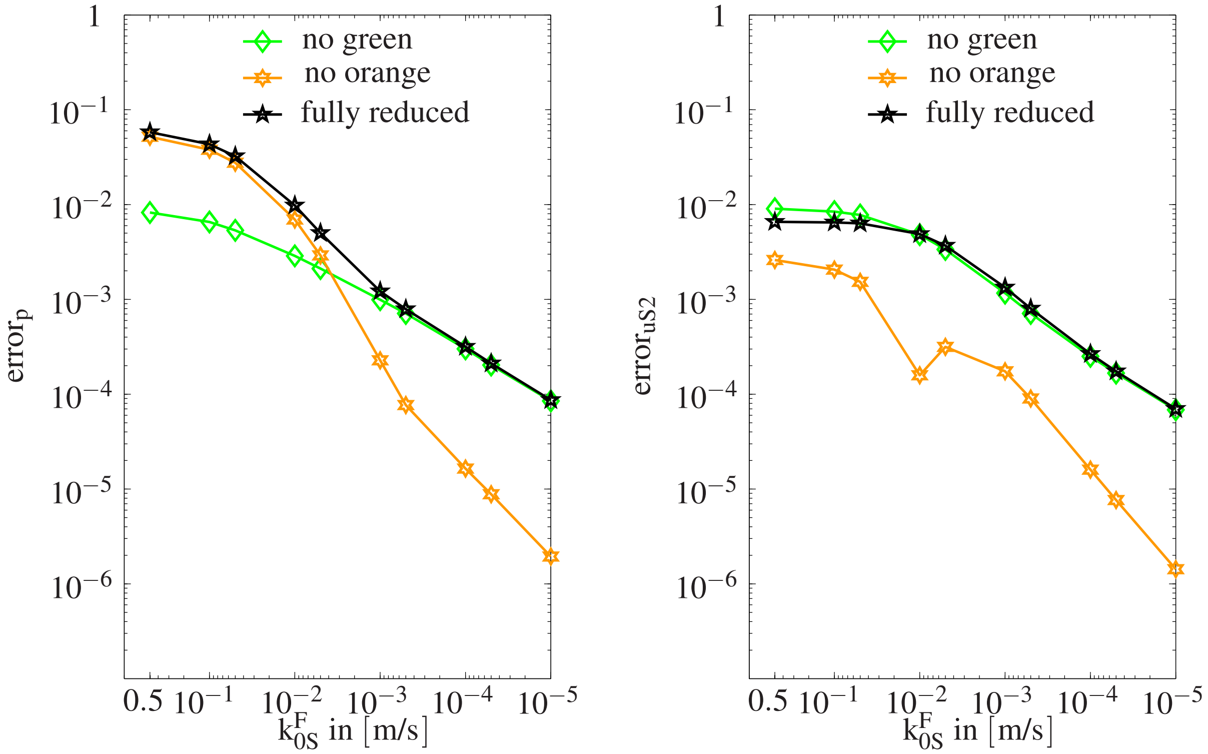


Figure 4.20: $\mathrm{error_p}$ and $\mathrm{error_{uS2}}$ computed using (4.4) at a point initially located at the axis of symmetry and 0.25 m below the top loaded boundary for $\kappa = 1$ and for the three reduced cases using $\mathbf{uw}p$(3)-TR-Q2/P1 method with $\Delta t = 2.5 \times 10^{-5}$ s and the anisotropic mesh level 5.

measured using the following relative error formula

$$error_p = \frac{\sum\limits_{i=1}^{n} |p(t_i) - p_r(t_i)|}{\sum\limits_{i=1}^{n} |p(t_i)|}, \tag{4.4}$$

where n denotes the number of time steps. By replacing p in (4.4) with u_{S2}, we obtain the formula for $error_{uS2}$. Based on Figure 4.20, we conclude that in case of strong coupling (i. e., moderately small k^F ($k^F \leq 10^{-5}$ [m/s]) and extremely small k^F ($k^F \leq 10^{-10}$ [m/s])), the convective terms become more or less unimportant and can be canceled out. For the case of weaker coupling corresponding to higher values of the permeability, the effect of the convective term on the displacement field $\mathbf{u}_S$ becomes remarkable. Thus, it is justified to neglect the convection (for $k^F_{0S} < 10^{-5}$ m/s) in order to simplify the system of equations.

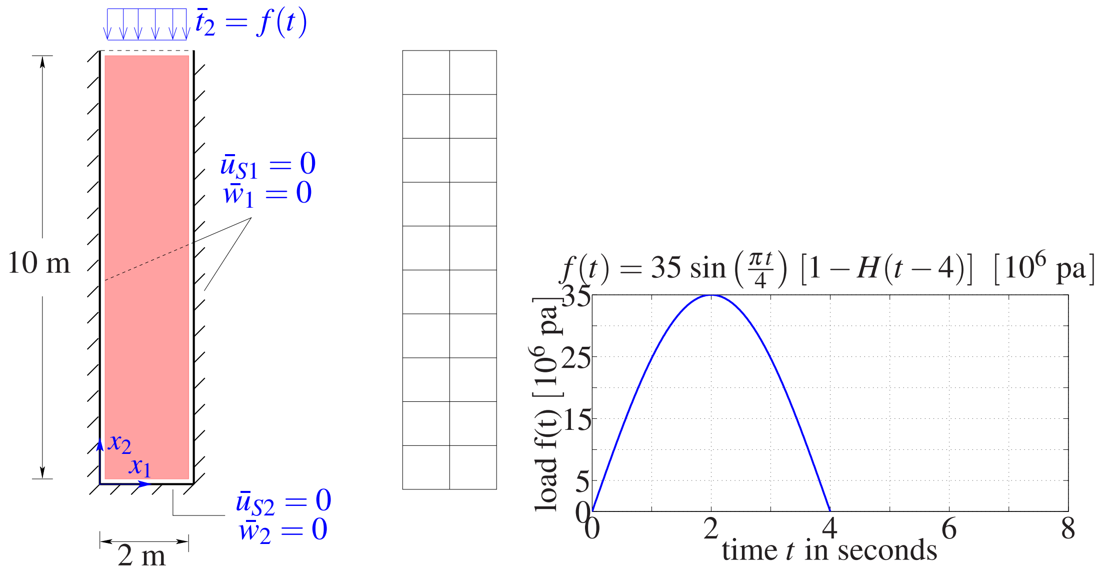


Figure 4.21: Geometry and boundary conditions (left) and Cartesian mesh level 1 with 1 element per meter (middle). The physical parameters are found in Table 4.11 and we set $k^F_{0S} = 10^{-4}\text{m/s}$.

4.2.3 Results III: Adaptive Time Stepping (ATS)

The convergence speed of the $\mathbf{uw}p$(3)-TR-Q2/P1 solver (Algorithm 2) is strongly influenced by the non-linear Cauchy extra stress $\mathbf{T}^S_E$, which, for our application, is inversely proportional to the minimum value of the deformation dependent n^F. This can be seen if we let $n^F \to 0$. Then by (2.122), $J_S \to n^S_{0S}$ and consequently $\tilde{h}(J_S)$ in (A.4) and (hence $\mathbf{T}^S_E$ in (A.1)) goes to infinity. Strictly speaking, if $\min\limits_{\mathbf{x}\in\Omega} n^F(\mathbf{x})$ gets smaller, then a corresponding local $\mathbf{T}^S_E$ becomes larger and more iterations are required. However, when integrating in time, the non-linear operator $\displaystyle\int_{\Omega(t)} \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}^S_E \,\mathrm{d}v^t$ is damped by small Δt, which we increase or decrease (in a suitable amount) according to the spatial maximum of $\mathbf{T}^S_E$. To demonstrate this fact, we will adopt the problem of a saturated poroelastic column (cf. Figure 4.1) under compression load, as for this specific problem, we know that

$$\min_{\mathbf{x}\in\Omega} n^F(\mathbf{x}(t)) = n^F(\mathbf{x}_{top}(t)) \quad \forall t \in [0\ 2]\ [\text{s}],$$

where $\mathbf{x}_{top}$ refers to the top surface. Note from Figure 4.22, we frequently have to decrease the time-step size Δt to adapt to the further increase in the non-linear $\mathbf{T}^S_E$, provoked by smaller $\mathrm{n}^{\mathrm{F}}(\mathbf{x}_{\mathrm{top}})$ in $t \in [0\ 2]$. But this reduction in Δt should be done carefully, so that the simulation is finished as fast as possible and within the desired accuracy. An excessive reduction in the time step size will prolong the CPU time, while poor reduction may slow down the speed of convergence of the ULP.

Since we do not yet have an excellent predictor for the best time step increase or decrease, we present an adaptive time stepping algorithm (see Algorithm 3), which differs from Algorithm 1 by the red statements. The algorithm uses the non-linear convergence rate ξ for iteration i,

$$\xi^i = \sqrt[i]{\frac{\|\mathbf{d}^i\|}{\|\mathbf{d}^0\|}},$$

as indicator for the strength of the non-linearity to adjust the time step size. Here, we set an upper bound (ξ_{max}) for ξ^i and if ξ^i happens to exceed ξ_{max}, the time step gets aborted (unless the solver accidental converged to the desired tolerance to avoid time wasting) and then reduced by $rat\%$ as described in the algorithm. Large values for percentage reduction $rat\%$ produce highly oscillating Δt's, while low values yield almost smooth Δt's as shown in Figure 4.23.

With regard to CPU timings, Table 4.12 shows similar values for $rat\% \in \{50\%, 5\%, 0.5\%\}$. Therefore, we shall switch to larger problems to find good combinations of ξ_{max} and $rat\%$. To do so, we will solve the problem illustrated in Figure 4.17, but with the following large impulse load

$$f(t) = 6.0\sin(25\pi t)\,[1 - H(t-0.04)]\;\;[10^6 \text{ pa}], \tag{4.5}$$

which generates large compression (and hence large local stresses $\mathbf{T}_E^S$) quickly, such that fixed time steps may fail (or become impractical) and ATS is pressingly needed. Next, we inspect the point positioned on the intersection of the top boundary with the axis of symmetry because the axis of symmetry experiences the largest settling (or minimum n^F) for the considered time interval t $\in [0\,0.2]$. From Figure 4.24, we notice larger settling (and hence larger local $\mathbf{T}_E^S$), which requires a smaller Δt and vice versa. Now, we come to our main issue, that is finding good combinations ξ_{max} and $rat\%$. From Table 4.13 and Table 4.14 and since all $\xi_{max} \leq 0.3$ give good accuracy, we conclude that $\xi_{max} \in [0.05\ 0.3]$ (in particular $\xi_{max} = 0.1$) together with $rat = 5\%$ are pretty good choices.

It remains to mention that in order to save some CPU time, it is recommended to start with an initial Δt large enough (for instance something around 1×10^{-3} [s]).

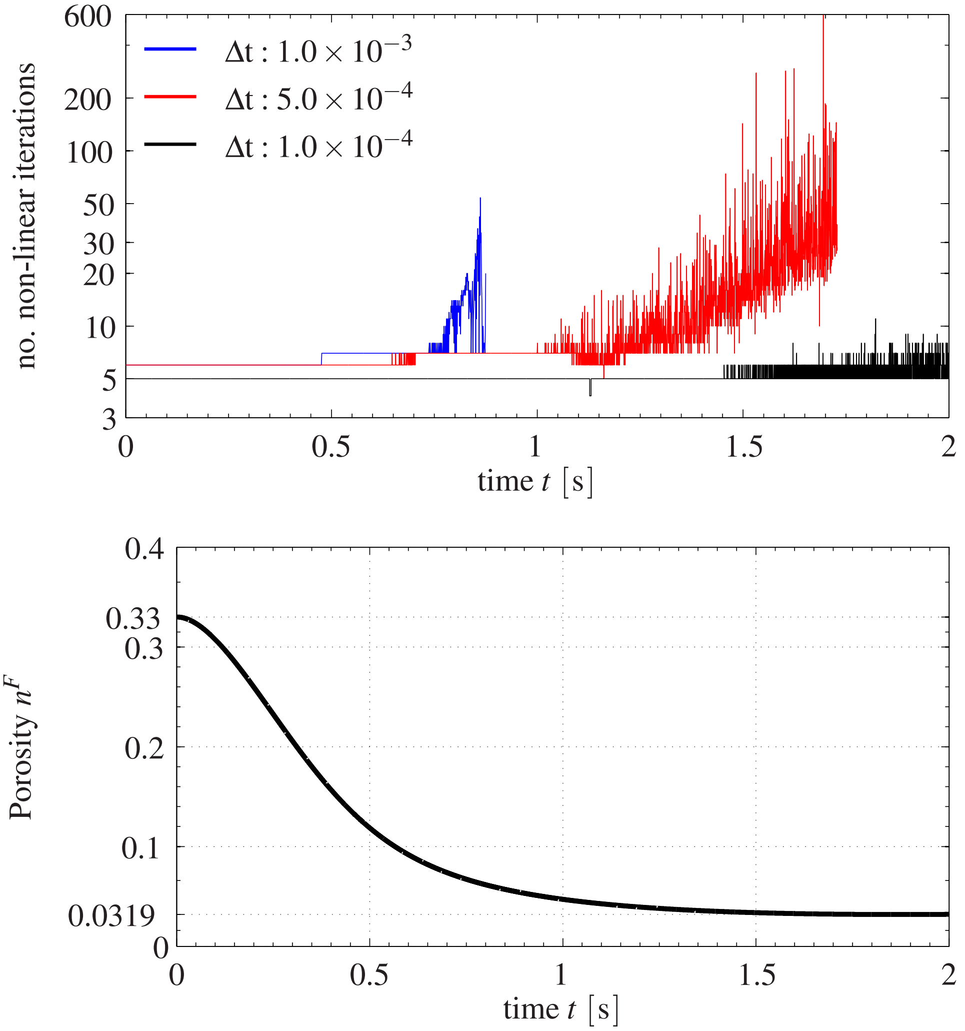


Figure 4.22: Number of non-linear iterations vs. time (top) and number of non-linear iterations vs. porosity n^F (bottom) for ATS-fully-reduced-**uw**p-TR-Q2/P1-ULP-UMFPACK, mesh level 1, tolerance = 10^{-5}.

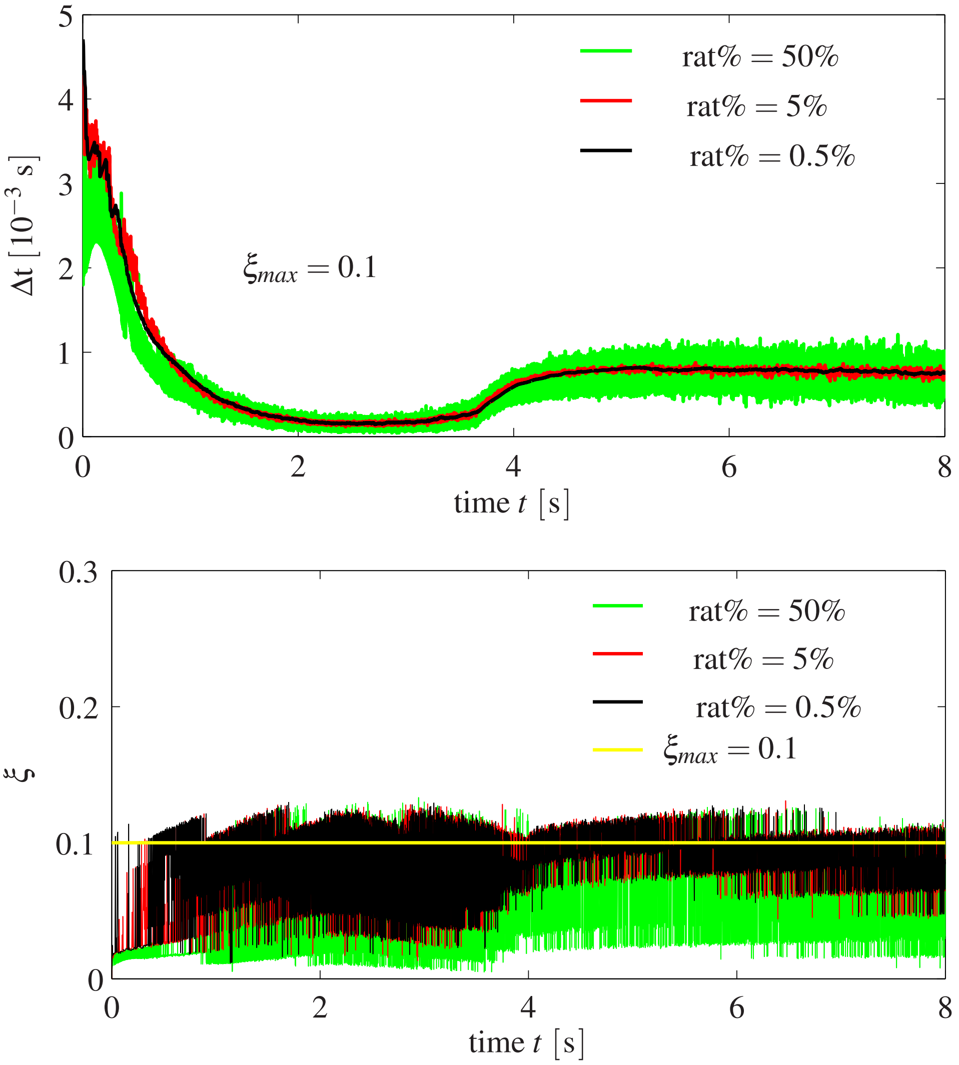


Figure 4.23: Time step Δt vs. time (top) and ξ vs. time (bottom) for fully-reduced-**uw**p-TR-Q2/P1, mesh level 1 and tolerance = 10^{-12}.

Table 4.12: Total number of steps (# Step), Total number of non-linear iterations (# Iter.) and total elapsed CPU time (CPU) in seconds for the described ULP solver for $t \in [0\ 2]$ **uw**p-TR-Q2/P1 + UMFPACK including the successful and the aborted steps.

$rat\% = 0.5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	209 346	782538	9332	338114	1598911	158058
0.05	27131	209857	2252	159781	1048473	99478
0.10	23788	243543	2585	138671	1135668	110425
0.30	16896	304417	3217	109966	1520808	143824
$rat\% = 5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	211895	785829	9188	340414	1610024	159398
0.05	26944	208853	2340	161031	1053202	99742
0.10	23623	239730	2531	139526	1134378	109241
0.30	15704	290501	2992	110735	1512872	140950
$rat\% = 50\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	256174	950533	11088	382596	1714069	170020
0.05	29073	203776	2204	184107	1136921	111305
0.10	26692	241811	2588	159349	1193064	107856
0.30	17711	289811	3013	127108	1509608	139835

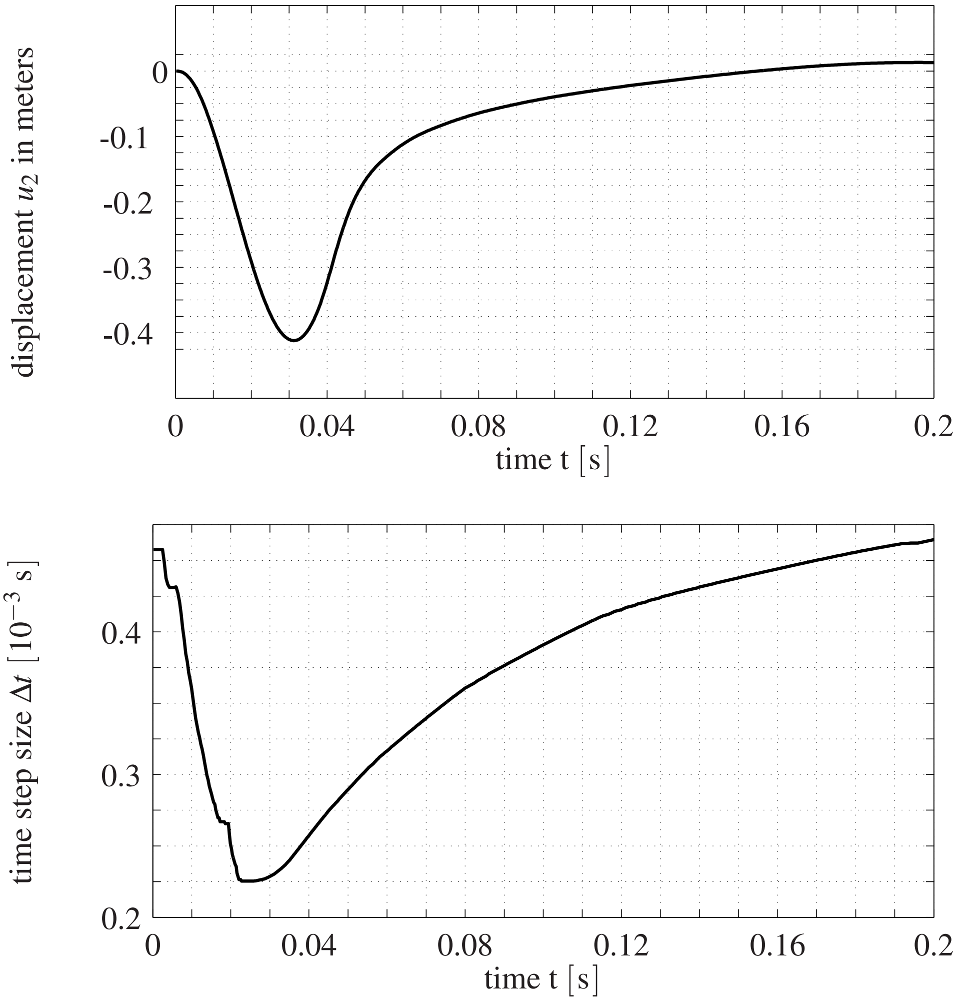


Figure 4.24: Time step Δt vs. time (bottom) Displacement of point A vs. time (top). The results were generated for **uw**p-TR-Q2/P1-ATS-ULP-UMFPACK solver with $rat\% = 0.5\%$ and $\xi_{max} = 0.1$, mesh level 2, tolerance = 10^{-12} and $k_{0S}^F = 10^{-2}$ [m/s].

Table 4.13: Total number of steps (# Step), Total number of non-linear iterations (# Iter.) and total elapsed CPU time (CPU) in seconds for the described ATS-ULP solver for t $\in [0\ 0.2]$ $\mathbf{uw}p$-TR-Q2/P1 + UMFPACK (as preconditioner) including the successful and the aborted steps. This table is related to the configuration of Figure 4.17 with load as defined in (4.5).

$k_{0S}^F = 10^{-2}$ [m/s]						
$rat\% = 0.5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	2651	12114	2616	4117	21537	22383
0.05	803	4508	960	1130	9641	9934
0.10	609	4142	879	854	9182	9418
0.30	413	3579	755	547	10240	10464
$rat\% = 5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	1525	9090	1950	2942	17911	18556
0.05	366	3526	741	712	7433	7664
0.10	257	3124	657	514	6783	6958
0.30	158	3244	674	314	7392	7551
$rat\% = 50\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	1375	8354	1794	3072	18996	19677
0.05	341	3527	745	741	7720	7928
0.10	242	3190	671	580	7347	7536
0.30	159	3522	734	381	8015	8188

Table 4.14: Total number of steps (# Step), Total number of non-linear iterations (# Iter.) and total elapsed CPU time (CPU) in seconds for the described ATS-ULP solver for $t \in [0\ 0.2]$ **uw**p-TR-Q2/P1 + UMFPACK (as preconditioner) including the successful and the aborted steps. This table is related to the configuration of Figure 4.17 with load as defined in (4.5).

$k_{0S}^{F} = 10^{-5}$ [m/s]						
$rat\% = 0.5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	1192	6505	1417	2257	12852	14433
0.05	666	4173	913	1127	8041	9107
0.10	516	3638	787	877	7149	8086
0.30	336	3484	744	582	6849	7697
$rat\% = 5\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	688	5102	1106	1618	11090	12415
0.05	318	3170	687	641	6602	7405
0.10	232	2853	609	453	5789	6508
0.30	140	2992	633	269	5861	6549
$rat\% = 50\%$						
ξ_{max}	level 1			level 2		
	# Step	# Iter	CPU	# Step	# Iter	CPU
0.01	660	5076	1098	1712	11990	13368
0.05	288	3094	685	619	6662	7498
0.10	228	2964	663	477	6272	7046
0.30	150	3240	709	334	6785	7591

5
Conclusion and Future work

5.1 Conclusion

In preparation for the numerical treatment, the governing equations of porous media dynamics were introduced first. This covers the porous media modeling approach, the corresponding kinematics as well as the equilibrium and the linear and non-linear elastic constitutive relations. Mathematically, this leads eventually to a set of four partial differential equations, which need to be solved: (1) the balance of momentum of the solid phase, (2) the balance of momentum of the fluid phase, (3) the volume balance of the mixture, (4) the solid velocity-displacement relation.

This set of PDE's together with the boundary values is called IBVP1. In addition, by replacing the fluid velocity with the Darcy velocity and the balance of momentum of the solid phase with the balance of momentum of the mixture, we obtain the so-called IBVP 2 (for linear constitutive model) or IBVP3 (for hyper-elastic constitutive model)

The forgoing three IBVP's were first discretizated in space within the mixed FEM by the well-known (non-parametric) Q2/P1 finite element pair, which belongs to the best choices for incompressible flow problems in terms of efficiency, accuracy and robustness, while the discretization in time has been carried out by the standard θ-scheme ($\theta = 1/2$), which leads to a fully implicit, monolithic treatment of all variables involved. The outcome is a discrete weak form that demands less regularity and allows Neumann boundary conditions that are more convenient to apply.

For the solution of the resulting (linear) systems of equations for each time step, the fast geometrical multigrid solver with special block Vanka smoother available in FEATFLOW has been used, which leads to convergence rates being independent of time step and mesh size, which is important particularly for large-scale problems. For the solution of the resulting nonlinear systems, the Picard iterative method combined with updated Lagrangian method were used to avoid the computation of the time consuming material tangent matrix.

For validation and evaluation the proposed procedures, canonical 1D and 2D examples were opted from the related literature and solved by our techniques, which we implemented into FEATFLOW. The result of FEATFLOW calculations showed (1) perfect matching with the published data with higher accuracy than the old techniques, (2) the problem of pressure instabilities or wrong solutions (used to show up in the old techniques for the examined problems) were not detected in our FEATFLOW results even in the worst cases, when using CN time integrator to solve IBVP2 in the case of strong coupling. For evaluation of the adopted special multigrid solver, we have chosen a well-known large-scale problem and solved IBVP1 and IBVP2 for isotropic and unstructured meshes for different values of permeability k_{0S}^{F} and observed the following: (1) the multigrid-IBVP1 solver is slightly faster than multigrid-IBVP2 in the regular meshes, (2) in case of irregular meshes and in particular for extremely low permeability multigrid-IBVP2 is superior. For the non-linear case and since we do not have yet published results for a rigorous quantitative benchmark to compare with, we first of all presented results for an analytical solution with no physical meaning but to validate our implementation by evaluating the L2- and H1-norms of the errors. We then solved a non-linear two-dimensional wave propagation IBVP where we noticed the negligible effect of convection for moderately and extremely low permeability and we also noticed that the proposed non-linear solver with fixed increment (time step) and due to the nature of the adopted special hyper-elastic model, may sometimes lead to a sudden increase in the non-linear iterations followed by stagnation of the solver in the next steps and then divergence in the later steps. To avoid such unfortunate situation, we introduced an adaptive time stepping procedure and specified the best control parameters through several numerical tests.

5.2 Future work

So far, only elastic materials have been discussed. For future works, the proposed non-linear solver needs to be compared with a full Newton solver and for more complicated material models. In addition, an extension of our implementation to 3D problems (with 27-node hexahedron elements for $\mathbf{u}_S$, $\mathbf{v}_S$, $\mathbf{v}_F$ and $\mathbf{w}$ and one node discontinuous pressure element with four degree of freedoms, p, $p_{,x}$, $p_{,y}$, $p_{,z}$) is necessary to open the avenue to more practically relevant applications.

Appendices

A

Computation of the internal energy

In the following, we show how to compute $\mathbf{K}_{\mathbf{uu}}$ and $\mathbf{h}$, defined in (2.228). Recall the Cauchy extra stress tensor $\mathbf{T}_E^S$ for the considered hyper-elastic material

$$\mathbf{T}_E^S = \frac{\mu^S}{J_S}\left(\mathbf{F}_S\mathbf{F}_S^T - \mathbf{I}\right) + \tilde{h}(J_S)\mathbf{I}, \tag{A.1}$$

where μ^S is the first lame parameter of the solid constituent, $\mathbf{F}_S$ is the solid deformation gradient and defined as follows

$$\mathbf{F}_S^{-1} = \mathbf{I} - \operatorname{grad}\mathbf{u}_S \tag{A.2}$$

and J_S is the solid Jacobian given by

$$\frac{1}{J_S} = \det\mathbf{F}_S^{-1} = 1 - \operatorname{div}\mathbf{u}_S + |\operatorname{grad}\mathbf{u}_S| \tag{A.3}$$

for the 2D problems. Additionally, the source term $\tilde{h}(J_S)$ is given by the following relation:

$$\tilde{h}(J_S) = \lambda^S\left(1 - n_{0S}^S\right)^2\left(\frac{1}{1 - n_{0S}^S} - \frac{1}{J_S - n_{0S}^S}\right). \tag{A.4}$$

Having $\mathbf{T}_E^S = \begin{pmatrix} T_{E_{11}}^S & T_{E_{12}}^S \\ T_{E_{21}}^S & T_{E_{22}}^S \end{pmatrix}$ and $\mathbf{u}_S = \begin{pmatrix} u_{S_1} \\ u_{S_2} \end{pmatrix}$, according to the previous four equations, it follows that

$$\begin{aligned}
T_{E_{11}}^S &= \mu^S\left(1 - u_{S_{2,2}}\right)u_{S_{1,1}} + \mu^S\left(J_S\,u_{S_{1,2}} + u_{S_{2,1}}\right)u_{S_{1,2}} \\
&\qquad + \mu^S\left(J_S\,u_{S_{2,2}} - 2J + 1\right)u_{S_{2,2}} + \overset{=h(J_S)}{\underbrace{\mu^S\left(J_S - 1\right) + \tilde{h}(J_S)}}, \\
T_{E_{12}}^S &= \mu^S\left(J_S - J_S\,u_{S_{1,1}}\right)u_{S_{1,2}} + \mu^S\left(J_S - J_S\,u_{S_{2,2}}\right)u_{S_{2,1}}, \\
T_{E_{21}}^S &= \mu^S\left(J_S - J_S\,u_{S_{1,1}}\right)u_{S_{1,2}} + \mu^S\left(J_S - J_S\,u_{S_{2,2}}\right)u_{S_{2,1}}, \\
T_{E_{22}}^S &= \mu^S\left(1 - u_{S_{1,1}}\right)u_{S_{2,2}} + \mu^S\left(J_S\,u_{S_{2,1}} + u_{S_{1,2}}\right)u_{S_{2,1}} \\
&\qquad + \mu^S\left(J_S\,u_{S_{1,1}} - 2J + 1\right)u_{S_{1,1}} + \overset{=h(J_S)}{\underbrace{\mu^S\left(J_S - 1\right) + \tilde{h}(J_S)}}.
\end{aligned} \tag{A.5}$$

Herein, $u_{S_{i,j}} = \frac{\partial u_{S_i}}{\partial x_j}$ are the spatial derivatives of the displacement and $\delta_{S_{i,j}} = \frac{\partial \delta S_i}{\partial x_j}$ are the spatial derivatives of the displacement test functions and they are both interpolated with the same basis function ϕ_i. We have

$$\int_{\Omega(t)} \operatorname{grad}\delta\mathbf{u}_S : \mathbf{T}_E^S \,\mathrm{d}v = \begin{pmatrix}\delta u_{S_1} & \delta u_{S_2}\end{pmatrix} \underbrace{\begin{pmatrix}\mathbf{K}_{11} & \mathbf{K}_{12}\\ \mathbf{K}_{21} & \mathbf{K}_{22}\end{pmatrix}}_{\mathbf{K_{uu}}} \begin{pmatrix}u_{S_1}\\ u_{S_2}\end{pmatrix} + \begin{pmatrix}\delta u_{S_1} & \delta u_{S_2}\end{pmatrix} \underbrace{\begin{pmatrix}h_1\\ h_2\end{pmatrix}}_{\mathbf{h}}, \tag{A.6}$$

where the element stiffness matrices are:

$$\begin{aligned}
(\mathbf{K}_{11})_{ij}^e &= \mu^S \int_{\Omega(t)} \left(1 - u_{S_{2,2}}\right)\phi_{i,1}\phi_{j,1} + \left(J_S u_{S_{1,2}} + u_{S_{2,1}}\right)\phi_{i,1}\phi_{j,2} + \left(J_S - J_S u_{S_{1,1}}\right)\phi_{i,2}\phi_{j,2}\ \mathrm{d}v,\\
(\mathbf{K}_{12})_{ij}^e &= \mu^S \int_{\Omega(t)} \left(J_S u_{S_{2,2}} - 2J_S + 1\right)\phi_{i,1}\phi_{j,2} + \left(J_S - J_S u_{S_{2,2}}\right)\phi_{i,2}\phi_{j,1}\ \mathrm{d}v,\\
(\mathbf{K}_{2,1})_{ij}^e &= \mu^S \int_{\Omega(t)} \left(J_S u_{S_{1,1}} - 2J_S + 1\right)\phi_{i,2}\phi_{j,1} + \left(J_S - J_S u_{S_{1,1}}\right)\phi_{i,1}\phi_{j,2}\ \mathrm{d}v,\\
(\mathbf{K}_{22})_{ij}^e &= \mu^S \int_{\Omega(t)} \left(1 - u_{S_{1,1}}\right)\phi_{i,2}\phi_{j,2} + \left(J_S u_{S_{2,1}} + u_{S_{1,2}}\right)\phi_{i,2}\phi_{j,1} + \left(J_S - J_S u_{S_{2,2}}\right)\phi_{i,1}\phi_{j,1}\ \mathrm{d}v
\end{aligned} \tag{A.7}$$

These element matrices are then assembled to give the global matrices $\mathbf{K}_{11}$, $\mathbf{K}_{12}$, $\mathbf{K}_{21}$ and $\mathbf{K}_{22}$. We also get extra source terms $\mathbf{h}$ due to the material model, which consists of two components:

$$\begin{aligned}
(h_1)_i^e &= \int_{\Omega(t)} \phi_{i,1}\cdot h(J_S)\ \mathrm{d}v = \mu^S \int_{\Omega(t)} \phi_{i,1}\cdot\left(J_S - 1 + \frac{\tilde{h}(J_S)}{\mu^S}\right)\ \mathrm{d}v,\\
(h_2)_i^e &= \int_{\Omega(t)} \phi_{i,2}\cdot h(J_S)\ \mathrm{d}v = \mu^S \int_{\Omega(t)} \phi_{i,2}\cdot\left(J_S - 1 + \frac{\tilde{h}(J_S)}{\mu^S}\right)\ \mathrm{d}v\,.
\end{aligned} \tag{A.8}$$

The element source terms are then assembled to give the global sources h_1 and h_2.

B
The inf-sup condition

B.0.1 The inf-sup condition (algebraic approach)

Here, we provide a superficial explanation for the case of finite dimensional problems for engineers who are not interested in deep mathematical analysis. The condition contains two terms, sup and inf. We first start with the source of sup. Recall that the magnitude (length or norm) of any vector in Euclidean space can be computed using the inner (scalar or dot) product with a unit vector $\mathbf{n}$ defined as follows

$$\mathbf{v}\cdot\mathbf{n} = ||\mathbf{n}||_E\ ||\mathbf{v}||_E \cos(\theta) = ||\mathbf{v}||_E \cos(\theta)\,, \tag{B.1}$$

where θ is the angle between $\mathbf{n}$ and $\mathbf{v}$. The above dot product yields exactly the magnitude of $\mathbf{v}$ if $\cos(\theta)$ is set to its maximum value 1. Hence, applying sup on both sides of (B.1), we obtain

$$||\mathbf{v}||_E = \sup_{\mathbf{n}=1} \mathbf{n}^T\mathbf{v} \tag{B.2}$$

or setting $\mathbf{n} = \frac{\mathbf{x}}{||\mathbf{x}||_E}$, the above equation can be written as

$$\begin{aligned}
||\mathbf{v}||_E &= \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{v}}{||\mathbf{x}||} \quad \text{or}\\
||\mathbf{v}||_E &= \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}\cdot\mathbf{v}}{||\mathbf{x}||} \quad \text{or}\\
||\mathbf{v}||_E &= \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\langle\mathbf{x},\mathbf{v}\rangle_E}{||\mathbf{x}||_E}\,,
\end{aligned} \tag{B.3}$$

so the supremum is required to evaluate the magnitude of the vector. What about the infimum? To answer this question, remember that the linear system

$$\mathbf{A}\mathbf{y} = \mathbf{b} \tag{B.4}$$

has a unique solution $\mathbf{y}$ if

$$\mathbf{A}\mathbf{y}_0 = \mathbf{0} \quad \Rightarrow \quad \mathbf{y}_0 = \mathbf{0}\,. \tag{B.5}$$

The above condition is commonly referred to as injectivity or one-to-one correspondence and holds if the linear operator $\mathbf{A}$ is bounding. Namely, if

$$||\mathbf{A}\mathbf{y}|| \geq c^+\,||\mathbf{y}|| \quad \text{and} \quad c^+ > 0\,, \tag{B.6}$$

since then

$$||\mathbf{A}\mathbf{y}|| = 0 \Rightarrow ||\mathbf{y}|| = 0 \quad \Leftrightarrow \mathbf{A}\mathbf{y} = \mathbf{0} \Rightarrow \mathbf{y} = \mathbf{0}\,. \tag{B.7}$$

Condition B.6 is indeed stronger than what we need, but it is necessary for the stability of the problem. If we ignore the trivial case $\mathbf{y} = \mathbf{0}$, then we can write B.6 as

$$\frac{||\mathbf{A}\mathbf{y}||}{||\mathbf{y}||} \geq c^+, \quad \text{where} \quad \mathbf{y} \neq \mathbf{0}\,. \tag{B.8}$$

B.8 is automatically satisfied if the infimum value of the left hand side is found to be greater than c^+. Therefore, B.8 is equivalent to

$$\inf_{\mathbf{y}\neq\mathbf{0}} \frac{||\mathbf{A}\mathbf{y}||}{||\mathbf{y}||} \geq c^+\,. \tag{B.9}$$

Next, using B.3 to evaluate the norm [1] $||\mathbf{A}\mathbf{y}||$, the above uniqueness condition reads

$$\inf_{\mathbf{y}\neq\mathbf{0}} \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||} \geq c^+\,. \tag{B.10}$$

So far we have only shown a sufficient condition for the uniqueness (injectivity) represented by B.5-B.10. For the existence (surjectivity or onto), we differentiate between two types of real matrices: (1) $\mathbf{A}$ is a square matrix, (2) $\mathbf{B}$ is not a square matrix. For the square $\mathbf{A}$, the injectivity and surjectivity are equivalent [2] and therefore both are met by B.10. For non-square $\mathbf{B}$, the surjectivity is equivalent to injectivity of $\mathbf{B}^T$ [3] and obtained by setting $\mathbf{A} = \mathbf{B}^T$ in B.10:

$$\inf_{\mathbf{p}\neq\mathbf{0}} \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{B}^T\mathbf{p}}{||\mathbf{x}||\,||\mathbf{p}||} \geq c^+\,. \tag{B.11}$$

[1] if we set $\mathbf{v} = \mathbf{A}\mathbf{y}$, then relation B.3 becomes a special case of the dual norm of the linear bounded functional $g(\mathbf{x}) = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle_E$, which is only equal to the euclidean norm $||\mathbf{A}\mathbf{y}||_E$, if the supremum is taken over $\mathbb{R}^n$. However, one may not be interested in whole $\mathbb{R}^n$ and the supremum will then be taken over a subset of $\mathbb{R}^n$. In bothe cases, $\mathbf{x} \neq \mathbf{0}$.

[2] The proof is extremely simple. See proof of corollary 3.1.3 in [6]

[3] See chapter 3 in [6] or read Banach closed range Theorem in page 214 of the mentioned reference.

Now we are ready to discuss our saddle point problem. The objective is to derive conditions for stability (and hence solvability) of the following problem:

$$\mathbf{Au} + \mathbf{B}^T\mathbf{p} = \mathbf{f}\,, \tag{B.12a}$$

$$\mathbf{Bu} = \mathbf{0}\,. \tag{B.12b}$$

Herein, **A** is square matrix and **B** is a rectangular matrix. The foregoing equations can be written in the following matrix-vector form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix}. \tag{B.13}$$

The above big matrix is obviously square matrix and therefore, it is sufficient for the uniqueness and existence to show that this big matrix is injective. To this end, we shall consider the homogeneous problem and derive conditions that ensure vanishing solutions. Namely, as stated before, to ensure the injectivity of the big matrix, we shall set $\mathbf{f} = \mathbf{0}$ and work on finding the conditions that yield $\mathbf{u} = \mathbf{0}$ and $\mathbf{p} = \mathbf{0}$. Our problem then reads

$$\mathbf{Au} + \mathbf{B}^T\mathbf{p} = \mathbf{0}\,, \tag{B.14a}$$

$$\mathbf{Bu} = \mathbf{0}\,. \tag{B.14b}$$

In particular, concerning **u** and due to (B.14b), we are only interested in those **u**'s that belong to $\ker(\mathbf{B})$. To make the problem easier, let us eliminate **p** and then find a condition for **u**. To eliminate **p**, we suitably scale every equation in (B.14a) and then add them up. The right multiplication (scaling) factors can be those components of any arbitrary $\mathbf{u}_0 \in \ker(\mathbf{B})$ and the foregoing scaling and additions can be expressed concisely (in compact form) as multiplication of $\frac{\mathbf{u}_0^T}{\|\mathbf{u}_0\|}$ with (B.14a) which results in

$$\frac{\mathbf{u}_0^T\mathbf{Au}}{\|\mathbf{u}_0\|} = 0 \quad \text{where} \quad \mathbf{u}_0 \in \ker(\mathbf{B})/\{\mathbf{0}\}\,. \tag{B.15}$$

Observe that the result of multiplication $\mathbf{u}_0^T\mathbf{Bp}$ is a scalar (1×1 matrix) and since the scalar is equal to its transpose, we have

$$\mathbf{u}_0^T\mathbf{B}^T\mathbf{p} = \mathbf{p}^T\underbrace{\mathbf{Bu}_0}_{=0,\text{ cf. B.12b}} = 0\,. \tag{B.16}$$

(B.15) is valid for any arbitrary $\mathbf{u}_0$ chosen from space $\ker(\mathbf{B})$. Hence, taking the supremum of (B.15), yields

$$\sup_{\mathbf{u}_0 \in \ker(\mathbf{B})} \frac{\mathbf{u}_0^T\mathbf{Au}}{\|\mathbf{u}_0\|} = \|\mathbf{Au}\| = 0 \quad \forall \mathbf{u} \in \ker(\mathbf{B}) \tag{B.17}$$

and we want to have $\mathbf{u} = \mathbf{0}$. This is exactly the condition B.7 which is satisfied by B.10. Hence, the condition

$$\inf_{\mathbf{u}\in\ker(\mathbf{B})/\{\mathbf{0}\}} \sup_{\mathbf{u}_0\in\ker(\mathbf{B})/\{\mathbf{0}\}} \frac{\mathbf{u}_0^T\mathbf{A}\mathbf{u}}{||\mathbf{u}_0||\,||\mathbf{u}||} \geq c^+ . \tag{B.18}$$

is what we want. If the above condition is satisfied, then $\mathbf{u} = \mathbf{0}$ and B.14a becomes

$$\mathbf{B}^T\mathbf{p} = \mathbf{0} \tag{B.19}$$

and we seek to have $\mathbf{p} = \mathbf{0}$. This is again the injectivity requirement on $\mathbf{B}^T$. As stated in B.11, we obtain

$$\inf_{\mathbf{p}\neq\mathbf{0}} \sup_{\mathbf{u}\neq\mathbf{0}} \frac{\mathbf{u}^T\mathbf{B}^T\mathbf{p}}{||\mathbf{u}||\,||\mathbf{p}||} \geq c^+ . \tag{B.20}$$

In the course of our derivations, we have picked $\mathbf{u}_0 \in \ker(\mathbf{B})$ (that is $\mathbf{u}_0$ which satisfies $\mathbf{B}\mathbf{u}_0 = \mathbf{0}$). How can we know that there 'exists ' such $\mathbf{u}_0$? in fact, the existence of a solution to $\mathbf{B}\mathbf{u}_0 = \mathbf{0}$ means that $\mathbf{B}$ must be surjective (or $\mathbf{B}^T$ is injective) which is fortunately guaranteed by B.20.

If we assume $\mathbf{u} = (\mathbf{u}_S, \mathbf{v}_S, \mathbf{w})$ and $\mathbf{B}^T$ as defined in 3.72 and 3.74, then the above condition is equivalent to

$$\inf_{\mathbf{p}\neq\mathbf{0}} \sup_{(\mathbf{v}_S,\mathbf{w})\neq\mathbf{0}} \frac{\mathbf{v}_S^T\mathbf{B}_S^T\mathbf{p} + \mathbf{w}^T\mathbf{B}_W^T\mathbf{p}}{(\|\mathbf{v}_S\| + \|\mathbf{w}\|)\,\|\mathbf{p}\|} \geq c^+ . \tag{B.21}$$

Notice the set $(\mathbf{v}_S, \mathbf{w}) \neq \mathbf{0}$ is a subset of the larger set $\mathbf{u} \neq \mathbf{0}$ and the superum over the larger set is larger (or equal). Hence, if B.21 is met, then B.20 will be automatically satisfied. In addition, for B.20, the supremum for the set $\mathbf{u} \neq \mathbf{0}$ locates only in the subset $(\mathbf{v}_S, \mathbf{w}) \neq \mathbf{0}$ since otherwise, the supremum over $(\mathbf{u}_S, \mathbf{0}, \mathbf{0}) \neq \mathbf{0}$ is equal to zero.

B.0.2 The inf-sup condition (analysis)

In the previous subsection we worked with special Hilbert space, finite dimensional Euclidean space with Euclidean norms, where the linear operators $\mathbf{A}$, $\mathbf{B}$ are usually obtained after discretization of the weak form and for every mesh refinement we obtain a new $\mathbf{A}$, $\mathbf{B}$ and therefore, the inf-sup condition needs to be checked again and again. In fact, this is not the right way to do it. For studying the problem of existence and uniqueness, we have to work with the infinite dimensional problem (that is before carrying out the FE discretization). In this section we shall show how this can be done shortly. In what follows, we will adopt the results of chapter 1 and chapter 2 of [45] and extend them to our porous media problem. As sample example, we shall consider the existence and uniqueness of IBVP 2, which we repeat here for the convenience of the reader

- Balance of momentum of the binary saturated mixture:

$$\rho_0\,(\mathbf{v}_S)'_S + \rho_0^{FR}\;(\mathbf{w})'_S - \operatorname{div}\mathbf{T}_E^S - \rho_0\mathbf{b} + \operatorname{grad}p = \mathbf{0}\,, \tag{B.22}$$

- Balance of momentum of the fluid phase:

$$\rho_0^{FR}\;(\mathbf{v}_S)'_S + \frac{\rho_0^{FR}}{n_{0S}^F}\,(\mathbf{w})'_S + \frac{\gamma^{FR}}{k^F}\,\mathbf{w} - \rho_0^{FR}\,\mathbf{b} + \operatorname{grad}p = \mathbf{0}\,, \tag{B.23}$$

- Volume balance of the binary saturated mixture:

$$\operatorname{div}(\mathbf{v}_S) + \operatorname{div}(\mathbf{w}) = 0\,, \tag{B.24}$$

- Velocity-displacement relationship ($\alpha = 1$ or any number. It used to see later if it can have any influence on convergence of the solver):

$$\big((\mathbf{u}_S)'_S = \mathbf{v}_S\big)\,\alpha\,. \tag{B.25}$$

We will deal with a purely homogeneous Dirichlet problem. In this case, the integration in time will not matter whether applied directly on the above strong forms or later on the weak forms as the traction loads (Neumann boundary values) are then not existent.

Next, let the scalar product $\langle\,\boldsymbol{\cdot}\,,\,\boldsymbol{\cdot}\,\rangle$ as defined in the footnote [4], integrate in time (for example using implicit backward Euler method), then multiply the discretized (in time) (B.22), (B.23), (B.24) and (B.25) with the test functions $\delta\mathbf{v}_S$, $\delta\mathbf{w}$, δp and $\delta\mathbf{u}_S$ and integrate over Ω, which (after some partial integrations) finally yield the following weak form (where $\mathrm{p} = \Delta t\,p$)

$$\Big\langle\operatorname{grad}\delta\mathbf{v}_S\,,\,\Delta t\mathbf{T}_E^S\Big\rangle + \big\langle\delta\mathbf{v}_S\,,\,\rho_0\mathbf{v}_S + \rho^{\mathrm{FR}}\mathbf{w}\big\rangle - \langle\mathrm{p}\,,\,\operatorname{div}\delta\mathbf{v}_S\rangle = \big\langle\delta\mathbf{v}_S\,,\,\mathbf{f}_{\delta\mathbf{v}_S}\big\rangle\,, \tag{B.26a}$$

$$\left\langle\delta\mathbf{w}\,,\,\rho^{\mathrm{FR}}\mathbf{v}_S + \left(\frac{\rho^{\mathrm{FR}}}{n_{0S}^F} + \frac{\gamma^{\mathrm{FR}}}{k_{0S}^F}\Delta t\right)\mathbf{w}\right\rangle - \langle\mathrm{p}\,,\,\operatorname{div}\delta\mathbf{w}\rangle = \langle\delta\mathbf{w}\,,\,\mathbf{f}_{\delta\mathbf{w}}\rangle\,, \tag{B.26b}$$

$$\langle\delta\mathbf{u}_S\,,\,\alpha\mathbf{u}_S - \Delta t\,\alpha\mathbf{v}_S\rangle = \big\langle\delta\mathbf{u}_S\,,\,\mathbf{f}_{\delta\mathbf{u}_S}\big\rangle\,, \tag{B.26c}$$

$$\langle\delta p\,,\,\operatorname{div}(\mathbf{v}_S + \mathbf{w})\rangle = 0\,. \tag{B.26d}$$

[4] For arbitrary scalars f and g, vectors $\mathbf{f}$ and $\mathbf{g}$ and second-order tensors $\mathbf{F}$ and $\mathbf{G}$, we define the following scalar products

$$\begin{aligned}\langle\mathbf{F},\mathbf{G}\rangle &= \int_\Omega \mathbf{F}:\mathbf{G}\,\mathrm{d}v\,,\\ \langle\mathbf{f},\mathbf{g}\rangle &= \int_\Omega \mathbf{f}\cdot\mathbf{g}\,\mathrm{d}v\,,\\ \langle f,g\rangle &= \int_\Omega fg\,\mathrm{d}v\,.\end{aligned}$$

If we add up (B.26a), (B.26b) and (B.26c), we arrive at

$$a(\delta\mathbf{u}_3,\mathbf{u}_3)+b(\mathrm{p},\delta\mathbf{u}_2)=\langle\delta\mathbf{u}_3\,,\,\mathbf{f}_{\delta\mathbf{u}}\rangle\,, \tag{B.27a}$$

$$b(\mathbf{u}_2,\delta p)=0\,. \tag{B.27b}$$

where $\mathbf{u}_2=(\mathbf{v}_S,\mathbf{w})\in\mathcal{U}_2=H_0^1(\Omega)^d\times H_0^1(\Omega)^d$, $\mathbf{u}_3=(\mathbf{u}_S,\mathbf{u}_2)\in\mathcal{U}_3=H_0^1(\Omega)^d\times H_0^1(\Omega)^d\times H_0^1(\Omega)^d$ and $p\in M=L_0^2(\Omega)$ (a closed subspace of $L^2(\Omega)$ which is the orthogonal complement of $\mathbb{R}$ in $L^2(\Omega)$) and the above (linear and bi-linear) operators are expressed as

$$\begin{aligned}a(\delta\mathbf{u}_3,\mathbf{u}_3)=&\left\langle\operatorname{grad}\delta\mathbf{v}_S\,,\,\Delta t\mathbf{T}_E^S\right\rangle+\left\langle\delta\mathbf{v}_S\,,\,\rho_0\mathbf{v}_S+\rho^{\mathrm{FR}}\mathbf{w}\right\rangle\\&+\left\langle\delta\mathbf{w}\,,\,\rho^{\mathrm{FR}}\mathbf{v}_S+\left(\frac{\rho^{\mathrm{FR}}}{n_{0S}^F}+\frac{\gamma^{\mathrm{FR}}}{k_{0S}^F}\Delta t\right)\mathbf{w}\right\rangle\\&+\langle\delta\mathbf{u}_S\,,\,\alpha\mathbf{u}_S-\Delta t\alpha\mathbf{v}_S\rangle\,,\end{aligned} \tag{B.28a}$$

$$b(\mathrm{p},\delta\mathbf{u}_2)=-\langle\mathrm{p}\,,\,\operatorname{div}(\delta\mathbf{v}_S+\delta\mathbf{w})\rangle\,, \tag{B.28b}$$

$$b(\mathbf{u}_2,\delta p)=-\langle\delta\mathrm{p}\,,\,\operatorname{div}(\mathbf{v}_S+\mathbf{w})\rangle\,, \tag{B.28c}$$

$$\left\langle\delta\mathbf{u}_3\,,\,\mathbf{f}_{\delta\mathbf{u}_3}\right\rangle=\left\langle\delta\mathbf{v}_S\,,\,\mathbf{f}_{\delta\mathbf{v}_S}\right\rangle+\langle\delta\mathbf{w}\,,\,\mathbf{f}_{\delta\mathbf{w}}\rangle+\left\langle\delta\mathbf{u}_S\,,\,\mathbf{f}_{\delta\mathbf{u}_S}\right\rangle\,. \tag{B.28d}$$

Obviously if (B.27) is uniquely solvable, then (B.26) will be so because any equation from (B.26) is nothing but a special case [5] of (B.27). Due to (B.27b), $\mathbf{u}_2$ must strictly belong to a closed subspace of $\mathcal{U}_2$ defined as

$$\mathcal{U}_2^\circ=\{\mathbf{u}_2\in\mathcal{U}_2\quad\text{s.t.}\quad b(\mathbf{u}_2,\delta p)=0\} \tag{B.29}$$

and consequently $\mathbf{u}_3\in\mathcal{U}_3^\circ=H_0^1(\Omega)^d\times\mathcal{U}_2^\circ$ [6] Following the mathematical analysis in the literature (see for example, chapter 4 of [6], chapters 1&2 of [45] or lecture notes by Endre Süli [7]), $\mathbf{u}_2$ will belong to $\mathcal{U}_2^\circ$ if there exists an independent positive constant c_b^+, such that

$$\inf_{\delta p\in M/\{0\}}\ \sup_{\mathbf{u}_2\in\mathcal{U}_2/\{\mathbf{0}\}}\frac{b(\mathbf{u}_2,\delta p)}{\|\mathbf{u}_2\|_{\mathcal{U}_2}\|\delta p\|_M}\geq c_b^+\,. \tag{B.31}$$

[5] If the equations are satisfied for arbitrary test functions, they will automatically be satisfied for some very special cases: For example, setting $\delta\mathbf{u}_3=(\delta\mathbf{u}_S,\delta\mathbf{v}_S,\delta\mathbf{w})=(\mathbf{0},\delta\mathbf{v}_S,\mathbf{0})$ in (B.27), we obtain (B.26a).

[6] In fact, for constant $n^S=n_0^S$ and according to (2.120), $\operatorname{div}\mathbf{v}_S=0$ and $\operatorname{div}\mathbf{v}_F=0$ and hence $\operatorname{div}\mathbf{w}=0$. Therefore, $\mathcal{U}_2^\circ$ can be made smaller and defined instead as

$$\mathcal{U}_2^\circ=\{\mathbf{u}_2\in\mathcal{U}_2\quad\text{s.t.}\quad b(\mathbf{v}_s,\delta p)=0\quad\text{and}\quad b(\mathbf{w},\delta p)=0\} \tag{B.30}$$

[7] The lectures are posted on the web and titled (A BRIEF EXCURSION INTO THE MATHEMATICAL THEORY OF MIXED FINITE ELEMENT METHODS).

If the above condition is enforced, then $\mathbf{u}_2$ has no choice but being only in $\mathcal{U}_2^\circ$ (and hence $\mathbf{u}_3 \in \mathcal{U}_3^\circ$). Now, let us assume that B.31 is met and pick $\delta\mathbf{u}_3 \in \mathcal{U}_3^\circ$, then B.27a boils down to

$$a(\delta\mathbf{u}_3, \mathbf{u}_3) = \langle \delta\mathbf{u}_3 \ , \ \mathbf{f}_{\delta\mathbf{u}} \rangle \ . \tag{B.32}$$

The above equation in a form ready for the application of the general version of the Lax-Milgram lemma (see page 7, lemma 1.2 in [45]). Therefore, we get a unique solution $\mathbf{u}_3$ if bounded bilinear operator a is such that

$$\inf_{\delta\mathbf{u}_3 \in \mathcal{U}_3^\circ/\{\mathbf{0}\}} \ \sup_{\mathbf{u}_3 \in \mathcal{U}_3^\circ/\{0\}} \frac{a(\delta\mathbf{u}_3, \mathbf{u}_3)}{\|\delta\mathbf{u}_3\|_{\mathcal{U}_3} \|\mathbf{u}_3\|_{\mathcal{U}_3}} \geq c^+_{unq} \, . \tag{B.33}$$

For uniqueness of $\mathbf{u}_3$ and for the existence, we need (as a is not symmetric) to fulfill

$$\inf_{\mathbf{u}_3 \in \mathcal{U}_3^\circ/\{\mathbf{0}\}} \ \sup_{\delta\mathbf{u}_3 \in \mathcal{U}_3^\circ/\{\mathbf{0}\}} \frac{a(\delta\mathbf{u}_3, \mathbf{u}_3)}{\|\delta\mathbf{u}_3\|_{\mathcal{U}_3} \|\mathbf{u}_3\|_{\mathcal{U}_3}} \geq c^+_{ext} \, . \tag{B.34}$$

Now by B.33 and B.34, we get a unique solution $\mathbf{u}_3$ and we can then shift a to the RHS of B.27 to obtain

$$b(\mathrm{p}, \delta\mathbf{u}_2) = \langle \delta\mathbf{u}_3 \ , \ \mathbf{f}_{\delta\mathbf{u}} \rangle \ - a(\delta\mathbf{u}_3, \mathbf{u}_3) = 0 \, . \tag{B.35}$$

Remark: Since $\mathbf{u}_3$ is fixed, the whole RHS is zero because B.32 is satisfied. Hence, the existence and uniqueness in the above special equation requires $p = 0$ which is nothing but the injectivity of the operator $\mathbb{B}^*$ induced by the above bilinear form. Since the adjoint operator $\mathbb{B}$ is surjective and has closed range (thanks to B.31), the closed range theorem tells us that $\mathbb{B}^*$ is already injective!

Using B.26d, equation B.31 can be expanded as follows:

$$\inf_{\delta p \in M/\{0\}} \ \sup_{(\mathbf{v}_S, \mathbf{w}) \in \mathcal{U}_2/\{\mathbf{0}\}} \frac{b(\mathbf{v}_S, \delta p) + b(\mathbf{w}, \delta p)}{(\|\mathbf{v}_S\|_{H^1} + \|\mathbf{w}\|_{H^1}) \, \|\delta p\|_M} \geq c^+_b \, . \tag{B.36}$$

The above equation and equations B.34 and B.33 are the three LBB stability conditions, associated with our IBVP 2.

Now, if we set $\mathbf{w} = \mathbf{0}$ or if we set $\mathbf{v}_S = \mathbf{0}$, we can easily see that the pressure element needs to be compatible with solid velocity as well as Darcy velocity (or fluid velocity), which implies that the QL option depicted in Figure 3.2 and adopted in [63] is unfortunately not LBB stable.

The elements of all the variables, that is $\mathbf{v}_F$, $\mathbf{v}_S$ and p need to be compatible. Therefore, for [63], we recommend the use of the Taylor-Hood-like element with biquadratic (Q) approximation for $\mathbf{v}_F$, $\mathbf{v}_S$ and $\mathbf{u}_S$ omitting the internal node (serendipity element) and continuous bilinear (L) approximations for p to see if the instability problem can be overcome.

Bibliography

[1] Myron B Allen. *Continuum Mechanics: The Birthplace of Mathematical Models*. John Wiley & Sons, 2015.

[2] Ralph Baierlein. *Thermal physics*. Cambridge University Press, 1999.

[3] A Bedford and Do S Drumheller. Theories of immiscible and structured mixtures. *International Journal of Engineering Science*, 21(8):863–960, 1983.

[4] A. W. Bishop. The effective stress principle. *Teknisk Ukeblad*, 39:859–863, 1959.

[5] Joachim Bluhm. *A consistent model for saturated and empty porous media*. na, 1997.

[6] Daniele Boffi, Franco Brezzi, Michel Fortin, et al. *Mixed finite element methods and applications*, volume 44. Springer, 2013.

[7] R. M. Bowen. Theory of mixtures. In A. C. Eringen, editor, *Continuum Physics*, volume III, pages 1–127. Academic Press, New York, 1976.

[8] R. M. Bowen. Incompressible porous media models by use of the theory of mixtures. *Int. J. Engng Sci.*, 18:1129–1148, 1980.

[9] R. M. Bowen. Compressible porous media models by use of the theory of mixtures. *Int. J. Engng Sci.*, 20:697–735, 1982.

[10] RM Bowen. Toward a thermodynamics and mechanics of mixtures. *Archive for Rational Mechanics and Analysis*, 24(5):370–403, 1967.

[11] Rebecca Brannon. *ROTATION: A review of useful theorems involving proper orthogonal matrices referenced to three dimensional physical space.* 2002.

[12] Rebecca Brannon. *Functional and Structured Tensor Analysis for Engineers*. 2003.

[13] Rebecca Brannon. *Kinematics: The mathematics of deformation*. 2008.

[14] S. Breuer. Quasi-static and dynamic behavior of saturated porous media with incompressible constituents. *Transp. Porous Media*, 34:285–303, 1999.

[15] Truesdell C. Sulle basi delle termomeccanica. *Rend. Lincei*, 22(196):158–166, 1957.

[16] J. Schotte C. Miehe, Schröder. Computational homogenization analysis in finite plasticity. simulation of texture development in polycrystalline materials. *Computer Methods in Applied Mechanics and Engineering*, 171:387–418, 1999.

[17] LR Calcóte. Mathematical preliminaries. *Introduction to Continuum Mechanics, D. Van Nostrand Company, Princeton, US*, 1968.

[18] Franco M Capaldi. *Continuum mechanics: constitutive modeling of structural and biological materials*. Cambridge University Press, 2012.

[19] B. D. Coleman and W. Noll. The thermodynamics of elastic materials with heat conduction and viscosity. *Arch. Rational Mech. Anal.*, 13:167–178, 1963.

[20] O. Coussy. *Mechanics of Porous Continua*. Wiley, Chichester, 1995.

[21] Olivier Coussy. *Poromechanics*. John Wiley & Sons, 2004.

[22] H. Damanik, J. Hron, A. Ouazzi, and S. Turek. Monolithic Newton–multigrid solution techniques for incompressible nonlinear flow models. *International Journal for Numerical Methods in Fluids*, Volume 71, Issue 2:208–222, 2012.

[23] R. de Boer. Highlights in the historical development of porous media theory: Toward a consistent macroscopic theory. *Applied Mechanics Review*, 49:201–262, 1996.

[24] R. de Boer. *Theory of Porous Media*. Springer-Verlag, Berlin, 2000.

[25] R De Boer and AK Didwania. The effect of uplift in liquid-saturated porous solids-karl terzaghi's contributions and recent findings. *Geotechnique*, 47(2):289–298, 1997.

[26] R. de Boer and W. Ehlers. The development of the concept of effective stresses. *Acta Mech.*, 83:77–92, 1990.

[27] R. de Boer, W. Ehlers, S. Kowalski, and J. Plischka. *Porous media – a survey of different approaches*. Forschungsberichte aus dem Fachbereich Bauwesen, Heft 54, Universität-GH-Essen, 1991.

[28] R. de Boer, W. Ehlers, and Z. Liu. One-dimensional wave propagation in fluid saturated incompressible porous media. *Arch. Appl. Mech.*, 63:59–72, 1993.

[29] S. Diebels. *Mikropolare Zweiphasenmodelle: Formulierung auf der Basis der Theorie Poröser Medien*. Habilitation, Bericht Nr. II-4 aus dem Institut für Mechanik (Bauwesen), Universität Stuttgart, 2000.

[30] S. Diebels, W. Ehlers, and B. Markert. Neglect of the fluid extra stresses in volumetrically coupled solid-fluid problems. *ZAMM*, 81:521 – 522, 2001.

[31] X. Du and M. Ostoja-Starzewski. on the size of representative volume element for darcy law in random media. In *Proc. R. Soc. A*, volume 426, pages 2949–2963, 2006.

[32] W Ehlers. On thermodynamics of elasto-plastic porous media. *Archives of Mechanics*, 41(1):73–93, 1989.

[33] W. Ehlers. *Poröse Medien – ein kontinuumsmechanisches Modell auf der Basis der Mischungstheorie*. Forschungsberichte aus dem Fachbereich Bauwesen, Heft 47, Universität-GH-Essen, 1989.

[34] W. Ehlers. Toward finite theories of liquid-saturated elasto-plastic porous media. *Int. J. Plast.*, 7:443–475, 1991.

[35] W. Ehlers. Compressible, incompressible and hybrid two-phase models in porous media theories. In Y. C. Angel, editor, *Anisotropy and Inhomogeneity in Elasticity and Plasticity*, AMD-Vol. 158, pages 25–38. The American Society of Mechanical Engineers, New York, 1993.

[36] W. Ehlers. Constitutive equations for granular materials in geomechanical context. In K. Hutter, editor, *Continuum Mechanics in Environmental Sciences and Geophysics*, CISM Courses and Lectures No. 337, pages 313–402. Springer-Verlag, Wien, 1993.

[37] W. Ehlers. Foundations of multiphasic and porous materials. In W. Ehlers and J. Bluhm, editors, *Porous Media: Theory, Experiments and Numerical Applications*, pages 3–86. Springer-Verlag, Berlin, 2002.

[38] Wolfgang Ehlers. Foundations of multiphasic and porous materials. In Wolfgang Ehlers and Joachim Bluhm, editors, *Porous Media: Theory, Experiments and Numerical Applications*, Contemporary Mathematics, pages 3–86. Springer-Verlag Berlin Heidelberg, 2002.

[39] Wolfgang Ehlers and Gernot Eipper. Finite elastic deformations in liquid-saturated and empty porous solids. *Transport in Porous Media*, 34(1–3):179–191, 1999.

[40] G. Eipper. *Theorie und Numerik finiter elastischer Deformationen in fluidgesättigten porösen Medien*. Dissertation, Bericht Nr. II-1 aus dem Institut für Mechanik (Bauwesen), Universität Stuttgart, 1998.

[41] A Cemal Eringen. *Nonlinear theory of continuous media*. McGraw-Hill, 1962.

[42] PJ Flory. Thermodynamic relations for high elastic materials. *Transactions of the Faraday Society*, 57:829–838, 1961.

[43] Daniel Frederick and Tien Sun Chang. *Continuum mechanics*. Allyn and Bacon Boston, Massachusetts, 1965.

[44] YC Fung. A first course in continuum mechanicsprentice-hall. *Englewood Cliffs, NJ*, 1969.

[45] Gabriel N Gatica. A simple introduction to the mixed finite element method. *Theory and Applications, Springer-Verlag, Berlin*, 2014.

[46] MA Goodman and SC Cowin. A continuum theory for granular materials. *Archive for Rational Mechanics and Analysis*, 44(4):249–266, 1972.

[47] S. M. Hassanizadeh and W. G. Gray. General conservation equations for multi-phase systems: 1. Averaging procedure. *Adv. Water Resources*, 2:131–144, 1979.

[48] S. M. Hassanizadeh and W. G. Gray. General conservation equations for multi-phase-systems: 2. mass, momenta, energy and entropy equations. *Adv. Water Resources*, 2:191–203, 1979.

[49] Y. Heider, O. Avci, B. Markert, and W. Ehlers. The dynamic response of fluid-saturated porous materials with application to seismically induced soil liquefaction. *Soil Dyn Earthq Eng*, 63:120–137, 2014.

[50] Y. Heider, B. Markert, and W. Ehlers. Dynamic wave propagation in infinite saturated porous media half spaces. *Computational Mechanics*, 49:319–336, 2012.

[51] G. A. Holzapfel. *Nonlinear solid mechanics: A continuum approach for engineering.* John Wiley and Sons, Chichester, UK, 2000.

[52] Sam Kennerly. A graphical derivation of the legendre transform. 2011.

[53] D.K. Kim and C.B. Yun. Time-domain soil-structure interaction analysis in two-dimensional medium based on analytical frequency-dependent infinite elements. *Int. J. Numer. Methods Eng.*, 47:1241 – 1261, 2000.

[54] Ragnar Larsson. Continuum mechanics of two-phase porous media. *Lecture notes*, 2006.

[55] DC Leigh. Nonlinear continuum mechanics. *McGrawHill, New York*, 1968.

[56] I-S. Liu. Method of Lagrange multipliers for exploitation of the entropy principle. *Arch. Rational Mech. Anal.*, 46:131–148, 1972.

[57] I-S. Liu and I. Müller. Thermodynamics of mixtures of fluids. In C. Truesdell, editor, *Rational Thermodynamics*, pages 264–285. Springer-Verlag, New York, 2nd edition, 1984.

[58] Lawrence E Malvern. *Introduction to the Mechanics of a Continuous Medium.* Number Monograph. 1969.

[59] B. Markert. *Porous Media Viscoelasticity with Application to Polymeric Foams*. Dissertation, Report No. II-12 of the Institute of Applied Mechanics (CE), Universität Stuttgart, Germany, 2005.

[60] B. Markert. A constitutive approach to 3-d nonlinear fluid flow through finite deformable porous continua. *Transp. Porous Med.*, 70:427–450, 2007.

[61] B. Markert. A biphasic continuum approach for viscoelastic high-porosity foams: Comprehensive theory, numerics, and application. *Arch. Comput. Methods Eng.*, 15:371–446, 2008.

[62] B. Markert. A survey of selected coupled multifield problems in computational mechanics. *J Coupled Syst Multiscale Dyn*, 27:22–48, 2013.

[63] B. Markert, Y. Heider, and W. Ehlers. Comparison of monolithic and splitting solution schemes for dynamic porous media problem. *Int. J. Numer. Meth. Eng.*, 82:1341–1383, 2010.

[64] N Mills. Incompressible mixtures of newtonian fluids. *International Journal of Engineering Science*, 4(2):97–112, 1966.

[65] M. Mooney. A theory of large elastic deformation. *J. Appl. Phys.*, 11:582–592, 1940.

[66] LW Morland. A simple constitutive theory for a fluid-saturated porous solid. *Journal of Geophysical Research*, 77(5):890–900, 1972.

[67] Ingo Müller. A new approach to thermodynamics of simple mixtures. *Zeitschrift für Naturforschung A*, 28(11):1801–1813, 1973.

[68] W. Noll. On the continuity of the solid and fluid states. *J. Rational Mech. Anal.*, 4:3–81, 1955.

[69] W. Noll. A mathematical theory of the mechanical behavior of continuous media. *Arch. Rational Mech. Anal.*, 2:197–226, 1958.

[70] JW Nunziato and SL Passman. Multiphase-mixture theory for fluid-saturated granular materials. In *Presented at Intern. Symp. on Mech. Behavior of Structured Media, Ottawa, 18 May 1981*, volume 1, 1981.

[71] Abdulrahman Obaid, Stefan Turek, Yousef Heider, and Bernd Markert. A new monolithic newton-multigrid-based fem solution scheme for large strain dynamic poroelasticity problems. *International Journal for Numerical Methods in Engineering*, 2016.

[72] R. W. Ogden. Large deformation isotropic elasticity – On the correlation of theory and experiment for compressible rubberlike solids. In *Proc. R. Soc. Lond. A*, volume 328, pages 323–338. 1972.

[73] R. W. Ogden. Large deformation isotropic elasticity – On the correlation of theory and experiment for incompressible rubberlike solids. In *Proc. R. Soc. Lond. A*, volume 326, pages 565–584. 1972.

[74] R. W. Ogden. Elastic deformations of rubberlike solids. In H. G. Hopkins and M. J. Sewell, editors, *Mechanics of Solids*, pages 499–537. Pergamon Press, 1982. The Rodney Hill 60th anniversary volume.

[75] R. W. Ogden. *Nonlinear elastic deformations*. Ellis Harwood Ltd., Chichester, 1984.

[76] SL Passman. Mixtures of granular materials. *International Journal of Engineering Science*, 15(2):117–129, 1977.

[77] Stephen L Passman, Jace W Nunziato, and Edward K Walsh. *A theory of multiphase mixtures*. Springer, 1984.

[78] J. Plischka. *Die Bedeutung der Durchschnittsbildungstheorie für dieTheorie poröser Medien*. PhD thesis, Universität-GH-Essen, Fachbereich Bauwesen, 1992.

[79] William Prager. *Introduction to mechanics of continua*. Courier Corporation, 1961.

[80] R. S. Rivlin. Large elastic deformations of isotropic materials. *Proc. Royal Soc. Lond. A*, 241:379–397, 1948.

[81] RS Rivlin. A note on the torsion of an incompressible highly-elastic cylinder. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 485–487. Cambridge Univ Press, 1949.

[82] Y. Saad and M.H. Schultz. A generalized minimal residual method for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7:856–869, 1986.

[83] J. Schröder. *Homogenisierungsmethoden der nichtlinearen Kontinuumsmechanik unter Beachtung von Stabilitätsproblemen*. Habilitation, Bericht Nr. I-7 aus dem Institut für Mechanik (Bauwesen), Universität Stuttgart, 2000.

[84] Louis Albert Scipio. *Principles of Continua with Applications*. John Wiley & Sons Inc, 1966.

[85] Leonid Ivanovich Sedov, Leonid Ivanovič Sedov, Mathematician Physicist, Soviet Union, Leonid Ivanovič Sedov, and Union Soviétique. *Introduction to the Mechanics of a Continuous Medium*. Addison-Wesley Reading, 1965.

[86] BR Seth. Generalized strain measure with applications to physical problems. Technical report, DTIC Document, 1961.

[87] M Scott Shell. *Thermodynamics and Statistical Mechanics: An Integrated Approach*. Cambridge University Press, 2015.

[88] A. W. Skempton. Significance of Terzaghi's concept of effective stress (Terzaghi's discovery of effective stress). In L. Bjerrum, A. Casagrande, R. B. Peck, and A. W. Skempton, editors, *From Theory to Practice in Soil Mechanics*, pages 42–53. Wiley, New York, 1960.

[89] Gilbert Strang. *Computational Science and Engineering*. Wellesley-Cambridge press, London, 2007.

[90] P. M. Suquet. Elements of homogenization for inelastic solid mechanics. In E. Sanchez-Palencia and A. Zaoui, editors, *Homogenization Techniques for Composite Media, Lecture Notes in Physics*, pages 193–277. Springer-Verlag, Berlin, 1987.

[91] K. Terzaghi. Die Berechnung der Durchlässigkeitsziffer des Tones aus dem Verlauf der hydrodynamischen Spannungserscheinungen. *Sitzungsber. Akad. Wiss. Wien, Math.-Naturwiss. Kl., Abt. II a*, 132:125–138, 1923.

[92] C. Truesdell. *A new definition of a fluid, II. The Maxwellian fluid*. U.S. Naval Res. Lab. Rep. No. P-3553, § 19, 1949.

[93] C. Truesdell. Thermodynamics of diffusion. In C. Truesdell, editor, *Rational Thermodynamics*, pages 81–98. McGraw-Hill, New York, 1st edition, 1969.

[94] C Truesdell. *Rational Mechanics*. Academic Press, New York, 1983.

[95] C. Truesdell. Thermodynamics of diffusion. In C. Truesdell, editor, *Rational Thermodynamics*, pages 219–236. Springer-Verlag, New York, 2nd edition, 1984.

[96] C. Truesdell and W. Noll. The non-linear field theories of mechanics. In S. Flügge, editor, *Handbuch der Physik*, volume III/3. Springer-Verlag, Berlin, 1965.

[97] C. Truesdell and W. Noll. In S. S. Antman, editor, *The Non-linear Field Theories of Mechanics*. Springer, Berlin, 3rd edition, 2004.

[98] C. Truesdell and R. A. Toupin. The classical field theories. In S. Flügge, editor, *Handbuch der Physik*, volume III/1, pages 226–902. Springer-Verlag, Berlin, 1960.

[99] Clifford Truesdell. A program toward rediscovering the rational mechanics of the age of reason. *Archive for history of exact sciences*, 1(1):1–36, 1960.

[100] Clifford Truesdell. *The principles of continuum mechanics*. Number 5. Field Research Laboratory, Socony Mobil Oil Company, 1961.

[101] Clifford Truesdell. *Rational thermodynamics*. Springer Science & Business Media, 2012.

[102] Clifford Truesdell and Richard Toupin. *The classical field theories*. Springer, 1960.

[103] Clifford Ambrose Truesdell. *Six lectures on modern natural philosophy*. Springer-Verlag, 2013.

[104] S. Turek. *Efficient Solvers for Incompressible Flow Problems: An Algorithmic and Computational Approach*. Springer-Verlag, 1999.

[105] S. Turek, J. Hron, M. Madlik, M. Razzaq, H. Wobker, and J. Acker. Numerical simulation and benchmarking of a monolithic multigrid solver for fluid–structure interaction problems with application to hemodynamics. In H. Bungartz, M. Mehl, and M. Schäfer, editors, *Fluid-Structure Interaction II: Modelling, Simulation, Optimisation*. Springer, 2010. doi 10.1007/978-3-642-14206-2.

[106] S. Turek, A. Obaid, and B. Markert. On a fully implicit, monolithic finite element method–multigrid solution approach for dynamic porous media problems. *Journal of Coupled Systems and Multiscale Dynamics*, 2013.

[107] H. Van der Vorst. Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 13:631–644, 1992.

[108] S. P. Vanka. Block-implicit multigrid solution of navier-stokes equations in primitive variables. *Journal of Computational Physics*, 65:138–158, 1986.

[109] O. von Estorff. Dynamic response of elastic blocks by time domain BEM and FEM. *Computers & Structures*, 38:289–300, 1991.

[110] O. von Estorff and M. Firuziaan. Coupled BEM/FEM approach for nonlinear soil/structure interaction. *Eng. Anal. Boundary Elem.*, 24:715–725, 2000.

[111] K. Wilmański and B. Albers. Acoustic waves in porous solid-fluid mixtures. In K. Hutter and N. Kirchner, editors, *Dynamic response of granular and porous materials under large and catastrophic deformations*, pages 285–313. Springer, Berlin, 2003.

[112] H. Wobker and S. Turek. Numerical studies of Vanka–type smoothers in computational solid mechanics. *Advances in Applied Mathematics and Mechanics*, 1(1):29–55, 2009.

[113] Royce KP Zia, Edward F Redish, and Susan R McKay. Making sense of the legendre transform. *American Journal of Physics*, 77(7):614–622, 2009.