POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione
Corso di Laurea Magistrale in Ingegneria Matematica

Poroelastic Computational Modeling of Biological Tissues

Application to the Mechanics of the Eye

Relatori: Prof. Riccardo SACCO
Dott.ssa Paola CAUSIN

Tesi di Laurea di:
Samuele TERRAGNI
Matr. 770420

Anno Accademico 2012-2013

Alla mia famiglia

Contents

List of Figures

List of Tables

Abstract

This thesis deals with the application of the theory of poroelasticity to biological tissues formed by an elastic porous medium and a fluid saturating the porous space. A relaxed fixed point iteration algorithm is proposed for the numerical solution of poroelastic problems. This algorithm, based on the finite element method for space discretization and the backward Euler method for time discretization, is implemented in MATLAB in both 1D and 2D cases and is analysed to investigate its convergence properties. Several numerical simulations are run in order to validate the code. An application to ocular biomechanics and hemodynamics is presented, describing a poroelastic model for the lamina cribrosa, a collagen structure in the optic nerve head which is of great interest for the study of several ocular diseases, such as glaucoma. Numerical results reveal that high values of intraocular pressure, which are the main risk factor for glaucoma, lead to high stresses and strains and to a decrease of porosity in the boundary regions of the lamina cribrosa, where the glaucomatous damage is known to start.

Introduzione

Uno degli aspetti più affascinanti della modellistica matematica consiste nella possibilità di descrivere fenomeni fisici apparentemente molto diversi tra loro attraverso le medesime equazioni. Un esempio di tale fatto è costituito dalla teoria della poroelasticità, nata intorno alla metà del secolo scorso grazie ai contributi di Biot e di altri studiosi nel contesto della geomeccanica, ma oggi ampiamente utilizzata in ambiti molto diversi da quello di origine, come, ad esempio, la descrizione di tessuti biologici. La teoria della poroelasticità descrive il comportamento di materiali porosi, formati nel caso più semplice da una componente solida elastica e da una componente fluida che occupa lo spazio lasciato dai pori. Queste componenti interagiscono tra di loro, dando luogo ad un problema meccanico e ad un problema fluidodinamico fortemente accoppiati. In virtù della sua generalità, la poroelasticità si presta ad essere applicata nei settori più svariati, dallo studio dei suoli e dei bacini petroliferi a quello della cartilagine articolare e dei tessuti nell'occhio umano.

Il nostro interesse nei confronti della poroelasticità è motivato dalle potenzialità espresse da tale disciplina nelle applicazioni biomatematiche. In modo particolare, proponiamo un modello poroelastico per lo studio dalla lamina cribrosa (LC), una struttura di collagene situata nella testa del nervo ottico e permeata da una rete di capillari. Da questo punto di vista, la tesi si inserisce in un progetto di ricerca più ampio, condotto dal Department of Mathematical Sciences della IUPUI (Indiana University and Purdue University at Indianapolis), in collaborazione con il Dipartimento di Matematica del Politecnico di Milano e il Dipartimento di Matematica dell'Università degli Studi di Milano. Tale progetto, che coinvolge oftalmologi e matematici, tra cui citiamo Giovanna Guidoboni della IUPUI e il suo gruppo di dottorandi, è volto allo studio dei meccanismi fisici che hanno luogo all'interno dell'occhio umano, al fine di investigare le possibili cause della patologia del glaucoma. Questa malattia è la seconda causa di cecità nel mondo ed è associata al danneggiamento degli assoni delle cellule gangliari retiniche, che trasportano il segnale visivo dalla retina al cervello, passando attraverso le fibre del nervo ottico. Un'ipotesi comune associa il danno assonale alle deformazioni subite dalla lamina cribrosa a causa di elevati valori di pressione intraoculare (IOP), che costituiscono il principale fattore di rischio per il glaucoma. Tuttavia, i meccanismi attraverso cui tale danno si manifesta non sono chiari.

In questo lavoro, ci proponiamo i seguenti obiettivi:

- studiare la teoria della poroelasticità, dalle ipotesi alla derivazione delle equazioni;
- analizzare i possibili trattamenti numerici delle equazioni poroelastiche e proporre uno schema iterativo per la loro risoluzione;
- implementare lo schema iterativo in MATLAB, scrivendo sia un codice 1D sia un codice 2D;
- applicare un modello poroelastico alla lamina cribrosa e simularne il comportamento usando il codice implementato.

La tesi risulta suddivisa in due parti, la prima di contenuto modellistico-numerico e la seconda di carattere più applicativo:

- la Parte I copre i primi tre obiettivi indicati in precedenza, comprendendo un'ampia descrizione della teoria della poroelasticità, presentando uno schema iterativo per la risoluzione numerica delle equazioni poroelastiche e mostrando una prima serie di risultati numerici, ottenuti con il nostro codice;
- la Parte II affronta l'ultimo obiettivo, proponendo un modello poroelastico per la lamina cribrosa e simulandone il comportamento.

Entrando maggiormente nel dettaglio, la Parte I comprende i primi tre capitoli della tesi. Nel capitolo 1, presentiamo la teoria della poroelasticità, limitandoci al caso lineare. Dopo aver analizzato le ipotesi alla base della teoria, deriviamo le equazioni per un materiale poroelastico lineare. Combinando insieme le equazioni costitutive per le componenti solida e fluida, le equazioni di bilancio del momento e della massa e l'equazione di Darcy per il flusso in un mezzo poroso, giungiamo ad un sistema di equazioni alle derivate parziali in due incognite, corrispondenti allo spostamento solido e alla pressione fluida. Particolare attenzione è posta a sottolineare la natura accoppiata dei fenomeni meccanici e fluidodinamici e a precisare il significato del concetto di incomprimibilità in un materiale poroelastico.

Nel capitolo 2, dopo aver distinto tra solutori monolitici e iterativi, proponiamo uno schema numerico iterativo per la risoluzione delle equazioni poroelastiche lineari. Ad ogni istante temporale, lo schema proposto risolve la parte meccanica e la parte fluida del problema poroelastico sequenzialmente e iterativamente, fino a raggiungere convergenza. La tecnica del rilassamento è utilizzata per favorire la convergenza del metodo. Tale schema numerico, basato sul metodo degli elementi finiti per la discretizzazione in spazio e sul metodo di Eulero implicito per la discretizzazione in tempo, è implementato in MATLAB sia nel caso 1D sia nel caso 2D. Nell'ultima

parte del capitolo, presentiamo un'analisi dello schema proposto, interpretandolo come metodo di punto fisso e studiando la dipendenza del grado di rilassamento dai parametri del modello fisico e della simulazione numerica.

Nel capitolo 3, validiamo il nostro codice, effettuando innanzitutto una analisi di convergenza su casi test con soluzione analitica nota. Mostriamo poi una prima applicazione biologica, descrivendo un modello poroelastico per la cartilagine articolare. Questo modello viene utilizzato per confermare l'analisi teorica dello schema iterativo presentata nel capitolo 2 e per trattare il problema del locking nella poroelasticità lineare.

Passando alla Parte II, essa comprende gli ultimi due capitoli della tesi. Nel capitolo 4, diamo alcune informazioni introduttive alla biomeccanica e all'emodinamica dell'occhio umano. Dopo aver descritto alcuni aspetti anatomici rilevanti per l'applicazione alla lamina cribrosa, facciamo una breve panoramica sulle conoscenze e le ipotesi relative all'eziologia del glaucoma e illustriamo le potenzialità della modellistica matematica e del calcolo scientifico in questo ambito.

Infine, nel capitolo 5, descriviamo un modello poroelastico originale per la lamina cribrosa, in cui la componente solida è rappresentata dalla matrice di collagene e quella fluida dal sangue che scorre nei capillari. La nonlinearità osservata nelle curve stress-strain sperimentali è riprodotta facendo variare i parametri elastici in funzione dello stato di sforzo, attraverso una relazione trilineare. Dai risultati delle simulazioni emerge come le regioni vicine al bordo della lamina cribrosa siano soggette a sforzi maggiori e ad una potenziale diminuzione di porosità. Questa situazione diventa più evidente al crescere della pressione intraoculare e suggerisce alcune considerazioni fisiologiche interessanti nello studio interpretativo della patologia del glaucoma.

Introduction

One of the most intriguing aspects of mathematical modeling is its ability to describe very different physical phenomena using the same equations. An example of this fact is given by the theory of poroelasticity, which was formulated by Biot and others around the middle of the last century for geomechanical applications, but is now widely used in contexts that are very different from the original one, such as the description of biological tissues. The theory of poroelasticity describes porous media which in the simplest case are formed by an elastic solid component and a fluid component saturating the porous space. These components significantly interact, thus originating a mechanical problem and a fluid dynamics problem that are strongly coupled. By virtue of its generality, poroelasticity is used in very different areas, from the study of soils and oilfields to the study of articular cartilage and ocular tissues.

Our interest in poroelasticity is in relation to biomathematical applications. In particular, we propose a poroelastic model for the lamina cribrosa (LC), a collagen structure located in the optic nerve head and pervaded by a capillary network. From this point of view, the thesis is part of a more comprehensive research project, carried out by the Department of Mathematical Sciences of IUPUI (Indiana University and Purdue University at Indianapolis), in collaboration with Dipartimento di Matematica of Politecnico di Milano and Dipartimento di Matematica of Università degli Studi di Milano. This project, which involves ophthalmologists and mathematicians, such as Giovanna Guidoboni and her Ph.D. students in IUPUI, aims at studying the physical mechanisms occurring in the human eye, in order to investigate the causes of the pathology of glaucoma. Such ocular disease is the second cause of blindness in the world and is associated with damage to the axons of the retinal ganglion cells, which transmit visual signals from the retina to the brain, through the fibres of the optic nerve. A common hypothesis associates the axonal damage with the deformations undergone by the lamina cribrosa because of elevated values of intraocular pressure (IOP), which are the main risk factor for glaucoma. However, the mechanisms leading to the axonal damage are not clear.

The objectives of this work are the following ones:

- to study the theory of poroelasticity, from the assumptions to the derivation of the equations;
- to analyse the possible numerical treatments of the poroelastic equations and to propose an iterative scheme for their solution;
- to implement the iterative scheme in MATLAB, developing both a 1D code and a 2D code;
- to apply a poroelastic model to the lamina cribrosa and to simulate its behaviour using our code.

The thesis is divided into two parts, the first one dealing with mathematical modeling and numerical analysis and the second one dealing with the physical application to which we are interested:

- Part I covers the first three objectives listed above, describing the theory of poroelasticity, presenting an iterative scheme for the numerical solution of the poroelastic equations and showing a first set of results, obtained by using our code;
- Part II deals with the last objective, proposing a poroelastic model for the lamina cribrosa and simulating its behaviour.

Part I includes the first three chapters of the thesis. In chapter 1, we present the theory of poroelasticity, limiting ourselves to the linear case. After discussing the assumptions on which the theory is based, we derive the equations for a linear poroelastic material. Gathering the constitutive equations for the solid and the fluid components, the momentum and the mass balance equations and Darcy's law describing the fluid flow in porous media, we end up with a system of partial differential equations whose unknowns are the solid displacement and the fluid pressure. Special attention is paid to the coupling between mechanical and fluid dynamics phenomena and to the meaning of incompressibility in poroelasticity.

In chapter 2, after making a distinction between monolithic and iterative solvers, we propose an iterative numerical scheme for the solution of the linear poroelastic equations. At each time level, the scheme solves the mechanical and the fluid parts of the poroelastic problem sequentially and iteratively, until convergence is reached. A relaxation technique is used to help convergence. This numerical scheme, based on the finite element method for spatial discretization and the backward Euler method for time discretization, is implemented in MATLAB both in the 1D case and in the 2D case. In the last part of the chapter, we interpret the scheme as a fixed point method and show how the degree of relaxation depends on physical model and discretization parameters.

In chapter 3, we validate our code by making a convergence analysis, using test cases endowed with analytical solutions. Then, we show a first

biological application, describing a poroelastic model for articular cartilage. We use this model to confirm the results of the theoretical analysis of the iterative scheme presented in chapter 2 and to deal with the problem of locking in linear poroelasticity.

Turning to Part II, it includes the last two chapters of the thesis. In chapter 4, we give some introductory information about ocular biomechanics and hemodynamics. After describing some anatomical details which are relevant to the application to the lamina cribrosa, we give a summary of the hypotheses about the etiology of glaucoma and illustrate the potential of mathematical modeling and scientific computing in this area.

Finally, in chapter 5, we describe an original poroelastic model for the lamina cribrosa, in which the solid component is represented by the collagen matrix and the fluid component by blood flowing in the capillary network. The material nonlinearity revealed by experimental stress-strain curves is kept into account by making the elastic parameters vary as a function of the stress state through a trilinear relation. The numerical simulations show that the regions near the edge of the lamina cribrosa are subject to higher stresses and can be affected by a decrease of porosity. This situation becomes more evident when the intraocular pressure increases and suggests some interesting physiological considerations that may be a useful starting step toward a deeper understanding of the causes of the pathology of glaucoma.

Part I

Poroelasticity: modeling and simulation

Chapter 1

Theory of linear poroelasticity

This chapter deals with the mathematical treatment of elastic porous media filled by a fluid, namely the theory of poroelasticity. After indicating some basic assumptions, we focus on the constitutive equations describing the material behaviour. Then, we present the momentum and mass balance equations, which together with the constitutive laws form the set of governing equations of the theory of poroelasticity.

1.1 General aspects and historical background

A porous medium is a material formed by several components. We consider the simplest case of a material formed by a solid and a fluid saturating the porous space. Poromechanics studies porous media whose mechanical behaviour is significantly influenced by the interstitial fluid. We talk about poroelasticity when the solid component has elastic properties, as we assume henceforth.

Poroelasticity involves a broad range of materials, from rocks and soils to gels and biological tissues. Accordingly, it is relevant to many different disciplines, such as geomechanics (see [8] and [44]) and biomechanics (see [7], [15], [41] and [45]). Despite their differences, all porous materials are subject to the same coupled processes, due to the interaction between the components. Two basic mechanisms play a key role in this interaction: firstly, an increase in fluid pressure induces a dilation of the porous medium; secondly, a compression of the porous medium causes a rise in fluid pressure, if the fluid is prevented from escaping or if compression is fast relative to the fluid flow rate.

The theory of poroelasticity was initially formulated to study the phenomenon of soil consolidation, the process which describes the gradual settlement of soil under an applied load. In 1923, a simple model to explain this phenomenon was first proposed by Terzaghi [46], who assumed that the grains forming soil are bound together by certain molecular forces and

constitute a porous material with elastic properties. The voids of this elastic skeleton are filled with water, so that a good example of such a model is a sponge saturated with water. A load applied to this system produces a gradual settlement, depending on the rate at which the water is being squeezed out. Terzaghi applied these concepts to the one-dimensional problem of a column of soil under a constant load and prevented from lateral expansion. In 1941, Biot [2] extended Terzaghi's theory, by considering the general three-dimensional case and an arbitrary load, variable with time. Biot is considered the father of poroelasticity, although this term appeared only in 1966, when it was coined by Geertsma in petroleum engineering. Essentially the same theory was reformulated several times by Biot himself and other scientists, thus offering different point of views on the description of the coupling between the solid and the fluid in porous media. It was also shown that the same equations characterizing Biot's theory can be obtained using the formalism of mixture theory. So, the theory of poroelasticity can also be addressed as two-component mixture theory or biphasic theory.

In the following, we will not make use of the formalism of mixture theory, although it is practised by some people for describing biological tissues (see, for example, the work of Frijns [15], who modeled cartilaginous tissues). Instead, we choose the formalism referable to the original theory, more suitable to explain the physical meaning of the parameters of the model and to deal with the concept of incompressibility in poroelasticity. In particular, we follow the work of Detournay and Cheng [13] for the presentation of constitutive equations and the work of Showalter [42] for momentum and mass balance equations. A more general derivation of constitutive equations, based on the principles of thermodynamics, can be found in [9].

1.2 Continuity hypothesis and basic assumptions

An elastic porous medium is composed of an elastic solid and an interconnected porous space filled by a fluid. The microscopic structure of such a material is too complicated to describe and often it is not even known in an exact manner. The continuity hypothesis remedies this problem considering the porous medium as the superposition of its components, each of them representing a fraction of the complete material at every point. By assuming that the description of each component considered separately does not differ from that of a monophasic continuous medium, the typical concepts of continuum mechanics can be applied to porous media.

The continuum approach is based on the concepts of representative elementary volume (REV) and volume fractions. Let the porous medium occupy the domain $\Omega \subset \mathbb{R}^3$ in a Cartesian coordinate system. At every point $\mathbf{x} \in \Omega$ and at every time t we define a REV, that is a material volume $V(\mathbf{x}, t)$ chosen so that it is large enough, compared to the pore size, that it may be

treated as homogeneous and, at the same time, it is small enough, compared to the typical scale of macroscopic phenomena, that it may be considered as infinitesimal in the mathematical treatment. The porous medium can be seen as the juxtaposition of these material volumes and the quantities used in the description of the material are assumed to be averaged over each REV. Denoting by $V_s(\mathbf{x}, t)$ and $V_f(\mathbf{x}, t)$ the volumes in $V(\mathbf{x}, t)$ occupied by the solid and the fluid, respectively, the quantity

$$\phi(\mathbf{x}, t) = \frac{V_f(\mathbf{x}, t)}{V(\mathbf{x}, t)}$$

is the volumetric fraction of the fluid component, also called the porosity, while the quantity

$$\frac{V_s(\mathbf{x}, t)}{V(\mathbf{x}, t)}$$

is the volumetric fraction of the solid component. We underline the dependence of the introduced quantities on space and time, so that we can say that there is a fraction of each component at every point $\mathbf{x}$ and at every time t.

Before presenting the constitutive laws and the balance equations describing the behaviour of porous media, we list the assumptions considered in our mathematical treatment:

- isotropy of the material;
- reversibility of the deformation process;
- small deformations and linearity of stress-strain relation;
- full saturation;
- isothermal processes;
- quasi-static coupled processes.

Isotropy means that the material exhibits the same physical properties in every direction, while reversibility means that no energy is dissipated during a closed loading cycle. Denoting by $\mathbf{u}$ the solid displacement, the assumption of small deformations corresponds to requiring that

$$\|\nabla \mathbf{u}\| \ll 1, \tag{1.1}$$

where $\|\cdot\|$ is the norm in a suitable function space that will be characterized more precisely in chapter 2. This condition leads to describe the deformation through the linearized strain tensor

$$\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T). \tag{1.2}$$

Furthermore, small deformations imply that the relation between stress and strain can be taken as linear in first approximation, the assumption of linearity being crucial in our treatment.

Remark 1 (From large deformations to small deformations). We want to give some more details on how the large deformation theory is reduced to the small deformation theory. At time $t = 0$, we consider an initial configuration for the porous medium, which occupies the volume Ω_0. In this configuration, a material particle is identified by its position vector $\mathbf{X}$. At time t, the porous medium has deformed due to external forces applied or changes in fluid pressure and lies in the current configuration, occupying the volume Ω_t. In this configuration, the particle whose initial position vector was $\mathbf{X}$ is now identified by the current position vector $\mathbf{x}$. Let $d\mathbf{X}$ be an infinitesimal material vector in the initial configuration, joining the particle located at $\mathbf{X}$ to the particle located at $\mathbf{X} + d\mathbf{X}$. After deformation, $d\mathbf{X}$ becomes $d\mathbf{x}$, joining the same particles in their new positions $\mathbf{x}$ and $\mathbf{x} + d\mathbf{x}$ (see figure 1.1). Introducing the deformation gradient

$$\mathbf{F} = \nabla_X \mathbf{x}, \tag{1.3}$$

where ∇_X indicates the gradient relative to the initial configuration, the material vector $d\mathbf{x}$ is obtained from $d\mathbf{X}$ according to the transport formula

$$d\mathbf{x} = \mathbf{F} d\mathbf{X}. \tag{1.4}$$

The displacement vector $\mathbf{u}(\mathbf{X}, t)$ of the particle whose initial and current

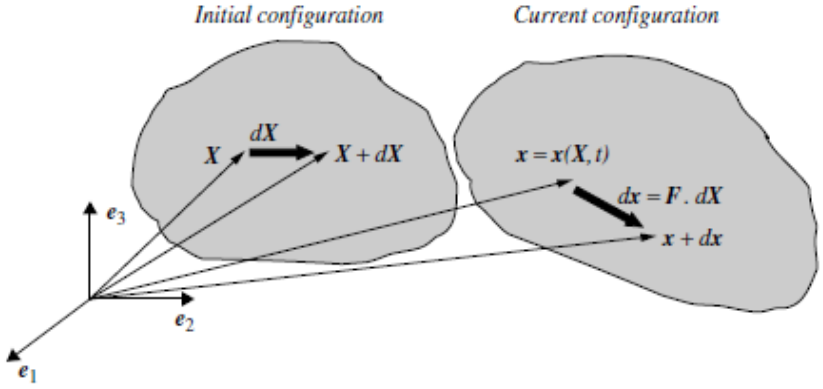


Figure 1.1: Deformation gradient $\mathbf{F}$ and transport of a material vector $d\mathbf{X}$ (taken from [9]).

positions are $\mathbf{X}$ and $\mathbf{x}$ is given by

$$\mathbf{x} = \mathbf{X} + \mathbf{u}, \tag{1.5}$$

so that we can express the deformation gradient in terms of the displacement vector, using (1.3) and (1.5):

$$\mathbf{F} = \mathbf{I} + \nabla_X \mathbf{u}, \tag{1.6}$$

where $\mathbf{I}$ is the identity tensor. Deformation induces changes in both the lengths of material vectors and the angles between them. A measure of these changes is expressed by the Green-Lagrange strain tensor $\boldsymbol{\Delta}$, which quantifies the variation of the scalar product of two material vectors $d\mathbf{X}$ and $d\mathbf{Y}$, transforming through the deformation into $d\mathbf{x}$ and $d\mathbf{y}$, according to

$$d\mathbf{x} \cdot d\mathbf{y} - d\mathbf{X} \cdot d\mathbf{Y} = 2d\mathbf{X} \cdot (\boldsymbol{\Delta} d\mathbf{Y}). \tag{1.7}$$

Substituting (1.4) into (1.7), we can express the Green-Lagrange strain tensor as a function of the deformation gradient:

$$\boldsymbol{\Delta} = \frac{1}{2}(\mathbf{F}^T\mathbf{F} - \mathbf{I}). \tag{1.8}$$

Substituting the expression (1.6) into (1.8), we write the Green-Lagrange strain tensor as a function of the displacement vector:

$$\boldsymbol{\Delta} = \frac{1}{2}(\nabla_X \mathbf{u} + \nabla_X \mathbf{u}^T + \nabla_X \mathbf{u}^T \nabla_X \mathbf{u}). \tag{1.9}$$

In the limit of infinitesimal deformations, the initial and the current configurations merge, so that $\nabla = \nabla_X = \nabla_x$. Therefore, under condition (1.1) the Green-Lagrange strain tensor $\boldsymbol{\Delta}$ reduces to the linearized strain tensor $\boldsymbol{\varepsilon}$:

$$\boldsymbol{\Delta} \simeq \boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T). \tag{1.10}$$

So, the tensor describing strains is $\boldsymbol{\Delta}$ in the large deformation theory and $\boldsymbol{\varepsilon}$ in the small deformation theory.

The assumption of full saturation means that the whole interconnected porous space is filled by the fluid and, in each REV, this fact is expressed by

$$V(\mathbf{x}, t) = V_s(\mathbf{x}, t) + V_f(\mathbf{x}, t). \tag{1.11}$$

Considering isothermal processes simply means that the temperature is assumed constant. Finally, the assumption of quasi-static coupled processes allows us to neglect inertial effects.

1.3 Constitutive equations

In this section, we present the constitutive equations describing the properties of a linear isotropic poroelastic material, by using the continuum approach. We pay particular attention to the volumetric response of such a material, in order to show the connection between continuum and micromechanical parameters. This connection is important to understand the meaning of incompressibility in the poroelastic framework.

1.3.1 Main variables and constitutive laws

In the description of a linear isotropic poroelastic material, composed by a solid and a fluid component, we make use of the following kinematic variables:

- the solid displacement $\mathbf{u}$, already introduced in the previous section, which tracks the movement of the solid with respect to an initial configuration;
- the specific discharge $\mathbf{v}$, defined as the rate of fluid volume crossing a unit area of porous medium, which describes the motion of the fluid relative to the solid.

Then, we use two variables associated with the concept of strain:

- the linearized strain tensor $\boldsymbol{\varepsilon}$, also called the small strain tensor;
- the variation of fluid content ζ, defined as the variation of fluid volume per unit volume of porous medium.

We follow the usual convention whereby strains are positive for extension and ζ is positive for a gain of fluid by the porous medium. Finally, we introduce the basic dynamic variables:

- the total stress tensor $\boldsymbol{\sigma}$, whose component σ_{ij} is the total force in the x_j direction per unit area whose normal is in the x_i direction;
- the fluid pressure p, also called the pore pressure, which in a material element is defined as the pressure in an hypothetical reservoir in equilibrium with the element (i.e. no fluid exchange takes place between the material element and the reservoir).

Positive stresses imply tensions, while the sign of p is defined in terms of excess pore pressure. Both $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$ are symmetric tensors. We note that stress and fluid pressure are the conjugate quantities of strain and variation of fluid content, respectively; in other words, the work increment associated with the increments $d\varepsilon_{ij}$ and $d\zeta$, in the presence of σ_{ij} and p, is

$$dW = \sigma_{ij} d\varepsilon_{ij} + p d\zeta. \tag{1.12}$$

Constitutive laws characterize the mechanical response of a porous material. Under the assumptions made in section 1.2, the variables $\boldsymbol{\varepsilon}$ and ζ are expressed as linear functions of the variables $\boldsymbol{\sigma}$ and p. The relation linking $\boldsymbol{\varepsilon}$ with $\boldsymbol{\sigma}$ and p is

$$\varepsilon_{ij} = \frac{1}{2G}\sigma_{ij} - \left(\frac{1}{6G} - \frac{1}{9K}\right)\sigma_{kk}\delta_{ij} + \frac{1}{3H'}p\delta_{ij}, \tag{1.13}$$

where $i, j = 1, 2, 3$ and summation over repeated indices is assumed. This relation is obtained by extending the known relation of the linear theory of

elasticity to include the effect of fluid pressure. Indeed, by taking $p = 0$ in (1.13), we get the standard Hooke's law for a linear isotropic elastic material. The effect of fluid pressure is introduced noting that, because of isotropy, p is not able to produce any shear strain and its effect must be the same on the diagonal components of strain. In (1.13), G and K are the shear modulus and the bulk modulus of the drained porous material and H' is an additional physical constant having the same units as G and K.

The most general relation linking ζ with $\boldsymbol{\sigma}$ and p takes the form

$$\zeta = a_1\sigma_{11} + a_2\sigma_{22} + a_3\sigma_{33} + a_4\sigma_{12} + a_5\sigma_{13} + a_6\sigma_{23} + a_7 p. \tag{1.14}$$

Because of isotropy, a change in sign of extra-diagonal components of the total stress does not affect the fluid content, therefore $a_4 = a_5 = a_6 = 0$. For the same reason, all directions must have equivalent properties, so that $a_1 = a_2 = a_3$. Therefore, relation (1.14) reduces to

$$\zeta = \frac{1}{3H''}\sigma_{kk} + \frac{1}{R}p, \tag{1.15}$$

where H'' and R are two physical constants.

Now, we want to show that $H' = H''$. The assumption of reversibility implies that the work increment

$$dW = \sigma_{ij}d\varepsilon_{ij} + pd\zeta = \varepsilon_{ij}d\sigma_{ij} + \zeta dp \tag{1.16}$$

is an exact differential, therefore the Euler conditions

$$\frac{\partial \varepsilon_{ij}}{\partial p} = \frac{\partial \zeta}{\partial \sigma_{ij}} \tag{1.17}$$

must be satisfied. Using relations (1.13) and (1.15) in the Euler conditions leads to $H' = H''$. Therefore, we can simply write H instead of H' or H'', so that the constitutive equations of a linear isotropic poroelastic material become:

$$\varepsilon_{ij} = \frac{1}{2G}\sigma_{ij} - \left(\frac{1}{6G} - \frac{1}{9K}\right)\sigma_{kk}\delta_{ij} + \frac{1}{3H}p\delta_{ij}, \tag{1.18}$$

$$\zeta = \frac{1}{3H}\sigma_{kk} + \frac{1}{R}p. \tag{1.19}$$

The first one describes the response of the solid component, while the second one that of the fluid, but they both present coupling terms. Four distinct physical constants are needed to characterize the response of the porous medium: two elastic constants (G and K) and two additional constants (H and R).

To interpret the constants H and R, let us consider a sample of porous material enclosed in a thin rubber bag, so that there is no stress applied. By

draining the fluid from the material through a thin tube, a negative pressure $-p_0$ is applied. From relation (1.19) we see that the change in fluid content is

$$\zeta = -\frac{p_0}{R}, \tag{1.20}$$

while the corresponding volume change is obtained from relation (1.18) and is given by

$$\varepsilon_{kk} = -\frac{p_0}{H}. \tag{1.21}$$

So, the coefficient $1/R$ is a measure of the change in fluid content due to a change in fluid pressure, while $1/H$ is a measure of the compressibility of the porous medium due to a change in fluid pressure.

1.3.2 Volumetric response

The constitutive equations shown before can be separated into a deviatoric response and a volumetric one. The study of the volumetric response allows us to explain the coupling between the solid and the fluid and to deduce the connection between continuum and micromechanical parameters. This link is important to understand the limiting behaviour of the porous medium, when one or both components are incompressible.

Introducing the total pressure P and the volumetric strain ε as the quantities defined by

$$P = -\frac{\sigma_{kk}}{3}, \tag{1.22}$$

$$\varepsilon = \varepsilon_{kk}, \tag{1.23}$$

we can express the deviatoric stress and strain, respectively, as

$$s_{ij} = \sigma_{ij} + P\delta_{ij}, \tag{1.24}$$

$$e_{ij} = \varepsilon_{ij} - \frac{\varepsilon}{3}\delta_{ij}. \tag{1.25}$$

Using (1.22) and (1.23) into (1.24) and (1.25), respectively, we immediately see that $s_{kk} = e_{kk} = 0$. Therefore, the constitutive equations (1.18) and (1.19) of a linear isotropic poroelastic material can be separated into the deviatoric response

$$e_{ij} = \frac{1}{2G}s_{ij} \tag{1.26}$$

and the volumetric one

$$\varepsilon = -\left(\frac{P}{K} - \frac{p}{H}\right), \tag{1.27}$$

$$\zeta = -\left(\frac{P}{H} - \frac{p}{R}\right). \tag{1.28}$$

It is clear that the deviatoric response is purely elastic, namely the fluid is not involved in it. On the contrary, the coupled effects associated with the constant H appear only in the volumetric response. This fact is a consequence of the assumption of isotropy and makes the volumetric response particularly interesting.

We can identify two kinds of volumetric response, corresponding to different limiting behaviours of the material:

- the drained response, characterized by $p = 0$ (the fluid can freely escape from the porous medium);
- the undrained response, characterized by $\zeta = 0$ (the fluid is trapped in the porous medium and prevented from escaping).

Under the drained condition $p = 0$, the volumetric strain is proportional to the total pressure (see (1.27)):

$$\varepsilon = -\frac{P}{K}. \tag{1.29}$$

Now it is clear why K represents the drained bulk modulus of the material. Substituting (1.29) into (1.28), we get

$$\zeta = \alpha\varepsilon, \tag{1.30}$$

where $\alpha = K/H$ is the Biot coefficient and represents the ratio of the volume of fluid gained or lost in a material element to the volume change of that element, when the fluid pressure is allowed to return to its initial value. Furthermore, α can not be larger than 1, because the volume of fluid gained or lost by an element can not be greater than the volume change of that element. So, $\alpha \in [0, 1]$.

Under the undrained condition $\zeta = 0$, from (1.28) we get

$$p = BP, \tag{1.31}$$

where $B = R/H$ is the Skempton coefficient and establishes a proportional relation between fluid pressure and total pressure, when the fluid is prevented from escaping. Substituting (1.31) into (1.27), we get

$$\varepsilon = -\frac{P}{K_u}, \tag{1.32}$$

where

$$K_u = K\left(1 + \frac{KR}{H^2 - KR}\right) \tag{1.33}$$

is the undrained bulk modulus of the material. Therefore, also in the undrained response the volumetric strain is proportional to the total pressure, but the undrained material is stiffer than the drained one ($K_u \in [K, +\infty)$).

A further interpretation of drained and undrained responses comes from the situation of a suddenly applied constant load. Just after the imposition of the load, the fluid has not had time to move between neighbouring material elements, so that we can consider $\zeta = 0$. After a sufficiently long time, the fluid pressure will equilibrate with the pressure imposed at the boundary; assuming this pressure to be equal to zero, we have $p = 0$ at equilibrium. In other words, the instantaneous response can be interpreted as undrained, while the long-term response as drained. Figure 1.2 gives a schematic representation of these concepts, describing the porous medium as an idealized system composed of a spring (the elastic solid component) and a container filled by fluid (the fluid component). The container has a

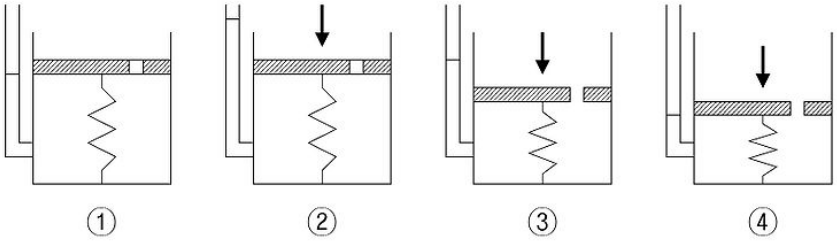


Figure 1.2: Schematic representation of drained and undrained conditions.

hole, which controls the fluid exchange with the outside. At stage 1, no load is applied and the hole is closed, so that the system is in equilibrium. At stage 2, a load is suddenly applied and the hole is still closed: only the fluid, prevented from escaping, resists the load, so that the fluid pressure rises. At stage 3, the hole is opened: the fluid starts to drain out, the fluid pressure decreases and the spring starts to shorten. Finally, at stage 4, the drainage of fluid no longer occurs: the system has reached a new equilibrium, where only the spring resists the load. In this description, stage 2 corresponds to undrained conditions, while stage 4 to drained conditions.

With little effort, we can rewrite relations (1.27) and (1.28) in terms of α, K and K_u instead of K, H and R:

$$\varepsilon = -\frac{1}{K}(P - \alpha p), \tag{1.34}$$

$$\zeta = -\frac{\alpha}{K}\left(P - \frac{p}{B}\right), \tag{1.35}$$

where

$$B = \frac{K_u - K}{\alpha K_u}. \tag{1.36}$$

We can also invert these relations, obtaining

$$P = \alpha M \zeta - K_u \varepsilon, \tag{1.37}$$

$$p = M(\zeta - \alpha \varepsilon), \tag{1.38}$$

where

$$M = \frac{K_u - K}{\alpha^2} \tag{1.39}$$

is the Biot modulus, which represents the inverse of a storage coefficient, defined as the increase of fluid content per unit material volume as a result of a unit increase of fluid pressure, when there is no volumetric strain.

1.3.3 Connection with micromechanical parameters

Now, we want to describe the volumetric response of a linear isotropic poroelastic material from a different point of view, that of a micromechanical approach. By doing so, we can establish a connection between continuum and micromechanical parameters, that is useful to deal with incompressible poroelastic materials.

The continuum approach, as widely discussed in section 1.2, describes the porous medium as a whole and continuum parameters are associated with a specific solid-fluid system. Therefore, it is not clear how these parameters are influenced by the coefficients related to the single components. The micromechanical approach looks into the micromechanics of the solid-fluid system, trying to deduce the contribution of each component and the dependence of the continuum parameters on those related to the components. It is useful to describe the micromechanical volumetric response both in terms of total pressure P and fluid pressure p and in terms of the so-called Terzaghi effective pressure $P' = P - p$ and Π-pressure $p' = p$, an alternative decomposition of the loading.

Let us consider a material volume V of the porous medium, containing an interconnected porous space of volume V_p. Denoting by V_s and V_f the volumes in V occupied by the solid and the fluid, respectively, we have that $V = V_s + V_p$ and, because of full saturation, $V_f = V_p$. The porosity is given by $\phi = V_p/V$.

The volumetric response of the porous material to the loading $\{P, p\}$ is described by the quantities $\Delta V/V$ and $\Delta V_p/V_p$. Recalling the assumption of linearity of stress-strain relationship, we can write

$$\frac{\Delta V}{V} = -\frac{1}{K}(P - \alpha p), \tag{1.40}$$

$$\frac{\Delta V_p}{V_p} = -\frac{1}{K_p}(P - \beta p). \tag{1.41}$$

Comparing (1.40) with (1.34), we note that K and α are exactly the coefficients introduced before, while K_p is the bulk modulus of the pore volumetric

strain and β is a dimensionless coefficient. By invoking the Betti-Maxwell reciprocal theorem, it can be shown that

$$K_p = \frac{\phi}{\alpha} K. \tag{1.42}$$

Considering the loading $\{P', p'\}$, we can express the volumetric response of the porous material as

$$\frac{\Delta V}{V} = -\frac{P'}{K} - \frac{p'}{K'_s}, \tag{1.43}$$

$$\frac{\Delta V_p}{V_p} = -\frac{P'}{K_p} - \frac{p'}{K''_s}, \tag{1.44}$$

where K'_s and K''_s are two bulk moduli, which in practice can be identified with a unique bulk modulus K_s for the solid component. In particular, the equivalence of the two moduli can be proved for a porous medium that is microscopically isotropic and homogeneous, but in practice K'_s and K''_s are always very similar when the porous space is interconnected. Comparing (1.43) and (1.44) with (1.40) and (1.41) shows that

$$\alpha = 1 - \frac{K}{K'_s}, \tag{1.45}$$

$$\beta = 1 - \frac{K_p}{K''_s}. \tag{1.46}$$

Now we consider the volumetric response related to the fluid component. Because of the assumption of full saturation, we have $\Delta V_f = \Delta V_p$. Moreover, the variation of fluid volume ΔV_f can be decomposed into two parts:

$$\Delta V_f = \Delta {V_f}^1 + \Delta {V_f}^2, \tag{1.47}$$

where $\Delta {V_f}^1$ is the component associated with compression or dilation of the interstitial fluid and $\Delta {V_f}^2$ is the component associated with fluid exchange between the material volume V and the outside. The component $\Delta {V_f}^1$ can be expressed in terms of fluid pressure, by introducing the bulk modulus of the fluid K_f:

$$\frac{\Delta {V_f}^1}{V_f} = -\frac{p}{K_f}. \tag{1.48}$$

The component $\Delta {V_f}^2$ is associated with the variation of fluid content ζ, so that

$$\zeta = \frac{\Delta {V_f}^2}{V} = \frac{\phi \Delta {V_f}^2}{V_f}. \tag{1.49}$$

Using (1.47), (1.44) and (1.48), relation (1.49) becomes

$$\zeta = -\frac{\phi}{K_p} \left(P - \left(1 - \frac{K_p}{K''_s} + \frac{K_p}{K_f} \right) p \right) \tag{1.50}$$

and comparing (1.50) with (1.35), recalling that relation (1.42) is valid, leads to

$$\frac{1}{B} = 1 - \frac{K_p}{K_s''} + \frac{K_p}{K_f}. \tag{1.51}$$

The correspondence between continuum and micromechanical parameters is made explicit by expressing the continuum coefficients appearing in (1.34), (1.35), (1.37) and (1.38) in terms of the micromechanical coefficients. In particular, we express α, K_u, B and M in terms of the micromechanical coefficients K_s', K_s'', and K_f, the drained bulk modulus K and the porosity ϕ. First, we remind that the Biot coefficient α is expressed by (1.45) as

$$\alpha = 1 - \frac{K}{K_s'}. \tag{1.52}$$

Then, using relations (1.51), (1.42) and (1.52), we obtain an expression for the Skempton coefficient B:

$$\begin{aligned} B &= \frac{1}{1 + \dfrac{K_p}{K_f} - \dfrac{K_p}{K_s''}} = 1 - \frac{\dfrac{K_p}{K_f} - \dfrac{K_p}{K_s''}}{1 + \dfrac{K_p}{K_f} - \dfrac{K_p}{K_s''}} = 1 - \frac{\dfrac{\phi}{\alpha} K \left(\dfrac{1}{K_f} - \dfrac{1}{K_s''} \right)}{1 + \dfrac{\phi}{\alpha} K \left(\dfrac{1}{K_f} - \dfrac{1}{K_s''} \right)} \\ &= 1 - \frac{\phi \left(\dfrac{K}{K_f} - \dfrac{K}{K_s''} \right)}{\left(1 - \dfrac{K}{K_s'} \right) + \phi \left(\dfrac{K}{K_f} - \dfrac{K}{K_s''} \right)}. \end{aligned} \tag{1.53}$$

Manipulating relation (1.36), we can express the undrained bulk modulus K_u in terms of the Biot coefficient α and the Skempton coefficient B, so that, using (1.52) and (1.53), we get

$$\begin{aligned} K_u = \frac{K}{1 - \alpha B} &= K \frac{\left(1 - \dfrac{K}{K_s'} \right) + \phi \left(\dfrac{K}{K_f} - \dfrac{K}{K_s''} \right)}{\dfrac{K}{K_s'} \left(1 - \dfrac{K}{K_s'} \right) + \phi \left(\dfrac{K}{K_f} - \dfrac{K}{K_s''} \right)} \\ &= K \left[1 + \frac{\left(1 - \dfrac{K}{K_s'} \right)^2}{\dfrac{K}{K_s'} \left(1 - \dfrac{K}{K_s'} \right) + \phi \left(\dfrac{K}{K_f} - \dfrac{K}{K_s''} \right)} \right]. \end{aligned} \tag{1.54}$$

Finally, recalling relation (1.39) and using (1.52) and (1.54), we express the

$\alpha = 1 - \dfrac{K}{K_s'}$
$K_u = K \left[1 + \dfrac{\left(1 - \dfrac{K}{K_s'}\right)^2}{\dfrac{K}{K_s'}\left(1 - \dfrac{K}{K_s'}\right) + \phi\left(\dfrac{K}{K_f} - \dfrac{K}{K_s''}\right)} \right]$
$B = 1 - \dfrac{\phi\left(\dfrac{K}{K_f} - \dfrac{K}{K_s''}\right)}{\left(1 - \dfrac{K}{K_s'}\right) + \phi\left(\dfrac{K}{K_f} - \dfrac{K}{K_s''}\right)}$
$\dfrac{1}{M} = \dfrac{K}{K_s'}\left(\dfrac{1}{K} - \dfrac{1}{K_s'}\right) + \phi\left(\dfrac{1}{K_f} - \dfrac{1}{K_s''}\right)$

Table 1.1: Correspondence between continuum and micromechanical parameters.

inverse of the Biot modulus M as

$$\begin{aligned} \frac{1}{M} = \frac{\alpha^2}{K_u - K} &= \frac{1}{K}\left[\frac{K}{K_s'}\left(1 - \frac{K}{K_s'}\right) + \phi\left(\frac{K}{K_f} - \frac{K}{K_s''}\right)\right] \\ &= \frac{1}{K_s'}\left(1 - \frac{K}{K_s'}\right) + \phi\left(\frac{1}{K_f} - \frac{1}{K_s''}\right) \\ &= \frac{K}{K_s'}\left(\frac{1}{K} - \frac{1}{K_s'}\right) + \phi\left(\frac{1}{K_f} - \frac{1}{K_s''}\right). \end{aligned} \tag{1.55}$$

We summarize the correspondence between continuum and micromechanical parameters in table 1.1. It is worth noting that these relations show also the dependence of the continuum parameters on the porosity.

We remark that poroelastic coefficients, like α, K and K_u, can be measured through appropriate laboratory tests, in which a sample of porous material is subjected to a small load increment under drained or undrained conditions. More details are given in [13].

1.3.4 Incompressibility in poroelasticity

Now, we deal with the concept of incompressibility in poroelasticity. As pointed out by Schanz and Pryl [40], in a porous material not only each component may be compressible, but also a structural compressibility exists. In a fully saturated porous medium, this structural compressibility is associated with the possibility of exchanging fluid with the outside. So, we may distinguish between the bulk modulus of the structure K and the bulk moduli of the components K_f and K_s (or its surrogates K_s' and K_s''). If the bulk modulus of a component is much larger than the bulk modulus of the porous medium, this component is assumed to be incompressible. Therefore, the condition for an incompressible solid component is

$$\frac{K}{K_s'} \ll 1 \quad \text{and} \quad \frac{K}{K_s''} \ll 1, \tag{1.56}$$

while that for an incompressible fluid component is

$$\frac{K}{K_f} \ll 1. \tag{1.57}$$

For example, in a soil saturated by water, both grains and water have bulk moduli much larger than that of the structure, so that each component is treated as incompressible, despite the fact that the structure is compressible. Three distinct situations must be considered.

- Only the solid is incompressible: $K/K_s' \ll 1$ and $K/K_s'' \ll 1$. The simplified expressions for the continuum poroelastic parameters are

$$\alpha \simeq 1, \tag{1.58}$$

$$K_u \simeq K\left(1 + \frac{K_f}{\phi K}\right), \tag{1.59}$$

$$B \simeq \frac{1}{1 + \dfrac{\phi K}{K_f}}, \tag{1.60}$$

$$M \simeq \frac{K_f}{\phi}. \tag{1.61}$$

- Only the fluid is incompressible: $K/K_f \ll 1$. The Biot coefficient α is unchanged and the other parameters simplify by neglecting the contribution of the ratios K/K_f and $1/K_f$.

- Both components are incompressible: $K/K_s' \ll 1$, $K/K_s'' \ll 1$ and $K/K_f \ll 1$. The simplified expressions for the continuum poroelastic

parameters are

$$\alpha \simeq 1, \tag{1.62}$$

$$K_u \to +\infty, \tag{1.63}$$

$$B \simeq 1, \tag{1.64}$$

$$M \to +\infty. \tag{1.65}$$

We conclude by making some considerations about the important case of a porous medium with incompressible components. In this case, all the poroelastic parameters assume their upper bound values. Furthermore, from relation (1.38) we obtain

$$\varepsilon = \zeta, \tag{1.66}$$

namely the volume change of the porous medium is equal to the fluid volume exchanged. As we shall see later, models for biological tissues usually consider incompressible components.

1.4 Balance equations

In the previous section we have discussed the constitutive laws describing the properties of a linear isotropic poroelastic material. In this section, we present two equations corresponding to basic physical principles, namely the momentum and the mass balance equations.

Recalling that the porous medium occupies the domain $\Omega \subset \mathbb{R}^3$, let $V \subset \Omega$ be an arbitrary open subset with boundary ∂V and a unit outward normal $\mathbf{n}$. Denoting by $\mathbf{f}$ the body force per unit volume of the porous medium, the momentum balance is given by

$$\int_{\partial V} \boldsymbol{\sigma}\mathbf{n}\, ds + \int_V \mathbf{f}\, dV = 0. \tag{1.67}$$

The first integral corresponds to the resultant of the internal surface forces acting on ∂V, while the second one corresponds to the resultant of the external body forces acting on V. The typical term with the second time derivative of $\mathbf{u}$ (accounting for inertial effects) does not appear in the above equation because of the quasi-static assumption. Applying the divergence theorem to equation (1.67), we get

$$\int_V \nabla \cdot \boldsymbol{\sigma}\, dV = -\int_V \mathbf{f}\, dV, \tag{1.68}$$

from which, thanks to the arbitrariness of V, we obtain the local momentum balance equation

$$\nabla \cdot \boldsymbol{\sigma} = -\mathbf{f}. \tag{1.69}$$

Turning to the mass balance equation, we consider again an arbitrary open subset $V \subset \Omega$ and denote by w a fluid source (or sink) term, that is the rate of injected (or sucked out) fluid volume per unit volume of the porous medium. The mass balance is expressed by

$$\frac{\partial}{\partial t}\int_V \zeta \, dV = -\int_{\partial V} \mathbf{v}\cdot\mathbf{n}\, ds + \int_V w \, dV. \tag{1.70}$$

The term on the left hand side of (1.70) represents the time rate of change of fluid content in V. There are two possible contributions to this change, corresponding to the integrals on the right hand side of (1.70): the first integral represents fluid exchange between V and the outside due to the fluid flux, while the second integral represents fluid gain (or loss) due to a given fluid source (or sink) inside the volume. Using the divergence theorem, equation (1.70) becomes

$$\frac{\partial}{\partial t}\int_V \zeta \, dV = -\int_V \nabla\cdot\mathbf{v}\, dV + \int_V w \, dV \tag{1.71}$$

and, recalling the arbitrariness of V, we get the local mass balance equation

$$\frac{\partial \zeta}{\partial t} = -\nabla\cdot\mathbf{v} + w. \tag{1.72}$$

The mass balance equation can be rewritten specifying the dependence of $\mathbf{v}$ on the fluid pressure p. It is known that Darcy's law is an empirical equation describing fluid flow in non-deformable porous media. Thanks to the assumption of small deformations, Darcy's law can be adopted here, so that

$$\mathbf{v} = -\mathbf{K}(\nabla p - \rho_f \mathbf{g}), \tag{1.73}$$

where ρ_f is the intrinsic fluid mass density, $\mathbf{g}$ is the gravitational acceleration and $\mathbf{K}$ is the permeability tensor. Because of isotropy, we have $\mathbf{K} = K\mathbf{I}$ and the permeability coefficient K is given by the ratio of the intrinsic permeability κ to the fluid viscosity μ_f: $K = \kappa/\mu_f$. The intrinsic permeability is usually a function of porosity ϕ and the particular form of this function depends on the porous space geometry. For example, the use of a packed array of spheres to represent the porous space geometry leads to the well known Carman-Kozeny formula, which is given by

$$\kappa(\phi) \div \frac{\phi^3}{(1-\phi)^2}. \tag{1.74}$$

Other models provide different laws. We note that the permeability coefficient could be influenced by the stress state of the porous medium, but we neglect this effect in order to avoid the introduction of a nonlinear relation in the model. In conclusion, we can substitute Darcy's law (1.73) into (1.72) and express the local mass balance equation in the form

$$\frac{\partial \zeta}{\partial t} = \nabla\cdot(\mathbf{K}(\nabla p - \rho_f \mathbf{g})) + w. \tag{1.75}$$

1.5 Governing equations of poroelasticity

Momentum and mass balance equations together with constitutive laws form the set of governing equations of the theory of poroelasticity, describing the coupled fluid flow and mechanical phenomena in linear isotropic elastic porous media. In this section, we put together the constitutive laws seen in section 1.3 and the balance equations seen in section 1.4, in order to get a system of partial differential equations amenable for mathematical treatment.

Let us introduce the drained and undrained Poisson ratios ν and ν_u, related to the drained and undrained bulk moduli K and K_u by the relations

$$\nu = \frac{3K - 2G}{2(3K + G)}, \tag{1.76}$$

$$\nu_u = \frac{3K_u - 2G}{2(3K_u + G)}. \tag{1.77}$$

For the limiting case of a porous medium with incompressible components, $K_u \to +\infty$, so that $\nu_u \to 0.5$, but this is not the trend of ν, which can assume values much smaller than 0.5. In fact, we have $\nu_u \in [\nu, 0.5]$. Making use of the relations between parameters illustrated so far, we can rewrite the constitutive equations (1.18) and (1.19) in terms of G, ν, α and M. The constitutive equation for the solid becomes

$$2G\varepsilon_{ij} = \sigma_{ij} - \frac{\nu}{1+\nu}\sigma_{kk}\delta_{ij} + \frac{\alpha(1-2\nu)}{1+\nu}p\delta_{ij}. \tag{1.78}$$

This relation can be inverted, thus obtaining

$$\sigma_{ij} = 2G\varepsilon_{ij} + \frac{2G\nu}{1-2\nu}\varepsilon\delta_{ij} - \alpha p\delta_{ij}, \tag{1.79}$$

where $\varepsilon = \varepsilon_{kk}$ according to (1.23). The quantity $\sigma'_{ij} = \sigma_{ij} + \alpha p\delta_{ij}$ is the standard stress tensor from the linear theory of elasticity and is called the effective stress in the field of poroelasticity. So, the total stress $\boldsymbol{\sigma}$ takes into account two contributions: the elastic stress in the material and the fluid pressure. An increase in fluid pressure tends to cause a structural expansion. On the other hand, the constitutive equation for the fluid becomes

$$\zeta = \frac{1}{M}p + \alpha\varepsilon. \tag{1.80}$$

We observe that an increment in fluid content can be due to an increment in fluid pressure or a structural expansion.

Again, we rewrite the constitutive equations in a different form, namely in terms of the Lamé's parameters

$$\mu = G \quad \text{and} \quad \lambda = \frac{2G\nu}{1-2\nu}. \tag{1.81}$$

From (1.79) and (1.80), reminding that $\boldsymbol{\varepsilon} = (\nabla\mathbf{u} + \nabla\mathbf{u}^T)/2$ and observing that $\varepsilon = \nabla \cdot \mathbf{u}$, we get

$$\boldsymbol{\sigma} = \mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla \cdot \mathbf{u}\mathbf{I} - \alpha p\mathbf{I}, \tag{1.82}$$

$$\zeta = \frac{1}{M}p + \alpha\nabla \cdot \mathbf{u}. \tag{1.83}$$

The complete set of governing equations is given by the momentum balance equation (1.69), the mass balance equation (1.72), Darcy's law (1.73) (where we neglect the effect of gravity, because it does not play a significant role in applications to biological tissues, which are the subject of our interest) and the constitutive equations (1.82) and (1.83):

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma} = -\mathbf{f}, \\ \quad \boldsymbol{\sigma} = \mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla \cdot \mathbf{u}\mathbf{I} - \alpha p\mathbf{I}, \\ \dfrac{\partial\zeta}{\partial t} = -\nabla \cdot \mathbf{v} + w, \\ \quad \mathbf{v} = -\mathbf{K}\nabla p, \\ \quad \zeta = \dfrac{1}{M}p + \alpha\nabla \cdot \mathbf{u}. \end{cases} \tag{1.84}$$

Substituting constitutive and Darcy's laws into balance equations, we can reduce (1.84) to a system of partial differential equations in Ω, whose unknowns are the displacement $\mathbf{u}$ and the fluid pressure p:

$$\begin{cases} \nabla \cdot (\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla \cdot \mathbf{u}\mathbf{I}) - \alpha\nabla p = -\mathbf{f}, \\ \quad \dfrac{1}{M}\dfrac{\partial p}{\partial t} + \alpha\dfrac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (\mathbf{K}\nabla p) = w. \end{cases} \tag{1.85}$$

The equations of this system are strongly coupled. The coupling term in the first equation is $\alpha\nabla p$, which represents the additional stress of the fluid pressure within the structure. The coupling term in the second equation is the time derivative of $\alpha\nabla \cdot \mathbf{u}$, which represents the additional variation of fluid content due to volume changes. Five material parameters, μ, λ, α, M and K, characterize a specific linear isotropic poroelastic material.

Since we are interested in biological tissues, we can simplify equations (1.85) by observing that both the solid and the fluid components of such materials are usually treated as incompressible, so that we can take $\alpha = 1$ and $M \to +\infty$. Considering these limiting values, we get

$$\begin{cases} \nabla \cdot (\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla \cdot \mathbf{u}\mathbf{I}) - \nabla p = -\mathbf{f}, \\ \quad \dfrac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (\mathbf{K}\nabla p) = w. \end{cases} \tag{1.86}$$

For an alternative derivation of equations (1.86), based on the concepts of mixture theory, see appendix A.

Proper initial and boundary conditions are required to obtain a well-posed mathematical problem. A poroelastic medium needs boundary conditions both for the solid component and for the fluid one. A generic initial condition is given by

$$\mathbf{u} = \mathbf{u}_0 \tag{1.87}$$

at $t = 0$. Then, we consider a generic set of boundary conditions for the solid component:

$$\begin{cases} \mathbf{u} = \mathbf{u}_D & \text{on } \Gamma_{s,D}, \\ (\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla\cdot\mathbf{u}\mathbf{I} - p\mathbf{I})\mathbf{n} = \mathbf{t}_N & \text{on } \Gamma_{s,N}, \end{cases} \tag{1.88}$$

where $\mathbf{n}$ denotes the outward normal to $\partial\Omega$ and $\Gamma_{s,D}$ and $\Gamma_{s,N}$ are open subsets of $\partial\Omega$, such that $\Gamma_{s,D} \cap \Gamma_{s,N} = \emptyset$ and $\overline{\Gamma}_{s,D} \cup \overline{\Gamma}_{s,N} = \partial\Omega$. The first condition sets the displacement and is a Dirichlet boundary condition, while the second one, equivalent to $\boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_N$ on $\Gamma_{s,N}$, sets the normal total stress and is a Neumann boundary condition. Finally, we consider a generic set of boundary conditions for the fluid component:

$$\begin{cases} p = p_D & \text{on } \Gamma_{f,D}, \\ -(\mathbf{K}\nabla p)\cdot\mathbf{n} = q_N & \text{on } \Gamma_{f,N}, \end{cases} \tag{1.89}$$

where $\Gamma_{f,D}$ and $\Gamma_{f,N}$ are open subsets of $\partial\Omega$, such that $\Gamma_{f,D} \cap \Gamma_{f,N} = \emptyset$ and $\overline{\Gamma}_{f,D} \cup \overline{\Gamma}_{f,N} = \partial\Omega$. The first condition sets the fluid pressure and is a Dirichlet boundary condition, while the second one, equivalent to $\mathbf{v}\cdot\mathbf{n} = q_N$ on $\Gamma_{f,N}$, sets the normal specific discharge and is a Neumann boundary condition. For simplicity, we assume that the Dirichlet and Neumann portions of $\partial\Omega$ do not depend on time.

To summarize, the mathematical problem for a linear isotropic poroelastic material with incompressible components is expressed by the following system of partial differential equations with initial and boundary conditions:

$$\begin{cases} \nabla\cdot(\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla\cdot\mathbf{u}\mathbf{I}) - \nabla p = -\mathbf{f} & \text{in } \Omega\times(0,+\infty), \\ \dfrac{\partial(\nabla\cdot\mathbf{u})}{\partial t} - \nabla\cdot(\mathbf{K}\nabla p) = w & \text{in } \Omega\times(0,+\infty), \\ \mathbf{u} = \mathbf{u}_D & \text{on } \Gamma_{s,D}\times(0,+\infty), \\ (\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla\cdot\mathbf{u}\mathbf{I} - p\mathbf{I})\mathbf{n} = \mathbf{t}_N & \text{on } \Gamma_{s,N}\times(0,+\infty), \\ p = p_D & \text{on } \Gamma_{f,D}\times(0,+\infty), \\ -(\mathbf{K}\nabla p)\cdot\mathbf{n} = q_N & \text{on } \Gamma_{f,N}\times(0,+\infty), \\ \mathbf{u} = \mathbf{u}_0 & \text{in } \Omega, \text{ at } t = 0. \end{cases} \tag{1.90}$$

This initial boundary value problem is well-posed, as proved by Showalter [42]. The next chapter is devoted to the numerical treatment of this mathematical problem.

Chapter 2

Numerical treatment of linear poroelasticity

In this chapter, we deal with the numerical approximation of the equations of linear poroelasticity. We propose a relaxed fixed point iteration algorithm, based on the backward Euler method for time discretization and the Galerkin finite element method for space discretization. The analysis of the iteration process allows us to deduce the maximum and optimal values for the relaxation parameter. The proposed algorithm has been implemented in MATLAB, a widely used numerical computing environment.

2.1 Monolithic and iterative solvers

In chapter 1, we have shown that the coupled fluid flow and mechanical phenomena in linear isotropic poroelastic media with incompressible components are described by the equations

$$\begin{cases} \nabla \cdot (\mu(\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \lambda \nabla \cdot \mathbf{u}\mathbf{I}) - \nabla p = -\mathbf{f}, \\ \dfrac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (\mathbf{K} \nabla p) = w, \end{cases} \tag{2.1}$$

expressed in terms of the unknowns $\mathbf{u}$ and p, namely the solid displacement and the fluid pressure. The first equation is associated with the mechanics of the material, while the second one with the fluid flow. Nevertheless, they are strongly coupled, because of the presence of both unknowns in both equations.

Several numerical schemes have been proposed in the literature for solving the PDE system (2.1) and can be divided into two main categories:

- monolithic solvers;
- iterative solvers.

Using a monolithic approach, equations (2.1) are solved simultaneously at each time level. Using an iterative approach, equations (2.1) are solved independently of one another and sequentially at each time level, so that the mechanical variable $\mathbf{u}$ and the fluid variable p are determined separately. Precisely, at each time level, the mechanical and fluid parts are solved iteratively until convergence is reached. Monolithic solvers are also referred to as coupled schemes, while iterative solvers as decoupled schemes.

Each kind of solver offers some advantages and some disadvantages. Referring to an approximation in the finite element framework, a monolithic solver leads to the solution of large linear algebraic systems, but avoids iterations at each time level. On the other hand, an iterative solver leads to smaller and more easily treatable linear algebraic systems, but the number of iterations at each time level may become unwieldy. Although the same results for monolithic and iterative solvers are expected at convergence, in an iterative scheme the accuracy of the solution depends on the convergence tolerance chosen, so that a monolithic scheme can be considered more robust. However, iterative schemes offer a higher degree of flexibility and allow also the usage of existing codes for the mechanical and fluid parts, which can be linked externally.

In this work, we propose an iterative algorithm that at each iteration of each time level solves first the mechanical part and then the fluid one. The algorithm is schematically described in figure 2.1, where n indicates the time level and k the internal iteration. The choice of an iterative algorithm

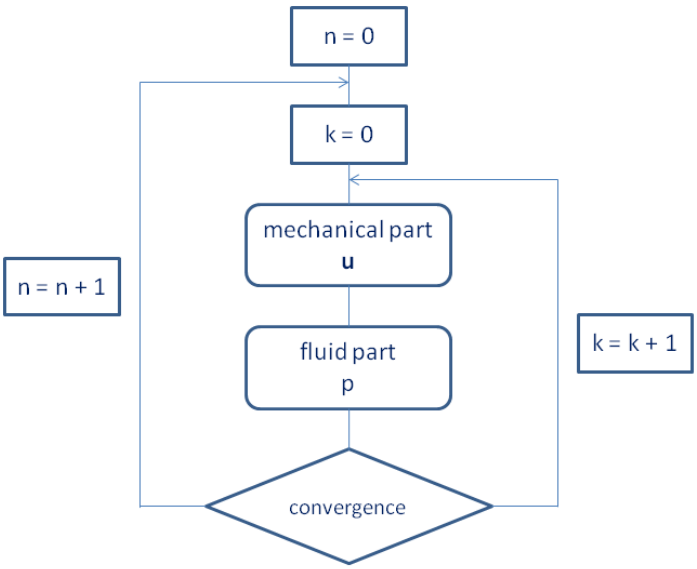


Figure 2.1: Workflow for the proposed iterative algorithm.

is motivated by three main reasons. Firstly, since our ultimate goal is the application of poroelastic equations in a new context, namely ocular biomechanics and hemodynamics, the flexibility offered by an iterative scheme is a desirable property. Secondly, the decoupling between the mechanic part and the fluid one allows us to implement the iterative algorithm starting from two existing codes, as we shall see later. Thirdly, we are interested in analysing an iterative method, studying its convergence properties and the need of relaxation, in order to identify the dependence of code performance on physical and simulation parameters.

Examples of the application of monolithic solvers to poroelastic problems can be found in [15] and [44]. An iterative solver is used in [7], where a model for large deformations is presented. Finally, we note that the idea to divide the original problem into subproblems of smaller size is widely applied in the different context of the numerical approximation of fluid-structure interaction problems. Indeed, these problems are often solved using partitioned algorithms, which treat the fluid subproblem and the structure subproblem independently and sequentially (see [5] and [25]).

2.2 The elastic pressure parameter

The equations of linear poroelasticity can be written in a different form by introducing the new dependent variable

$$\mathcal{P} = -\lambda \nabla \cdot \mathbf{u}, \tag{2.2}$$

which is known as the elastic pressure parameter. Replacing (2.2) into system (2.1) and reminding that $\boldsymbol{\varepsilon}(\mathbf{u}) = (\nabla \mathbf{u} + \nabla \mathbf{u}^T)/2$, we get

$$\begin{cases} \nabla \cdot (2\mu \boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I}) - \nabla p = -\mathbf{f}, \\ \dfrac{\mathcal{P}}{\lambda} + \nabla \cdot \mathbf{u} = 0, \\ -\dfrac{1}{\lambda}\dfrac{\partial \mathcal{P}}{\partial t} - \nabla \cdot (\mathbf{K} \nabla p) = w. \end{cases} \tag{2.3}$$

This is a system of five scalar differential equations in five unknowns. The first four equations and the variables $\mathbf{u}$ and $\mathcal{P}$ are associated with the mechanical part of the problem, while the last equation and the variable p are associated with the fluid part. We refer to system (2.3) both for the description of the iterative algorithm and for its implementation. The name given to $\mathcal{P}$ is inherited from the linear theory of elasticity and the reasons for the introduction of such a parameter will be clarified when we deal with space discretization.

Initial and boundary conditions for equations (2.3) are those introduced before, but must be rewritten in terms of the elastic pressure parameter.

Therefore, from (1.88) we get the boundary conditions for the solid component

$$\begin{cases} \mathbf{u} = \mathbf{u}_D & \text{on } \Gamma_{s,D}, \\ (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I} - p\mathbf{I})\mathbf{n} = \mathbf{t}_N & \text{on } \Gamma_{s,N}, \end{cases} \tag{2.4}$$

while from (1.89) we get the boundary conditions for the fluid component

$$\begin{cases} p = p_D & \text{on } \Gamma_{f,D}, \\ -(\mathbf{K}\nabla p)\cdot\mathbf{n} = q_N & \text{on } \Gamma_{f,N}. \end{cases} \tag{2.5}$$

Finally, we point out that using (2.2) into (1.82) we can express the total stress tensor $\boldsymbol{\sigma}$ as

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I} - p\mathbf{I}, \tag{2.6}$$

namely as a function of the unknowns $\mathbf{u}$, $\mathcal{P}$ and p. We note that the contribution of the elastic pressure parameter $\mathcal{P}$ to the total stress tensor is similar to that of the fluid pressure p.

2.3 Time discretization: backward Euler method

For expository purposes, we deal with time discretization of equations (2.3) before presenting the general structure of the iterative algorithm. We assume that we have to solve equations (2.3) from $t = 0$ to $t = T$ and denote by Δt the time step. We write $t_n = n\Delta t$, where $n \in \mathbb{N}$. Let $\mathbf{u}_n = \mathbf{u}(t_n)$, $\mathcal{P}_n = \mathcal{P}(t_n)$ and $p_n = p(t_n)$. We adopt the backward Euler method to advance from t_n to t_{n+1}: given $\mathbf{u}_n$, $\mathcal{P}_n$ and p_n, find $\mathbf{u}_{n+1}$, $\mathcal{P}_{n+1}$ and p_{n+1} such that

$$\begin{cases} \nabla\cdot(2\mu\boldsymbol{\varepsilon}(\mathbf{u}_{n+1}) - \mathcal{P}_{n+1}\mathbf{I}) - \nabla p_{n+1} = -\mathbf{f}_{n+1} & \text{in } \Omega, \\ \dfrac{\mathcal{P}_{n+1}}{\lambda} + \nabla\cdot\mathbf{u}_{n+1} = 0 & \text{in } \Omega, \\ -\dfrac{1}{\lambda}\dfrac{\mathcal{P}_{n+1}}{\Delta t} - \nabla\cdot(\mathbf{K}\nabla p_{n+1}) = w_{n+1} - \dfrac{1}{\lambda}\dfrac{\mathcal{P}_n}{\Delta t} & \text{in } \Omega, \end{cases} \tag{2.7}$$

where $\mathbf{f}_{n+1} = \mathbf{f}(t_{n+1})$ and $w_{n+1} = w(t_{n+1})$. It is worth recalling that the backward Euler method is an unconditionally stable first-order scheme.

2.4 A relaxed fixed point iteration algorithm

At each time level t_n, we propose a relaxed fixed point iteration algorithm, that is an iterative method endowed with a relaxation parameter to help convergence. Since we have implemented this algorithm in both 1D and 2D cases, henceforth we assume that $\Omega \subset \mathbb{R}^2$. We also assume that convergence is reached when, for each dependent variable, the difference between two

consecutive iterations (measured in the $L^\infty(\Omega)$ norm) is smaller than a fixed tolerance, where for any $f \in L^\infty(\Omega)$ we set

$$\|f\|_{L^\infty(\Omega)} = \inf\{C \geq 0 : |f(\mathbf{x})| \leq C \text{ for almost every } \mathbf{x} \in \Omega\}. \tag{2.8}$$

Let ω be the relaxation parameter and ϵ_{u_1}, ϵ_{u_2}, $\epsilon_{\mathcal{P}}$ and ϵ_p the convergence tolerances related to the variables, where u_1 and u_2 denote the components of the displacement vector $\mathbf{u}$. The relaxed fixed point iteration algorithm advances from t_n to t_{n+1} in the following way:

1. let $k = 0$, $\mathbf{u}^{(0)} = \mathbf{u}_n$, $\mathcal{P}^{(0)} = \mathcal{P}_n$ and $p^{(0)} = p_n$;
2. set $k = k + 1$;
3. find $\mathbf{u}^{(k)}$ and $\mathcal{P}^{(k)}$ such that

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) - \mathcal{P}^{(k)}\mathbf{I}) = -\mathbf{f}_{n+1} + \nabla p^{(k-1)} & \text{in } \Omega, \\ \dfrac{\mathcal{P}^{(k)}}{\lambda} + \nabla \cdot \mathbf{u}^{(k)} = 0 & \text{in } \Omega, \\ \mathbf{u}^{(k)} = \mathbf{u}_D(t_{n+1}) & \text{on } \Gamma_{s,D}, \\ (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) - \mathcal{P}^{(k)}\mathbf{I} - p^{(k-1)}\mathbf{I})\mathbf{n} = \mathbf{t}_N(t_{n+1}) & \text{on } \Gamma_{s,N}; \end{cases}$$

4. find $\tilde{p}^{(k)}$ such that

$$\begin{cases} -\nabla \cdot (\mathbf{K}\nabla\tilde{p}^{(k)}) = w_{n+1} + \dfrac{1}{\lambda\Delta t}(\mathcal{P}^{(k)} - \mathcal{P}_n) & \text{in } \Omega, \\ \tilde{p}^{(k)} = p_D(t_{n+1}) & \text{on } \Gamma_{f,D}, \\ -(\mathbf{K}\nabla\tilde{p}^{(k)}) \cdot \mathbf{n} = q_N(t_{n+1}) & \text{on } \Gamma_{f,N}; \end{cases}$$

5. set $p^{(k)} = \omega\tilde{p}^{(k)} + (1 - \omega)p^{(k-1)}$;
6. if

$$\begin{aligned} \frac{\|u_1^{(k)} - u_1^{(k-1)}\|_{L^\infty(\Omega)}}{\|u_1^{(k-1)}\|_{L^\infty(\Omega)}} &< \epsilon_{u_1}, \\ \frac{\|u_2^{(k)} - u_2^{(k-1)}\|_{L^\infty(\Omega)}}{\|u_2^{(k-1)}\|_{L^\infty(\Omega)}} &< \epsilon_{u_2}, \\ \frac{\|\mathcal{P}^{(k)} - \mathcal{P}^{(k-1)}\|_{L^\infty(\Omega)}}{\|\mathcal{P}^{(k-1)}\|_{L^\infty(\Omega)}} &< \epsilon_{\mathcal{P}}, \\ \frac{\|p^{(k)} - p^{(k-1)}\|_{L^\infty(\Omega)}}{\|p^{(k-1)}\|_{L^\infty(\Omega)}} &< \epsilon_p \end{aligned}$$

are all verified, set $\mathbf{u}_{n+1} = \mathbf{u}^{(k)}$, $\mathcal{P}_{n+1} = \mathcal{P}^{(k)}$ and $p_{n+1} = p^{(k)}$, otherwise return to step 2.

Therefore, the original problem is divided into two subproblems, which are solved iteratively at each time level. These subproblems correspond to the mechanical part (step 3) and the fluid part (step 4). When we solve the mechanical subproblem at step 3, we use the fluid pressure computed at the previous iteration, so that its gradient can be seen as an additional body force per unit volume of the porous medium. When we solve the fluid subproblem at step 4, we use the elastic pressure parameter computed at step 3, so that the term with the approximation of the time derivative of $\mathcal{P}$ can be seen as an additional fluid source or sink. In other words, step 3 can be reduced to the solution of a standard problem of linear elasticity, while step 4 requires the solution of a linear Poisson problem with the fluid pressure as unknown. The proposed algorithm makes use of the backward Euler method described in the previous section for time discretization, while space discretization will be discussed in the next section, where the Galerkin finite element method is applied to the mechanical and fluid subproblems. Further remarks about the proposed algorithm are discussed below.

Remark 2 (Fixed point iteration). At each time level, the algorithm described above corresponds to a fixed point iteration method of the form

$$p^{(k)} = \mathcal{B}(p^{(k-1)}), \tag{2.9}$$

where $\mathcal{B}$ is the linear differential operator resulting from the algorithm. An extensive analysis of the fixed point iteration will be shown later.

Remark 3 (Relaxation). Step 5 of the proposed algorithm is the so-called relaxation step. It helps convergence and avoids ending up with a large number of iterations for particular choices of physical and simulation parameters. A small value of ω means a high degree of relaxation. We use a static relaxation technique, in which the relaxation parameter is kept constant for all iterations. This parameter has to be small enough to keep the iteration from diverging, but at the same time large enough in order to use as much of the new solution as possible and to avoid unnecessary iterations. Therefore, the choice of ω requires particular attention. We will show how its maximum and optimal values can be estimated in poroelastic problems. Details about more general dynamic relaxation techniques, such as the Aitken's method, can be found in [25].

Remark 4 (Convergence test). The convergence test corresponding to step 6 of the algorithm requires special care in choosing the numerical values of the tolerances associated with the variables, because the accuracy of the solution depends on such a choice. Moreover, a maximum number of iterations has to be fixed, in order to avoid infinitely long loops.

2.5 Spatial discretization: finite element method

In this section, we deal with the spatial discretization of poroelastic equations through the Galerkin finite element method (GFEM). Let us introduce the Hilbert spaces

$$L^2(\Omega) = \{q : \Omega \to \mathbb{R} \,|\, \int_\Omega |q|^2 \, dV < +\infty\}, \tag{2.10}$$

$$[L^2(\Omega)]^2 = \{\mathbf{q} : \Omega \to \mathbb{R}^2 \,|\, \int_\Omega |\mathbf{q}|^2 \, dV < +\infty\}, \tag{2.11}$$

$$H^1(\Omega) = \{r \in L^2(\Omega) \,|\, \nabla r \in [L^2(\Omega)]^2\}, \tag{2.12}$$

$$H^1_{f,D}(\Omega) = \{r \in H^1(\Omega) \,|\, r = 0 \text{ on } \Gamma_{f,D}\}, \tag{2.13}$$

$$[H^1(\Omega)]^2 = \{\mathbf{v} : \Omega \to \mathbb{R}^2 \,|\, v_i \in H^1(\Omega),\, i = 1, 2\}, \tag{2.14}$$

$$[H^1_{s,D}(\Omega)]^2 = \{\mathbf{v} \in [H^1(\Omega)]^2 \,|\, \mathbf{v} = \mathbf{0} \text{ on } \Gamma_{s,D}\}. \tag{2.15}$$

For the sake of simplicity, in the remainder of this chapter we consider homogeneous Dirichlet boundary conditions both for the mechanical part and for the fluid one, namely $\mathbf{u}_D = \mathbf{0}$ and $p_D = 0$.

2.5.1 Weak formulation

First, we consider the problem defined at step 3 of the algorithm, where we take $\mathbf{u}_D = \mathbf{0}$:

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) - \mathcal{P}^{(k)}\mathbf{I}) = -\mathbf{f}_{n+1} + \nabla p^{(k-1)} & \text{in } \Omega, \\ \dfrac{\mathcal{P}^{(k)}}{\lambda} + \nabla \cdot \mathbf{u}^{(k)} = 0 & \text{in } \Omega, \\ \mathbf{u}^{(k)} = \mathbf{0} & \text{on } \Gamma_{s,D}, \\ (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) - \mathcal{P}^{(k)}\mathbf{I} - p^{(k-1)}\mathbf{I})\mathbf{n} = \mathbf{t}_N(t_{n+1}) & \text{on } \Gamma_{s,N}. \end{cases} \tag{2.16}$$

Let V and Q be the function spaces for displacement and elastic pressure parameter, respectively. We set $V = [H^1_{s,D}(\Omega)]^2$ and $Q = L^2(\Omega)$. Multiplying the first equation of (2.16) by a test function $\mathbf{v} \in V$, we get

$$\int_\Omega (\nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) - \mathcal{P}^{(k)}\mathbf{I})) \cdot \mathbf{v} \, dV = -\int_\Omega \mathbf{f}_{n+1} \cdot \mathbf{v} \, dV + \int_\Omega \nabla p^{(k-1)} \cdot \mathbf{v} \, dV.$$

Integrating by parts in order to take into account the boundary conditions, we have

$$\begin{aligned} &\int_{\partial\Omega} (2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)})\mathbf{n}) \cdot \mathbf{v} \, ds - \int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) : \boldsymbol{\varepsilon}(\mathbf{v}) \, dV - \int_{\partial\Omega} \mathcal{P}^{(k)}\mathbf{n} \cdot \mathbf{v} \, ds \\ &+ \int_\Omega \mathcal{P}^{(k)} \nabla \cdot \mathbf{v} \, dV = -\int_\Omega \mathbf{f}_{n+1} \cdot \mathbf{v} \, dV + \int_{\partial\Omega} p^{(k-1)}\mathbf{n} \cdot \mathbf{v} \, ds \\ &- \int_\Omega p^{(k-1)} \nabla \cdot \mathbf{v} \, dV. \end{aligned}$$

Finally, using the boundary conditions, we obtain

$$\int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dV - \int_\Omega \mathcal{P}^{(k)} \nabla \cdot \mathbf{v}\, dV = \int_\Omega \mathbf{f}_{n+1} \cdot \mathbf{v}\, dV$$
$$+ \int_{\Gamma_{s,N}} \mathbf{t}_N(t_{n+1}) \cdot \mathbf{v}\, ds + \int_\Omega p^{(k-1)} \nabla \cdot \mathbf{v}\, dV. \tag{2.17}$$

Multiplying the second equation of (2.16) by a test function $q \in Q$, we get

$$\int_\Omega q \nabla \cdot \mathbf{u}^{(k)}\, dV + \frac{1}{\lambda} \int_\Omega \mathcal{P}^{(k)} q\, dV = 0. \tag{2.18}$$

Now, we consider the problem defined at step 4 of the algorithm, where we take $p_D = 0$:

$$\begin{cases} -\nabla \cdot (\mathbf{K} \nabla \tilde{p}^{(k)}) = w_{n+1} + \dfrac{1}{\lambda \Delta t}(\mathcal{P}^{(k)} - \mathcal{P}_n) & \text{in } \Omega, \\ \tilde{p}^{(k)} = 0 & \text{on } \Gamma_{f,D}, \\ -(\mathbf{K} \nabla \tilde{p}^{(k)}) \cdot \mathbf{n} = q_N(t_{n+1}) & \text{on } \Gamma_{f,N}. \end{cases} \tag{2.19}$$

Let R be the function space for the fluid pressure. We set $R = H^1_{f,D}(\Omega)$. Multiplying the equation of problem (2.19) by a test function $r \in R$, reminding that $\mathbf{K} = K\mathbf{I}$ and assuming that K is a positive constant, we get

$$-\int_\Omega \nabla \cdot (K \nabla \tilde{p}^{(k)}) r\, dV = \int_\Omega w_{n+1} r\, dV + \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}^{(k)} r\, dV - \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}_n r\, dV.$$

Integrating by parts where necessary, we have

$$-\int_{\partial\Omega} K \nabla \tilde{p}^{(k)} \cdot \mathbf{n} r\, ds + \int_\Omega K \nabla \tilde{p}^{(k)} \cdot \nabla r\, dV = \int_\Omega w_{n+1} r\, dV$$
$$+ \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}^{(k)} r\, dV - \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}_n r\, dV.$$

Finally, using the boundary conditions, we obtain

$$\int_\Omega K \nabla \tilde{p}^{(k)} \cdot \nabla r\, dV = \int_\Omega w_{n+1} r\, dV - \int_{\Gamma_{f,N}} q_N(t_{n+1}) r\, ds$$
$$+ \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}^{(k)} r\, dV - \frac{1}{\lambda \Delta t} \int_\Omega \mathcal{P}_n r\, dV. \tag{2.20}$$

To summarize, the weak formulation of problem (2.16), corresponding to the mechanical part, reads: find $\mathbf{u}^{(k)} \in V$ and $\mathcal{P}^{(k)} \in Q$ such that

$$\begin{cases} \displaystyle\int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}^{(k)}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dV - \int_\Omega \mathcal{P}^{(k)} \nabla \cdot \mathbf{v}\, dV = \int_\Omega \mathbf{f}_{n+1} \cdot \mathbf{v}\, dV \\ \displaystyle + \int_{\Gamma_{s,N}} \mathbf{t}_N(t_{n+1}) \cdot \mathbf{v}\, ds + \int_\Omega p^{(k-1)} \nabla \cdot \mathbf{v}\, dV & \forall \mathbf{v} \in V, \\ \displaystyle\int_\Omega q \nabla \cdot \mathbf{u}^{(k)}\, dV + \frac{1}{\lambda} \int_\Omega \mathcal{P}^{(k)} q\, dV = 0 & \forall q \in Q. \end{cases} \tag{2.21}$$

On the other hand, the weak formulation of problem (2.19), corresponding to the fluid part, reads: find $\tilde{p}^{(k)} \in R$ such that

$$\int_\Omega K\nabla\tilde{p}^{(k)} \cdot \nabla r \, dV = \int_\Omega w_{n+1} r \, dV - \int_{\Gamma_{f,N}} q_N(t_{n+1}) r \, ds$$
$$+\frac{1}{\lambda\Delta t}\int_\Omega \mathcal{P}^{(k)} r \, dV - \frac{1}{\lambda\Delta t}\int_\Omega \mathcal{P}_n r \, dV \qquad \forall r \in R. \tag{2.22}$$

2.5.2 GFEM for linear isotropic elasticity

The weak formulation (2.21) of the mechanical subproblem can be reduced to that of the Herrmann formulation of linear isotropic elasticity (see [20]). We want to show this analogy and give some details about the GFEM applied to a problem of elasticity. It is well known that linear isotropic elasticity is described by the Navier-Lamé system

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla\cdot\mathbf{u}\mathbf{I}) = -\mathbf{f} & \text{in } \Omega, \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma_{s,D}, \\ (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla\cdot\mathbf{u}\mathbf{I})\mathbf{n} = \mathbf{t}_N & \text{on } \Gamma_{s,N}, \end{cases} \tag{2.23}$$

where the sole unknown is the displacement $\mathbf{u}$. The application of the GFEM to the Navier-Lamé system is subject to difficulties when the material becomes incompressible (see [21]). Differently from what occurs in poroelasticity, in elasticity an incompressible material is characterized by $\lambda \to +\infty$ and $\nabla \cdot \mathbf{u} = 0$, so that the product $\lambda\nabla\cdot\mathbf{u}$ gives rise to an undetermined form. In this case, the numerical solution is affected by the so-called volumetric locking, which leads to unphysical results. In order to overcome this problem, the elastic pressure parameter $\mathcal{P} = -\lambda\nabla\cdot\mathbf{u}$ is introduced, leading to the Herrmann formulation of linear isotropic elasticity

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I}) = -\mathbf{f} & \text{in } \Omega, \\ \dfrac{\mathcal{P}}{\lambda} + \nabla\cdot\mathbf{u} = 0 & \text{in } \Omega, \\ \mathbf{u} = \mathbf{0} & \text{on } \Gamma_{s,D}, \\ (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I})\mathbf{n} = \mathbf{t}_N & \text{on } \Gamma_{s,N}. \end{cases} \tag{2.24}$$

This formulation is well defined for every $\lambda \in (0, +\infty]$, so that it can be considered a unified formulation of linear isotropic elasticity. It is a two-field model and its numerical discretization requires particular care, because of the choice of suitable finite dimensional spaces for the approximate solutions $\mathbf{u}_h$ and $\mathcal{P}_h$. Proceeding as before, we obtain the weak formulation of the

Herrmann model (2.24): find $\mathbf{u} \in V$ and $\mathcal{P} \in Q$ such that

$$\begin{cases} \displaystyle\int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dV - \int_\Omega \mathcal{P}\nabla\cdot\mathbf{v}\, dV = F(\mathbf{v}) & \forall \mathbf{v} \in V, \\ \displaystyle\int_\Omega q\nabla\cdot\mathbf{u}\, dV + \frac{1}{\lambda}\int_\Omega \mathcal{P} q\, dV = 0 & \forall q \in Q, \end{cases} \tag{2.25}$$

where

$$F(\mathbf{v}) = \int_\Omega \mathbf{f}\cdot\mathbf{v}\, dV + \int_{\Gamma_{s,N}} \mathbf{t}_N \cdot \mathbf{v}\, ds. \tag{2.26}$$

The function spaces V and Q are exactly those defined before. We observe that the weak formulation (2.21) coincides with (2.25), except for the presence of an additional term on the right hand side of the first equation, but this term can be integrated into the definition of the linear functional $F : V \to \mathbb{R}$, so that we can solve the mechanical part of the proposed algorithm using the standard methods applied to the Herrmann model of elasticity.

In order to deal with the well-posedness and the finite element approximation of the Herrmann formulation, we consider $\lambda \to +\infty$ and $\Gamma_{s,N} = \emptyset$. Problem (2.24) becomes

$$\begin{cases} \nabla\cdot(2\mu\boldsymbol{\varepsilon}(\mathbf{u}) - \mathcal{P}\mathbf{I}) = -\mathbf{f} & \text{in } \Omega, \\ \nabla\cdot\mathbf{u} = 0 & \text{in } \Omega, \\ \mathbf{u} = \mathbf{0} & \text{on } \partial\Omega. \end{cases} \tag{2.27}$$

Since we have a Dirichlet boundary condition on the whole boundary $\partial\Omega$, we choose $V = [H_0^1(\Omega)]^2$, where

$$[H_0^1(\Omega)]^2 = \left\{ \mathbf{v} \in [H^1(\Omega)]^2 \,|\, \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega \right\}. \tag{2.28}$$

Then, we note that if $\mathcal{P}^\star$ is a solution of (2.27), also $\mathcal{P}^\star + C$ is a solution, C being an arbitrary constant. Therefore, we choose $Q = L_0^2(\Omega)$, where

$$L_0^2(\Omega) = \left\{ q \in L^2(\Omega) \,|\, \int_\Omega q\, dV = 0 \right\}, \tag{2.29}$$

so that the further requirement on the elastic pressure parameter to have zero mean over Ω allows to fix C in a unique manner. The weak formulation of problem (2.27) reads: find $\mathbf{u} \in V$ and $\mathcal{P} \in Q$ such that

$$\begin{cases} \displaystyle\int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dV - \int_\Omega \mathcal{P}\nabla\cdot\mathbf{v}\, dV = \int_\Omega \mathbf{f}\cdot\mathbf{v}\, dV & \forall \mathbf{v} \in V, \\ \displaystyle -\int_\Omega q\nabla\cdot\mathbf{u}\, dV = 0 & \forall q \in Q. \end{cases} \tag{2.30}$$

Introducing the bilinear forms $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and $b(\cdot,\cdot) : V \times Q \to \mathbb{R}$, defined by

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dV, \tag{2.31}$$

$$b(\mathbf{u},q) = -\int_\Omega q\nabla\cdot\mathbf{u}\, dV, \tag{2.32}$$

and the linear functional $F(\cdot) : V \to \mathbb{R}$, defined by

$$F(\mathbf{v}) = \int_\Omega \mathbf{f}\cdot\mathbf{v}\, dV, \tag{2.33}$$

the weak formulation (2.30) can be rewritten in the following form: find $\mathbf{u} \in V$ and $\mathcal{P} \in Q$ such that

$$\begin{cases} a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},\mathcal{P}) = F(\mathbf{v}) & \forall \mathbf{v} \in V, \\ b(\mathbf{u},q) = 0 & \forall q \in Q. \end{cases} \tag{2.34}$$

This is the general form of a saddle-point problem. Its study is a source of interest not only in the field of elasticity and poroelasticity, but also in fluid dynamics, because of its close analogy with the stationary Stokes equations. A fundamental result is expressed by the following theorem of functional analysis.

Theorem 1 (Unique solvability of a saddle-point problem)**.** *If $a(\cdot,\cdot)$ is a continuous and coercive bilinear form, $b(\cdot,\cdot)$ is a continuous bilinear form, $F(\cdot)$ is a continuous linear functional and there exists a positive constant γ such that*

$$\forall q \in Q\ \exists \mathbf{v} \in V \text{ such that } \int_\Omega q\nabla\cdot\mathbf{v}\, dV \geq \gamma\|q\|_Q\|\mathbf{v}\|_V, \tag{2.35}$$

then the saddle-point problem (2.34) *admits a unique solution.*

It is immediate to prove that the bilinear form $a(\cdot,\cdot)$ defined in (2.31) is continuous and coercive and the bilinear form $b(\cdot,\cdot)$ defined in (2.32) is continuous. Furthermore, by virtue of our choices of V and Q, it is possible to show that the compatibility condition (2.35) is satisfied. Therefore, we have the following result.

Theorem 2 (Well-posedness of the Herrmann formulation)**.** *Let $\mathbf{f} \in [L^2(\Omega)]^2$. Then, the weak problem* (2.30) *admits a unique solution pair $(\mathbf{u},\mathcal{P}) \in V \times Q$. Moreover,*

$$\|\mathbf{u}\|_V \leq C_1\|\mathbf{f}\|_{[L^2(\Omega)]^2}, \tag{2.36}$$

$$\|\mathcal{P}\|_Q \leq C_2\|\mathbf{f}\|_{[L^2(\Omega)]^2}, \tag{2.37}$$

where C_1 and C_2 are suitable positive constants.

The compatibility condition (2.35) is also called inf-sup condition, because it can be equivalently reformulated as

$$\inf_{q\in Q, q\neq 0} \sup_{\mathbf{v}\in V, \mathbf{v}\neq \mathbf{0}} \frac{\int_\Omega q\nabla\cdot\mathbf{u}\,dV}{\|q\|_Q\|\mathbf{v}\|_V} \geq \gamma, \tag{2.38}$$

or LBB condition, from the names of the mathematicians Ladyzhenskaya, Babuska and Brezzi.

Now, we apply the GFEM to the weak problem (2.30). Let $\mathcal{T}_h$ be a triangulation of the domain Ω made of triangular elements K, whose diameters are denoted by h_K, so that $h = \max_{K\in\mathcal{T}_h} h_K$. Then, we consider two finite dimensional subspaces of polynomial functions $V_h \subset V$ and $Q_h \subset Q$. The Galerkin problem associated with the weak problem (2.30) reads: find $\mathbf{u}_h \in V_h$ and $\mathcal{P}_h \in Q_h$ such that

$$\begin{cases} \displaystyle\int_\Omega 2\mu\boldsymbol{\varepsilon}(\mathbf{u}_h) : \boldsymbol{\varepsilon}(\mathbf{v}_h)\,dV - \int_\Omega \mathcal{P}_h\nabla\cdot\mathbf{v}_h\,dV = \int_\Omega \mathbf{f}\cdot\mathbf{v}_h\,dV & \forall \mathbf{v}_h \in V_h, \\ \displaystyle -\int_\Omega q_h\nabla\cdot\mathbf{u}_h\,dV = 0 & \forall q_h \in Q_h. \end{cases} \tag{2.39}$$

The linear algebraic system emanating from (2.39) is obtained by introducing two sets of polynomial basis functions: $\{\boldsymbol{\varphi}_1,\ldots,\boldsymbol{\varphi}_{N_h}\}$ for V_h and $\{\psi_1,\ldots,\psi_{M_h}\}$ for Q_h. Therefore, we have $\dim V_h = N_h$ and $\dim Q_h = M_h$. Since $\mathbf{u}_h \in V_h$ and $\mathcal{P}_h \in Q_h$, we can write

$$\mathbf{u}_h(\mathbf{x}) = \sum_{j=1}^{N_h} u_j\boldsymbol{\varphi}_j(\mathbf{x}), \qquad \mathcal{P}_h(\mathbf{x}) = \sum_{j=1}^{M_h} \mathcal{P}_j\psi_j(\mathbf{x}), \tag{2.40}$$

where $\{u_1,\ldots,u_{N_h}\}$ and $\{\mathcal{P}_1,\ldots,\mathcal{P}_{M_h}\}$ are the sets of degrees of freedom for $\mathbf{u}_h$ and $\mathcal{P}_h$, respectively. Substituting relations (2.40) into (2.39) and using the basis functions of V_h and Q_h as test functions, we get

$$\begin{cases} \displaystyle\sum_{j=1}^{N_h} A_{ij}u_j + \sum_{j=1}^{M_h} B_{ji}\mathcal{P}_j = F_i & \forall i = 1,\ldots,N_h, \\ \displaystyle\sum_{j=1}^{N_h} B_{ij}u_j = 0 & \forall i = 1,\ldots,M_h, \end{cases} \tag{2.41}$$

where A_{ij} and B_{ij} are the entries of the matrices $A \in \mathbb{R}^{N_h\times N_h}$ and $B \in \mathbb{R}^{M_h\times N_h}$ given by

$$A_{ij} = \int_\Omega 2\mu\boldsymbol{\varepsilon}(\boldsymbol{\varphi}_j) : \boldsymbol{\varepsilon}(\boldsymbol{\varphi}_i)\,dV, \qquad i,j = 1,\ldots,N_h, \tag{2.42}$$

$$B_{ij} = -\int_\Omega \psi_i\nabla\cdot\boldsymbol{\varphi}_j\,dV, \qquad i = 1,\ldots,M_h,\ j = 1,\ldots,N_h, \tag{2.43}$$

while F_i are the entries of the vector $\mathbf{F} \in \mathbb{R}^{N_h}$ given by

$$F_i = \int_\Omega \mathbf{f} \cdot \boldsymbol{\varphi}_i \, dV, \qquad i = 1, \ldots, N_h. \tag{2.44}$$

Denoting by $\mathbf{U}$ and $\mathbf{P}_{el}$ the vectors whose entries are the degrees of freedom for $\mathbf{u}_h$ and $\mathcal{P}_h$, respectively, we obtain the linear algebraic system

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{P}_{el} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix}, \tag{2.45}$$

corresponding to the Galerkin problem (2.39). We observe that A is a symmetric and positive definite matrix, while the global matrix of the system is symmetric, but indefinite. Therefore, problem (2.27) is reduced to the solution of the linear algebraic system (2.45).

The finite element spaces V_h and Q_h have to be properly chosen, in order to ensure the existence and uniqueness of the solution of the Galerkin problem.

Theorem 3 (Well-posedness of the discrete Herrmann formulation)**.** *Let $\mathbf{f} \in [L^2(\Omega)]^2$ and assume that there exists a positive constant $\gamma^\star$, independent of h, such that*

$$\forall q_h \in Q_h \; \exists \mathbf{v}_h \in V_h \text{ such that } \int_\Omega q_h \nabla \cdot \mathbf{v}_h \, dV \geq \gamma^\star \|q_h\|_Q \|\mathbf{v}_h\|_V. \tag{2.46}$$

Then, the Galerkin problem (2.39)*, or equivalently* (2.45)*, admits a unique solution pair $(\mathbf{u}_h, \mathcal{P}_h) \in V_h \times Q_h$. Moreover,*

$$\|\mathbf{u}_h\|_V \leq C_1 \|\mathbf{f}\|_{[L^2(\Omega)]^2}, \tag{2.47}$$

$$\|\mathcal{P}_h\|_Q \leq C_3 \|\mathbf{f}\|_{[L^2(\Omega)]^2}, \tag{2.48}$$

where C_1 and C_3 are suitable positive constants.

We point out that C_1 is the same constant appearing in Theorem 2, while C_3 differs from C_2 because it depends on $\gamma^\star$ rather than γ. Condition (2.46) is the so-called discrete inf-sup condition. If it is not satisfied, then there may exist a function $\mathcal{P}_h^\star \in Q_h$, with $\mathcal{P}_h^\star \neq 0$, such that

$$\int_\Omega \mathcal{P}_h^\star \nabla \cdot \mathbf{v}_h \, dV = 0 \qquad \forall \mathbf{v}_h \in V_h. \tag{2.49}$$

The identity (2.49) tells us that if $\mathcal{P}_h$ is a solution of (2.39), also $\mathcal{P}_h + \mathcal{P}_h^\star$ is a solution, namely the solution of (2.39) is no longer unique. Thus, the discrete inf-sup condition filters out the occurrence of such solutions $\mathcal{P}_h^\star$, called spurious pressure modes, ensuring that the numerical solution is not affected by unphysical instabilities. We can say that the discrete inf-sup

condition requires the use of an approximation space for the displacement with the property of being richer than that for the elastic pressure parameter; in fact, it can be proved that a necessary condition for (2.46) to hold is that

$$\dim V_h \geq \dim Q_h. \tag{2.50}$$

In particular, we can not use equal-order polynomial spaces for both V_h and Q_h. Among the pairs (V_h, Q_h) satisfying the discrete inf-sup condition, we mention the one used in the code, given by

$$V_h = \{\mathbf{v}_h \in [\mathcal{C}^0(\overline{\Omega})]^2 : \mathbf{v}_h|_K \in [\mathbb{P}_2(K)]^2 \ \forall K \in \mathcal{T}_h,\ \mathbf{v}_h = \mathbf{0} \text{ on } \partial\Omega\}, \tag{2.51}$$

$$Q_h = \{q_h \in \mathcal{C}^0(\overline{\Omega}) : q_h|_K \in \mathbb{P}_1(K)\ \forall K \in \mathcal{T}_h\}. \tag{2.52}$$

This choice is known as the Taylor-Hood element and the corresponding

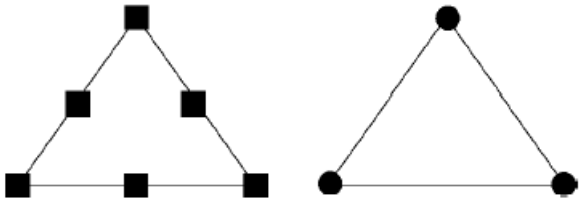


Figure 2.2: Degrees of freedom for the Taylor-Hood element: dofs for the displacement to the left and dofs for the elastic pressure parameter to the right (taken from [38]).

degrees of freedom are depicted in figure 2.2. Finally, we give a result about convergence of the GFEM when using the Taylor-Hood element.

Theorem 4 (Convergence of the Taylor-Hood element method)**.** *Let V_h and Q_h be as in* (2.51) *and* (2.52) *and let $\mathbf{u} \in [H^3(\Omega)]^2$ and $\mathcal{P} \in H^2(\Omega)$. Then, the approximate solution $(\mathbf{u}_h, \mathcal{P}_h)$ of* (2.39) *converges to the exact solution $(\mathbf{u}, \mathcal{P})$ of* (2.30)*. Moreover,*

$$\|\mathbf{u} - \mathbf{u}_h\|_{[H^1(\Omega)]^2} + \|\mathcal{P} - \mathcal{P}_h\|_{L^2(\Omega)} \leq \widetilde{C} h^2 (\|\mathbf{u}\|_{[H^3(\Omega)]^2} + \|\mathcal{P}\|_{H^2(\Omega)}), \tag{2.53}$$

where $\widetilde{C}$ is a positive constant independent of h and $(\mathbf{u}, \mathcal{P})$.

2.5.3 GFEM for the Poisson problem

The fluid subproblem (2.19) is a standard Poisson problem, which can be written in the simplified form

$$\begin{cases} -\nabla \cdot (K \nabla p) = w & \text{in } \Omega, \\ p = 0 & \text{on } \Gamma_{f,D}, \\ -K \nabla p \cdot \mathbf{n} = q_N & \text{on } \Gamma_{f,N}, \end{cases} \tag{2.54}$$

having integrated the terms related to the elastic pressure parameter into the source term w. The corresponding simplified version of the weak formulation (2.22) reads: find $p \in R$ such that

$$\int_\Omega K\nabla p \cdot \nabla r \, dV = \int_\Omega wr \, dV - \int_{\Gamma_{f,N}} q_N r \, ds \qquad \forall r \in R. \tag{2.55}$$

The function space R is exactly that defined before. Introducing the bilinear form $d(\cdot,\cdot) : R \times R \to \mathbb{R}$, defined by

$$d(p,r) = \int_\Omega K\nabla p \cdot \nabla r \, dV, \tag{2.56}$$

and the linear functional $W(\cdot) : R \to \mathbb{R}$, defined by

$$W(r) = \int_\Omega wr \, dV - \int_{\Gamma_{f,N}} q_N r \, ds, \tag{2.57}$$

we rewrite the weak formulation (2.55) in the form: find $r \in R$ such that

$$d(p,r) = W(r) \qquad \forall r \in R. \tag{2.58}$$

Theorem 5 (Well-posedness of the Poisson problem). *Let $w \in L^2(\Omega)$ and $q_N \in L^2(\Gamma_{f,N})$. Then, the weak problem (2.55) admits a unique solution $p \in R$. Moreover,*

$$\|p\|_R \leq C_4(\|w\|_{L^2(\Omega)} + \|q_N\|_{L^2(\Gamma_{f,N})}), \tag{2.59}$$

where C_4 is a suitable positive constant.

The proof of this result is a straightforward application of the Lax-Milgram lemma to the weak problem and consists in showing that the bilinear form $d(\cdot,\cdot)$ is continuous and coercive and the linear functional $W(\cdot)$ is continuous.

Now, we apply the GFEM to the weak problem (2.55). Let $\mathcal{T}_h$ be the triangulation of Ω introduced in section 2.5.2 and $R_h \subset R$ be a finite dimensional subspace of polynomial functions. The Galerkin problem associated with the weak problem (2.55) reads: find $p_h \in R_h$ such that

$$\int_\Omega K\nabla p_h \cdot \nabla r_h \, dV = \int_\Omega wr_h \, dV - \int_{\Gamma_{f,N}} q_N r_h \, ds \qquad \forall r_h \in R_h. \tag{2.60}$$

Introducing a set of polynomial basis functions $\{\eta_1, \ldots, \eta_{L_h}\}$ for R_h, so that $\dim R_h = L_h$, and noting that $p_h \in R_h$, we can write

$$p_h(\mathbf{x}) = \sum_{j=1}^{L_h} p_j \eta_j(\mathbf{x}), \tag{2.61}$$

where $\{p_1, \ldots, p_{L_h}\}$ is the set of degrees of freedom for p_h. Substituting (2.61) into (2.60) and using the basis functions of R_h as test functions, we get

$$\sum_{j=1}^{L_h} D_{ij} p_j = W_i \qquad \forall i = 1, \ldots, L_h, \tag{2.62}$$

where D_{ij} are the entries of the matrix $D \in \mathbb{R}^{L_h \times L_h}$ given by

$$D_{ij} = \int_\Omega K \nabla \eta_j \cdot \nabla \eta_i \, dV, \qquad i, j = 1, \ldots, L_h, \tag{2.63}$$

while W_i are the entries of the vector $\mathbf{W} \in \mathbb{R}^{L_h}$ given by

$$W_i = \int_\Omega w \eta_i \, dV - \int_{\Gamma_{f,N}} q_N \eta_i \, ds, \qquad i = 1, \ldots, L_h. \tag{2.64}$$

Denoting by $\mathbf{P}$ the vector whose entries are the degrees of freedom for p_h, we obtain the linear algebraic system

$$D\mathbf{P} = \mathbf{W}, \tag{2.65}$$

corresponding to the Galerkin problem (2.60). We note that D is a symmetric and positive definite matrix.

The Lax-Milgram lemma ensures also the existence and uniqueness of the solution of the Galerkin problem (2.60).

Theorem 6 (Well-posedness of the discrete Poisson problem)**.** *Let $w \in L^2(\Omega)$ and $q_N \in L^2(\Gamma_{f,N})$. Then, the Galerkin problem* (2.60)*, or equivalently* (2.65)*, admits a unique solution $p_h \in R_h$. Moreover,*

$$\|p_h\|_R \leq C_4(\|w\|_{L^2(\Omega)} + \|q_N\|_{L^2(\Gamma_{f,N})}), \tag{2.66}$$

where C_4 is the same constant as in Theorem 5.

Finally, taking

$$R_h = \{r_h \in \mathcal{C}^0(\overline{\Omega}) : r_h|_K \in \mathbb{P}_1(K) \; \forall K \in \mathcal{T}_h, \; r_h = 0 \text{ on } \Gamma_{f,D}\} \tag{2.67}$$

in accordance with what implemented in the code, we give the following result about convergence of the GFEM for the Poisson problem.

Theorem 7 (Convergence of the GFEM for the Poisson problem)**.** *Let R_h be as in* (2.67) *and let $p \in H^2(\Omega)$. Then, the approximate solution p_h of* (2.60) *converges to the exact solution p of* (2.55)*. Moreover,*

$$\|p - p_h\|_{H^1(\Omega)} \leq \widehat{C} h \|p\|_{H^2(\Omega)}, \tag{2.68}$$

$$\|p - p_h\|_{L^2(\Omega)} \leq \overline{C} h^2 \|p\|_{H^2(\Omega)}, \tag{2.69}$$

where $\widehat{C}$ and $\overline{C}$ are positive constants independent of h and p.

2.5.4 Implementation

We have shown how the mechanical and fluid subproblems can be solved by means of the GFEM. It is clear that the solution of the mechanical part is more difficult, because it corresponds to a saddle-point problem. On the other hand, the fluid part requires the solution of a standard Poisson problem, which is a simpler task. Now, using the formalism introduced so far, we want to write the algebraic counterparts of the Galerkin problems obtained by discretization of (2.21) and (2.22).

Let us consider the finite dimensional spaces V_h, Q_h and R_h given by (2.51), where we replace $\partial\Omega$ by $\Gamma_{s,D}$, (2.52) and (2.67), respectively. We note that the only difference between Q_h and R_h lies in the portion of $\partial\Omega$ on which the space functions vanish. Therefore, since in the code the Dirichlet boundary conditions are imposed after the assembly of the matrices, we can treat these spaces as identical and indicate by $\{\psi_1, \ldots, \psi_{M_h}\}$ a set of basis functions. Recalling the definitions of the matrices $A \in \mathbb{R}^{N_h \times N_h}$, $B \in \mathbb{R}^{M_h \times N_h}$ and $D \in \mathbb{R}^{M_h \times M_h}$ and introducing the matrix $C \in \mathbb{R}^{M_h \times M_h}$, we have

$$A_{ij} = \int_\Omega 2\mu\boldsymbol{\varepsilon}(\boldsymbol{\varphi}_j) : \boldsymbol{\varepsilon}(\boldsymbol{\varphi}_i)\, dV, \qquad i,j = 1, \ldots, N_h, \tag{2.70}$$

$$B_{ij} = -\int_\Omega \psi_i \nabla \cdot \boldsymbol{\varphi}_j\, dV, \qquad i = 1, \ldots, M_h,\ j = 1, \ldots, N_h, \tag{2.71}$$

$$C_{ij} = -\int_\Omega \frac{1}{\lambda}\psi_j\psi_i\, dV, \qquad i,j = 1, \ldots, M_h, \tag{2.72}$$

$$D_{ij} = \int_\Omega K\nabla\psi_j \cdot \nabla\psi_i\, dV, \qquad i,j = 1, \ldots, M_h. \tag{2.73}$$

Then, we define the vectors $\mathbf{F}_{n+1} \in \mathbb{R}^{N_h}$ and $\mathbf{W}_{n+1} \in \mathbb{R}^{M_h}$ as

$$F_i = \int_\Omega \mathbf{f}_{n+1} \cdot \boldsymbol{\varphi}_i\, dV + \int_{\Gamma_{s,N}} \mathbf{t}_N(t_{n+1}) \cdot \boldsymbol{\varphi}_i\, ds, \qquad i = 1, \ldots, N_h, \tag{2.74}$$

$$W_i = \int_\Omega w_{n+1}\psi_i\, dV - \int_{\Gamma_{f,N}} q_N(t_{n+1})\psi_i\, ds, \qquad i = 1, \ldots, M_h. \tag{2.75}$$

Finally, we denote by $\mathbf{U}^{(k)} \in \mathbb{R}^{N_h}$, $\mathbf{P}_{el}^{(k)} \in \mathbb{R}^{M_h}$, $\mathbf{P}_{el,n} \in \mathbb{R}^{M_h}$, $\mathbf{P}^{(k)} \in \mathbb{R}^{M_h}$ and $\widetilde{\mathbf{P}}^{(k)} \in \mathbb{R}^{M_h}$ the vectors whose entries are the degrees of freedom of the discrete counterparts of $\mathbf{u}^{(k)}$, $\mathcal{P}^{(k)}$, $\mathcal{P}_n$, $p^{(k)}$ and $\tilde{p}^{(k)}$, respectively.

Therefore, the mechanical subproblem (2.21) is reduced by the GFEM to the solution of the linear algebraic system

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \begin{bmatrix} \mathbf{U}^{(k)} \\ \mathbf{P}_{el}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{n+1} - B^T\mathbf{P}^{(k-1)} \\ \mathbf{0} \end{bmatrix}. \tag{2.76}$$

On the other hand, the fluid subproblem (2.22) is reduced to the solution of the linear algebraic system

$$D\widetilde{\mathbf{P}}^{(k)} = \mathbf{W}_{n+1} + \frac{1}{\Delta t} C\mathbf{P}_{el,n} - \frac{1}{\Delta t} C\mathbf{P}_{el}^{(k)}. \tag{2.77}$$

Furthermore, the relaxation step of the proposed iterative algorithm can be written in algebraic form as

$$\mathbf{P}^{(k)} = \omega\widetilde{\mathbf{P}}^{(k)} + (1-\omega)\mathbf{P}^{(k-1)}. \tag{2.78}$$

At each time level, the algorithm requires to solve systems (2.76) and (2.77) and to perform the relaxation step (2.78) iteratively. We point out the importance of using the same finite elements for the approximation of the elastic pressure parameter and the fluid pressure: this allows us to have the same matrix B in the left and right hand sides of system (2.76). Moreover, we can finally justify the introduction of the elastic pressure parameter in our model. Firstly, it ensures a robust treatment of the mechanical part, avoiding numerical difficulties when $\lambda \to +\infty$. Such limiting value for λ, corresponding to $\nu = 0.5$, characterizes an incompressible material in linear elasticity, while in poroelasticity a medium with incompressible components not necessarily exhibits $\lambda \to +\infty$ (or $\nu = 0.5$). Secondly, the introduction of the elastic pressure parameter replaces the time derivative of the divergence of the displacement by the time derivative of the elastic pressure parameter in the fluid subproblem, thus avoiding the numerical approximation of such a divergence, which is cause of degradation in accuracy.

The relaxed fixed point iteration algorithm has been implemented in MATLAB. We have developed a 2D code for linear poroelasticiy, starting from two existing codes (available at the link in [38]), one for linear elasticity (based on the Herrmann formulation) and the other for elliptic problems. At each time level, the code solves (2.76) and (2.77) and computes (2.78) iteratively. As regards the initialization, at $t = 0$ only $\mathbf{u} = \mathbf{u}_0$ is given, while $\mathcal{P}_0$ is obtained by $\mathbf{u}_0$ according to the discrete counterpart of

$$\int_\Omega \mathcal{P}_0 q \, dV = -\lambda \int_\Omega q \nabla \cdot \mathbf{u}_0 \, dV \qquad \forall q \in Q. \tag{2.79}$$

On the other hand, we can take any guess for p_0. When performing simulations, particular attention has to be paid in choosing the grid size h, the time step Δt, the convergence tolerances for the variables and the relaxation parameter ω. We use the $\mathbb{P}_2 - \mathbb{P}_1$ finite element pair for the mechanical part and the $\mathbb{P}^1$ finite elements for the fluid one, although the code provides a large variety of finite element spaces ($\mathbb{P}_0$, $\mathbb{P}_1$, $\mathbb{P}_2$, $\mathbb{P}_3$ and $\mathbb{P}_1^b$, namely $\mathbb{P}_1$ finite elements enriched by bubble functions). The validation of the code, together with some important numerical results, will be presented in chapter 3. In addition to the 2D code for linear poroelasticity, we have also developed a 1D

code. The 1D code does not introduce the elastic pressure parameter, but offers the possibility to solve a poroelastic problem by using both an iterative and a monolithic solver. Therefore, it is an important tool to analyse differences and similarities between 1D and 2D cases and between monolithic and iterative solvers. The 1D code uses $\mathbb{P}^1$ finite elements for both displacement and fluid pressure.

2.6 Analysis of the relaxed iterative algorithm

In this section, we analyse the fixed point iteration scheme and the role played by relaxation. We start our analysis from the 1D case without the elastic pressure parameter. Then, we turn to the 2D case with the elastic pressure parameter.

2.6.1 The 1D case without the elastic pressure parameter

We consider a 1D linear poroelastic problem, defined on the interval $(0, L)$ from $t = 0$ to $t = T$. In the 1D case, all the variables appearing in the model are scalar functions. Since the total stress σ is given by

$$\sigma = 2\mu \frac{\partial u}{\partial x} + \lambda \frac{\partial u}{\partial x} - p, \tag{2.80}$$

it is useful to introduce the aggregate modulus $H_A = 2\mu + \lambda$, so that

$$\sigma = H_A \frac{\partial u}{\partial x} - p. \tag{2.81}$$

A 1D linear poroelastic problem is expressed by the equations

$$\begin{cases} \dfrac{\partial \sigma}{\partial x} = -f & \text{in } (0, L) \times (0, T), \\ \sigma = H_A \dfrac{\partial u}{\partial x} - p & \text{in } (0, L) \times (0, T), \\ \dfrac{\partial}{\partial t} \dfrac{\partial u}{\partial x} + \dfrac{\partial v}{\partial x} = w & \text{in } (0, L) \times (0, T), \\ v = -K \dfrac{\partial p}{\partial x} & \text{in } (0, L) \times (0, T). \end{cases} \tag{2.82}$$

We complete the problem by adding a generic set of boundary conditions for the solid part

$$\begin{cases} u(0, t) = 0, \\ \sigma(L, t) = \sigma_N(t), \end{cases} \tag{2.83}$$

and a generic set of boundary conditions for the fluid part

$$\begin{cases} v(0, t) = v_N(t), \\ p(L, t) = 0. \end{cases} \tag{2.84}$$

Substituting the second and fourth equations of (2.82) into the first and third equations, respectively, we can express the poroelastic problem in the compact form

$$\begin{cases} \dfrac{\partial}{\partial x}\left(H_A \dfrac{\partial u}{\partial x}\right) - \dfrac{\partial p}{\partial x} = -f \quad \text{in } (0,L)\times(0,T), \\ \dfrac{\partial}{\partial t}\dfrac{\partial u}{\partial x} - \dfrac{\partial}{\partial x}\left(K\dfrac{\partial p}{\partial x}\right) = w \quad \text{in } (0,L)\times(0,T). \end{cases} \tag{2.85}$$

Now, we introduce the function spaces

$$V = H^1_{s,D}(0,L) = \{v \in H^1(0,L) \,|\, v(0) = 0\}, \tag{2.86}$$

$$R = H^1_{f,D}(0,L) = \{r \in H^1(0,L) \,|\, r(L) = 0\} \tag{2.87}$$

and multiply the first equation of (2.85) by a test function $v \in V$ and the second equation by a test function $r \in R$, so that, after integration by parts, we get the weak problem: find $u \in V$ and $p \in R$ such that

$$\begin{cases} \displaystyle\int_0^L H_A \frac{\partial u}{\partial x}\frac{\partial v}{\partial x}\,dx - \int_0^L p\frac{\partial v}{\partial x}\,dx = \int_0^L fv\,dx + \sigma_N v(L) \quad \forall v \in V, \\ \displaystyle\frac{\partial}{\partial t}\int_0^L \frac{\partial u}{\partial x} r\,dx + \int_0^L K\frac{\partial p}{\partial x}\frac{\partial r}{\partial x}\,dx = \int_0^L wr\,dx + v_N r(0) \quad \forall r \in R. \end{cases} \tag{2.88}$$

Denoting by $V_h \subset V$ and $R_h \subset R$ two finite dimensional subspaces and using the GFEM for discretization in space, we obtain the Galerkin problem: find $u_h \in V_h$ and $p_h \in R_h$ such that

$$\begin{cases} \displaystyle\int_0^L H_A \frac{\partial u_h}{\partial x}\frac{\partial v_h}{\partial x}\,dx - \int_0^L p_h\frac{\partial v_h}{\partial x}\,dx = \int_0^L fv_h\,dx + \sigma_N v_h(L) \quad \forall v_h \in V_h, \\ \displaystyle\frac{\partial}{\partial t}\int_0^L \frac{\partial u_h}{\partial x} r_h\,dx + \int_0^L K\frac{\partial p_h}{\partial x}\frac{\partial r_h}{\partial x}\,dx = \int_0^L wr_h\,dx + v_N r_h(0) \quad \forall r_h \in R_h. \end{cases} \tag{2.89}$$

Since we use $\mathbb{P}_1$ finite elements for both displacement and fluid pressure, recalling that the imposition of the Dirichlet boundary conditions occurs after the assembly of the matrices, we can use a unique set of basis functions for V_h and R_h, which we denote by $\{\varphi_1, \ldots, \varphi_{N_h}\}$. Introducing the matrices $A \in \mathbb{R}^{N_h \times N_h}$, $B \in \mathbb{R}^{N_h \times N_h}$ and $D \in \mathbb{R}^{N_h \times N_h}$, defined by

$$A_{ij} = \int_0^L H_A \frac{\partial \varphi_j}{\partial x}\frac{\partial \varphi_i}{\partial x}\,dx, \qquad i,j = 1,\ldots,N_h, \tag{2.90}$$

$$B_{ij} = -\int_0^L \varphi_i \frac{\partial \varphi_j}{\partial x}\,dx, \qquad i,j = 1,\ldots,N_h, \tag{2.91}$$

$$D_{ij} = \int_0^L K\frac{\partial \varphi_j}{\partial x}\frac{\partial \varphi_i}{\partial x}\,dx, \qquad i,j = 1,\ldots,N_h, \tag{2.92}$$

and using the backward Euler method to advance from $t = t_n$ to $t = t_{n+1}$, we can write the linear algebraic system emanating from (2.89) as

$$\begin{bmatrix} A & B^T \\ -\dfrac{1}{\Delta t}B & D \end{bmatrix} \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{P}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{n+1} \\ \mathbf{W}_{n+1} - \dfrac{1}{\Delta t}B\mathbf{U}_n \end{bmatrix}, \tag{2.93}$$

where $\mathbf{U}_{n+1}$ and $\mathbf{P}_{n+1}$ denote the vectors whose entries are the degrees of freedom for u_h and p_h, respectively, at time $t = t_{n+1}$, while $\mathbf{F}_{n+1}$ and $\mathbf{W}_{n+1}$ are the vectors taking into account the source terms and the Neumann boundary conditions, computed at time $t = t_{n+1}$.

A monolithic approach solves system (2.93) to advance from $t = t_n$ to $t = t_{n+1}$. However, we are interested in an iterative approach as alternative. Applying the basic idea of iterative schemes directly to system (2.93) leads to the iterative solution of the following steps, till convergence is reached.

- Mechanical part:

$$A\mathbf{U}^{(k)} = \mathbf{F}_{n+1} - B^T\mathbf{P}^{(k-1)}. \tag{2.94}$$

- Fluid part:

$$D\widetilde{\mathbf{P}}^{(k)} = \mathbf{W}_{n+1} - \frac{1}{\Delta t}B\mathbf{U}_n + \frac{1}{\Delta t}B\mathbf{U}^{(k)}. \tag{2.95}$$

- Relaxation:

$$\mathbf{P}^{(k)} = \omega\widetilde{\mathbf{P}}^{(k)} + (1-\omega)\mathbf{P}^{(k-1)}. \tag{2.96}$$

Since A is symmetric and positive definite, from (2.94) we can write

$$\mathbf{U}^{(k)} = A^{-1}(\mathbf{F}_{n+1} - B^T\mathbf{P}^{(k-1)}). \tag{2.97}$$

Substituting (2.97) into (2.95) and noting that also D is symmetric and positive definite, we get

$$\widetilde{\mathbf{P}}^{(k)} = D^{-1}(\mathbf{W}_{n+1} - \frac{1}{\Delta t}B\mathbf{U}_n) + \frac{1}{\Delta t}D^{-1}BA^{-1}(\mathbf{F}_{n+1} - B^T\mathbf{P}^{(k-1)}). \tag{2.98}$$

Finally, substituting (2.98) into (2.96), we obtain

$$\mathbf{P}^{(k)} = B_\omega\mathbf{P}^{(k-1)} + \mathbf{g}_\omega, \tag{2.99}$$

where

$$B_\omega = (1-\omega)I - \frac{\omega}{\Delta t}D^{-1}BA^{-1}B^T, \tag{2.100}$$

$$\mathbf{g}_\omega = \omega D^{-1}\left(\frac{1}{\Delta t}BA^{-1}\mathbf{F}_{n+1} + \mathbf{W}_{n+1} - \frac{1}{\Delta t}B\mathbf{U}_n\right), \tag{2.101}$$

where I is the identity matrix. We have expressed the iterative method in the abstract form (2.9). The matrix B_ω is the iteration matrix of the relaxed fixed point iteration algorithm. It is well known that (2.99) converges if and only if the spectral radius of the iteration matrix (namely the maximum of the absolute values of the eigenvalues of the matrix) is strictly less than 1:

$$\rho(B_\omega) < 1. \tag{2.102}$$

Condition (2.102) involves both physical and simulation parameters, such as H_A, K, h, Δt and ω. In particular, it allows us to determine the degree of relaxation needed by a specific problem to reach convergence. Since the matrix $M = D^{-1}BA^{-1}B^T$ has real positive eigenvalues μ_i, with $i = 1, \ldots, N_h$ (as confirmed by numerical tests), we can order them in the following way:

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{N_h} > 0. \tag{2.103}$$

Condition (2.102) can be written as

$$\max_{i=1,\ldots,N_h} \left|1 - \omega - \frac{\omega}{\Delta t}\mu_i\right| < 1, \tag{2.104}$$

which is equivalent to

$$-1 < 1 - \omega - \frac{\omega}{\Delta t}\mu_i < 1 \qquad \forall i = 1, \ldots, N_h. \tag{2.105}$$

Simple calculations lead to

$$0 < \omega < \frac{2}{1 + \dfrac{\mu_i}{\Delta t}} \qquad \forall i = 1, \ldots, N_h. \tag{2.106}$$

In particular, from (2.106) we deduce that the relaxation parameter has to be smaller than a maximum value, in order to have convergence:

$$0 < \omega < \omega_{max} = \frac{2}{1 + \dfrac{\mu_1}{\Delta t}}. \tag{2.107}$$

We observe that if we reduce Δt, we have to use a higher degree of relaxation (namely a smaller value of ω). The dependence of ω_{max} on the other parameters is hidden in μ_1 and can be specified considering the definitions of the matrices A, B and D. The results of our analysis will be verified through appropriate numerical tests illustrated in chapter 3. Using a uniform grid of size h on the interval $(0, L)$ and the Lagrange basis functions for $\mathbb{P}_1$ finite elements, before enforcing the Dirichlet boundary conditions we have

$$A = \frac{H_A}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}, \qquad B = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & \\ 1 & 0 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & -1 \\ & & & 1 & -1 \end{bmatrix},$$

$$D = \frac{K}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}. \tag{2.108}$$

Therefore, we can express the matrix M as

$$M = \frac{h^2}{H_A K} T^{-1} B T^{-1} B^T, \tag{2.109}$$

where

$$T = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}. \tag{2.110}$$

We note that $T^{-1}BT^{-1}B^T$ is a dimensionless matrix, while the scalar factor appearing in (2.109) has the units of a time:

$$\left[\frac{h^2}{H_A K}\right] = \frac{\mathrm{cm}^2}{\dfrac{\mathrm{g}}{\mathrm{cm\ s}^2}\dfrac{\mathrm{cm}^3\ \mathrm{s}}{\mathrm{g}}} = \mathrm{s}. \tag{2.111}$$

It is useful to carry out the scaling of equations (2.85), where we take $f = 0$ and $w = 0$ for simplicity. Introducing the dimensionless quantities

$$\hat{x} = \frac{x}{L}, \qquad \hat{t} = \frac{t}{L^2} H_A K, \qquad \hat{u} = \frac{u}{L}, \qquad \hat{p} = \frac{p}{H_A}, \tag{2.112}$$

after some calculations, we get the dimensionless equations

$$\begin{cases} \dfrac{\partial}{\partial \hat{x}}\left(\dfrac{\partial \hat{u}}{\partial \hat{x}}\right) - \dfrac{\partial \hat{p}}{\partial \hat{x}} = 0, \\ \dfrac{\partial}{\partial \hat{t}}\dfrac{\partial \hat{u}}{\partial \hat{x}} - \dfrac{\partial}{\partial \hat{x}}\left(\dfrac{\partial \hat{p}}{\partial \hat{x}}\right) = 0. \end{cases} \tag{2.113}$$

Repeating the above analysis for the dimensionless equations (2.113) leads to the iteration matrix

$$\hat{B}_\omega = (1-\omega)I - \frac{\omega}{\hat{\Delta t}} \hat{D}^{-1} B \hat{A}^{-1} B^T, \tag{2.114}$$

where

$$\hat{A} = \hat{D} = \frac{1}{\hat{h}} T, \tag{2.115}$$

so that

$$\hat{M} = \hat{D}^{-1} B \hat{A}^{-1} B^T = \hat{h}^2 T^{-1} B T^{-1} B^T. \tag{2.116}$$

The method converges if and only if $\rho(\hat{B}_\omega) < 1$, which leads to $0 < \omega < \omega_{max}$, where

$$\omega_{max} = \frac{2}{1 + \dfrac{\mu_{max}(\hat{M})}{\hat{\Delta}t}} = \frac{2}{1 + \dfrac{\hat{h}^2}{\hat{\Delta}t}\mu_{max}(T^{-1}BT^{-1}B^T)}$$
$$= \frac{2}{1 + \dfrac{h^2}{H_A K \Delta t}\mu_{max}(T^{-1}BT^{-1}B^T)}, \tag{2.117}$$

having denoted by μ_{max} the maximum eigenvalue of the matrix indicated in parentheses at each occurrence. Considering boundary conditions as in (2.83) and (2.84), a numerical test run in MATLAB reveals that

$$\lim_{\hat{h}\to 0} \hat{h}^2 \mu_{max}(T^{-1}BT^{-1}B^T) = \hat{c} \simeq 0.025, \tag{2.118}$$

as shown in figure 2.3. Therefore, we have

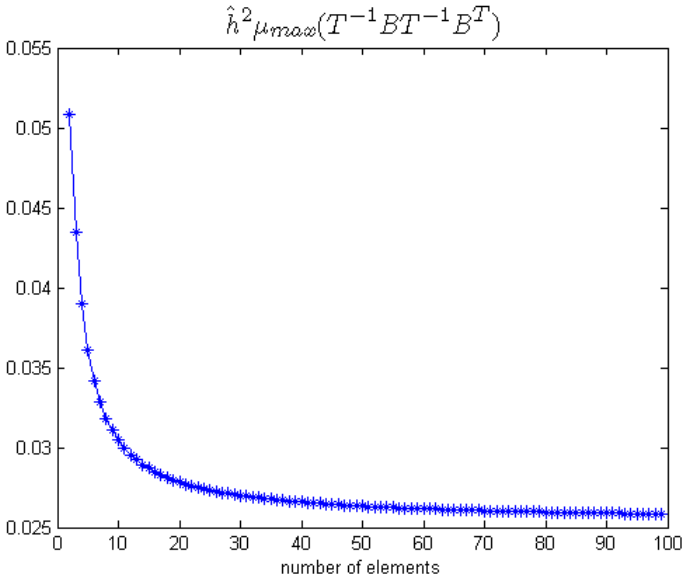


Figure 2.3: Numerical test for the analysis of ω_{max}.

$$\lim_{\hat{h}\to 0} \omega_{max} = \frac{2}{1 + \dfrac{\hat{c}}{\hat{\Delta}t}} = \frac{2}{1 + \dfrac{c}{H_A K \Delta t}}, \tag{2.119}$$

where $c = \hat{c}L^2$. From (2.119) we deduce that ω_{max} is independent of h, but depends on H_A, K and Δt. We observe that a reduction in H_A or K, as well

as a reduction in Δt, leads to a higher degree of relaxation. Furthermore, we note that

$$[H_A K \Delta t] = \frac{\text{g}}{\text{cm s}^2} \frac{\text{cm}^3 \text{ s}}{\text{g}} \text{s} = \text{cm}^2, \tag{2.120}$$

so that we can interpret $H_A K \Delta t$ as the square of a diffusion length L_{diff} and distinguish between the two following limiting cases:

- if $L_{diff}/L \gg \sqrt{\hat{c}}$, then $\omega_{max} \to 2^-$;
- if $L_{diff}/L \ll \sqrt{\hat{c}}$, then $\omega_{max} \to 0^+$.

Finally, we get an expression for the optimal value of the relaxation parameter. We say that ω is optimal if it minimizes the spectral radius of the iteration matrix. Since

$$\rho_{opt} = \min_{\omega \in (0,1]} \max_{i=1,\ldots,N_h} \left| 1 - \omega - \frac{\omega}{\Delta t} \mu_i \right|, \tag{2.121}$$

we can determine the value of ω_{opt} using the graphical analysis shown in figure 2.4. We see that ω_{opt} satisfy

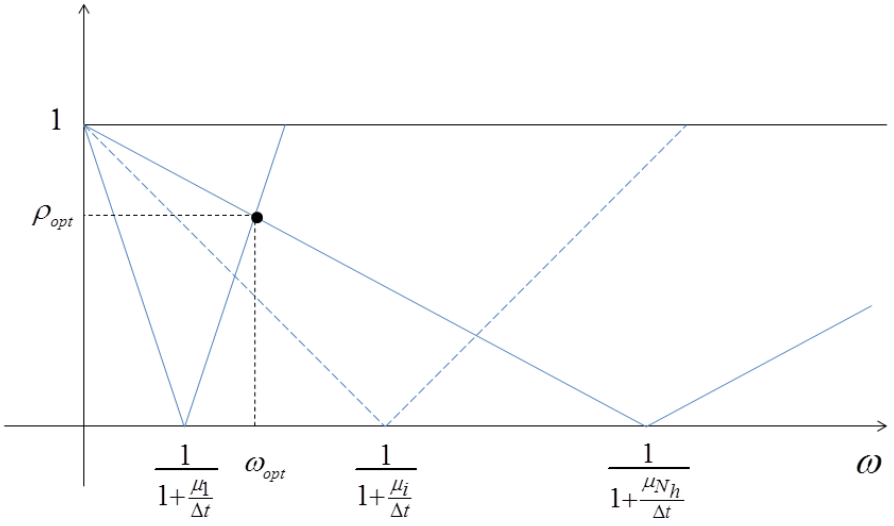


Figure 2.4: Optimal value for the relaxation parameter.

$$1 - \omega_{opt} \left(1 + \frac{\mu_{N_h}}{\Delta t} \right) = \omega_{opt} \left(1 + \frac{\mu_1}{\Delta t} \right) - 1, \tag{2.122}$$

from which we obtain

$$\omega_{opt} = \frac{1}{1 + \dfrac{\mu_1 + \mu_{N_h}}{2\Delta t}}. \tag{2.123}$$

2.6.2 The 2D case with the elastic pressure parameter

We repeat here the above analysis in the 2D case with the elastic pressure parameter. At each time level, the proposed algorithm requires the iterative solution of the following steps (see (2.76), (2.77) and (2.78)), till convergence is reached.

- Mechanical part:

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \begin{bmatrix} \mathbf{U}^{(k)} \\ \mathbf{P}_{el}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{n+1} - B^T \mathbf{P}^{(k-1)} \\ \mathbf{0} \end{bmatrix}. \tag{2.124}$$

- Fluid part:

$$D\widetilde{\mathbf{P}}^{(k)} = \mathbf{W}_{n+1} + \frac{1}{\Delta t} C\mathbf{P}_{el,n} - \frac{1}{\Delta t} C\mathbf{P}_{el}^{(k)}. \tag{2.125}$$

- Relaxation:

$$\mathbf{P}^{(k)} = \omega \widetilde{\mathbf{P}}^{(k)} + (1-\omega)\mathbf{P}^{(k-1)}. \tag{2.126}$$

System (2.124) can be rewritten as

$$\begin{cases} A\mathbf{U}^{(k)} + B^T \mathbf{P}_{el}^{(k)} = \mathbf{F}_{n+1} - B^T \mathbf{P}^{(k-1)}, \\ B\mathbf{U}^{(k)} + C\mathbf{P}_{el}^{(k)} = \mathbf{0}. \end{cases} \tag{2.127}$$

Since C is symmetric and negative definite, from the second equation of (2.127) we have

$$\mathbf{P}_{el}^{(k)} = -C^{-1} B \mathbf{U}^{(k)}. \tag{2.128}$$

Substituting (2.128) into the first equation of (2.127) and noting that the matrix $A - B^T C^{-1} B$ is symmetric and positive definite, we get

$$\mathbf{U}^{(k)} = (A - B^T C^{-1} B)^{-1} (\mathbf{F}_{n+1} - B^T \mathbf{P}^{(k-1)}). \tag{2.129}$$

Now, substituting (2.129) into (2.128) leads to

$$\mathbf{P}_{el}^{(k)} = -C^{-1} B (A - B^T C^{-1} B)^{-1} (\mathbf{F}_{n+1} - B^T \mathbf{P}^{(k-1)}). \tag{2.130}$$

Using this expression into (2.125) and noting that D is symmetric and positive definite, we have

$$\begin{aligned} \widetilde{\mathbf{P}}^{(k)} =& D^{-1}(\mathbf{W}_{n+1} + \frac{1}{\Delta t} C\mathbf{P}_{el,n}) \\ &+ \frac{1}{\Delta t} D^{-1} B (A - B^T C^{-1} B)^{-1} (\mathbf{F}_{n+1} - B^T \mathbf{P}^{(k-1)}). \end{aligned} \tag{2.131}$$

Finally, substituting (2.131) into (2.126), we obtain

$$\mathbf{P}^{(k)} = B_\omega \mathbf{P}^{(k-1)} + \mathbf{g}_\omega, \tag{2.132}$$

where the iteration matrix B_ω and the vector $\mathbf{g}_\omega$ are given by

$$B_\omega = (1-\omega)I - \frac{\omega}{\Delta t} D^{-1} B (A - B^T C^{-1} B)^{-1} B^T, \tag{2.133}$$

$$\mathbf{g}_\omega = \omega D^{-1} \left(\frac{1}{\Delta t} B (A - B^T C^{-1} B)^{-1} \mathbf{F}_{n+1} + \mathbf{W}_{n+1} + \frac{1}{\Delta t} C \mathbf{P}_{el,n} \right). \tag{2.134}$$

The iterative method converges if and only if condition (2.102) is satisfied. Since the matrix $D^{-1}B(A - B^T C^{-1} B)^{-1} B^T$ has real and positive eigenvalues (as confirmed by numerical tests), the largest of which is denoted by λ_1, we proceed as in the 1D case, obtaining

$$0 < \omega < \omega_{max} = \frac{2}{1 + \dfrac{\lambda_1}{\Delta t}}. \tag{2.135}$$

Similar comments to those made for the 1D case in section 2.6.1 can be repeated for the current analysis, in order to show how the degree of relaxation depends on physical and simulation parameters. We conclude by pointing out that the difference between the two cases studied does not lie in the dimensionality of the problem (1D or 2D), but in the presence or absence of the elastic pressure parameter in the model.

Chapter 3

Numerical tests

In this chapter, we illustrate the results emerging from the numerical simulation of linear poroelastic problems. Firstly, we intend to validate our 2D code by performing a convergence analysis of the solution provided by the relaxed fixed point iteration algorithm. Secondly, we show that the maximum value of the relaxation parameter predicted by theory is confirmed by numerical tests, using both the 1D code and the 2D one. Thirdly, we deal with the problem of locking in linear poroelasticity, specifying the differences with respect to linear elasticity. For the first goal, we use two test cases which have no immediate physical interpretation, but are useful to set up the convergence analysis. Instead, for the second and third goals, we refer to a significant biological application, namely a linear poroelastic model for articular cartilage. All the dimensional quantities are expressed using the CGS system.

3.1 Convergence analysis

In order to carry out a convergence analysis to validate our 2D code, we construct two linear poroelastic problems with analytical solutions. We consider the equations of linear poroelasticity for a material with incompressible components, namely

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla\cdot\mathbf{u}\mathbf{I}) - \nabla p = -\mathbf{f} & \text{in } \Omega\times(0,T), \\ \dfrac{\partial(\nabla\cdot\mathbf{u})}{\partial t} - \nabla\cdot(K\nabla p) = w & \text{in } \Omega\times(0,T), \end{cases} \tag{3.1}$$

defined on the rectangular domain

$$\Omega = (0,L)\times(-d/2,d/2)\subset\mathbb{R}^2, \tag{3.2}$$

whose boundary is divided into four portions, as depicted in figure 3.1, so that

$$\partial\Omega = \overline{\Sigma}_{d/2}\cup\overline{\Sigma}_L\cup\overline{\Sigma}_{-d/2}\cup\overline{\Sigma}_0. \tag{3.3}$$

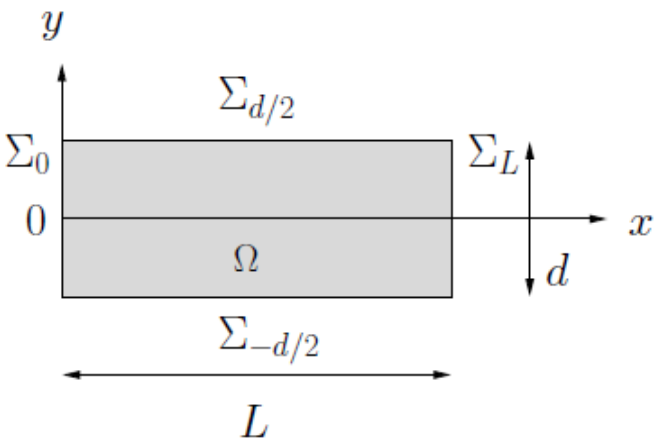


Figure 3.1: Geometrical domain for the convergence analysis.

We take into account the following boundary conditions for the solid component and the fluid one, where the quantities $\mathbf{t}_t$, $\mathbf{t}_r$, $\mathbf{t}_b$, P_r and P_l are given functions of $\mathbf{x}$ and t.

- On $\Sigma_{d/2} \times (0,T)$:

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_t, \\ \nabla p \cdot \mathbf{n} = 0. \end{cases} \tag{3.4}$$

- On $\Sigma_L \times (0,T)$:

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_r, \\ p = P_r. \end{cases} \tag{3.5}$$

- On $\Sigma_{-d/2} \times (0,T)$:

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_b, \\ \nabla p \cdot \mathbf{n} = 0. \end{cases} \tag{3.6}$$

- On $\Sigma_0 \times (0,T)$:

$$\begin{cases} \mathbf{u} = \mathbf{0}, \\ p = P_l. \end{cases} \tag{3.7}$$

The initial condition $\mathbf{u}_0$ is obtained by substituting $t = 0$ in the analytical solution of the problems. On the other hand, since we can take any guess for p_0, we generally set $p_0 = 0$.

We propose two test cases, one for the study of convergence with respect to space discretization and the other for the study of convergence with respect to time discretization.

3.1.1 Convergence with respect to space discretization

Dealing with convergence with respect to space discretization, it is useful to consider a stationary solution of linear poroelastic equations. We take

$$\mathbf{u} = \begin{bmatrix} 0 \\ (a_u x^2 + b_u x)e^{-d_u x} \end{bmatrix}, \tag{3.8}$$

$$p = (a_p x^2 + b_p x + c_p)e^{d_p x}, \tag{3.9}$$

where

$$b_u = \frac{a_u L(d_u L - 2)}{1 - d_u L}, \tag{3.10}$$

$$c_p = -a_p L^2 - b_p L. \tag{3.11}$$

The corresponding forcing terms are given by

$$\mathbf{f} = \begin{bmatrix} (a_p d_p x^2 + (2a_p + b_p d_p)x + b_p + c_p d_p)e^{d_p x} \\ -\mu(a_u d_u^2 x^2 + (b_u d_u^2 - 4a_u d_u)x + 2a_u - 2b_u d_u)e^{-d_u x} \end{bmatrix}, \tag{3.12}$$

$$w = -K(a_p d_p^2 x^2 + (4a_p d_p + b_p d_p^2)x + 2a_p + 2b_p d_p + c_p d_p^2)e^{d_p x}, \tag{3.13}$$

while the corresponding boundary conditions are those introduced in (3.4)-(3.7), where we take

$$\mathbf{t}_t = \begin{bmatrix} -\mu(a_u d_u x^2 - (2a_u - b_u d_u)x - b_u)e^{-d_u x} \\ -(a_p x^2 + b_p x + c_p)e^{d_p x} \end{bmatrix}, \tag{3.14}$$

$$\mathbf{t}_r = \mathbf{0}, \tag{3.15}$$

$$\mathbf{t}_b = \begin{bmatrix} \mu(a_u d_u x^2 - (2a_u - b_u d_u)x - b_u)e^{-d_u x} \\ (a_p x^2 + b_p x + c_p)e^{d_p x} \end{bmatrix}, \tag{3.16}$$

$$P_r = 0, \tag{3.17}$$

$$P_l = c_p. \tag{3.18}$$

We run simulations from $t = 0$ to $t = 20$ on the domain defined by $L = 0.075$ and $d = 0.02$, considering $a_u = 1$, $d_u = 20$, $a_p = 1$, $b_p = -1$, $dp = 10$ and the physical and simulation parameters reported in tables 3.1 and 3.2. The choice of the domain size is in agreement with the application to the lamina cribrosa, that we will describe later. The parameter E in table 3.1 denotes the Young modulus of the poroelastic material. It is associated with the Lamé's parameter λ and the Poisson ratio ν by

$$\lambda = \frac{\mu(E - 2\mu)}{3\mu - E}, \tag{3.19}$$

$$\nu = \frac{E}{2\mu} - 1. \tag{3.20}$$

Parameters	Values	Units
μ	$4 \cdot 10^6$	[dyn/cm^2]
E	$9.6 \cdot 10^6$	[dyn/cm^2]
K	$1.6 \cdot 10^{-4}$	[cm^3 s/g]

Table 3.1: Values of physical parameters for the convergence analysis.

Parameters	Values
ϵ_{u_1}, ϵ_{u_2}, $\epsilon_{\mathcal{P}}$ and ϵ_p	10^{-6}
maximum number of iterations	1000
ω	1

Table 3.2: Values of some simulation parameters for the convergence analysis.

Using the values of μ and E in the table, we have $\lambda \simeq 2.667 \cdot 10^6$ and $\nu \simeq 0.200$. From table 3.2, we note that the same convergence tolerance is used for all the variables. Furthermore, we observe that we set $\omega = 1$, that is no relaxation is performed in the present example. Since we have a stationary solution, we are not forced to use a small value of Δt, so that we consider $\Delta t = 2$. The numerical solution at $t = 20$, computed by the 2D code implementing the relaxed fixed point iteration algorithm on a given mesh, is shown in figures 3.2 and 3.3, which refer to the solid displacement and the fluid pressure, respectively.

In order to estimate the order of convergence with respect to the grid size h, we simulate the problem on increasingly refined grids and compute the numerical errors associated with the variables in both H^1 and L^2 norms (errors refer to the numerical solution at the final time). Each grid is obtained by meshing the domain with squares of size h, which are then divided into two triangles. Numerical errors are plotted versus h using logarithmic scales for both the axes and a linear fitting of the data thus obtained provides the order of convergence looked for. The results of this analysis are shown in tables 3.3 and 3.4 and in figures 3.4 and 3.5. We see that the estimate of the order of convergence with respect to h for the approximation of the displacement $\mathbf{u}$ is very close to 2, when computing errors in the H^1 norm. On the other hand, the estimate of the order of convergence with respect to h for the approximation of the fluid pressure p is very close to 1, when computing errors in the H^1 norm, and very close to 2, when computing errors in the L^2 norm. There is an excellent agreement between these results and Theorems 4 and 7 about convergence of the Galerkin finite element method.

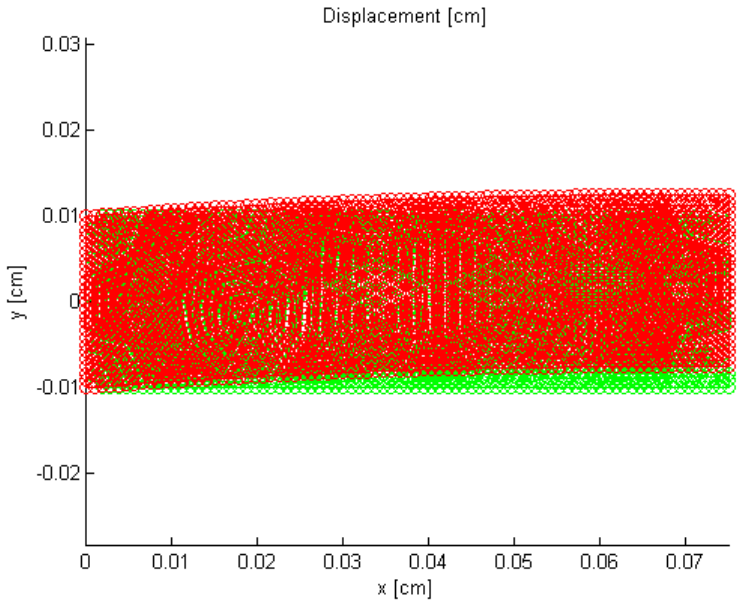


Figure 3.2: Approximation of the displacement $\mathbf{u}$ at $t = 20$ for the first test case, shown in terms of initial (green) and deformed (red) configurations.

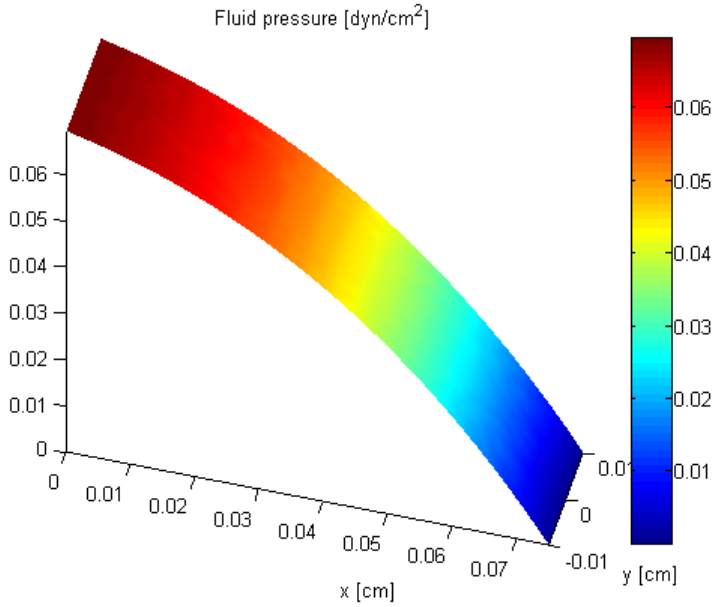


Figure 3.3: Approximation of the fluid pressure p at $t = 20$ for the first test case.

h	$\|\mathbf{u}-\mathbf{u}_h\|_{[H^1(\Omega)]^2}$	$\|\mathbf{u}-\mathbf{u}_h\|_{[L^2(\Omega)]^2}$
0.005	$4.66507 \cdot 10^{-7}$	$5.05680 \cdot 10^{-10}$
0.0025	$1.18552 \cdot 10^{-7}$	$5.15478 \cdot 10^{-11}$
0.00125	$2.98621 \cdot 10^{-8}$	$5.96577 \cdot 10^{-12}$
0.000625	$7.49230 \cdot 10^{-9}$	$7.29954 \cdot 10^{-13}$

Table 3.3: Numerical errors for the approximation of the displacement $\mathbf{u}$ for the first test case.

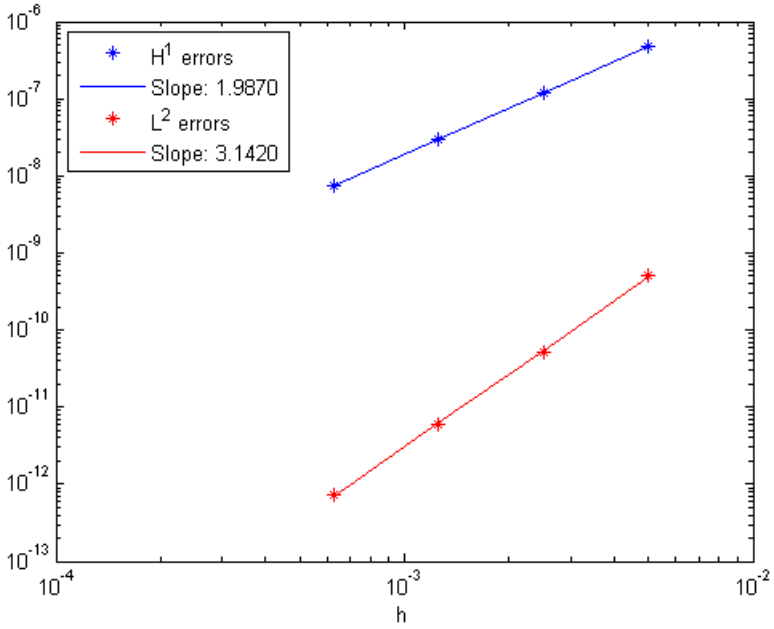


Figure 3.4: Estimates of the orders of convergence with respect to h for the approximation of the displacement $\mathbf{u}$ for the first test case.

h	$\|p - p_h\|_{H^1(\Omega)}$	$\|p - p_h\|_{L^2(\Omega)}$
0.005	$1.16105 \cdot 10^{-3}$	$1.83875 \cdot 10^{-6}$
0.0025	$5.80690 \cdot 10^{-4}$	$4.59901 \cdot 10^{-7}$
0.00125	$2.90366 \cdot 10^{-4}$	$1.14989 \cdot 10^{-7}$
0.000625	$1.45186 \cdot 10^{-4}$	$2.87481 \cdot 10^{-8}$

Table 3.4: Numerical errors for the approximation of the fluid pressure p for the first test case.

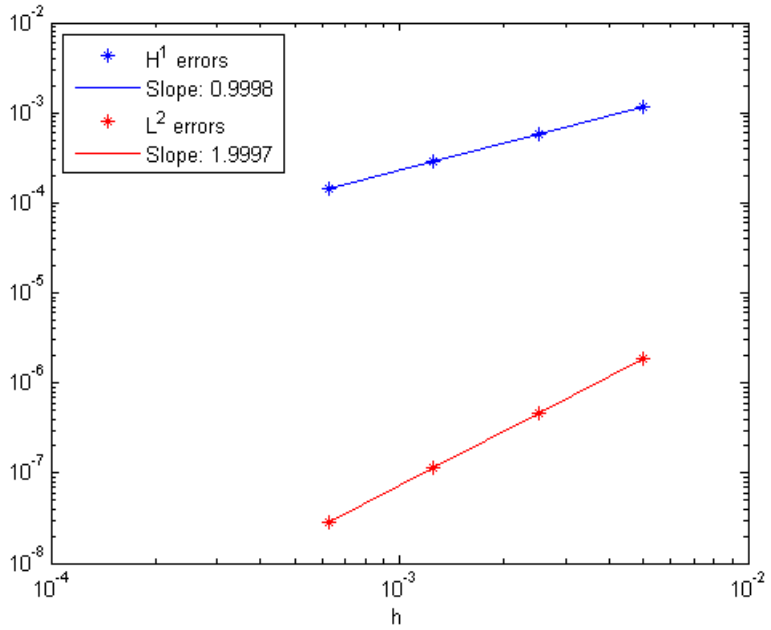


Figure 3.5: Estimates of the orders of convergence with respect to h for the approximation of the fluid pressure p for the first test case.

3.1.2 Convergence with respect to time discretization

Dealing with convergence with respect to time discretization, we consider a time-dependent solution of linear poroelastic equations. We take

$$\mathbf{u} = \begin{bmatrix} a_x x \sin(2\pi t) \\ (a_y y + b_y) x \sin(2\pi t) \end{bmatrix}, \tag{3.21}$$

$$p = (a_p x^2 + b_p x + c_p) \cos(2\pi t). \tag{3.22}$$

The corresponding forcing terms are given by

$$\mathbf{f} = \begin{bmatrix} -(\mu + \lambda) a_y \sin(2\pi t) + (2 a_p x + b_p) \cos(2\pi t) \\ 0 \end{bmatrix}, \tag{3.23}$$

$$w = 2\pi (a_y x + a_x) \cos(2\pi t) - 2K a_p \cos(2\pi t), \tag{3.24}$$

while the corresponding boundary conditions are defined by

$$\mathbf{t}_t = \begin{bmatrix} \mu (0.5 a_y d + b_y) \sin(2\pi t) \\ ((2\mu + \lambda) a_y x + \lambda a_x) \sin(2\pi t) - (a_p x^2 + b_p x + c_p) \cos(2\pi t) \end{bmatrix}, \tag{3.25}$$

$$\mathbf{t}_r = \begin{bmatrix} ((2\mu + \lambda) a_x + \lambda a_y L) \sin(2\pi t) - (a_p L^2 + b_p L + c_p) \cos(2\pi t) \\ \mu (a_y y + b_y) \sin(2\pi t) \end{bmatrix}, \tag{3.26}$$

$$\mathbf{t}_b = \begin{bmatrix} \mu (0.5 a_y d - b_y) \sin(2\pi t) \\ -((2\mu + \lambda) a_y x + \lambda a_x) \sin(2\pi t) + (a_p x^2 + b_p x + c_p) \cos(2\pi t) \end{bmatrix}, \tag{3.27}$$

$$P_r = (a_p L^2 + b_p L + c_p) \cos(2\pi t), \tag{3.28}$$

$$P_l = c_p \cos(2\pi t). \tag{3.29}$$

We run simulations from $t = 0$ to $t = 0.05$ on the same domain as in the first test case, considering $a_x = a_y = b_y = a_p = b_p = c_p = 0.1$ and the physical and simulation parameters reported in tables 3.1 and 3.2. We adopt a fixed unstructured mesh for all the simulations of the second test case. The numerical solution at $t = 0.05$, computed by the 2D code using $\Delta t = 10^{-4}$, is shown in figures 3.6 and 3.7.

In order to estimate the order of convergence with respect to the time step Δt, we simulate the problem using increasingly smaller time steps and analyse the numerical errors (referred to the final time) as done for the first test case. The results are shown in tables 3.5 and 3.6 and in figures 3.8 and 3.9, where we see that the order of convergence with respect to Δt is estimated to be very close to 1 for all the variables, computing errors in any norm, as we expect because of the use of the backward Euler method. Therefore, we conclude that the results of the simulation of the first and second test cases validate the 2D code implementing the iterative algorithm. Similar results for both cases are obtained when choosing the mechanical parameters so that ν is close to 0.5 (this corresponds to taking very large values of λ), thus confirming the robustness of the algorithm in dealing with such limiting case.

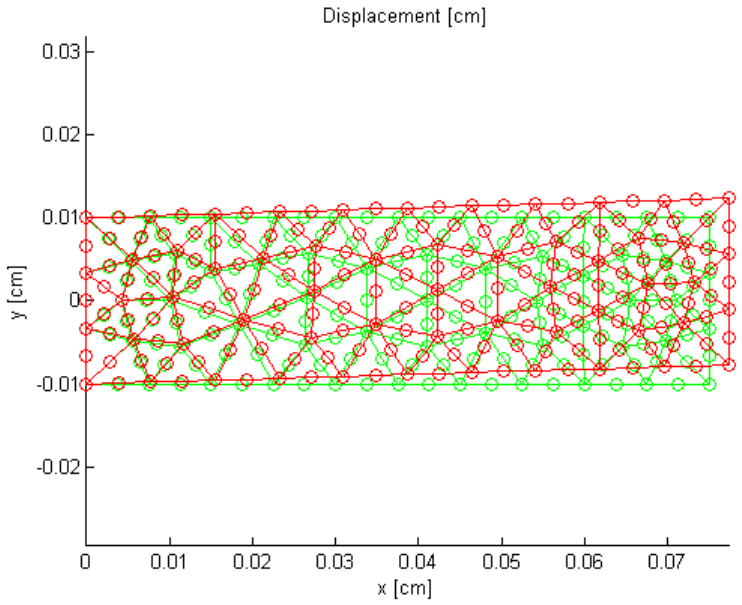


Figure 3.6: Approximation of the displacement $\mathbf{u}$ at $t = 0.05$ for the second test case, shown in terms of initial (green) and deformed (red) configurations.

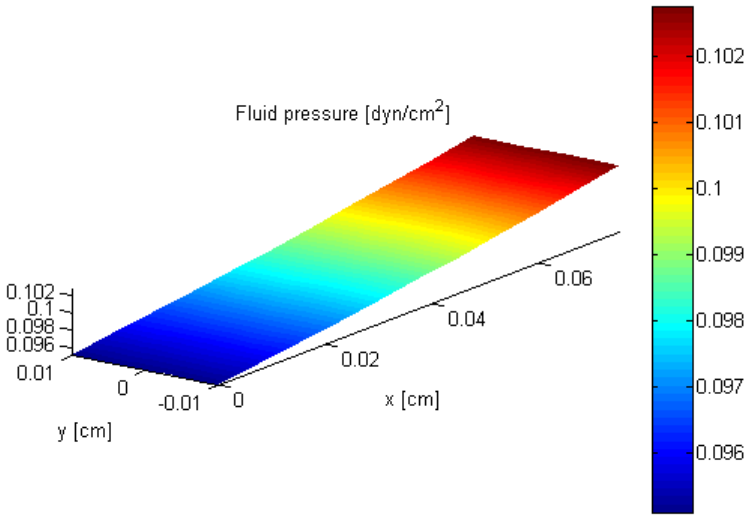


Figure 3.7: Approximation of the fluid pressure p at $t = 0.05$ for the second test case.

Δt	$\|\mathbf{u}-\mathbf{u}_h\|_{[H^1(\Omega)]^2}$	$\|\mathbf{u}-\mathbf{u}_h\|_{[L^2(\Omega)]^2}$
0.001	$8.44437 \cdot 10^{-12}$	$2.45266 \cdot 10^{-13}$
0.0005	$4.23461 \cdot 10^{-12}$	$1.22988 \cdot 10^{-13}$
0.00025	$2.11946 \cdot 10^{-12}$	$6.15520 \cdot 10^{-14}$
0.000125	$1.05930 \cdot 10^{-12}$	$3.07576 \cdot 10^{-14}$
0.0000625	$5.28592 \cdot 10^{-13}$	$1.53429 \cdot 10^{-14}$

Table 3.5: Numerical errors for the approximation of the displacement $\mathbf{u}$ for the second test case.

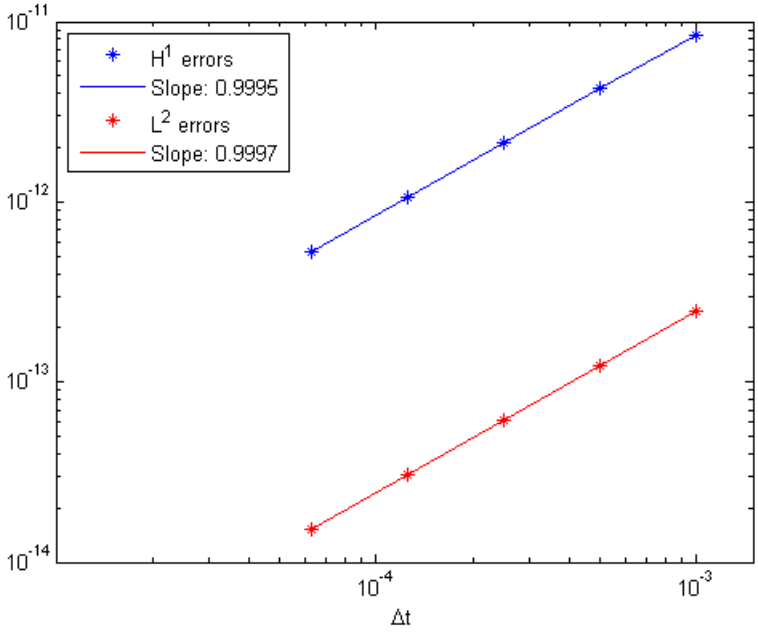


Figure 3.8: Estimate of the order of convergence with respect to Δt for the approximation of the displacement $\mathbf{u}$ for the second test case.

Δt	$\|p - p_h\|_{H^1(\Omega)}$	$\|p - p_h\|_{L^2(\Omega)}$
0.001	$3.28463 \cdot 10^{-3}$	$7.76336 \cdot 10^{-5}$
0.0005	$1.64769 \cdot 10^{-3}$	$3.89309 \cdot 10^{-5}$
0.00025	$8.25258 \cdot 10^{-4}$	$1.94849 \cdot 10^{-5}$
0.000125	$4.13125 \cdot 10^{-4}$	$9.73835 \cdot 10^{-6}$
0.0000625	$2.06972 \cdot 10^{-4}$	$4.85916 \cdot 10^{-6}$

Table 3.6: Numerical errors for the approximation of the fluid pressure p for the second test case.

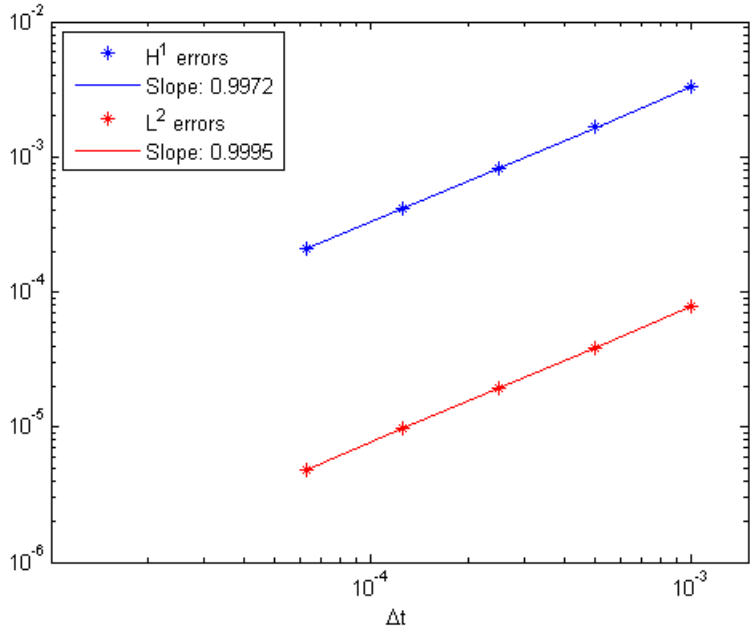


Figure 3.9: Estimate of the order of convergence with respect to Δt for the approximation of the fluid pressure p for the second test case.

3.2 Analysis of the relaxed iterative algorithm

In this section, we analyse the dependence of the relaxation parameter on physical and simulation parameters. We aim at showing that the maximum value of the relaxation parameter predicted by the analysis of chapter 2 is confirmed by numerical tests, both in the 1D case without the elastic pressure parameter and in the 2D case with the elastic pressure parameter.

3.2.1 An application to articular cartilage

Although the theory of poroelasticity arose in the context of geomechanics, over the last years it has been applied to the study of several biological tissues. An important application is that to articular cartilage, which is a soft hydrated tissue particularly suitable to be described by a poroelastic model. In fact, cartilage is formed by a collagen-proteoglycan matrix and an interstitial fluid, which consists mainly of water. We adopt the 1D linear poroelastic model proposed by Soltz and Ateshian [45] for the description of a confined compression creep experiment.

In this experiment, a cartilage specimen is subjected to a constant load and prevented from lateral expansion, so that we can consider the problem as one-dimensional. The 1D linear poroelastic equations

$$\begin{cases} \dfrac{\partial \sigma}{\partial x} = 0 & \text{in } (0,L)\times(0,+\infty), \\ \sigma = H_A \dfrac{\partial u}{\partial x} - p & \text{in } (0,L)\times(0,+\infty), \\ \dfrac{\partial}{\partial t}\dfrac{\partial u}{\partial x} + \dfrac{\partial v}{\partial x} = 0 & \text{in } (0,L)\times(0,+\infty), \\ v = -K\dfrac{\partial p}{\partial x} & \text{in } (0,L)\times(0,+\infty) \end{cases} \tag{3.30}$$

are completed by the boundary conditions for the solid component

$$\begin{cases} u(0,t) = 0, \\ \sigma(L,t) = -P_A, \end{cases} \tag{3.31}$$

those for the fluid component

$$\begin{cases} v(0,t) = 0, \\ p(L,t) = 0, \end{cases} \tag{3.32}$$

and the initial condition $u(x,0) = 0$. These conditions express the presence of a fixed impermeable interface to the left end and of a free-draining porous interface subject to a constant compressive load to the right end. It is possible to check that this 1D linear poroelastic problem can be reduced to the

following equivalent form

$$\begin{cases} \dfrac{\partial u}{\partial t} - H_A K \dfrac{\partial^2 u}{\partial x^2} = 0 & \text{in } (0,L)\times(0,+\infty), \\ u(0,t) = 0 & \text{for } t \in (0,+\infty), \\ H_A \dfrac{\partial u}{\partial x}(L,t) = -P_A & \text{for } t \in (0,+\infty), \\ u(x,0) = 0 & \text{for } x \in (0,L), \end{cases} \tag{3.33}$$

whose exact solution is

$$u(x,t) = -\frac{P_A}{H_A}\left[x - \frac{2L}{\pi^2}\sum_{n=0}^{+\infty}\frac{(-1)^n}{(n+0.5)^2}\sin\left[(n+0.5)\frac{\pi x}{L}\right]e^{-\frac{H_A K}{L^2}(n+0.5)^2\pi^2 t}\right]. \tag{3.34}$$

From the first equation of (3.30), we see that the total stress σ does not depend on x. Then, since $\sigma(L,t) = -P_A$ for all $t \in (0,+\infty)$, we have $\sigma = -P_A$ for all $x \in (0,L)$ and $t \in (0,+\infty)$. Using the second equation of (3.30) and the analytical solution for the displacement (3.34), we get the analytical solution for the fluid pressure

$$p(x,t) = \frac{2P_A}{\pi}\sum_{n=0}^{+\infty}\frac{(-1)^n}{n+0.5}\cos\left[(n+0.5)\frac{\pi x}{L}\right]e^{-\frac{H_A K}{L^2}(n+0.5)^2\pi^2 t}. \tag{3.35}$$

We consider the physical parameters given in [45] and reported in table 3.7. We run the 1D code implementing the relaxed fixed point iteration algo-

Parameters	Values	Units
L	0.1	[cm]
H_A	$0.97 \cdot 10^7$	[dyn/cm^2]
K	$2.9 \cdot 10^{-13}$	[cm^3 s/g]
P_A	$60 \cdot 10^4$	[dyn/cm^2]

Table 3.7: Values of physical parameters for the simulation of the confined compression creep experiment.

rithm, in order to simulate the confined compression creep experiment from $t = 0$ to $t = 10^4$, a sufficiently large time to reach a final equilibrium state. In figures 3.10, 3.11 and 3.12, we compare the analytical solution given by (3.34) and (3.35) with the numerical solution obtained using a mesh consisting of 30 elements, $\Delta t = 200$, $\epsilon_u = \epsilon_p = 10^{-6}$, $\omega = 0.1$ and a maximum number of iterations equal to 100. We see that there is a very good agreement between the exact solution and that computed by the 1D code. Furthermore, almost identical results are obtained by running the monolithic version of

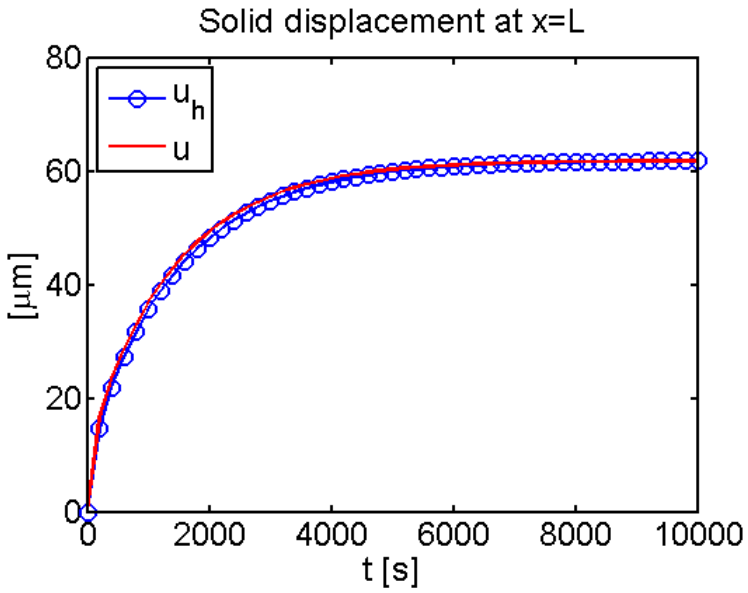


Figure 3.10: Solid displacement at $x = L$ in the confined compression creep experiment.

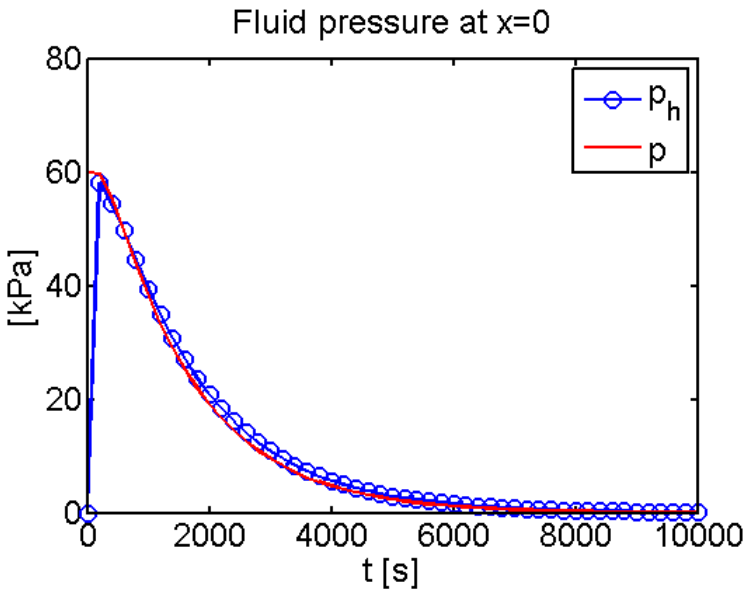


Figure 3.11: Fluid pressure at $x = 0$ in the confined compression creep experiment.

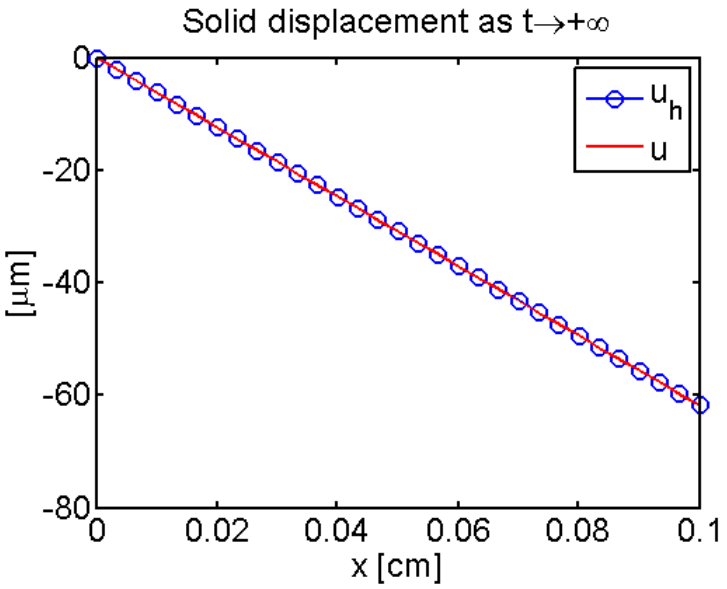


Figure 3.12: Solid displacement as $t \to +\infty$ in the confined compression creep experiment.

the 1D code, thus confirming that the same results for monolithic and iterative solvers have to be expected at convergence. From figure 3.11, we note that the fluid pressure plays a fundamental role in articular cartilage, supporting the great majority of the applied load in the early stage of the time response of the tissue. When this pressure subsides, the tissue reaches its equilibrium deformation, where all of the applied load is supported by the collagen-proteoglycan matrix. A very low permeability delays the loss of interstitial fluid pressurization by retarding the rate of fluid flow away from the pressurized region.

3.2.2 1D computations

Now, we want to show that the values of ω_{max} predicted by the theory developed in chapter 2 are verified by numerical tests. We start from the 1D case without the elastic pressure parameter. We consider the test case corresponding to the confined compression creep experiment and run the 1D code from $t = 0$ to $t = 10^4$. In all simulations we take $\epsilon_u = \epsilon_p = 10^{-6}$ and a maximum number of iterations equal to 1000. We summarize the results of our analysis in tables 3.8, 3.9, 3.10 and 3.11. In each table, only one parameter is allowed to vary, in order to characterize the dependence of ω_{max} on that specific parameter. The theoretical value of ω_{max} is computed by estimating the maximum eigenvalue μ_1 appearing in (2.107), while the numerical value is computed by observing when convergence is reached (more

precisely, we report a range in which the numerical value falls).

We see that the correctness of the theoretical analysis for the 1D case without the elastic pressure parameter is confirmed by numerical tests, which show that ω_{max} does not depend on h, but depends on H_A, K and Δt. In particular, a decrease of H_A, K or Δt leads to a decrease of ω_{max} and this means that more relaxation is needed to reach convergence.

H_A	theoretical ω_{max}	numerical ω_{max}
$0.5 \cdot 10^7$	0.1341	$0.13 - 0.14$
$0.97 \cdot 10^7$	0.2447	$0.24 - 0.25$
$2 \cdot 10^7$	0.4465	$0.44 - 0.45$

Table 3.8: Analysis of the dependence of ω_{max} on H_A (fixed parameters: $K = 2.9 \cdot 10^{-13}$, $h = 0.01$, $\Delta t = 200$).

K	theoretical ω_{max}	numerical ω_{max}
$1.5 \cdot 10^{-13}$	0.1345	$0.13 - 0.14$
$2.9 \cdot 10^{-13}$	0.2447	$0.24 - 0.25$
$5.8 \cdot 10^{-13}$	0.4360	$0.43 - 0.44$

Table 3.9: Analysis of the dependence of ω_{max} on K (fixed parameters: $H_A = 0.97 \cdot 10^7$, $h = 0.01$, $\Delta t = 200$).

h	theoretical ω_{max}	numerical ω_{max}
0.005	0.2440	$0.24 - 0.25$
0.01	0.2447	$0.24 - 0.25$
0.02	0.2474	$0.24 - 0.25$

Table 3.10: Analysis of the dependence of ω_{max} on h (fixed parameters: $H_A = 0.97 \cdot 10^7$, $K = 2.9 \cdot 10^{-13}$, $\Delta t = 200$).

Δt	theoretical ω_{max}	numerical ω_{max}
100	0.1303	$0.12 - 0.14$
200	0.2447	$0.24 - 0.25$
400	0.4360	$0.43 - 0.44$

Table 3.11: Analysis of the dependence of ω_{max} on Δt (fixed parameters: $H_A = 0.97 \cdot 10^7$, $K = 2.9 \cdot 10^{-13}$, $h = 0.01$).

3.2.3 2D computations

Now, we turn to the 2D case with the elastic pressure parameter. In order to consider the same test case corresponding to the confined compression creep experiment, we reformulate the problem as if it were a 2D problem defined on the rectangular domain $\Omega = (0, L) \times (-d/2, d/2)$, where $L = 0.1$ and we set $d = 0.02$. The corresponding boundary conditions on the left side become

$$\begin{cases} \mathbf{u} = \mathbf{0}, \\ \nabla p \cdot \mathbf{n} = 0, \end{cases} \tag{3.36}$$

on the right side

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \begin{bmatrix} -P_A \\ 0 \end{bmatrix}, \\ p = 0, \end{cases} \tag{3.37}$$

on the upper side

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_t, \\ \nabla p \cdot \mathbf{n} = 0, \end{cases} \tag{3.38}$$

on the lower side

$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = \mathbf{t}_b, \\ \nabla p \cdot \mathbf{n} = 0, \end{cases} \tag{3.39}$$

where $\mathbf{t}_t$ and $\mathbf{t}_b$ are determined considering (3.34) and (3.35). We run the 2D code to simulate this problem from $t = 0$ to $t = 10^4$, using $\epsilon_{u_1} = \epsilon_{u_2} = \epsilon_{\mathcal{P}} = \epsilon_p = 10^{-6}$ and a maximum number of iterations equal to 1000. We proceed as done for the 1D case and summarize the results in tables 3.12, 3.13, 3.14, 3.15 and 3.16. The computation of the theoretical value of ω_{max} is based on the estimate of the maximum eigenvalue λ_1 appearing in (2.135). All simulations are run using the same unstructured mesh, except those related to table 3.15, in which we use structured meshes. In the 2D case the aggregate modulus H_A is replaced by the Lamé's parameters μ and λ.

The theoretical analysis proves to be correct also for the 2D case with the elastic pressure parameter. Relaxation does not depend on the mesh, while a decrease of μ, λ, K or Δt leads to a decrease of ω_{max}.

μ	theoretical ω_{max}	numerical ω_{max}
$1.5 \cdot 10^6$	0.1404	$0.13 - 0.15$
$3 \cdot 10^6$	0.1783	$0.17 - 0.18$
$6 \cdot 10^6$	0.2494	$0.24 - 0.25$

Table 3.12: Analysis of the dependence of ω_{max} on μ (fixed parameters: $\lambda = 3.7 \cdot 10^6$, $K = 2.9 \cdot 10^{-13}$, h, $\Delta t = 200$).

λ	theoretical ω_{max}	numerical ω_{max}
$1.8 \cdot 10^6$	0.1317	$0.13 - 0.14$
$3.7 \cdot 10^6$	0.1783	$0.17 - 0.18$
$7.4 \cdot 10^6$	0.2624	$0.26 - 0.27$

Table 3.13: Analysis of the dependence of ω_{max} on λ (fixed parameters: $\mu = 3 \cdot 10^6$, $K = 2.9 \cdot 10^{-13}$, h, $\Delta t = 200$).

K	theoretical ω_{max}	numerical ω_{max}
$1.5 \cdot 10^{-13}$	0.0964	$0.09 - 0.1$
$2.9 \cdot 10^{-13}$	0.1783	$0.17 - 0.18$
$5.8 \cdot 10^{-13}$	0.3274	$0.32 - 0.33$

Table 3.14: Analysis of the dependence of ω_{max} on K (fixed parameters: $\mu = 3 \cdot 10^6$, $\lambda = 3.7 \cdot 10^6$, h, $\Delta t = 200$).

mesh	theoretical ω_{max}	numerical ω_{max}
20×8	0.1779	$0.17 - 0.18$
15×6	0.1781	$0.17 - 0.18$
10×4	0.1785	$0.17 - 0.18$

Table 3.15: Analysis of the dependence of ω_{max} on h (fixed parameters: $\mu = 3 \cdot 10^6$, $\lambda = 3.7 \cdot 10^6$, $K = 2.9 \cdot 10^{-13}$, $\Delta t = 200$).

Δt	theoretical ω_{max}	numerical ω_{max}
100	0.0933	$0.09 - 0.1$
200	0.1783	$0.17 - 0.18$
400	0.3274	$0.32 - 0.33$

Table 3.16: Analysis of the dependence of ω_{max} on Δt (fixed parameters: $\mu = 3 \cdot 10^6$, $\lambda = 3.7 \cdot 10^6$, $K = 2.9 \cdot 10^{-13}$, h).

3.3 The problem of locking

Finite element error estimates become unreliable when the constants appearing in them become large in correspondence of a particular combination of the parameters of the problem. When this happens, the numerical solution can be very different from the exact solution. This is known as the problem of locking. In linear elasticity, locking leads to some problems with the approximation of the displacement and the stress. In linear poroelasticity, it manifests itself in spurious, unphysical fluid pressure oscillations. In order to understand the reasons of locking in linear elasticity and poroelasticity, we follow the heuristic approach used by Phillips and Wheeler [33].

First, we consider the linear elastic case, described by

$$\nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla \cdot \mathbf{u}\mathbf{I}) = -\mathbf{f}. \tag{3.40}$$

In linear elasticity, locking is associated with incompressibility, characterized by $\lambda \to +\infty$ and $\nabla \cdot \mathbf{u} \to 0$. In this case, the use of continuous linear finite elements is problematic, because the only divergence-free vectors belonging to that space are constant vectors, but the boundary conditions may require a non-constant displacement field. Among the possible solutions to the problem of locking in linear elasticity, there is the use of discontinuous finite elements, as proposed in [33], since they constitute a space endowed with non-constant divergence-free vectors. Another possible remedy is the adoption of the Herrmann formulation, as widely discussed in chapter 2.

Turning to linear poroelasticity, the equation governing the fluid flow is given by

$$\frac{1}{M}\frac{\partial p}{\partial t} + \alpha\frac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (K\nabla p) = w. \tag{3.41}$$

In linear poroelasticity, locking is not associated with the parameter λ, but occurs when there is no fluid source term ($w = 0$), both components are incompressible ($M \to +\infty$ and $\alpha = 1$) and K and Δt are sufficiently small. Under these conditions, taking $K = \mathcal{O}(\epsilon)$ for some very small ϵ, time discretization of (3.41) reads

$$\frac{\nabla \cdot \mathbf{u}_{n+1} - \nabla \cdot \mathbf{u}_n}{\Delta t} - \epsilon = 0, \tag{3.42}$$

so that

$$\nabla \cdot \mathbf{u}_{n+1} = \nabla \cdot \mathbf{u}_n + \Delta t\epsilon. \tag{3.43}$$

If Δt is small and $\mathbf{u}_0 = \mathbf{0}$, then

$$\nabla \cdot \mathbf{u}_1 \simeq 0, \tag{3.44}$$

namely the displacement is driven to a divergence-free state and locking occurs. This is the same numerical problem found in linear elasticity. Therefore, discontinuous finite elements allow to overcome the problem of locking

also in linear poroelasticity, as shown again in [33]. Instead, the introduction of the elastic pressure parameter is not an effective remedy in linear poroelasticity, since locking is not associated with the parameter λ. We point out that, since the deformation becomes more pronounced over time, from a certain time level $\nabla \cdot \mathbf{u}_n$ is no longer negligible and $\nabla \cdot \mathbf{u}_{n+1} \neq 0$, so that the problem of locking in linear poroelasticity is limited to the early stages of the time evolution of the system.

We run the 1D code (which uses continuous linear finite elements for both displacement and fluid pressure) in order to monitor the occurrence of spurious fluid pressure oscillations when using very small values of permeability and time step. We consider the confined compression creep experiment of section 3.2 with the physical parameters reported in table 3.7. We fix the permeability K and use increasingly small time steps. Table 3.17 shows that locking occurs only when we use a very small time step. The corresponding spurious fluid pressure oscillations are shown in figure 3.13. The same results are obtained by running the monolithic and the iterative versions of the 1D code. It is important to point out that oscillations occur only for very small Δt, so that our code is not subject to significant limitations. Moreover, we see that oscillations disappear over time (for example, the fluid pressure at $t = 40$ has no oscillations).

Δt	locking
200	no
100	no
50	no
10	no
1	yes

Table 3.17: Numerical test for the problem of locking.

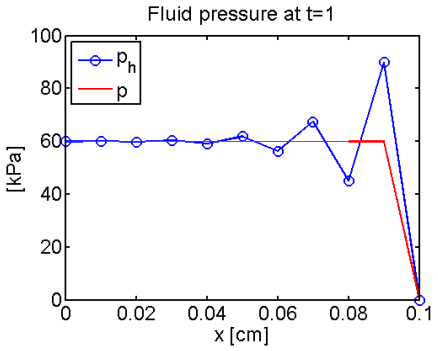


Figure 3.13: Spurious fluid pressure oscillations at $t = 1$ when using $\Delta t = 1$.

Part II

Application to ocular biomechanics

Chapter 4

Introduction to ocular biomechanics and hemodynamics

The study of the relationship between ocular mechanics and hemodynamics is an intriguing and challenging task. A better understanding of physical mechanisms which take place in the human eye is needed, in order to develop new therapeutic concepts for treating ocular diseases, such as glaucoma. Mathematical models can play a significant role in the interpretation of laboratory and clinical data and in the detection of the interplay between ocular mechanics and hemodynamics. In this chapter, we give a brief description of ocular physiology, dwelling upon the vascular network of the retina and the anatomy of the optic nerve head. Then, we illustrate the most important features of glaucoma, analysing different possible risk factors for this pathology. Finally, we give a survey of possible uses of mathematics in the study of the human eye.

4.1 Anatomy of the retina and the optic nerve head

Retina plays a crucial role in the mechanism of vision. It is a light-sensitive tissue lining the inner surface of the human eye. Light enters the eye passing through the cornea, a transparent structure in the front of the eye. Behind the cornea, there is a a coloured membrane (the iris) with an adjustable circular opening (the pupil) at its centre, which can expand or contract depending on the amount of light entering the eye. The space between the cornea and the iris is filled with a clear fluid, called the aqueous humor. The crystalline lens lies behind the iris and is surrounded by the ciliary muscles, which regulate the lens allowing to focus close and far objects. The interior chamber of the eye is filled with a clear gel, called the vitreous humor, and light has to pass through it before reaching the retina. Light striking the

retina initiates a cascade of chemical and electrical events that trigger nerve impulses, which are sent to the brain through the fibres of the optic nerve. Figure 4.1 shows the principal components of the human eye.

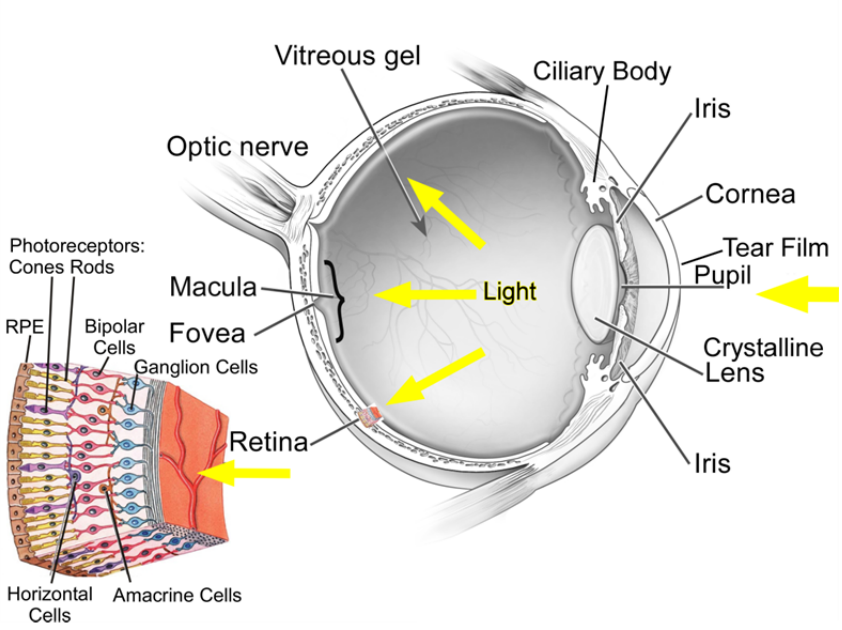


Figure 4.1: Cross-section of a left human eye, showing the components of the eye and those of the retina.

Retina is composed of several layers: the inner layers contain the ganglion cells, which are the output neurons of the retina, while the outer layers contain the photoreceptors, which are of two kinds (rods and cones). A chemical change in the rods and cones sends an electrical signal to the ganglion cells, whose axons take it to the brain through the fibres of the optic nerve. A schematic three-dimensional representation of retinal layers is contained again in figure 4.1.

Retinal vascular network is quite complex. We can distinguish between two sources of blood supply: the choroidal vessels and the central retinal artery (CRA). The choroid is the middle layer between the retina and the sclera, which is the outermost shell of the eye, and its vessels carry blood to the outer layers of the retina, containing the photoreceptors. On the contrary, the CRA nourishes the inner layers of the retina, containing the ganglion cells, through a lot of branches which originate from it. Blood supplies the retinal tissue with nutrients, especially oxygen, allowing the retinal cells to live and work correctly. The transport of nutrients takes place at the level of the capillaries, embedded in the retinal tissue.

As previously stated, the retinal ganglion cells are responsible for the

transmission of visual information from the retina to the brain, via the optic nerve. At the centre of the optic nerve head, there is a thin structure, the lamina cribrosa (LC), through which retinal ganglion cell axons pass. The lamina cribrosa is a collagen structure, pierced by the central retinal artery (CRA) and the central retinal vein (CRV) approximately in its centre. It is anchored to the sclera and maintains a pressure gradient between the intraocular space and the optic nerve canal. So, the lamina cribrosa is subject to the following tensile contributions (see figure 4.2):

- the intraocular pressure (IOP), which is the pressure inside the eye globe (baseline value 12-15 mmHg, as reported in [18]);
- the retrolaminar tissue pressure (RLT_p), which is the pressure inside the optic nerve canal (baseline value 7-10 mmHg, as reported in [18]);
- the scleral tension.

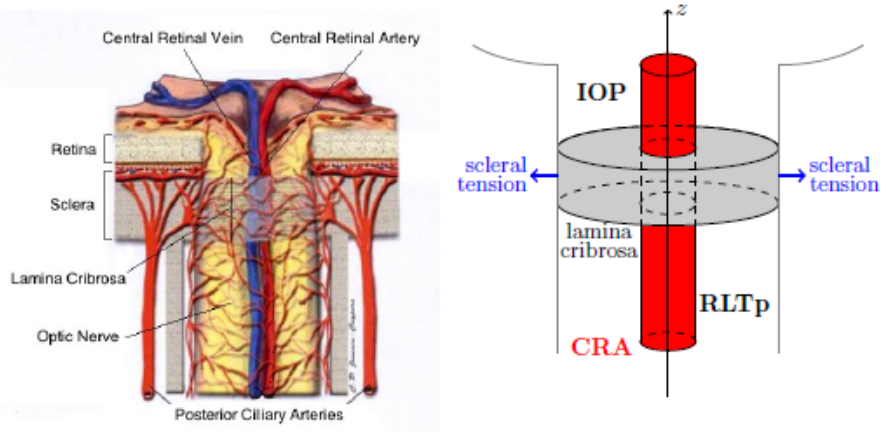


Figure 4.2: Left: anatomy of the optic nerve head. Right: schematic representation of the mechanical environment of the lamina cribrosa (taken from [18]).

The lamina cribrosa deforms under the combined action of IOP, RLT_p and scleral tension and the effects of its deformation are crucial in the study of glaucoma, as will be shown later. The pressure gradient originated by the difference between IOP and RLT_p may cause bowing of the lamina cribrosa toward the site of lower pressure. The scleral tension, exerted on the edge of the lamina cribrosa, helps to limit this bowing.

Figure 4.2 shows also which blood vessels supply blood to the lamina cribrosa, providing the axons passing through it with nutrients. The CRA, though passing through the lamina cribrosa, ensures blood supply only for

the retinal tissue. Instead, the lamina cribrosa is nourished by the capillary bed which originates from the posterior ciliary arteries, which also supply the choroid. Therefore, blood reaches the lamina cribrosa from its boundary and is drained by the central retinal vein in its centre. The arteries surrounding the optic nerve head are sometimes called the circle of Zinn-Haller.

Figure 4.3 shows the collagen structure of the lamina cribrosa, allowing the passage of retinal ganglion cell axons and central retinal vessels.

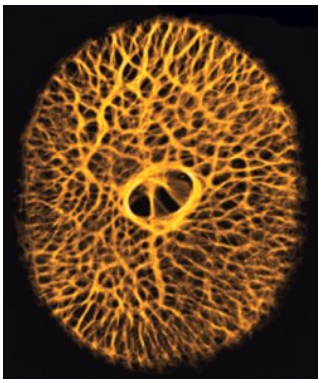


Figure 4.3: Scanning electron micrograph showing the collagen structure of the lamina cribrosa.

4.2 Glaucoma and its risk factors

Glaucoma is a family of optic neuropathies having in common a characteristic cupping of the optic nerve head, progressive retinal ganglion cell death and irreversible visual field loss. Glaucoma is the second cause of blindness in the world and medical treatment can only slow or stop its progression. There are different kinds of glaucoma. For example, open-angle glaucoma (OAG), which affects 1% of the population over 40 and more than 10% over 80, tends to progress at a slow rate and patients initially may not notice they have visual impairment. On the contrary, angle-closure glaucoma, which is less common, usually appears suddenly and in a painful way.

The role of increased IOP as a risk factor for glaucoma is now well established and current treatment of glaucoma aims at reducing IOP. A rise in IOP is associated with an imbalance between the production of aqueous humor and its drainage through the trabecular meshwork, a drainage system located around the base of the cornea, near the ciliary body. Administration of eye drops or other medications and surgery, such as trabeculectomy, aim at reducing IOP by correcting such imbalance. Elevated IOP is generally associated with glaucoma, because it is thought to cause deformation of the

optic nerve head, that is the characteristic cupping of the lamina cribrosa (see figure 4.4). In turn, retrodisplacement of the lamina cribrosa is thought to cause structural damage to the axons passing through it, but the relationship between lamina cribrosa deformation and axonal damage is not yet fully understood.

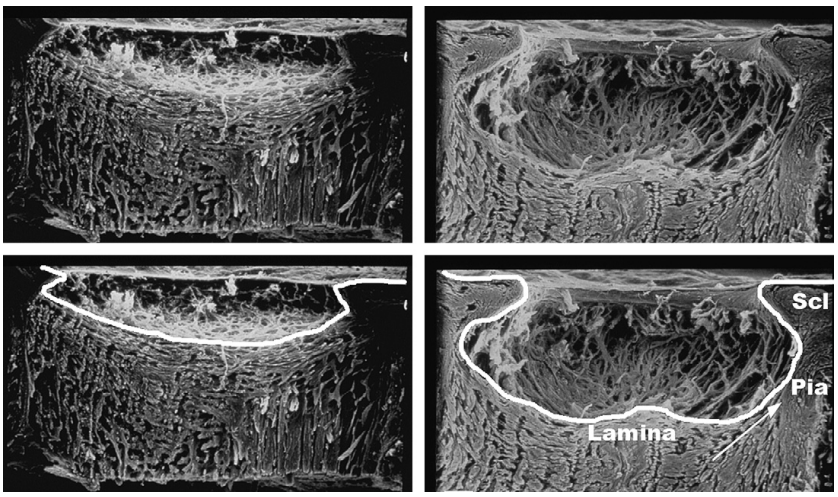


Figure 4.4: Scanning electron micrographs of optic nerve heads from normal (left) and advanced glaucoma (right) human eyes (taken from [11]).

Although to date elevated IOP is the only treatable risk factor for glaucoma, there is significant evidence that other factors might be involved in the disease. In fact, many individuals with elevated IOP never develop OAG, while many patients continue to progress to blindness despite normal values of IOP (this is the case of the so-called normal-tension glaucoma). In recent years, vascular factors have attracted the attention of the scientific community. Several studies suggest that alterations in ocular hemodynamics may play a significant role in the OAG pathology. In [4], Caprioli and Coleman provided a critical review of the relationships between blood pressure, blood flow and glaucoma, on the basis of selected literature, clinical measurements and population-based studies. They identified low blood pressure and low ocular perfusion pressure (OPP) as possible risk factors for glaucoma. The OPP is the driving force for ocular blood flow and is defined as

$$\mathrm{OPP} = \frac{2}{3}\,\mathrm{MAP} - \mathrm{IOP},$$

where MAP denotes the mean arterial pressure measured at the level of the brachial artery, so that a multiplicative factor is introduced in order to consider the hydrostatic column effect and obtain an estimated pressure

in the arteries of the eye. However, it is not clear whether low OPP is independent of the sum of the two separate risk factors of low MAP and high IOP.

A central point that is presently debated is to understand if ocular hemodynamic alterations are primary or secondary to optic nerve damage. In other words, there are two main hypotheses that try to explain the relationship between hemodynamic alterations and glaucoma:

- the first hypothesis is that the compression exerted by the lamina cribrosa on the CRA and the optic nerve head capillaries under elevated IOP causes an increase in the resistance to flow, leading to a reduction in blood flow; this reduction in blood supply causes the ischemic death of retinal ganglion cells;

- the second hypothesis is that retinal ganglion cell death is a direct consequence of the deformation of the optic nerve head under elevated IOP, which causes axonal damage; reduced blood flow is a response to the decrease in metabolic demand for oxygen and nutrients secondary to retinal ganglion cell death.

Clearly, if the first hypothesis will be proved, the development of therapeutic treatments aimed at modifying ocular hemodynamics will be of utmost importance.

Besides elevated IOP and vascular factors, there are also other possible risk factors for glaucoma. We want to underline the role played by the anatomy of ocular structures. Radius and thickness of the lamina cribrosa, for example, seem to be important in the response to IOP elevation. Structural differences could explain why certain individuals are more susceptible to glaucoma than others, even though they experience the same level of IOP. Furthermore, anatomical differences between individuals of different ethnicities could explain the increased incidence and prevalence of OAG in people of African descent compared to those of European descent.

4.3 The role of mathematical modeling

As seen in the previous section, more efforts are required to fully understand the complex phenomena associated with incidence and progression of ocular diseases, such as glaucoma. A deep knowledge of what occurs in the human eye is fundamental, in view of the important objective of considering new effective therapeutic strategies for these pathologies. Clinical, population-based and animal studies are the primary approaches implemented by the scientific and medical communities, but these approaches can be supported by another powerful tool, that is mathematical modeling. Mathematical models can aid the interpretation of data obtained in labs and clinics and, mostly, can be used as a virtual lab to simulate and elucidate the complex

relationships between different phenomena, such as ocular mechanics and hemodynamics.

In [19], Harris et al. identified four steps interpretable as potential directions of research involving mathematics:

- the description of the mechanical response of the optic nerve head and the LC to variations in IOP, RLT_p and scleral tension;
- the explanation of the basic hemodynamic principles governing blood flow in retinal, choroidal and optic nerve head vascular beds;
- the analysis of the mechanisms of blood flow autoregulation in retinal, choroidal and optic nerve head vascular beds;
- the identification of the relationship between glaucoma progression and hemodynamic alterations.

Over the last two decades, mathematical models describing the optic nerve head mechanical behaviour were presented, while mathematical models dealing with ocular hemodynamics and autoregulation are more recent. The fourth step is yet to be investigated.

For what concerns the first step, some models were presented to explain the LC mechanical behaviour. For example, Dongqi and Zeqin [14] proposed a linear elastic model for the LC and studied its response to elevation of the IOP. Newson and El-Sheikh [30] compared six idealized modeling techniques based on different assumptions, in order to identify the importance of each considered factor (such as the presence of a pre-stress due to scleral expansion and the degree of fixity offered by the sclera) in a model for the LC aimed to be realistic. Burgoyne et al. proposed several finite element models with the purpose of describing the biomechanical response of the LC (see [10] and [43]). Grytz et al. [17] tried to illustrate the micro-architecture of the collagen fibrils in the LC, using a computational remodeling approach. The mechanical response of the LC to variations in RLT_p has not been studied to the same extent as IOP; it is expected that IOP elevations and RLT_p reductions do not have equivalent effects on the optic nerve head deformation, because variations in IOP also affect the scleral tension. In fact, elevated IOP causes both bowing of the LC and scleral expansion, so that the consequent rise of scleral tension helps the LC to limit its bowing (see figure 4.5).

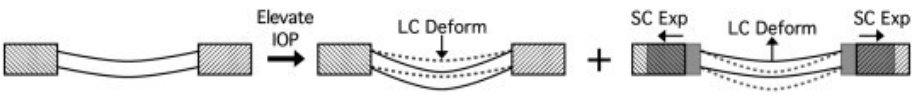


Figure 4.5: Effects of elevated IOP on the LC (taken from [11]).

With regard to the second step, mathematical modeling is still preliminary and has been mainly focused on retinal vasculature. Malgaroli, in her thesis [28], proposed a mathematical model for retinal vasculature, which distinguishes between the different districts of arterioles, capillaries and tissue. Guidoboni et al. [18] developed a model that combines the mechanical action of IOP and the blood flow in the CRA, to elucidate how IOP elevation affects the CRA hemodynamics. In the latter work, the LC is modeled as a nonlinear homogeneous isotropic elastic circular plate of finite thickness and the CRA is modeled as a fluid-structure interaction system, where blood flow is modeled as the stationary Stokes flow of a Newtonian incompressible viscous fluid and the arterial wall is modeled as a linear elastic cylindrical thin shell. This model provided the first evidence that the constriction of the CRA caused by the LC deformation (as effect of IOP elevation) leads to an increase in the CRA resistance to flow and to a reduction in blood velocity and flow. Models describing hemodynamics in the choroidal and optic nerve head vascular beds have not yet been developed.

The third step deals with blood flow autoregulation, that is the ability of vascular beds to maintain relatively constant blood flow over a large range of perfusion pressure. Autoregulation is a process involving several mechanisms of different kind and some studies showed correlations between impairment of blood flow regulation and OAG progression. Mathematical models trying to describe autoregulation in ocular vascular beds are quite recent (see [1]).

Finally, the fourth step requires to combine models describing mechanical and vascular factors with models describing metabolic factors and cellular functions, in order to study how the availability of oxygen and nutrients associated to a specific mechanical and vascular pattern affects the functionality of retinal ganglion cells. The coupling of these aspects has not yet been treated by mathematical models.

The reliability of predictions of a mathematical model depends on the correctness of the assumptions on which the model is based, the precision of the parameter values adopted and the accuracy of the solutions obtained. So, mathematical modeling is not a stand-alone investigative tool, but it has to be combined with laboratory experiments and clinical studies.

In this thesis work, based on the analysis carried out in chapters 1, 2 and 3, we present a poroelastic model for the LC, which is treated as a porous collagen structure filled with blood. It is worth specifying that, in our model, the porous space is not associated with the space devoted to axon passage through the LC, but is intended as the space occupied by the capillary network originating from the posterior ciliary arteries and pervading the LC. In this way, our model deals with the first and the second steps discussed before. In fact, by describing both the collagen matrix and blood in the capillary network, it allows us to study both the mechanical response of the LC to variations in IOP and $\mathrm{RLT_p}$ and some aspects regarding optic nerve head hemodynamics.

Chapter 5

A poroelastic model for the lamina cribrosa

In this chapter, we propose a poroelastic model for the lamina cribrosa, aimed at investigating the response of such biological structure to several load conditions. Firstly, we discuss the biomechanical features of the model with a comment on its adequacy in describing the behaviour of the considered tissue. Secondly, we illustrate the results of numerical simulations aimed at making a comparison with experimental data and some physical considerations about what observed in correspondence of different levels of IOP.

5.1 Motivation of the model

A poroelastic model is adopted here for describing the lamina cribrosa tissue. The solid component is represented by the collagen matrix and the fluid component by blood flowing in the capillary network. By virtue of the continuity hypothesis, the poroelastic model treats these components as two overlapping and interacting continua, as shown in figure 5.1.

We want to discuss the validity of the assumptions listed at the beginning of chapter 1 for a model aimed at describing the LC. There is no doubt that the assumptions of full saturation, isothermal conditions and quasi-static processes are completely reasonable. Also the assumption of reversibility, often used for describing living tissues, does not create particular problems. On the contrary, the hypotheses of isotropy, small deformations and linearity require a few more words. Although Grytz et al. [17] revealed anisotropic attributes in the collagen fibril architecture of the LC, using a computational remodeling approach, isotropic models proposed in the literature obtained good results. For example, Guidoboni et al. [18] proposed an isotropic elastic model for the LC and observed a very satisfactory agreement between model predictions and experimental data. Moreover, isotropy naturally applies to a capillary network (see [12] and [28]), so that we consider an isotropic model in

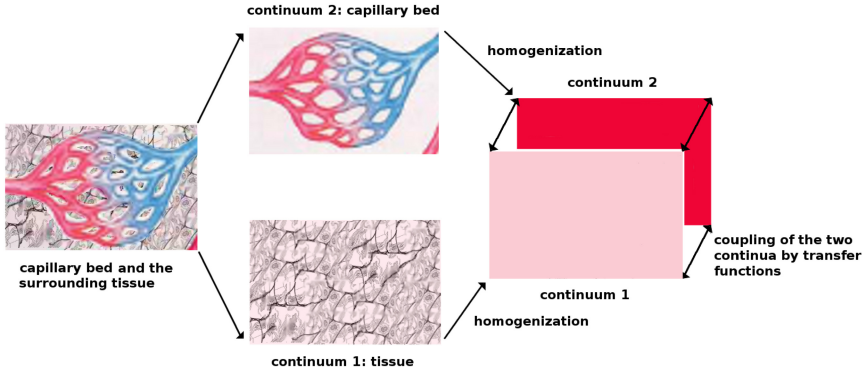


Figure 5.1: Continuity hypothesis applied to the lamina cribrosa.

which isotropy refers to both the collagen matrix and the porous space filled by blood. Then, since Newson and El-Sheikh [30] showed that the central displacement of the LC computed adopting small and large deformations did not reveal significant differences, we choose to consider small deformations. Finally, as regards linearity, Woo et al. [47] obtained a nonlinear stress-strain curve for the optic nerve head from laboratory measurements and modeled this nonlinearity using a trilinear relation, so that the values assumed by the Lamé's parameters depend on the stress state of the material. We take into account this material nonlinearity, which is crucial when considering high values of IOP, adopting the trilinear relation proposed in [47].

To summarize, we consider a trilinear isotropic poroelastic model for the LC. Since both the solid and the fluid components are incompressible and there are no forcing terms, the equations characterizing the model are given by

$$\begin{cases} \nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda \nabla \cdot \mathbf{u}\mathbf{I}) - \nabla p = \mathbf{0}, \\ \dfrac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (K \nabla p) = 0, \end{cases} \tag{5.1}$$

where μ and λ are functions of the stress state.

We remark that, to the best of our knowledge, it is the first time that a poroelastic model considering the collagen matrix and blood in the capillary network is applied to the LC. This allows us to describe both the mechanics and the hemodynamics of the LC. So, the choice of a trilinear isotropic model can be seen also as the first step of a new research project, in which the poroelastic theory is used for describing the optic nerve head. The following possible steps, such as the introduction of anisotropy or the consideration of laminar pre-stress, represent matter of future work.

5.2 Numerical results

The LC is approximately a cylindrical structure located in the optic nerve head (see figure 5.2). When using cylindrical coordinates, the natural assumption of axially symmetric solutions leads to consider a simple rectangular domain, corresponding to a section of the cylinder. Since we have implemented a 2D code, we consider such a rectangular domain in which a Cartesian coordinate system is adopted. This simplifying assumption can be regarded as a first step towards a more rigorous description of the physical problem.

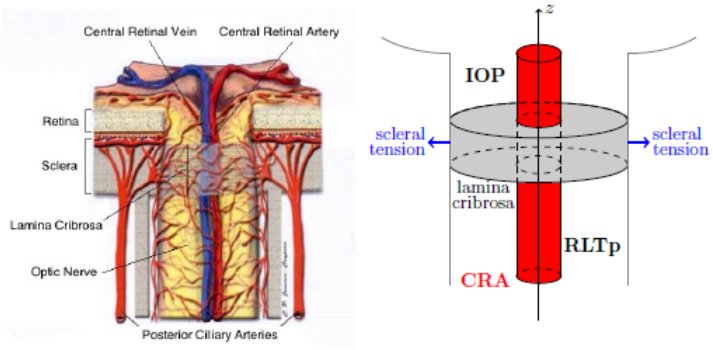


Figure 5.2: Left: anatomy of the optic nerve head. Right: schematic representation of the mechanical environment of the lamina cribrosa (taken from [18]).

Referring to the rectangular domain depicted in figure 5.3, the left side corresponds to the interface with the sclera, the right side corresponds to the interface with the central retinal vein, the upper side faces the interior of the eye globe and the lower side faces the optic nerve canal. Therefore, denoting

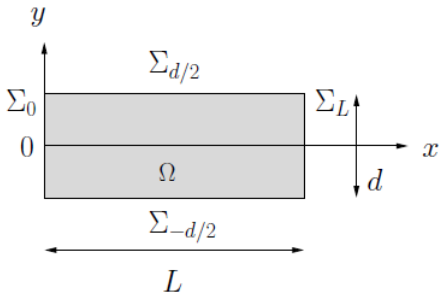


Figure 5.3: Geometrical domain for the application to the LC.

by P_a the blood pressure in the arteries supplying blood to the LC and by P_v the blood pressure in the central retinal vein, we complete equations (5.1) with the following boundary conditions.

- On $\Sigma_{d/2} \times (0,T)$:
$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = -\mathrm{IOP}\mathbf{n}, \\ \nabla p \cdot \mathbf{n} = 0. \end{cases} \tag{5.2}$$

- On $\Sigma_L \times (0,T)$:
$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = -P_v\mathbf{n}, \\ p = P_v. \end{cases} \tag{5.3}$$

- On $\Sigma_{-d/2} \times (0,T)$:
$$\begin{cases} \boldsymbol{\sigma}\mathbf{n} = -\mathrm{RLT_p}\mathbf{n}, \\ \nabla p \cdot \mathbf{n} = 0. \end{cases} \tag{5.4}$$

- On $\Sigma_0 \times (0,T)$:
$$\begin{cases} \mathbf{u} = \mathbf{0}, \\ p = P_a. \end{cases} \tag{5.5}$$

From the mechanical boundary conditions we note that the sclera provides the LC with a clamped edge and the LC is subject to the pressure gradient originating from the difference between the IOP and the $\mathrm{RLT_p}$. On the other hand, the boundary conditions for the fluid part reveal the architecture of the vascular bed in the LC, with blood reaching the LC from its boundary and being drained by the central retinal vein in its centre.

In order to run simulations, we choose physiologically representative values for the geometry of the domain and the permeability K. These values are listed in table 5.1 together with the bibliographical references we have used. We point out that L and d must be considered as average values,

Parameters	Values	Units	Sources
L	0.075	[cm]	[22], [18]
d	0.02	[cm]	[37], [18]
K	$1.6 \cdot 10^{-8}$	[cm^3 s/g]	[12]

Table 5.1: Values of some physical parameters for the application to the LC.

because of the variability of radius and thickness of the LC in human eyes. Furthermore, we remind that the parameter K depends in general on the blood viscosity and the material porosity, but we consider here a constant value for simplicity.

Regarding the choice of the elastic parameters, we adopt the trilinear relation proposed by Woo et al. [47], whereby the shear modulus μ varies with the effective stress. We recall that the effective stress $\boldsymbol{\sigma}'$ is the elastic part of the total stress $\boldsymbol{\sigma}$, namely the total stress deprived of the contribution of the fluid pressure. Evaluating the magnitude of the effective stress through the quantity

$$\sigma'_e = \sqrt{\frac{(\sigma'_{11} - \sigma'_{22})^2 + (\sigma'_{11} - \sigma'_{33})^2 + (\sigma'_{22} - \sigma'_{33})^2 + 6(\sigma'^2_{12} + \sigma'^2_{13} + \sigma'^2_{23})}{2}}, \tag{5.6}$$

the trilinear relation describes the variation of the shear modulus μ as a function of σ'_e, as reported in table 5.2. Following [18] and [30], once determined

μ [dyn/cm^2]	Range of σ'_e [dyn/cm^2]
$1.2 \cdot 10^6$	$0.00 \cdot 10^6 < \sigma'_e < 0.08 \cdot 10^6$
$2.2 \cdot 10^6$	$0.08 \cdot 10^6 < \sigma'_e < 0.15 \cdot 10^6$
$6.1 \cdot 10^6$	$0.15 \cdot 10^6 < \sigma'_e$

Table 5.2: Values of the shear modulus μ as a function of σ'_e.

the value of μ, the value of λ is computed through (1.81), by taking $\nu = 0.49$. Therefore, the use of the trilinear relation implies that the Lamé's parameters can take different values at different points of the domain, depending on the magnitude of the effective stress to which the points are subject. We have implemented the trilinear relation in the 2D code.

5.2.1 Setting of the simulation parameters

Our main objectives are to analyse the mechanical response of the LC to different levels of IOP and to study the effect of IOP-induced LC deformation on the blood flow through the tissue. Before carrying out this study, we need to set the simulation parameters appropriately. Since equations (5.1) have no forcing terms and we typically run simulations considering boundary conditions not depending on time, we expect the solution to reach a final equilibrium state after a certain time. In 1D problems, the time constant

$$\tau_{1D} = \frac{L^2}{H_A K} \tag{5.7}$$

gives information about the order of magnitude of the time taken by the solution to reach the final equilibrium state. In the confined compression creep experiment described in chapter 3, using the values in table 3.7, we get $\tau_{1D} \simeq 3555$ s. In 2D problems, the time constant corresponding to (5.7) is given by

$$\tau_{2D} = \frac{L^2}{(2\mu + \lambda)K}. \tag{5.8}$$

Choosing L and K as in table 5.1 and the central value of μ in table 5.2, we get $\tau_{2D} \simeq 0.003$ s. Such a significant difference between the time constant of the confined compression creep experiment and that of the simulation of the LC is due to the large difference in the corresponding values of permeability. We point out that, differently from what occurs in the case of the confined compression creep experiment, the time constant computed for the simulation of the LC has no physical meaning, but its relevance is only numerical, because we simply run numerical tests without reproducing the conditions of a specific experiment.

In order to verify the correctness of the considerations above, we run a numerical test imposing a step in the IOP as in figure 5.4, in such a way that the IOP varies between 15 mmHg and 25 mmHg, while we take

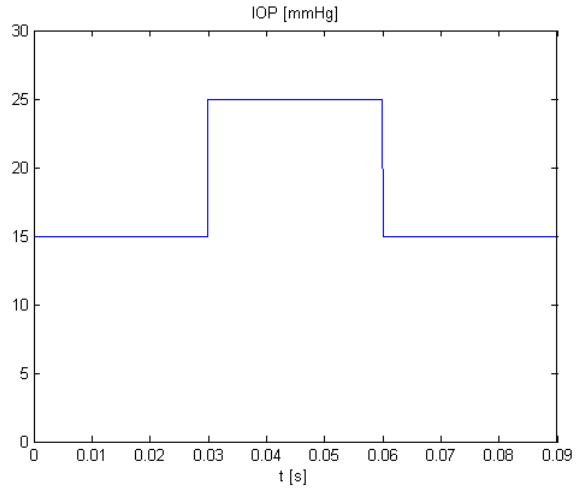


Figure 5.4: Step in the IOP imposed in the numerical test.

$\mathrm{RLT_p} = 7$ mmHg, $P_a = 30$ mmHg and $P_v = 20$ mmHg. Figure 5.6 shows the fluid pressure at a point P (located at $x = 0.0266$ and $y = 3.6398 \cdot 10^{-4}$ and shown in figure 5.5) as a function of time. The time taken by the fluid

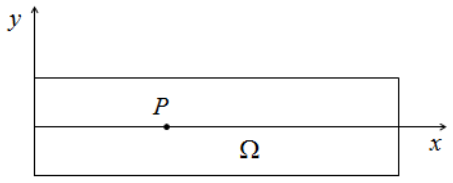


Figure 5.5: The point P located at $x = 0.0266$ and $y = 3.6398 \cdot 10^{-4}$.

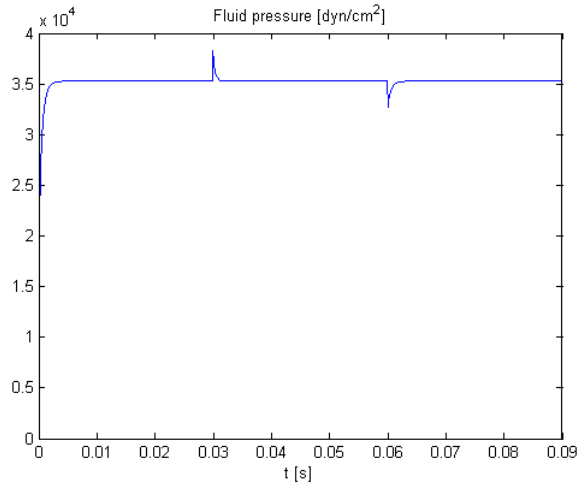


Figure 5.6: Fluid pressure at the point P in the response to the step in the IOP.

pressure to reach an equilibrium state at the beginning and after each change in the IOP is similar and never exceeds 0.006 s. Similar results are obtained considering other locations in the computational domain, thus confirming the estimate (5.8). Figure 5.6 shows also the tendency of the blood pressure to redistribute in the same manner after each change in the IOP.

The above analysis allows us to determine the length of the time interval characterizing the simulations. Several numerical tests have been run in order to choose the grid and the time step used in the simulations. We adopt an unstructured grid consisting of 336 triangular elements and take $\Delta t = 10^{-4}$. The use of finer grids or smaller time steps leads to similar results. Regarding the relaxation, the choice of such a time step implies to take $\omega = 0.25$ according to relation (2.135). Finally, we set $\epsilon_{u_1} = \epsilon_{u_2} = \epsilon_{\mathcal{P}} = \epsilon_p = 10^{-6}$.

5.2.2 Validation of the model

Though our study has to be considered preliminary with respect to the more rigorous description of the physical problem by means of cylindrical coordinates, it is useful to make a comparison between model predictions and experimental data, in order to understand the degree of significance of our analysis. We consider the work by Yan et al. [48], who mounted three enucleated human eyes on a specially designed experimental apparatus, which allowed them to sequentially increase the IOP from 5 mmHg to 15, 30 and 50 mmHg. Topographic images of the optic nerve head were taken at each pressure using a scanning laser tomographer (Heidelberg Retina Tomograph - HRT). These images were analysed using selected standard parameters

computed by the HRT software, such as the mean effective depth (MED).

We make a comparison between the values of the IOP-induced increments of MED measured experimentally and those predicted by our model. For a given IOP, the corresponding increment of MED is defined as the difference between the value of MED in correspondence of that IOP and the value of MED in correspondence of IOP equal to 5 mmHg. We do not extend the comparison to the case in which IOP = 50 mmHg, because our model assumes small deformations. $\mathrm{RLT_p}$ is taken equal to a minimum value corresponding to 2.92 mmHg, since the experiment makes use of enucleated eyes (see [18] to have more details).

The results of the comparison between model predictions and experimental data are shown in table 5.3 and figure 5.7. We see that the increments of MED predicted by the model fall within the range of values measured experimentally. This suggests that our model for the LC provides physiologically reasonable results when comparing the response to different levels of IOP.

IOP [mmHg]	Eye A	Eye B	Eye C	Numerical
5	0	0	0	0
15	16	29	9	12
30	27	37	22	31

Table 5.3: IOP-induced increments of MED measured by Yan et al. [48] and predicted by the model (values are expressed in μm).

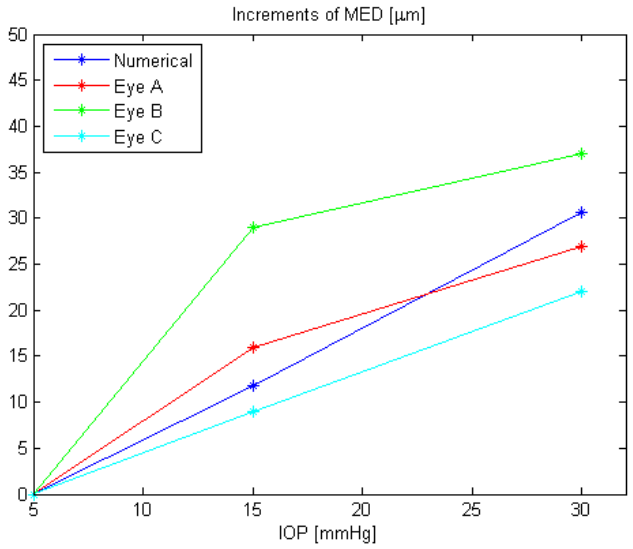


Figure 5.7: Comparison between model predictions and experimental data.

5.2.3 Comparison between different levels of IOP

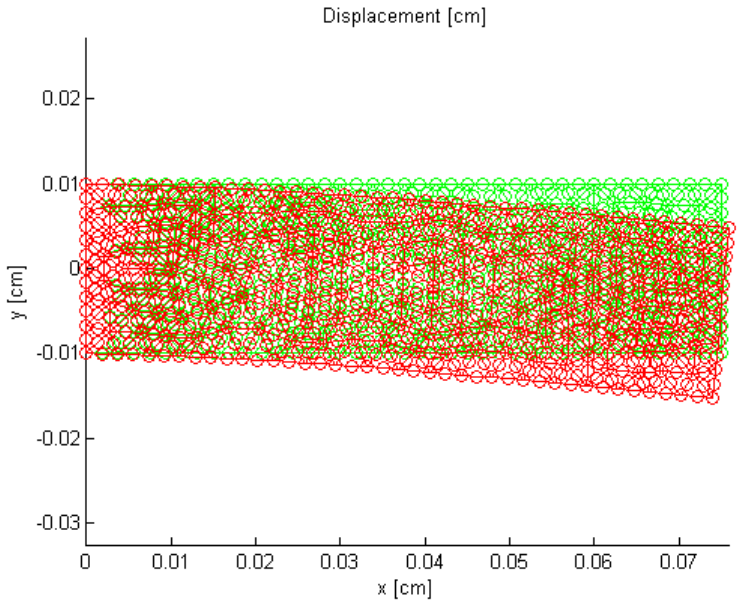
Now, we analyse the response of the LC to different levels of IOP. As discussed in chapter 4, elevated IOP is the main risk factor for glaucoma, though there are some patients exhibiting low values of IOP and some healthy people exhibiting high values of IOP. Values of IOP between 10 mmHg and 21 mmHg are considered normal, while higher values are associated with an increasing risk of developing glaucoma. We run simulations referring to the following representative cases:

- normal IOP, whereby we refer to IOP = 15 mmHg;
- elevated IOP, whereby we refer to IOP = 25 mmHg.

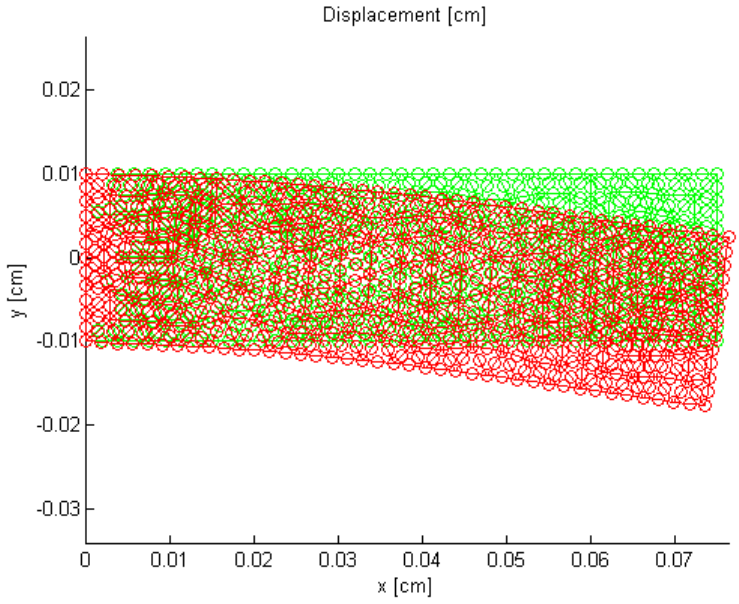
Each case is simulated from $t = 0$ s to $t = 0.05$ s, using the physical and simulation parameters indicated before and taking $\mathrm{RLT_p} = 7$ mmHg, $P_a = 30$ mmHg and $P_v = 20$ mmHg. Figures 5.8, 5.9 and 5.10 show the variables computed by the 2D code implementing the relaxed fixed point iteration algorithm, revealing the differences between the two cases corresponding to normal and elevated IOP. As we expect, the deformation undergone by the LC is more pronounced in the case of elevated IOP (see figure 5.8), in which the LC is subject to a greater pressure gradient. Figure 5.9 shows the elastic pressure parameter and gives additional information about the difference between the two cases. On the contrary, the distribution of blood pressure does not reveal significant differences between the two cases (see figure 5.10), so that it seems not to be affected by small increases in IOP.

Starting from the solution provided by the code, namely the finite element approximation of the displacement $\mathbf{u}$, the elastic pressure parameter $\mathcal{P}$ and the fluid pressure p, in the barycentre of each triangle of the mesh we compute the approximation of the specific discharge $\mathbf{v}$, the linearized strain tensor $\boldsymbol{\varepsilon}$, the effective stress $\boldsymbol{\sigma}'$ and the total stress $\boldsymbol{\sigma}$. We note that the computation of the quantity σ'_e in (5.6) is based on the approximation of the components of the effective stress $\boldsymbol{\sigma}'$. The specific discharge, which is the relative velocity of the solid with respect to the fluid weighted by the porosity, is computed from the fluid pressure using $\mathbf{v} = -K\nabla p$, so that we do not see significant differences between the two cases under analysis. Therefore, we only report the specific discharge computed in the case corresponding to IOP = 15 mmHg: the x-component is shown in figure 5.11, where it turns out to be almost constant, while the y-component is not shown because it is almost equal to zero.

Turning to stresses and strains, they are significantly higher near the left side of the rectangular domain, namely along the edge of the LC. This is a very interesting fact, since it is known that the loss of vision in glaucomatous eyes starts along the boundaries and grows towards the centre and the axons transmitting the vision signals along the boundaries cross the LC near its

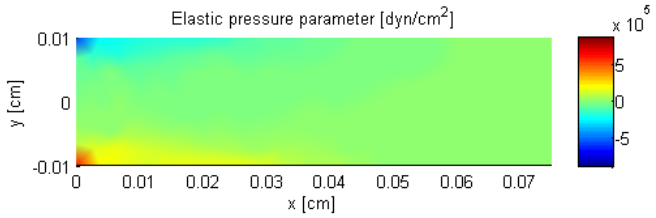


(a) Normal IOP (IOP = 15 mmHg).

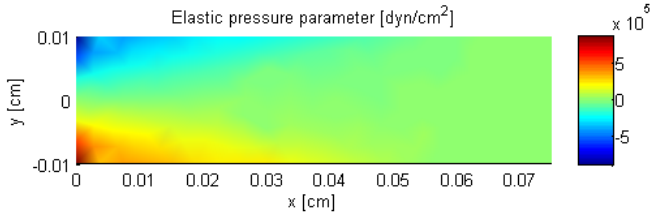


(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.8: Approximation of the displacement $\mathbf{u}$, shown in terms of initial (green) and deformed (red) configurations.

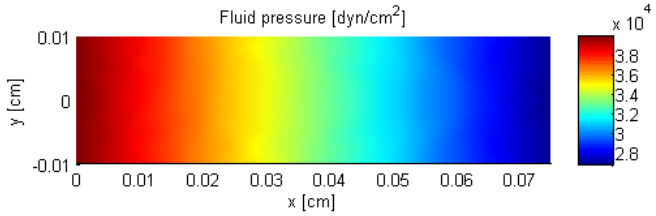


(a) Normal IOP (IOP = 15 mmHg).

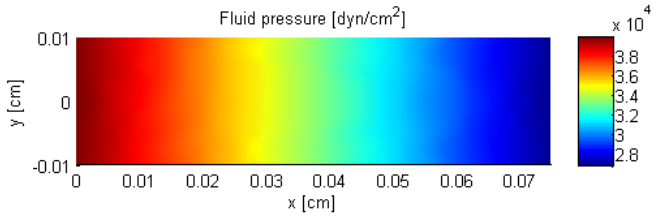


(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.9: Approximation of the elastic pressure parameter $\mathcal{P}$.



(a) Normal IOP (IOP = 15 mmHg).



(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.10: Approximation of the fluid pressure p.

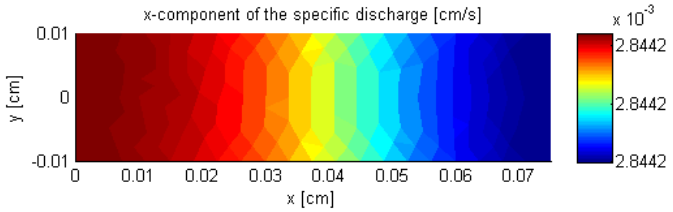
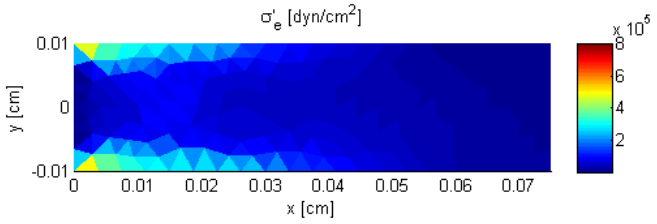
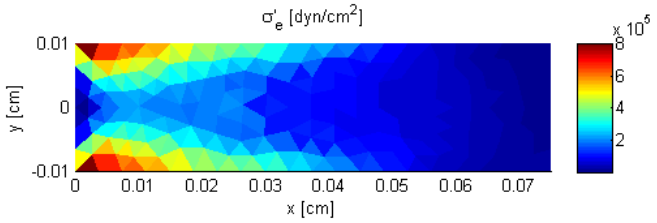


Figure 5.11: Approximation of the x-component of the specific discharge in the case of normal IOP (IOP = 15 mmHg).

edge. Therefore, the high stresses and strains in the boundary regions of the LC, increasing with IOP, could be responsible for damage to those axons, causing the loss of vision along the boundaries (see also [17] and [30]). Instead of showing each component of stress and strain, we report the quantity σ'_e and the volumetric strain $\varepsilon = \nabla \cdot \mathbf{u}$, which clearly confirm the above considerations (see figures 5.12 and 5.14). We show also the distribution of the shear modulus μ, closely linked to that of σ'_e, which highlights the importance of using a nonlinear relation to model the elastic properties of the material (see figure 5.13).

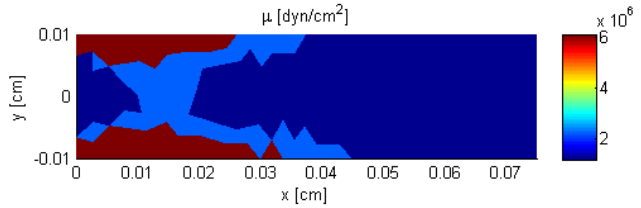


(a) Normal IOP (IOP = 15 mmHg).

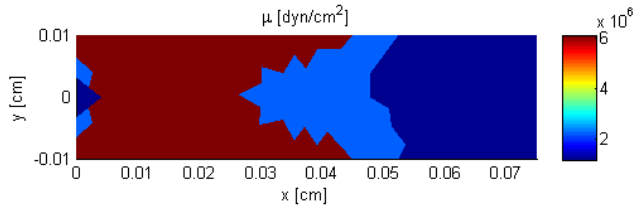


(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.12: Approximation of the quantity σ'_e.

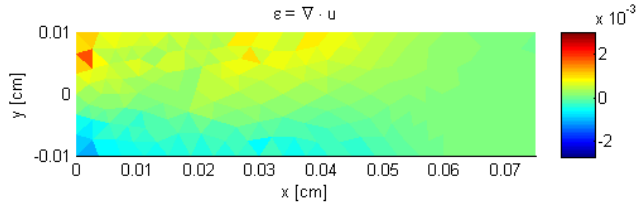


(a) Normal IOP (IOP = 15 mmHg).

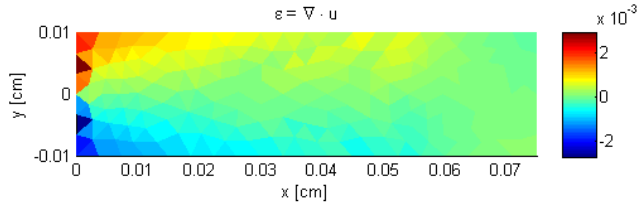


(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.13: Computation of the shear modulus μ.



(a) Normal IOP (IOP = 15 mmHg).



(b) Elevated IOP (IOP = 25 mmHg).

Figure 5.14: Approximation of the volumetric strain $\varepsilon = \nabla \cdot \mathbf{u}$.

Figure 5.14, obtained by reconstructing the divergence of the displacement from the elastic pressure parameter using $\mathcal{P} = -\lambda \nabla \cdot \mathbf{u}$, suggests another important consideration, that comes from having used a poroelastic model for the LC. In a poroelastic material with incompressible components, the change in porosity is given by

$$\phi = \phi_0 + \varepsilon = \phi_0 + \nabla \cdot \mathbf{u}, \tag{5.9}$$

where ϕ_0 denotes the porosity in the initial configuration (see [9]). Figure 5.14 shows that near the edge of the LC there are regions where the divergence of the displacement is negative and these regions become wider in space as the IOP increases. From (5.9) we see that negative values of the divergence of the displacement cause a decrease in the porosity associated with the capillary network nourishing the LC. For very negative values of volumetric strain, we could think that the porosity gets closer to zero, that is some capillaries may become closed, causing damage to the axons crossing the LC in the outer regions, because of a lower supply of nutrients by means of blood flow. This is an intriguing new hypothesis, but to date we lack experimental measurements to verify its validity. More efforts both in clinical studies and in mathematical modeling are needed in order to increase knowledge of the physical mechanisms occurring in the human eye. This is of utmost importance for the development of new therapeutic concepts for several ocular pathologies, such as glaucoma.

Conclusions and future works

This thesis aimed at showing the potential use of the theory of poroelasticity for the description of biological tissues, dealing with the numerical treatment of the poroelastic equations and proposing an original poroelastic model for the lamina cribrosa.

We have presented the theory of linear poroelasticity from the assumptions to the equations, showing that the behaviour of a linear poroelastic material can be described by a system of partial differential equations, whose unknowns are the solid displacement and the fluid pressure. We have proposed a relaxed fixed point iteration algorithm for the numerical solution of the linear poroelastic equations. Such algorithm solves the mechanical and the fluid parts of a poroelastic problem sequentially and iteratively, until convergence is reached. It is based on the finite element method for spatial discretization and the backward Euler method for time discretization and has been implemented in MATLAB, both in the 1D case and in the 2D case. The introduction of the elastic pressure parameter provides greater robustness to the numerical algorithm. We have analysed the proposed algorithm, showing that its convergence properties depend on the elastic parameters of the material, the permeability associated with the fluid in the porous space and the time step used in simulations. Then, we have validated our code by making a convergence analysis and we have confirmed the theoretical analysis of the iterative algorithm using a poroelastic model for articular cartilage proposed in the literature. Finally, we have presented the possible use of mathematical modeling to investigate the physiological mechanisms occurring in the human eye, in order to clarify the causes of some ocular diseases, such as glaucoma. In this context, we have proposed a poroelastic model for the lamina cribrosa and made some interesting physiological considerations, suggested by the results of the numerical simulations.

Applying mathematical models to living tissues is always a hard challenge, so that there are still many issues that can be addressed in the description of the lamina cribrosa. We list some possible steps for future works, aiming at obtaining an increasingly realistic model for the lamina cribrosa.

- The first logical step is to switch to cylindrical coordinates, whereby the real 3D problem is properly reduced to a 2D one and a different

set of boundary conditions is imposed, considering the scleral tension. Such a switch implies considerable changes to the code.

- We have seen in chapter 1 that the permeability K depends on the porosity ϕ, but for simplicity we have considered a constant value of K in all the numerical simulations. Another step is to model the dependence of the permeability on the porosity, identifying a functional relation between these quantities and implementing such relation in the code.

- Again, some further aspects, such as anisotropy and the presence of pre-stress, can be considered in the model to complete the physical description of the lamina cribrosa.

- Finally, the poroelastic model based on small deformations can be extended to consider large deformations, in order to simulate the behaviour of the lamina cribrosa when it is subject to very high levels of IOP, which may occur in advanced glaucoma.

Appendix A

Linear poroelastic equations with the formalism of mixture theory

In this appendix, we show how equations (1.86) for linear isotropic poroelastic media with incompressible components can be alternatively derived using the typical formalism of mixture theory. We follow [15] and use the subscripts s and f to refer to the solid component and the fluid one, respectively. Denoting by $\boldsymbol{\sigma}_s$ and $\boldsymbol{\sigma}_f$ the partial stress tensors (namely the stress tensors referred to the components) and by $\boldsymbol{\pi}_s$ and $\boldsymbol{\pi}_f$ the vectors representing the interaction between the components, we have the momentum balance equations for solid and fluid components

$$\nabla \cdot \boldsymbol{\sigma}_s + \boldsymbol{\pi}_s = \mathbf{0}, \tag{A.1}$$

$$\nabla \cdot \boldsymbol{\sigma}_f + \boldsymbol{\pi}_f = \mathbf{0}, \tag{A.2}$$

where we neglect body forces. The partial stress tensors are given by

$$\boldsymbol{\sigma}_s = -(1-\phi)p\mathbf{I} + \boldsymbol{\sigma}', \tag{A.3}$$

$$\boldsymbol{\sigma}_f = -\phi p\mathbf{I}, \tag{A.4}$$

where the effective stress $\boldsymbol{\sigma}'$ is defined as before by

$$\boldsymbol{\sigma}' = \mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) + \lambda\nabla \cdot \mathbf{u}\mathbf{I}. \tag{A.5}$$

Let ρ_s and ρ_f be the intrinsic mass densities and $\mathbf{v}_s$ and $\mathbf{v}_f$ the velocities. Mass balances are expressed by the continuity equations

$$\frac{\partial((1-\phi)\rho_s)}{\partial t} + \nabla \cdot ((1-\phi)\rho_s\mathbf{v}_s) = 0, \tag{A.6}$$

$$\frac{\partial(\phi\rho_f)}{\partial t} + \nabla \cdot (\phi\rho_f\mathbf{v}_f) = 0, \tag{A.7}$$

where we neglect fluid sources or sinks.

Substituting relations (A.3) and (A.4) into equations (A.1) and (A.2), summing these equations and observing that $\boldsymbol{\pi}_s + \boldsymbol{\pi}_f = \mathbf{0}$ because of the action-reaction principle, we get

$$\nabla \cdot (\mu(\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \lambda \nabla \cdot \mathbf{u}\mathbf{I}) - \nabla p = \mathbf{0}. \tag{A.8}$$

On the other hand, summing equations (A.6) and (A.7) and reminding that ρ_s and ρ_f are constants due to incompressibility, we obtain

$$\nabla \cdot \mathbf{v}_s + \nabla \cdot (\phi(\mathbf{v}_f - \mathbf{v}_s)) = 0. \tag{A.9}$$

Furthermore, observing that

$$\mathbf{v}_s = \frac{\partial \mathbf{u}}{\partial t} \tag{A.10}$$

and defining the specific discharge as

$$\mathbf{v} = \phi(\mathbf{v}_f - \mathbf{v}_s), \tag{A.11}$$

we get

$$\frac{\partial(\nabla \cdot \mathbf{u})}{\partial t} + \nabla \cdot \mathbf{v} = 0. \tag{A.12}$$

Applying Darcy's law to equation (A.12), we finally obtain

$$\frac{\partial(\nabla \cdot \mathbf{u})}{\partial t} - \nabla \cdot (\mathbf{K} \nabla p) = 0. \tag{A.13}$$

The above analysis shows that equations (1.86) defining the poroelastic model for biological tissues can be reconstructed starting from the different perspective of mixture theory.

Bibliography

[1] J. Arciero, A. Harris, B. Siesky, A. Amireskandari, V. Gershuny, A. Pickrell, and G. Guidoboni. Theoretical analysis of vascular regulatory mechanisms contributing to retinal blood flow autoregulation. *Invest. Ophthalmol. Vis. Sci.*, 54(8):5584–5593, 2013.

[2] M. A. Biot. General Theory of Three-Dimensional Consolidation. *J. Appl. Phys.*, 12(2):155–164, 1941.

[3] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, 1991.

[4] J. Caprioli and A. L. Coleman. Blood pressure, perfusion pressure, and glaucoma. *Am. J. Ophthalmol.*, 149(5):704–712, 2010.

[5] P. Causin, J.-F. Gerbeau, and F. Nobile. Added-mass effect in the design of partitioned algorithms for fluid-structure problems. *Comput. Methods Appl. Mech. Engrg.*, 194(42-44):4506–4527, 2005.

[6] P. Causin and R. Sacco. A computational model for biomass growth simulation in tissue engineering. *Commun. Appl. Ind. Math.*, 2(1):e–370, 20, 2011.

[7] D. Chapelle, J.-F. Gerbeau, J. Sainte-Marie, and I. E. Vignon-Clementel. A poroelastic model valid in large strains with applications to perfusion in cardiac modeling. *Comput. Mech.*, 46(1):91–101, 2010.

[8] H. Chen, R. E. Ewing, S. L. Lyons, G. Qin, T. Sun, and D. P. Yale. A numerical algorithm for single phase fluid flow in elastic porous media. In *Numerical treatment of multiphase flows in porous media (Beijing, 1999)*, volume 552 of *Lecture Notes in Phys.*, pages 80–92. Springer, Berlin, 2000.

[9] O. Coussy. *Poromechanics.* John Wiley & Sons, 2004.

[10] J. Crawford Downs, M. D. Roberts, C. F. Burgoyne, and R. T. Hart. Multiscale finite element modeling of the lamina cribrosa microarchitecture in the eye. In *Engineering in Medicine and Biology Society, 2009.*

EMBC 2009. Annual International Conference of the IEEE, pages 4277–4280, 2009.

[11] J. Crawford Downs, M. D. Roberts, and I. A. Sigal. Glaucomatous cupping of the lamina cribrosa: a review of the evidence for active progressive remodeling as a mechanism. *Exp. Eye Res.*, 93(2):133–140, 2011.

[12] C. D'Angelo. *Multiscale modelling of metabolism and transport phenomena in living tissues.* PhD thesis, EPFL, Switzerland, 2007.

[13] E. Detournay and A. H.-D. Cheng. Fundamentals of Poroelasticity. In C. Fairhurst, editor, *Comprehensive Rock Engineering: Principles, Practice and Projects*, volume II, chapter 5, pages 113–171. Pergamon Press, 1993.

[14] H. Dongqi and R. Zeqin. A biomathematical model for pressure-dependent lamina cribrosa behavior. *J. Biomech.*, 32(6):579–584, 1999.

[15] A. J. H. Frijns. *A four-component mixture theory applied to cartilaginous tissues: Numerical modelling and experiments.* ProQuest LLC, Ann Arbor, MI, 2000. Thesis (Dr.ir.)–Technische Universiteit Eindhoven (The Netherlands).

[16] J.-F. Gerbeau and M. Vidrascu. A quasi-Newton algorithm based on a reduced model for fluid-structure interaction problems in blood flows. *M2AN Math. Model. Numer. Anal.*, 37(4):631–647, 2003.

[17] R. Grytz, G. Meschke, and J. B. Jonas. The collagen fibril architecture in the lamina cribrosa and peripapillary sclera predicted by a computational remodeling approach. *Biomech. Model. Mechanobiol.*, 10(3):371–382, 2011.

[18] G. Guidoboni, A. Harris, L. Carichino, Y. Arieli, and B. A. Siesky. Effect of intraocular pressure on the hemodynamics of the central retinal artery: a mathematical model. *Math. Biosci. Eng.* Accepted for publication.

[19] A. Harris, G. Guidoboni, J. C. Arciero, A. Amireskandari, L. A. Tobe, and B. A. Siesky. Ocular hemodynamics and glaucoma: the role of mathematical modeling. *Eur. J. Ophthalmol.*, 23(2):139–146, 2013.

[20] L. R. Herrmann. Elasticity equations for incompressible and nearly incompressible materials by a variational theorem. *AIAA Journal*, 3(10):1896–1900, 1965.

[21] T. J. R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis.* Prentice-Hall, 1987.

[22] J. B. Jonas, C. Y. Mardin, U. Schlötzer-Schrehardt, and G. O. Naumann. Morphometry of the human lamina cribrosa surface. *Invest. Ophthalmol. Vis. Sci.*, 32(2):401–405, 1991.

[23] S. M. Klisch, R. L. Sah, and A. Hoger. A cartilage growth mixture model for infinitesimal strains: solutions of boundary-value problems related to in vitro growth experiments. *Biomech. Model. Mechanobiol.*, 3(4):209–223, 2005.

[24] R. K. Korhonen, M. S. Laasanen, J. Töyräs, R. Lappalainen, H. J. Helminen, and J. S. Jurvelin. Fibril reinforced poroelastic model predicts specifically mechanical behavior of normal, proteoglycan depleted and collagen degraded articular cartilage. *J. Biomech.*, 36(9):1373–1379, 2003.

[25] U. Küttler and W. A. Wall. Fixed-point fluid-stucture interaction solvers with dynamic relaxation. *Comput. Mech.*, 43(1):61–72, 2008.

[26] M. E. Levenston, S. R. Eisenberg, and A. J. Grodzinsky. A variational formulation for coupled physicochemical flows during finite deformations of charged porous media. *Int. J. Solids Structures*, 35(34):4999–5019, 1998.

[27] L. P. Li, J. Soulhat, M. D. Buschmann, and A. Shirazi-Adl. Nonlinear analysis of cartilage in unconfined ramp compression using a fibril reinforced poroelastic model. *Clin. Biomech.*, 14(9):673–682, 1999.

[28] F. Malgaroli. Multiscale Models and Numerical Simulation of Retinal Microcirculation: Blood Flow and Mass Transport Phenomena. Master's thesis, Università degli Studi di Milano, 2013.

[29] W. H. Morgan, B. C. Chauhan, D.-Y. Yu, S. J. Cringle, V. A. Alder, and P. H. House. Optic disc movement with variations in intraocular and cerebrospinal fluid pressure. *Invest. Ophthalmol. Vis. Sci.*, 43(10):3236–3242, 2002.

[30] T. Newson and A. El-Sheikh. Mathematical modeling of the biomechanics of the lamina cribrosa under elevated intraocular pressures. *J. Biomech. Eng.*, 128(4):496–504, 2006.

[31] F. Nobile. *Numerical approximation of fluid-structure interaction problems with application to haemodynamics.* PhD thesis, EPFL, Switzerland, 2001.

[32] P. J. Phillips and M. F. Wheeler. A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity. *Comput. Geosci.*, 12(4):417–435, 2008.

[33] P. J. Phillips and M. F. Wheeler. Overcoming the problem of locking in linear elasticity and poroelasticity: an heuristic approach. *Comput. Geosci.*, 13(1):5–12, 2009.

[34] M. Preisig and J. H. Prévost. Stabilization procedures in coupled poromechanics problems: A critical assessment. *Int. J. Numer. Anal. Meth. Geomech.*, 35(11):1207–1225, 2011.

[35] A. Quarteroni. *Modellistica Numerica per Problemi Differenziali.* Springer-Verlag, IV edition, 2008.

[36] A. Quarteroni, R. Sacco, and F. Saleri. *Matematica Numerica.* Springer-Verlag, III edition, 2008.

[37] R. Ren, N. Wang, B. Li, L. Li, F. Gao, X. Xu, and J. B. Jonas. Lamina cribrosa and peripapillary sclera histomorphometry in normal and advanced glaucomatous Chinese eyes with various axial length. *Invest. Ophthalmol. Vis. Sci.*, 50(5):2175–2184, 2009.

[38] R. Sacco. Numerical methods for civil engineering. Notes of the course, available at the link `https://beep.metid.polimi.it`.

[39] S. Salsa. *Equazioni a Derivate Parziali. Metodi, Modelli e Applicazioni.* Springer-Verlag, II edition, 2010.

[40] M. Schanz and D. Pryl. Dynamic fundamental solutions for compressible and incompressible modeled poroelastic continua. *Int. J. Solids Structures*, 41(15):4047–4073, 2004.

[41] B. G. Sengers, C. W. J. Oomens, and F. P. T. Baaijens. An integrated finite-element approach to mechanics, transport and biosynthesis in tissue engineering. *J. Biomech. Eng.*, 126(1):82–91, 2004.

[42] R. E. Showalter. Diffusion in poro-elastic media. *J. Math. Anal. Appl.*, 251(1):310–340, 2000.

[43] I. A. Sigal, H. Yang, M. D. Roberts, C. F. Burgoyne, and J. Crawford Downs. IOP-induced lamina cribrosa displacement and scleral canal expansion: an analysis of factor interactions using parameterized eye-specific models. *Invest. Ophthalmol. Vis. Sci.*, 52(3):1896–1907, 2011.

[44] M. M. Silbernagel. Modeling coupled fluid flow and geomechanical and geophysical phenomena within a finite element framework. Master's thesis, Colorado School of Mines, 2007.

[45] M. A. Soltz and G. A. Ateshian. Experimental verification and theoretical prediction of cartilage interstitial fluid pressurization at an impermeable contact interface in confined compression. *J. Biomech.*, 31(10):927–934, 1998.

[46] K. Terzaghi. *Erdbaumechanik auf bodenphysikalischer Grundlage.* Deuticke, 1925.

[47] S. L.-Y. Woo, A. S. Kobayashi, W. A. Schlegel, and C. Lawrence. Nonlinear material properties of intact cornea and sclera. *Exp. Eye Res.*, 14(1):29–39, 1972.

[48] D. B. Yan, J. G. Flanagan, T. Farra, G. E. Trope, and C. R. Ethier. Study of regional deformation of the optic nerve head using scanning laser tomography. *Curr. Eye Res.*, 17(9):903–916, 1998.

Ringraziamenti

Giunto al termine di questo intenso e appassionante lavoro di tesi, vorrei ringraziare le persone che mi hanno accompagnato lungo questo cammino, contribuendo in maniera differente ma determinante al raggiungimento della meta.

Desidero ringraziare innanzitutto il Prof. Riccardo Sacco e la Dott.ssa Paola Causin per avermi sempre seguito con grande disponibilità e attenzione, per avermi trasmesso il loro entusiasmo e per avermi fatto sentire concretamente la loro fiducia. Lavorare insieme a loro è stata un'importante occasione di crescita sotto tanti punti di vista. Vorrei poi ringraziare la Prof.ssa Giovanna Guidoboni per i suoi preziosi consigli e per aver avvicinato due mondi apparentemente lontani, come la matematica e la medicina. Ringrazio anche Daniele Prada, dottorando presso la IUPUI negli USA, per il periodo in cui abbiamo lavorato insieme e per lo stimolante e utile scambio di idee degli ultimi mesi. Un ringraziamento va anche all'Ing. Marco Restelli che, pur non essendo coinvolto in prima persona in questo lavoro, mi ha svelato alcuni misteri dell'informatica e della programmazione con la sua consueta disponibilità.

Grazie a tutti gli amici del Poli, Gio, Luca, Fabio, Marco, Laura, Silvia, Teo, Usu, Sara, Dinho, Gian, Giudy, Vincenza, per avere reso questi anni indimenticabili. Grazie anche per le serate insieme e per le partite di calcetto, compresa quella che deve essere ancora organizzata da qualcuno!

Grazie agli amici di sempre, quelli su cui sai di poter contare in ogni momento, e agli amici più recenti, che portano un vento di novità. Grazie ad Anti, Ricky, Fede, Silvia, Eli, Fra, Marta, Laura, Lauretta, Jack, Caro, Andre, Fra, Ali, Vale, Marci e altri ancora, per l'affetto, i momenti di svago e tutto quello che abbiamo combinato insieme!

Grazie ai miei parenti, in particolare ai miei cugini Jacopo e Niky e ai miei nonni, che da Lassù hanno sicuramente fatto il tifo per me, soprattutto quando ne ho avuto più bisogno.

Grazie a Veronica per avermi sempre sostenuto con dolcezza e attenzione, per avermi saputo ascoltare anche nei momenti in cui devo essere stato davvero insopportabile e per avermi fatto capire quali sono le cose veramente importanti.

Infine, grazie alla mia famiglia, a cui questa tesi è dedicata. Grazie a

mamma e papà per tutto quello che mi hanno insegnato, per il loro supporto incondizionato e per essere stati un appoggio sicuro lungo tutti questi anni. Grazie a mio fratello Filippo per l'affetto, la stima reciproca e tutto ciò che ci lega nonstante la distanza fisica che ci separa.

Samuele