VERLAG DER TECHNISCHEN UNIVERSITÄT GRAZ

Matthias Gsell

Mortar Domain Decomposition Methods for Quasilinear Problems and Applications

CES 30

MONOGRAPHIC SERIES TU GRAZ
COMPUTATION IN ENGINEERING AND SCIENCE

Matthias Gsell

Mortar Domain Decomposition Methods for Quasilinear Problems and Applications

Monographic Series TU Graz

Computation in Engineering and Science

Monographic Series TU Graz

Computation in Engineering and Science   Volume 30

Matthias Gsell

Mortar Domain Decomposition Methods for Quasilinear Problems and Applications

This work is based on the dissertation *“Mortar Domain Decomposition Methods for Quasilinear Problems and Applications”*, presented by Matthias Gsell at Graz University of Technology, Institute of Computational Mathematics (Math D) in Nov. 2016.
Assessor internal: Olaf Steinbach (Graz University of Technology)
Assessor external: Ralf Kornhuber (Berlin Freie Universität)
Assessor external: Barbara Wohlmuth (Technical University of Munich)

Cover photo	Vier-Spezies-Rechenmaschine by courtesy of the Gottfried Wilhelm Leibniz Bibliothek – Niedersächsische Landesbibliothek Hannover
Layout	Christina Fraueneder, TU Graz Stefan Schleich, TU Graz
Printed	Printservice, TU Graz

Verlag der Technischen Universität Graz
www.ub.tugraz.at/Verlag

Print
ISBN 978-3-85125-522-5

E-Book
ISBN 978-3-85125-523-2
DOI 10.3217/ 978-3-85125-522-5

CONTENTS

1 Introduction **1**

2 Modeling **5**
2.1 Homogeneous Soil . . . 9
2.2 Heterogeneous Soil . . . 9
2.3 Boundary and Initial Conditions . . . 10

3 Mathematical Preliminaries **13**
3.1 Elementary Functional Analysis . . . 13
3.1.1 Liner Operator Equations . . . 18
3.1.2 Nonliner Operator Equations . . . 20
3.2 Function Spaces . . . 21
3.2.1 Classical Spaces . . . 21
3.2.2 Lebesgue Spaces . . . 23
3.2.3 Sobolev Spaces . . . 25
3.3 Superposition Operators . . . 31

4 Variational Formulation **37**
4.1 Time Discrete Variational Formulation . . . 38
4.2 Solvability and Uniqueness . . . 41
4.3 Kirchhoff Transformation in Homogeneous Soil . . . 59
4.4 Kirchhoff Transformation in Heterogeneous Soil . . . 68
4.4.1 The Primal Hybrid Framework . . . 71
4.4.2 The Primal Hybrid Formulation . . . 78

5 Linearization, Discretization and Implementation **89**
5.1 Linearization . . . 93
5.1.1 The Newton Method . . . 93
5.1.2 Newton Linearization . . . 95
5.2 Space Discretization . . . 96
5.3 Solvability and Uniqueness . . . 104
5.4 Implementation . . . 119

6 Numerical Experiments **125**
6.1 Experiment in 2D 125
6.2 Experiment in 3D 131

7 Conclusion **139**

1 INTRODUCTION

The equation of interest in this thesis is the Richards equation, which describes the saturated–unsaturated flow of fluid (water) through a porous medium and which was introduced in 1931 by the American physicist Lorenzo Adolph Richards, see [52]. In Figure 1.1 one can see Lorenzo Adolph Richards in front of an experimental

Figure 1.1: Lorenzo Adolph Richards.

setup. The Richards equation is an elliptic–parabolic partial differential equation and since the equation involves two nonlinear terms, straightforward approximation methods have to be handled with care or are not applicable at all. Therefore, new strategies and efficient methods are necessary for solving such kind of problems. In [12] a monotone multigrid method is considered in the case of homogeneous soil. In the case of heterogeneous soil, nonlinear Dirichlet–Neumann methods as well as nonlinear Robin type domain decomposition methods are discussed.

In this thesis a different approach is considered. As already mentioned, the Richards equation involves two nonlinearities, one to describe the saturation and one to describe the permeability of the soil. In [12, 57] the so called Kirchhoff transformation is used to shift the nonlinear behaviour of the diffusion coefficient from the domain to the boundary. However, this transformation can only be applied in a homogeneous setting. To carry this idea over to the heterogeneous case, a different formulation is needed to ensure compatibility with the Kirchhoff transformation. The primal hybrid formulation, see [19, 51], is used to derive such kind of formulation. After applying local Kirchhoff transformations a coupled system of equations with a linear capacity coefficient within the subdomains and nonlinear coupling conditions is derived.

The analogy of the resulting continuous formulation to the discrete mortar finite element method was decisive for its application as approximation method. The mortar finite element method was introduced as a nonconforming approach for several approximation methods, see for example [11]. A lot of work was done in the field of the mortar finite element method and several articles were published, the most crucial ones for this work are [15, 16, 41, 70, 71]. Beside these citations, many more publications on that topic are available. In view of efficiency, domain decomposition methods for the mortar finite element method are of special interest, see for example [59]. The precise outline of this thesis is listed in the following section.

Outline

In Chapter 2 the Richards equation will be derived starting with the principle of mass balance. Several laws from hydrology are used to obtain the pressure formulation of the Richards equation which is of interest in this thesis. Furthermore, two cases are distinguished, in the first case a homogeneous soil is considered and in the second case a general heterogeneous soil is assumed.

Afterwards, mathematical preliminaries are summarized. In Section 3.1 basics from functional analysis are repeated, especially statements about linear and nonlinear operator equations in Banach spaces. Section 3.2 is about function spaces and fundamental theorems in those spaces. Finally, superposition operators and related properties are considered in Section 3.3.

The main focus of Chapter 4 is on the derivation of a variational formulation which corresponds to the Richards equation. In Section 4.1 the variational problem will be analyzed in view of well posedness. Solvability as well as uniqueness results will be presented. As already mentioned, one has to use the primal hybrid formulation to obtain a Kirchhoff transformation compatible formulation. This is done in Section 4.2. Finally, in Section 4.3, the Kirchhoff transformation is applied to the primal hybrid formulation.

Discretization and linearization strategies for the transformed variational formulation derived in Chapter 4 are discussed in Chapter 5. As done for the continuous formulation, solvability and uniqueness are investigated and open problems in the discrete setting are pointed out. Furthermore, a brief description on some implementational details is done.

Finally, in Chapter 6, numerical experiments in two and three space dimension are presented and discussed.

The main new result of this thesis is the derived system of local acting partial differential equations coupled via nonlinear coupling conditions and the analysis of the corresponding variational problem. Even though the analysis is only done for the Richards equation, it is rather easy to extend the theory to general quasilinear partial differential equations. The application of the mortar finite element method to compute the approximate solution of the nonlinear transmission problem is also a rather new approach. Although we were not able to answer all the questions concerning the stability of the derived discrete problem, the numerical experiments show promising results.

2 MODELING

Basis of this work is the Richards equation which describes the flow of water in saturated–unsaturated ground. This equation was first published in [52]. In order to derive the Richards equation we use fundamental physical laws and laws from hydrology.

Starting point is the principle of mass balance. Therefore, we consider an arbitrary time dependent control volumina $\omega(t) \subset \mathbb{R}^3$ for $t \geq 0$. Let $\omega(t)$ move with the fluid and let $\tilde{\mathbf{v}}$ be the velocity field describing the speed of the movement. The velocity field $\tilde{\mathbf{v}}$ is called microscopic velocity. We know, that the change of mass in the control volumina $\omega(t)$ is balanced by sources and sinks within $\omega(t)$. Let f prescribe these sources and sinks within $\omega(t)$, therefore we can write the balance equation as

$$\frac{\mathrm{d}}{\mathrm{d}t} \int\limits_{\omega(t)} \tilde{\varrho} \, \mathrm{d}\mathbf{x} = \int\limits_{\omega(t)} \varrho \, f \, \mathrm{d}\mathbf{x}$$

with the mass density function $\tilde{\varrho}$ and the constant water density ϱ. By the application of the Reynolds transport theorem, see [30, Theorem 5.4], we obtain

$$\int\limits_{\omega(t)} \left[\frac{\partial \tilde{\varrho}}{\partial t} + \nabla \cdot (\tilde{\varrho} \, \tilde{\mathbf{v}}) \right] \mathrm{d}\mathbf{x} = \int\limits_{\omega(t)} \varrho f \, \mathrm{d}\mathbf{x} \tag{2.1}$$

for all control volumina $\omega(t)$ and for all $t \geq 0$.

The variational principle applied on $\omega(t)$ in equation (2.1) implies

$$\frac{\partial \tilde{\varrho}}{\partial t} + \nabla \cdot (\tilde{\varrho} \, \tilde{\mathbf{v}}) = \varrho \, f \tag{2.2}$$

which is the starting point for further considerations.

In the field of hydrology the mass density is given as $\tilde{\varrho} = \varrho \, \theta \, n$, θ describes the saturation of the soil and n is the porosity of the soil. The porosity is a quantity which depends on the soil only, that is $n = n(\mathbf{x})$. It is defined as the ratio of the void volume and the bulk volume and hence always positive.

By setting the macroscopic velocity field $\mathbf{v}$ as $\mathbf{v} := \theta \, n \, \tilde{\mathbf{v}}$, we can rewrite equation (2.2) and obtain

$$\frac{\partial (\varrho \, \theta \, n)}{\partial t} + \nabla \cdot (\varrho \, \mathbf{v}) = \varrho \, f.$$

Since $\varrho \equiv$ const and n is just a soil parameter depending on $\mathbf{x}$ and independent of t, we get

$$n\,\frac{\partial\theta}{\partial t} + \nabla\cdot\mathbf{v} = f$$

after cancelling ϱ

The next important result from hydrology is the law of Darcy, see [52]. This law is given by the equation

$$\mathbf{v} = -C\,\nabla h$$

and describes the relation between the macroscopic velocity field and the pressure of the water. Here C is the hydraulic conductivity and is a scalar function if we consider the flow in an isotropic medium. The quantity h is called piezometric head and can be interpreted as the groundwater level at a point $\mathbf{x}\in\mathbb{R}^3$ with $\mathbf{x}=(x,y,z)$, see [7]. Its relation to the pressure p is given by the identity

$$h = \frac{p-d}{\varrho\,g}$$

where g is the gravitational constant and $d(\mathbf{x}) = d(x,y,z) = \varrho\,g\,z$ where z is the component of the coordinate system pointing downwards in the direction of gravity. The pressure p is the difference of the pressure of water p_w and the constant pressure of air p_a, that is $p = p_w - p_a$. The quantity $p/(\varrho\,g)$ is known as the pressure head ψ and comes from a hydrostatic pressure if p is positive or from a capillary pressure or a suction if p is negative.

If the soil is fully saturated, the hydraulic conductivity C is given by the expression

$$C = \frac{\varrho\,g}{\mu}\,K$$

where μ is the viscosity of water. With K we denote the permeability of the soil, which is just a function depending on $\mathbf{x}\in\mathbb{R}^3$ and independent of the fluid. It describes the ability of a porous medium to allow fluids to pass through it and is a positive quantity.

In the unsaturated case, we can describe C in dependency on θ as

$$C(\theta) = k(\theta)\,\frac{\varrho\,g}{\mu}\,K$$

with the so called relative permeability k. The relative permeability k can be prescribed by a function mapping the interval $[\theta_{min},\theta_{max}]$ to the interval $[0,1]$ in a monotonically increasing way. Furthermore, the saturation θ can be written as a monotonically increasing function mapping the hydrostatic pressure p to the interval $[\theta_{min},\theta_{max}]$.

If we put these relations into the equation (2.2) we obtain

$$n\,\frac{\partial\theta(p)}{\partial t} - \nabla\cdot\left(\frac{K}{\mu}\,k\big(\theta(p)\big)\,\nabla\big(p-d\big)\right) = f \tag{2.3}$$

which is known as the pressure formulation of the Richards equation. As one can see, the Richards equation is a quasilinear elliptic–parabolic equation. If the soil is fully saturated, that is $\theta(p) \equiv \theta_{max} \equiv$ const, we obtain a linear elliptic equation and in the unsaturated case, $\theta(p) \neq$ const, we obtain a quasilinear parabolic equation.

There are several possibilities how one can choose the parameter functions for the saturation θ and for the relative permeability k. One possibility is the model based on the work of Brooks and Corey [21, 22] or the model by Van Genuchten [65]. In this thesis we consider the model introduced by Brooks and Corey. In their work they introduced the soil–water retention curve Θ as

$$\Theta(p) := \frac{\theta(p)-\theta_{min}}{\theta_{max}-\theta_{min}} = \begin{cases} \left(\frac{p}{p_b}\right)^{-\lambda}, & p \le p_b, \\ 1, & p > p_b \end{cases}$$

with the so–called bubbling pressure $p_b < 0$ and the pore size distribution factor $\lambda > 0$. Then, the relative permeability k can be expressed in terms of Θ as

$$k(\theta) = \Theta^{e(\lambda)} = \left(\frac{\theta-\theta_{min}}{\theta_{max}-\theta_{min}}\right)^{e(\lambda)} \tag{2.4}$$

with exponent

$$e(\lambda) := \begin{cases} 3+\frac{2}{\lambda} & \text{due to Burdine,} \\ \frac{5}{2}+\frac{2}{\lambda} & \text{due to Mualem,} \end{cases}$$

see [65]. The exponent $e(\lambda)$ due to Burdine was used in the work of Brooks and Corey, therefore we obtain $k(\theta) = \Theta^{3+2/\lambda}$, see Figure 2.1. Within this framework we obtain the following representation for the saturation

$$\theta(p) := \begin{cases} \left(\frac{p}{p_b}\right)^{-\lambda}\,\big(\theta_{max}-\theta_{min}\big)+\theta_{min} & \text{for } p \le p_b, \\ \theta_{max} & \text{for } p > p_b, \end{cases} \tag{2.5}$$

see Figure 2.2a. The relative permeability can then be written in terms of the pressure p as

$$k(\theta(p)) := \begin{cases} \left(\frac{p}{p_b}\right)^{-\lambda\,e(\lambda)} = \left(\frac{p}{p_b}\right)^{-3\lambda-2} & \text{for } p \le p_b, \\ 1 & \text{for } p > p_b, \end{cases} \tag{2.6}$$

see Figure 2.2b.

So far, we have derived a partial differential equation describing the flow of water in porous media and we fixed the choice of the nonlinear parameter functions, which

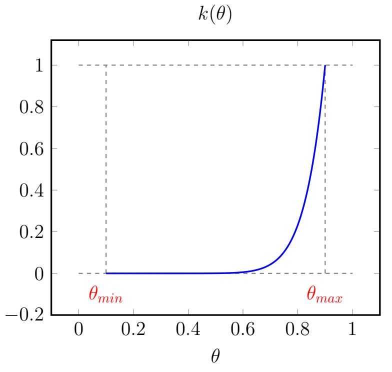
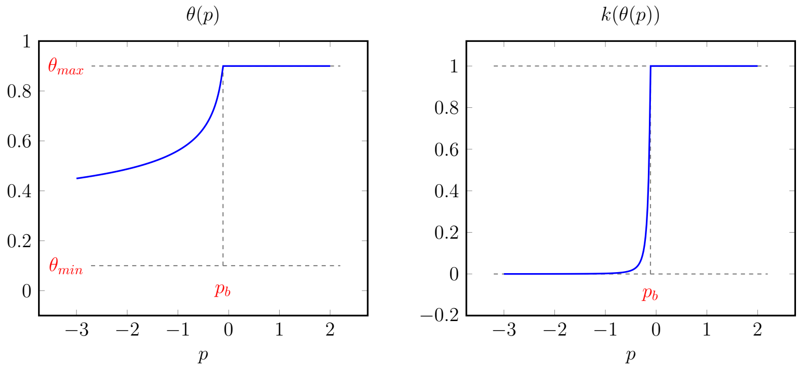


Figure 2.1: Relative permeability, $\theta \mapsto k(\theta)$.



(a) Saturation, $p \mapsto \theta(p)$. (b) Composition, $p \mapsto k(\theta(p))$.

Figure 2.2: Saturation and composition.

occur in the Richards equation (2.3). All the considerations we made assume a single soil type in an isotropic medium. In the following two sections we will briefly discuss two different cases. In the first section we will consider a homogeneous soil type whereas in the second section a heterogeneous soil type is of interest.

2.1 Homogeneous Soil

In this section we discuss a homogeneous soil type. Therefore, we consider a domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, and one specific soil type. See for example Figure 2.3 and assume, that the domain behaves like a sand–type soil.

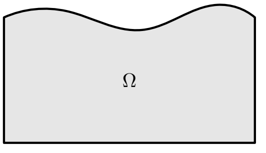


Figure 2.3: Illustration of a homogeneous soil.

Then the characteristic parameter functions k and θ are uniquely determined by the equations (2.4) and (2.5). The nonlinear parameter functions θ and k just depend on the unknown pressure p. If we write equation (2.3) in dependency on $\mathbf{x} \in \Omega$ and $t > 0$ we obtain the equation

$$n(\mathbf{x})\,\frac{\partial\theta(p(\mathbf{x},t))}{\partial t} - \nabla\cdot\left(\frac{K(\mathbf{x})}{\mu}\,k\Big(\theta(p(\mathbf{x},t))\Big)\,\nabla\Big(p(\mathbf{x},t) - d(\mathbf{x})\Big)\right) = f(\mathbf{x},t)$$

for the unknown pressure p.

Next, we want to extend this consideration to the heterogeneous case.

2.2 Heterogeneous Soil

In this section we will consider a heterogeneous soil. Therefore, we consider a domain $\Omega \subset \mathbb{R}^d$ with different layers and we assume that each layer behaves like a different soil type. See for example Figure 2.4 and assume L_1 to behave like a sand–type soil, L_2 behaves like a sandy loam–type soil and L_3 is assumed to behave like a loam–type soil.

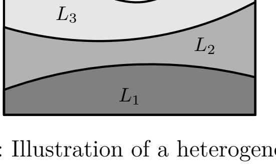


Figure 2.4: Illustration of a heterogeneous soil.

Within each layer $L_i \subset \Omega$, $i = 1, \ldots, N_L$, the corresponding characteristic parameter functions k_i and θ_i are uniquely determined by the equations (2.4) and (2.5) and they just depend on the unknown pressure p. We can define a global parameter function $\theta : \Omega \times \mathbb{R} \to \mathbb{R}$ by defining

$$\theta(\mathbf{x}, p) := \theta_i(p) \qquad \text{for } \mathbf{x} \in L_i \text{ and } p \in \mathbb{R}, \tag{2.7}$$

as well as a global function $k : \Omega \times \mathbb{R} \to \mathbb{R}$ defined by

$$k(\mathbf{x}, \theta) := k_i(\theta) \qquad \text{for } \mathbf{x} \in L_i \text{ and } \theta \in \mathbb{R}. \tag{2.8}$$

With this representation, the composite function $k \circ \theta$ can be written as

$$k(\mathbf{x}, \theta(\mathbf{x}, p)) := k_i(\theta_i(p)) \qquad \text{for } \mathbf{x} \in L_i \tag{2.9}$$

and $p \in \mathbb{R}$.

If we write equation (2.3) in dependency on $\mathbf{x} \in L_i$ and $t > 0$ we obtain the equation

$$n(\mathbf{x}) \, \frac{\partial \theta_i(p(\mathbf{x}, t))}{\partial t} - \nabla \cdot \left(\frac{K(\mathbf{x})}{\mu} \, k_i\Big(\theta_i(p(\mathbf{x}, t))\Big) \, \nabla\Big(p(\mathbf{x}, t) - d(\mathbf{x})\Big) \right) = f(\mathbf{x}, t)$$

for the unknown pressure p in each layer L_i. To complete the set of equations, we assume continuity of the pressure and of the conormal derivative across the interfaces.

The Richards equation for a homogeneous soil coincides with the Richards equation for a heterogeneous soil with just one layer, that is $N_L = 1$. In this thesis we will mainly restrict ourselves to the more general heterogeneous case.

2.3 Boundary and Initial Conditions

The purpose of this section is to understand the hydrological meaning of different boundary conditions. Thus, let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a bounded Lipschitz domain (see Section 3.2.2) with boundary $\partial\Omega$. Then, for all $t > 0$ the Richards equation reads

$$n \, \frac{\partial \theta(p)}{\partial t} - \nabla \cdot \left(\frac{K}{\mu} \, k\Big(\theta(p)\Big) \, \nabla\Big(p - d\Big) \right) = f \quad \text{ in } \Omega$$

in terms of the unknown pressure p. In view of readability, we neglect the dependencies on $\mathbf{x}$ and t. In this thesis, we consider Dirichlet boundary conditions and Neumann boundary conditions. For a fixed time $t > 0$, we denote the Dirichlet boundary by $\Gamma_D \subset \partial\Omega$ and the Neumann boundary by $\Gamma_N \subset \partial\Omega$.

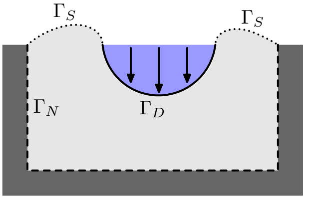


Figure 2.5: Boundary conditions.

In hydrology, the Dirichlet boundary condition prescribe a hydrostatic pressure which is given at the Dirichlet boundary. We denote the given hydrostatic pressure by g_D and so the Dirichlet condition can be written as

$$p = g_D \quad \text{on } \Gamma_D.$$

Dirichlet boundary conditions can appear from surface water like lakes or rivers, see Figure 2.5. Another type of boundary conditions are Neumann type boundary conditions. These boundary conditions prescribe the flow of water into or out of the domain. This can occur due to water movement around Ω or for example rain. Homogeneous Neumann boundary conditions can be used to simulate a impermeable material attached to Ω at Γ_N, see Figure 2.5. We denote the flux across Γ_N by g_N and write

$$\frac{K}{\mu}\, k\Big(\theta(p)\Big)\, \nabla\Big(p - d\Big) \cdot \mathbf{n} = g_N \quad \text{on } \Gamma_N$$

where $\mathbf{n}$ denotes the outer unit normal, see Section 3.2.

It is also possible to take a linear combination of Dirichlet and Neumann boundary conditions, this would lead to a Robin type boundary condition. Apart from this more or less standard boundary conditions, Signorini boundary conditions can also be considered in the context of hydrology. Signorini boundary conditions are well known for contact problems, see [42]. In the hydrological framework, Signorini boundary conditions usually appear in the case of dam problems, see [12].

To obtain a well posed initial boundary value problem, we have to prescribe an initial condition for $t = 0$. We write

$$p = p_0 \quad \text{in } \Omega$$

where p_0 is the given initial condition describing a certain ground state of the pressure p.

To summarize, in this chapter we have derived the Richards equation to describe the flow in porous media and we briefly discussed the case of a homogeneous soil in Section 2.1 and the case of a heterogeneous soil in Section 2.2. In Section 2.3 we discussed the hydrological meaning of Dirichlet and Neumann boundary conditions. In the next chapter we will make some preliminary considerations on Banach spaces, function spaces and operators acting in those spaces.

3 MATHEMATICAL PRELIMINARIES

In this chapter we recall the mathematical tools we use in this thesis. The focus of Section 3.1 is on functional analytic basics. We continue with the introduction of function spaces in Section 3.2 and we will state the needed definitions and theorems. The last part, Section 3.3, is about superposition operators acting in Lebesgue spaces as well as in Sobolev spaces.

3.1 Elementary Functional Analysis

In this section we recall elementary definitions and tools from functional analysis which we need later in this thesis. For more details on this topic we refer to [19, 43, 58, 68, 74], and [75]. In the first part we recall basic definitions on Banach spaces and Hilbert spaces and on operators acting in those spaces.

Let V be a Banach space. The space of all bounded and linear functionals from V to $\mathbb{R}$ is denoted by V' and is called dual space of V. For $u \in V$ and $f \in V'$ the duality pairing

$$\langle f, u \rangle_{V' \times V} := f(u)$$

is defined as the application of f to u. Using the duality pairing the dual norm can therefore be written as

$$\|f\|_{V'} := \sup_{0 \neq u \in V} \frac{\left| \langle f, u \rangle_{V' \times V} \right|}{\|u\|_V}$$

for all $f \in V'$. Consequently, there holds the inequality

$$\langle f, u \rangle_{V' \times V} \leq \|f\|_{V'} \|u\|_V$$

for all $f \in V'$ and $u \in V$. Equipped with the dual norm, the dual space is again a Banach space.

If V is a Hilbert space, we denote by $(u, v)_V$ the inner product in V for $u, v \in V$. We write

$$\|u\|_V := \sqrt{(u, u)_V}$$

for the induced norm on V.

The bidual space V'' of V is defined as the dual space of V'. It is easy to see, that for each $u \in V$ we can construct $U \in V''$ as $U(f) := \langle f, u \rangle_{V' \times V}$ with

$$\|U\|_{V''} = \sup_{0 \neq f \in V'} \frac{|U(f)|}{\|f\|_{V'}} = \sup_{0 \neq f \in V'} \frac{\left|\langle f, u \rangle_{V' \times V}\right|}{\|f\|_{V'}} = \|u\|_V$$

for all $u \in V$, see [74, Section 21.5]. If we set

$$\iota(u) := U \tag{3.1}$$

we obtain a linear injective mapping $\iota : V \to V''$ and $\iota(V) \subset V''$. Using this mapping, we can introduce the concept of reflexivity.

Definition 3.1 (Reflexive Banach space)**.** *Let V be a Banach space. We say V is reflexive iff the mapping $\iota : V \to V''$ is surjective.*

A reflexive Banach space V is therefore normisomorph to V'', we write $V \cong V''$. Due to the Fréchet-Riesz Theorem, [68, Theorem V.3.6], each Hilbert space is a reflexive Banach space. Another important property of Banach spaces is separability.

Definition 3.2 (Separable Banach space)**.** *A Banach space V is separable iff there exists a dense subset $W \subset V$ which is at most countable.*

In contrast to the finite dimensional case, there exists a further concept of convergence in the case of infinite dimensional spaces.

Definition 3.3 (Weak convergence)**.** *Let $\{u_n\} \subset V$ be a sequence in a Banach space V. We say the sequence converges weakly to $u \in V$ iff*

$$\lim_{n \to \infty} \langle f, u_n \rangle_{V' \times V} = \langle f, u \rangle_{V' \times V}$$

for all $f \in V'$. We write $u_n \rightharpoonup u$.

As the name suggests, weak convergence is a generalization of the usual strong convergence. In other words, $u_n \to u$ implies $u_n \rightharpoonup u$ in V.

Since many partial differential equations can be written as abstract operator equations in Banach or Hilbert spaces, we will recall some of the basic definitions and notations on operators.

Definition 3.4 (Linear and nonlinear operator)**.** *Let V, W be two Banach spaces over $\mathbb{R}$. The mapping $A : V \to W$ is called a linear operator iff*

$$A(\eta\, u + \omega\, v) = \eta\, A(u) + \omega\, A(v)$$

holds in W for all $\eta, \omega \in \mathbb{R}$ and $u, v \in V$. Otherwise, the operator is called nonlinear. For a linear operator we will neglect the parentheses and write Au instead of $A(u)$.

Definition 3.5 (Bounded operator)**.** *Let V, W be two Banach spaces and let $A : V \to W$ be an operator mapping V to W. The operator is called bounded, iff the set $A(U) := \{A(u) \mid u \in U\}$ is bounded in W for each bounded set $U \subset V$.*

For operators in Banach spaces there are several different concepts of continuity, see for example [75]. We will just repeat four different types of continuity.

Definition 3.6 (Continuous operator)**.** *Let V, W be two Banach spaces and let $A : V \to W$ be an operator mapping V to W. The operator is called continuous at $u \in V$, if for any sequence $\{u_n\} \subset V$ which converges to u in V, the sequence $\{A(u_n)\} \subset W$ converges to $A(u)$ in W. If the operator is continuous at all $u \in V$, then A is called continuous from V to W.*

The above definition of continuity is based on strong convergent sequences. Using the concept of the more general weak convergence, we can generalize the definition of continuity.

Definition 3.7 (Demicontinuous operator)**.** *Let V, W be two Banach spaces and let $A : V \to W$ be an operator mapping V to W. The operator is called demicontinuous, if $u_n \to u$ in V implies $A(u_n) \rightharpoonup A(u)$ in W as $n \to \infty$.*

In the definition of continuity, we used strong convergence in the domain and in the codomain. In the definition of demicontinuity, we used strong convergence in the domain and weak convergence in the codomain. It is also possible to interchange the concepts of convergence, which leads to the following definition.

Definition 3.8 (Strongly continuous operator)**.** *Let V, W be two Banach spaces and let $A : V \to W$ be an operator mapping V to W. The operator is called strongly continuous, if $u_n \rightharpoonup u$ in V implies $A(u_n) \to A(u)$ in W as $n \to \infty$.*

The last type of continuity is called hemicontinuity and is defined in the following way.

Definition 3.9 (Hemicontinuous operator)**.** *Let V be a real Banach space and let $A : V \to V'$. The operator A is called hemicontinuous, if the mapping*

$$t \mapsto \langle A(u + tv), w \rangle_{V' \times V}$$

is continuous on $[0, 1]$ for all $u, v, w \in V$.

For a real Banach space V and an arbitrary operator $A : V \to V'$ it can be easily checked that strong continuity implies continuity which implies demicontinuity which finally implies hemicontinuity.

The next proposition shows how the boundedness and the continuity of an operator are related, see [58, Proposition 2.1].

Proposition 3.10. *Let V, W be two Banach spaces and let $A : V \to W$ be a linear operator. Then the following statements are equivalent.*

(1) A is continuous at $0 \in V$.
(2) A is a continuous operator from V to W.
(3) A is a bounded operator.

The space of all linear and bounded operators mapping a Banach space V to a Banach space W is denoted by $\mathcal{L}(V, W)$. Equipped with the norm

$$\|A\|_{\mathcal{L}(V,W)} := \sup_{0 \neq u \in V} \frac{\|Au\|_W}{\|u\|_V} = \sup_{\substack{u \in V \\ \|u\|_V = 1}} \|Au\|_W$$

this space is again a Banach space, see [58, Proposition 2.2].

Definition 3.11 (Adjoint operator)**.** *Let V, W be two Banach spaces and let $A : V \to W$ be a linear operator. The adjoint operator $A' : W' \to V'$ is then defined by*

$$\langle A'g, u \rangle_{V' \times V} := \langle g, Au \rangle_{W' \times W}$$

for all $u \in V$ and $g \in W'$.

In order to show results on nonlinear operator equations, we need the following definitions.

Definition 3.12 (Coercive operator)**.** *Let $A : V \to V'$ be an operator mapping a Banach space V to its dual space V'. If*

$$\lim_{\|u\|_V \to \infty} \frac{\langle A(u), u \rangle_{V' \times V}}{\|u\|_V} = \infty$$

then A is called coercive.

Definition 3.13 (V–elliptic operator)**.** *Let $A : V \to V'$ be a continuous and linear operator mapping a Banach space V to its dual space V'. If there exists a constant $\alpha_A > 0$ such that*

$$\langle Au, u \rangle_{V' \times V} \geq \alpha_A \|u\|_V^2 \qquad \text{for all } u \in V,$$

then A is called V–elliptic.

In order to state solvability and uniqueness results of nonlinear operator equations in Banach spaces, the concept of monotonicity is fundamental.

Definition 3.14 (Monotone operator)**.** *Let $A : V \to V'$ be an operator mapping a Banach space V to its dual space V'.*

(1) The operator A is called monotone iff

$$\langle A(u) - A(v), u - v \rangle_{V' \times V} \geq 0 \qquad \text{for all } u, v \in V.$$

(2) The operator A is called strictly monotone iff

$$\langle A(u) - A(v), u - v \rangle_{V' \times V} > 0 \qquad \text{for all } u, v \in V, u \neq v.$$

(3) The operator A is called strongly monotone iff there is a positive constant $\alpha_A > 0$ such that

$$\langle A(u) - A(v), u - v \rangle_{V' \times V} > \alpha_A \|u - v\|_V^2 \qquad \text{for all } u, v \in V.$$

(4) Let V be a real reflexive Banach space. The operator A is called pseudomonotone iff $u_n \rightharpoonup u$ as $n \to \infty$ and

$$\limsup_{n\to\infty} \langle A(u_n), u_n - u \rangle_{V' \times V} \leq 0$$

implies

$$\langle A(u), u - w \rangle_{V' \times V} \leq \liminf_{n\to\infty} \langle A(u_n), u_n - w \rangle_{V' \times V}$$

for all $w \in V$.

It is easy to verify that strong monotonicity implies strict monotonicity which implies monotonicity. In order to show that an operator is pseudomonotone, the following lemma is of importance.

Lemma 3.15. *Let $A : V \to V'$ be an operator mapping a real reflexive Banach space V to its dual. Assume that A satisfies the representation*

$$A(u) = \widetilde{A}(u, u)$$

with $\widetilde{A} : V \times V \to V'$. If $\widetilde{A}$ satisfies

(1) $\widetilde{A}(\cdot, v)$ is hemicontinuous and bounded for all $v \in V$,
(2) $\widetilde{A}(u, \cdot)$ is hemicontinuous for all $u \in V$,
(3) $\widetilde{A}(u, \cdot)$ is monotone,
(4) if $u_n \rightharpoonup u$ in V and $\langle \widetilde{A}(u_n, u_n) - \widetilde{A}(u_n, u), u_n - u \rangle_{V' \times V} \to 0$ implies $\widetilde{A}(u_n, v) \rightharpoonup \widetilde{A}(u, v)$ in V' for all $v \in V$,
(5) if $u_n \rightharpoonup u$ in V and $\widetilde{A}(u_n, v) \rightharpoonup b$ in V' implies $\langle \widetilde{A}(u_n, v), u_n \rangle_{V' \times V} \to \langle b, u \rangle_{V' \times V}$,

then the operator A is pseudomonotone.

Proof. See [32, 6.10 Lemma]. ∎

After we have repeated some of the basic definitions on operators in Banach spaces, we can state the main results on linear and nonlinear operator equations in Banach and Hilbert spaces.

3.1.1 Liner Operator Equations

Let V, W be two Hilbert spaces and let $A : V \to W'$ be a continuous and linear operator. For a given $g \in W'$ consider the linear operator equation

$$Au = g \tag{3.2}$$

in W'. To answer the question of solvability the following theorem is an essential tool.

Theorem 3.16. *Let V, W be two Hilbert spaces and let $A : V \to W'$ be a continuous and linear operator. Then, for some $\alpha_A > 0$ the following statements are equivalent.*

(1) For all $v \in W$ there holds the inf–sup–condition

$$\inf_{0 \neq v \in W} \sup_{0 \neq u \in V} \frac{\langle Au, v \rangle_{W' \times W}}{\|u\|_V \|v\|_W} \geq \alpha_A. \tag{3.3}$$

(2) There exists a $A^\dagger \in \mathcal{L}(W', V)$ such that $A \circ A^\dagger = I$ on W' and

$$\left\| A^\dagger \right\|_{\mathcal{L}(W',V)} \leq {}^1\!/\!\alpha_A.$$

Proof. For a proof see [18, Theorem 0.1]. ∎

Theorem 3.16 plays a major role in the theory of saddle point problems, it ensures solvability of the operator equation (3.2) in Hilbert spaces. The question of uniqueness can be answered by the following theorem.

Theorem 3.17 (Generalized Lax–Milgram–Lemma)**.** *Let V, W be two Hilbert spaces and let $A : V \to W'$ be a continuous and linear operator. Assume there exists a constant $\alpha_A > 0$ such that*

$$\inf_{0 \neq u \in V} \sup_{0 \neq u \in W} \frac{\langle Au, v \rangle_{W' \times W}}{\|u\|_V \|v\|_W} \geq \alpha_A \tag{3.4a}$$

holds and suppose that for all $0 \neq v \in W$ *we have*

$$\sup_{0 \neq u \in V} \langle Au, v \rangle_{W' \times W} \neq 0, \tag{3.4b}$$

then $A : V \to W'$ *is an isomorphism and* $\|A^{-1}\|_{\mathcal{L}(W',V)} \leq {}^{1}/_{\alpha_A}$ *holds.*

Proof. For a proof see [14, Theorem 3.6] or [55, Section 4.2.1]. ■

Theorem 3.17 ensures unique solvability of the equation (3.2).

Corollary 3.18. *Theorem 3.17 remains true if the adjoint conditions are assumed, that is there exists a constant* $\alpha_{A'} > 0$ *such that*

$$\inf_{0 \neq v \in W} \sup_{0 \neq u \in V} \frac{\langle Au, v \rangle_{W' \times W}}{\|u\|_V \|v\|_W} \geq \alpha_{A'} \tag{3.5a}$$

holds and

$$\sup_{0 \neq v \in W} \langle Au, v \rangle_{W' \times W} \neq 0 \tag{3.5b}$$

for all $0 \neq u \in V$. *To be precise, there holds the equivalence of* (3.4) *and* (3.5).

Proof. For a proof see [55, Theorem 2.1.44, Remark 2.1.45]. ■

For $W = V$ one can use the following theorem to show unique solvability of the operator equation (3.2).

Theorem 3.19 (Lax–Milgram–Lemma)**.** *Let* V *be a Hilbert space and let* $A : V \to V'$ *be a continuous and linear operator. Assume there exists a constant* $\alpha_A > 0$ *such that* A *is* V*–elliptic, that is*

$$\langle Au, u \rangle_{V' \times V} \geq \alpha_A \|u\|_V^2 \tag{3.6}$$

for all $u \in V$. *Then* $A : V \to V'$ *is an isomorphism and* $\|A^{-1}\|_{\mathcal{L}(V',V)} \leq {}^{1}/_{\alpha_A}$ *holds.*

Proof. For a proof see [55, Lemma 2.1.51]. ■

Unfortunately these theorems are only applicable to linear operators, but there are also results on the solvability and uniqueness of nonlinear operator equations in Banach spaces.

3.1.2 Nonliner Operator Equations

Let V be a Banach space and let $A : V \to V'$ be a nonlinear operator. For a given $g \in V'$ consider the nonlinear operator equation to find $u \in V$ such that

$$A(u) = g \tag{3.7}$$

in V'.

Theorem 3.20 (Main theorem on monotone operators)**.** *Let V be a real, reflexive and separable Banach space. Let $A : V \to V'$ be a monotone, coercive and hemicontinuous operator. Then for each $g \in V'$ there exists a $u \in V$ such that $A(u) = g$ in V' and the solution set is bounded, convex and closed. If, in addition, A is strictly monotone, then the solution is unique.*

Proof. For a proof see [75, Theorem 26.A]. ∎

The main theorem on monotone operators provides results on solvability and uniqueness of nonlinear operator equations of the form (3.7). Unfortunately, in many cases it is hard to prove that a given operator is strictly monotone or even monotone. Sometimes it is just possible to show pseudomonotonicity, in this case the following theorem is of importance.

Theorem 3.21 (Main theorem on pseudomonotone operators)**.** *Let V be a real, reflexive and separable Banach space of infinite dimension. Let $A : V \to V'$ be a pseudomonotone, bounded and coercive operator. Then for each $g \in V'$ there exists at least one $u \in V$ such that $A(u) = g$ in V'.*

Proof. For a proof see [75, Theorem 27.A]. ∎

We have recalled the most useful theorems on operator equations in Banach spaces. Since we are dealing with partial differential equations, we have to introduce suitable function spaces which allow an application of the abstract theory done in this section. The introduction is done in the next section and we start with classical function spaces. Then we will repeat the Lebesgue spaces and we finish the section with Sobolev spaces.

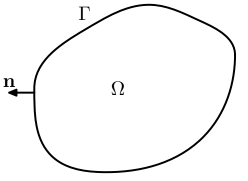


Figure 3.1: Domain (bounded) in $\mathbb{R}^2$.

3.2 Function Spaces

In this section we will introduce some of the most used function spaces. This section is based on [1, 2, 3, 17, 36, 48, 69, 76]. We say $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a domain if Ω is open and connected. We denote its boundary by $\Gamma = \partial\Omega$ and by $\mathbf{n}$ we denote the outer unit normal of Γ.

We begin with the classical spaces of continuous functions.

3.2.1 Classical Spaces

For $d \in \mathbb{N}$ we call $(k_1, k_2, \dots, k_d) = \mathbf{k} \in \mathbb{N}_0^d$ a multi index with absolute value $|\mathbf{k}| = k_1 + k_2 + \ldots + k_d$ and factorial $\mathbf{k}! = k_1!\,k_2!\,\ldots\,k_d!$. Let u be a sufficient smooth function, $u : \Omega \to \mathbb{R}$, we write

$$D^{\mathbf{k}} u(\mathbf{x}) := \frac{\partial^{k_1}}{\partial x_1^{k_1}} \frac{\partial^{k_2}}{\partial x_2^{k_2}} \cdots \frac{\partial^{k_d}}{\partial x_d^{k_d}}\, u(x_1, x_2, \dots, x_d)$$

for the partial derivative of order $|\mathbf{k}|$.

Moreover, we denote by

$$\operatorname{supp} u := \overline{\{\mathbf{x} \in \Omega \mid u(\mathbf{x}) \neq 0\}}$$

the support of u.

Definition 3.22 ($\mathcal{C}^k(\Omega)$–Spaces)**.** *The space $\mathcal{C}^0(\Omega) \equiv \mathcal{C}(\Omega)$ consists of all functions $u : \Omega \to \mathbb{R}$ which are continuous on Ω, that is*

$$\mathcal{C}(\Omega) := \{u : \Omega \to \mathbb{R} \mid u \text{ is continuous }\}.$$

For $k \in \mathbb{N}$ the space $\mathcal{C}^k(\Omega)$ consists of all functions $u : \Omega \to \mathbb{R}$ such that u is differentiable for all multi indices $\mathbf{k} \in \mathbb{N}_0^d$ with $|\mathbf{k}| \leq k$, that is

$$\mathcal{C}^k(\Omega) := \left\{u : \Omega \to \mathbb{R} \mid D^{\mathbf{k}} u \in \mathcal{C}(\Omega)\ \forall\, \mathbf{k} \in \mathbb{N}_0^d \text{ with } |\mathbf{k}| \leq k\right\}.$$

For $k = \infty$, the space $\mathcal{C}^\infty(\Omega)$ is defined as

$$\mathcal{C}^\infty(\Omega) := \bigcap_{k \in \mathbb{N}_0} \mathcal{C}^k(\Omega).$$

Definition 3.23 ($\mathcal{C}^k(\overline{\Omega})$–Spaces)**.** *The space $\mathcal{C}^0(\overline{\Omega}) \equiv \mathcal{C}(\overline{\Omega})$ consists of all functions $u \in \mathcal{C}(\Omega)$ such that u has a continuous extension to $\overline{\Omega}$. For $k \in \mathbb{N}$ the space $\mathcal{C}^k(\overline{\Omega})$ consists of all functions $u \in \mathcal{C}^k(\Omega)$ such that $D^{\mathbf{k}} u \in \mathcal{C}(\overline{\Omega})$ for all $\mathbf{k} \in \mathbb{N}_0^d$ with $|\mathbf{k}| \leq k$. The space $\mathcal{C}^\infty(\overline{\Omega})$ is defined in the same manner as in Definition 3.22.*

With $\|u\|_{\infty,\Omega} := \sup_{\mathbf{x} \in \Omega} |u(\mathbf{x})|$ we can define a norm

$$\|u\|_{\mathcal{C}^k(\Omega)} := \max_{\substack{\mathbf{k} \in \mathbb{N}_0^d \\ |\mathbf{k}| \leq k}} \left\| D^{\mathbf{k}} u \right\|_{\infty,\Omega}$$

on $\mathcal{C}^k(\overline{\Omega})$. Equipped with this norm the space $\mathcal{C}^k(\overline{\Omega})$ is a Banach space.

Definition 3.24 ($\mathcal{C}_0^k(\Omega)$–Spaces)**.** *For $k \in \mathbb{N}_0 \cup \{\infty\}$ we define the space $\mathcal{C}_0^k(\Omega)$ as*

$$\mathcal{C}_0^k(\Omega) := \left\{ u \in \mathcal{C}^k(\Omega) \mid \operatorname{supp} u \Subset \Omega \right\}.$$

This is the space of all functions in $\mathcal{C}^k(\Omega)$ with compact support in Ω. Equipped with the norm $\|\cdot\|_{\mathcal{C}^k(\Omega)}$ the spaces $\mathcal{C}_0^k(\Omega)$ form Banach spaces.

Another important space is the class of Hölder continuous functions. A function $u : \Omega \to \mathbb{R}$ is called Hölder continuous, if there exists positive constants $c_L, \gamma \in \mathbb{R}$ with $\gamma \in (0, 1]$, such that

$$|u(\mathbf{x}) - u(\mathbf{y})| \leq c_L \, |\mathbf{x} - \mathbf{y}|^\gamma$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$. The space of functions $u \in \mathcal{C}(\overline{\Omega})$ satisfying this condition is designated by $\mathcal{C}^{0,\gamma}(\overline{\Omega})$. For $u \in \mathcal{C}^{0,\gamma}(\overline{\Omega})$ we can define the Hölder quotient $[u]_\gamma : \Omega \times \Omega \to \mathbb{R}$ as

$$[u]_\gamma (\mathbf{x}, \mathbf{y}) := \frac{|u(\mathbf{x}) - u(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^\gamma}$$

for $\gamma \in (0, 1]$ and $u \in \mathcal{C}^{0,\gamma}(\overline{\Omega})$.

Definition 3.25 ($\mathcal{C}^{k,\gamma}(\overline{\Omega})$–Spaces)**.** *For $k \in \mathbb{N}$ and $\gamma \in (0, 1]$ we define $\mathcal{C}^{k,\gamma}(\overline{\Omega})$ as*

$$\mathcal{C}^{k,\gamma}(\overline{\Omega}) := \left\{ u \in \mathcal{C}^k(\overline{\Omega}) \mid D^{\mathbf{k}} u \in \mathcal{C}^{0,\gamma}(\overline{\Omega}) \; \forall \, \mathbf{k} \in \mathbb{N}_0^d \; \textit{with} \; |\mathbf{k}| \leq k \right\}.$$

Equipped with the norm

$$\|u\|_{\mathcal{C}^{k,\gamma}(\Omega)} := \|u\|_{\mathcal{C}^{k}(\Omega)} + |u|_{\mathcal{C}^{k,\gamma}(\Omega)} = \|u\|_{\mathcal{C}^{k}(\Omega)} + \max_{\substack{\mathbf{k}\in\mathbb{N}_0^d \\ |\mathbf{k}|=k}} \left\| \left[D^{\mathbf{k}} u \right]_\gamma \right\|_{\infty,\Omega\times\Omega}$$

the space $\mathcal{C}^{k,\gamma}(\overline{\Omega})$ is a Banach space.

For further information about continuous functions and the corresponding function spaces, see [3, Chapter 1], which is the main reference of this subsection. In the next subsection we want to consider a class of more general function spaces, the spaces of integrable functions.

3.2.2 Lebesgue Spaces

In this subsection we assume that $\Omega \subset \mathbb{R}^d$ is a bounded domain, that is, there exists a constant $K > 0$ such that $|\Omega|_d \leq K < \infty$.

Before we introduce the Lebesgue space we need the following definition.

Definition 3.26 ($\mathcal{M}(\Omega)$–Space)**.** *By $\mathcal{M}(\Omega)$ we denote the space of all functions $u : \Omega \to \mathbb{R}$, such that u is measurable, that is*

$$\mathcal{M}(\Omega) := \{u : \Omega \to \mathbb{R} \mid u \textit{ measurable}\}.$$

We introduce the Lebesgue spaces as subsets of $\mathcal{M}(\Omega)$ in the following way.

Definition 3.27 ($L_p(\Omega)$–Spaces)**.** *For $1 \leq p < \infty$ the space $L_p(\Omega)$ is defined as the space of all functions $u : \Omega \to \mathbb{R}$, such that the p-th power of the absolute value is integrable, that is*

$$L_p(\Omega) := \left\{ u : \Omega \to \mathbb{R} \mid u \in \mathcal{M}(\Omega) \textit{ and } \|u\|_{L_p(\Omega)} < \infty \right\}$$

with the norm

$$\|u\|^p_{L_p(\Omega)} := \int_\Omega |u|^p \, \mathrm{d}\mathbf{x}.$$

The Lebesgue space $L_p(\Omega)$ is a Banach space, see [3, 1.21 Satz von Fischer-Risz].

For $p = \infty$ we define $\|u\|_{L_\infty(\Omega)} := \operatorname{ess\,sup}_{\mathbf{x}\in\Omega} |u(\mathbf{x})|$ and the space $L_\infty(\Omega)$ as

$$L_\infty(\Omega) := \left\{ u : \Omega \to \mathbb{R} \mid u \in \mathcal{M}(\Omega) \textit{ and } \|u\|_{L_\infty(\Omega)} < \infty \right\}.$$

The space $L_\infty(\Omega)$ is again a Banach space, see [3, 1.17 Lemma].

For $1 \leq p < \infty$ the dual space of $L_p(\Omega)$ is normisomorph to $L_q(\Omega)$, i.e. $L_p(\Omega)' \cong L_q(\Omega)$, with ${}^1\!/_p + {}^1\!/_q = 1$. In other words, to each functional $f \in L_p(\Omega)'$ there exists an element $u \in L_q(\Omega)$ with $\|f\|_{L_p(\Omega)'} = \|u\|_{L_q(\Omega)}$ such that

$$\langle f, v\rangle_{L_p(\Omega)' \times L_p(\Omega)} = \int\limits_{\Omega} u\, v\, \mathrm{d}\mathbf{x}$$

for all $v \in L_p(\Omega)$, see [68, Satz II.2.4].

Of special interest is the Lebesgue space for $p = 2$. In this case the space $L_2(\Omega)$ is a Hilbert space with the inner product

$$(u, v)_{L_2(\Omega)} := \int\limits_{\Omega} u\, v\, \mathrm{d}\mathbf{x}$$

defined for all $u, v \in L_2(\Omega)$.

The inequalities given below are of special interest since they are frequently used in this thesis.

Theorem 3.28 (Hölder inequality)**.** *Let $p, q \in \mathbb{R}$ with $1 \leq p, q \leq \infty$ and ${}^1\!/_p + {}^1\!/_q = 1$. For $u \in L_p(\Omega)$ and $v \in L_q(\Omega)$ we have $u\, v \in L_1(\Omega)$ and there holds*

$$\int\limits_{\Omega} u\, v\, \mathrm{d}\mathbf{x} \leq \|u\|_{L_p(\Omega)} \|v\|_{L_q(\Omega)}.$$

Proof. For a proof see [2, Theorem 2.4]. ■

The Hölder inequality can be extended to more than two functions. For $n \in \mathbb{N}$ and $i = 1, \ldots, n$ let $u_i \in L_{p_i}(\Omega)$ with $\sum\limits_{i=1}^{n} {}^1\!/_{p_i} = 1$. Then there holds

$$\int\limits_{\Omega} \prod_{i=1}^{n} u_i\, \mathrm{d}\mathbf{x} \leq \prod_{i=1}^{n} \|u_i\|_{L_{p_i}(\Omega)}.$$

This statement can easily be proven by a recursive application of the above theorem.

Theorem 3.29 (Minkowski inequality)**.** *Let $1 \leq p \leq \infty$. For $u, v \in L_p(\Omega)$ there holds*

$$\|u + v\|_{L_p(\Omega)} \leq \|u\|_{L_p(\Omega)} + \|v\|_{L_p(\Omega)}.$$

Proof. For a proof see [3, 1.20 Lemma]. ■

Notation 3.30. *Let* $1 \leq p \leq \infty$*, we write* $[L_p(\Omega)]^d$ *as* $\mathbf{L}_p(\Omega)$ *and* $\|\mathbf{u}\|_{[L_p(\Omega)]^d}$ *as* $\|\mathbf{u}\|_{\mathbf{L}_p(\Omega)}$ *with*

$$\|\mathbf{u}\|^p_{\mathbf{L}_p(\Omega)} := \sum_{i=1}^{d} \|u_i\|^p_{L_p(\Omega)}$$

for all $\mathbf{u} \in \mathbf{L}_p(\Omega)$*. The Hölder inequality and the Minkowski inequality also hold in the* $\mathbf{L}_p(\Omega)$*–spaces.*

To introduce the Sobolev spaces in a proper way we need the concept of weak derivatives. Therefore, let $u \in L_1(\Omega)$ and $\mathbf{k} \in \mathbb{N}_0^d$. We say $v \in L_1(\Omega)$ is the $|\mathbf{k}|$–th weak derivative of u iff

$$\int_\Omega v\varphi \,\mathrm{d}\mathbf{x} = (-1)^{|k|} \int_\Omega u D^{\mathbf{k}} \varphi \,\mathrm{d}\mathbf{x}$$

holds for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$. We write $D^{\mathbf{k}} u = v$. There is even a more general definition of the weak derivative using the space $L_1^{loc}(\Omega)$, see [28, Definition 5.3], but for our purpose the above definition is sufficient.

Definition 3.31. *Using the definition of the weak derivative, we can introduce the weak gradient and the weak divergence. For* $u \in L_1(\Omega)$ *we call* $\mathbf{v} \in \mathbf{L}_1(\Omega)$ *weak gradient if*

$$\int_\Omega \mathbf{v} \cdot \varphi \,\mathrm{d}\mathbf{x} = (-1) \int_\Omega u \,\nabla \cdot \varphi \,\mathrm{d}\mathbf{x}$$

is satisfied for all $\varphi \in [\mathcal{C}_0^\infty(\Omega)]^d$*, we write* $\nabla u = \mathbf{v}$*. Conversely we call* $v \in L_1(\Omega)$ *weak divergence of* $\mathbf{u} \in \mathbf{L}_1(\Omega)$ *if*

$$\int_\Omega v\varphi \,\mathrm{d}\mathbf{x} = (-1) \int_\Omega \mathbf{u} \cdot \nabla\varphi \,\mathrm{d}\mathbf{x}$$

holds for all $\varphi \in \mathcal{C}_0^\infty(\Omega)$*, we write* $\nabla \cdot \mathbf{u} = v$*.*

3.2.3 Sobolev Spaces

For the introduction of the Sobolev spaces we assume that Ω is a bounded Lipschitz domain, see [37, Definition 1.2.1.1], [60, Definition 2.1] or [40, Definition 3.3.1].

Definition 3.32 ($W_p^k(\Omega)$–Spaces)**.** *Let* $k \in \mathbb{N}_0$ *and* $1 \leq p \leq \infty$*. We define the Sobolev space* $W_p^k(\Omega)$ *as*

$$W_p^k(\Omega) := \Big\{ u \in L_p(\Omega) \mid D^{\mathbf{k}} u \in L_p(\Omega) \;\forall\, \mathbf{k} \in \mathbb{N}_0^d \text{ with } |\mathbf{k}| \leq k \Big\}.$$

The corresponding norm is defined as

$$\|u\|^p_{W^k_p(\Omega)} := \sum_{\substack{\mathbf{k}\in\mathbb{N}_0^d \\ |\mathbf{k}|\le k}} \left\|D^{\mathbf{k}}u\right\|^p_{L_p(\Omega)}$$

for $u \in W^k_p(\Omega)$. *Equipped with this norm the* $W^k_p(\Omega)$ *spaces form Banach spaces, see [28, Satz 5.10].*

The norm can equivalently be written as

$$\|u\|^p_{W^k_p(\Omega)} := \|u\|^p_{W^{k-1}_p(\Omega)} + |u|^p_{W^k_p(\Omega)} = \|u\|^p_{W^{k-1}_p(\Omega)} + \sum_{\substack{\mathbf{k}\in\mathbb{N}_0^d \\ |\mathbf{k}|= k}} \left\|D^{\mathbf{k}}u\right\|^p_{L_p(\Omega)}$$

where $|u|^p_{W^k_p(\Omega)}$ denotes the corresponding seminorm.

In [49] it was proven, that the Sobolev spaces can also be characterized as the closure of $\mathcal{C}^\infty(\Omega)$ with respect to the $W^k_p(\Omega)$–norm for $1 \le p < \infty$. For Lipschitz domains $\Omega \subset \mathbb{R}^d$ this statement can be sharpened to

$$W^k_p(\Omega) = \overline{\mathcal{C}^l(\overline{\Omega})}^{\|\cdot\|_{W^k_p(\Omega)}}$$

for all $l \ge k$. The space $W^k_{p,0}(\Omega)$ is defined by

$$W^k_{p,0}(\Omega) := \overline{\mathcal{C}^k_0(\Omega)}^{\|\cdot\|_{W^k_p(\Omega)}}$$

and is the closure of all functions $u \in \mathcal{C}^\infty(\Omega)$ with compact support in Ω. Furthermore, it is a closed subspace of $W^k_p(\Omega)$.

As for Lebesgue spaces, the case $p = 2$ is again of special interest. The Sobolev space $W^k_2(\Omega)$ is a Hilbert space with the inner product

$$(u, v)_{W^k_2(\Omega)} := \sum_{\substack{\mathbf{k}\in\mathbb{N}_0^d \\ |\mathbf{k}|\le k}} \int_\Omega (D^{\mathbf{k}}u)\,(D^{\mathbf{k}}v)\,\mathrm{d}\mathbf{x}$$

for $u, v \in W^k_2(\Omega)$. We identify $W^k_2(\Omega)$ with $H^k(\Omega)$ and $W^k_{2,0}(\Omega)$ with $H^k_0(\Omega)$ where the latter are defined via Fourier transformations, see [48, Chapter 3].

In the space $H^k_0(\Omega)$ the functional $|\cdot|_{H^k(\Omega)}$ defines a norm and is equivalent to the standard norm $\|\cdot\|_{H^k(\Omega)}$. Thus, there holds the norm equivalence

$$c_N \|u\|_{H^k(\Omega)} \le |u|_{H^k(\Omega)} \le \|u\|_{H^k(\Omega)}$$

for some $0 < c_N < 1$ and all $u \in H_0^k(\Omega)$, see [36, Theorem 1.1], we write $\|u\|_{H^k(\Omega)} \simeq |u|_{H^k(\Omega)}$.

It is also possible to introduce real order Sobolev spaces, this is done in the following definition.

Definition 3.33 ($W_p^s(\Omega)$–Spaces)**.** *Let $1 \leq p \leq \infty$, $s \in \mathbb{R}$ with $s = k + \sigma$ where $k \in \mathbb{N}_0$ and $\sigma \in (0,1)$. We define the Sobolev space $W_p^s(\Omega)$ as*

$$W_p^s(\Omega) := \left\{ u \in W_p^k(\Omega) \mid \left[D^{\mathbf{k}} u \right]_\gamma \in L_p(\Omega \times \Omega) \; \forall \, \mathbf{k} \in \mathbb{N}_0^d \text{ with } |\mathbf{k}| \leq k \right\}$$

with $\gamma := \frac{d}{p} + \sigma$ and the norm

$$\begin{aligned} \|u\|^p_{W_p^s(\Omega)} &:= \|u\|^p_{W_p^k(\Omega)} + |u|^p_{W_p^\sigma(\Omega)} = \|u\|^p_{W_p^k(\Omega)} + \sum_{\substack{\mathbf{k} \in \mathbb{N}_0^d \\ |\mathbf{k}| = k}} \left\| \left[D^{\mathbf{k}} u \right]_\gamma \right\|^p_{L_p(\Omega \times \Omega)} \\ &= \|u\|^p_{W_p^k(\Omega)} + \sum_{\substack{\mathbf{k} \in \mathbb{N}_0^d \\ |\mathbf{k}| = k}} \int_\Omega \int_\Omega \frac{\left| D^{\mathbf{k}} u(\mathbf{x}) - D^{\mathbf{k}} u(\mathbf{y}) \right|^p}{|\mathbf{x} - \mathbf{y}|^{d + \sigma p}} \, \mathrm{d}\mathbf{y} \, \mathrm{d}\mathbf{x} \end{aligned}$$

which is known as the Sobolev–Slobodetskii norm. Equipped with this norm the $W_p^s(\Omega)$ spaces form Banach spaces, see [37, Section 1.3].

In the next step we define Sobolev spaces on manifolds. For this reason consider $\Gamma = \partial\Omega$, the boundary of the bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$.

Definition 3.34 ($W_p^s(\Gamma)$–Spaces)**.** *Let $s \in (0,1)$ and $1 \leq p < \infty$. Then the space $W_p^s(\Gamma)$ is defined as*

$$W_p^s(\Gamma) := \left\{ u \in L_p(\Gamma) \mid [u]_\gamma \in L_p(\Gamma \times \Gamma) \right\}$$

with $\gamma := \frac{(d-1)}{p} + s$. Equipped with the norm

$$\begin{aligned} \|u\|^p_{W_p^s(\Gamma)} &:= \|u\|^p_{L_p(\Gamma)} + |u|^p_{W_p^s(\Gamma)} := \|u\|^p_{L_p(\Gamma)} + \left\| [u]_\gamma \right\|^p_{L_p(\Gamma \times \Gamma)} \\ &= \|u\|^p_{L_p(\Gamma)} + \int_\Gamma \int_\Gamma \frac{|u(\mathbf{x}) - u(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{(d-1) + s p}} \, \mathrm{d}s_{\mathbf{y}} \, \mathrm{d}s_{\mathbf{x}} \end{aligned}$$

the fractional order Sobolev space $W_p^s(\Gamma)$ is a Banach space, see [1, Theorem 7.51].

For $p = 2$ the Sobolev space $W_2^s(\Gamma)$ is a Hilbert space with the inner product

$$(u,v)_{W_2^s(\Gamma)} := \int\limits_\Gamma u\,v\,\mathrm{d}s_{\mathbf{x}} + \int\limits_\Gamma\int\limits_\Gamma \frac{\big(u(\mathbf{x})-u(\mathbf{y})\big)\big(v(\mathbf{x})-v(\mathbf{y})\big)}{|\mathbf{x}-\mathbf{y}|^{(d-1)+2s}}\,\mathrm{d}s_{\mathbf{y}}\,\mathrm{d}s_{\mathbf{x}}.$$

for $u, v \in W_2^s(\Gamma)$. We identify $W_2^s(\Gamma)$ with $H^s(\Gamma)$.

Consider for a moment the space $\mathcal{C}^\infty(\overline{\Omega})$, then we can define the operator γ_Γ^0 with $\gamma_\Gamma^0 : \mathcal{C}^\infty(\overline{\Omega}) \to \mathcal{C}^\infty(\Gamma)$ as the restriction of u to the boundary, that is

$$\gamma_\Gamma^0\, u := u_{|\Gamma}$$

for all $u \in \mathcal{C}^\infty(\overline{\Omega})$. Since $\mathcal{C}^\infty(\overline{\Omega})$ is dense in $W_p^1(\Omega)$, the operator γ_Γ^0 can be extended in a continuous way, see the following theorem.

Theorem 3.35 (Trace Theorem)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with boundary $\Gamma = \partial\Omega$. For $1/2 < s < 3/2$ the operator γ_Γ^0 can be extended to a linear operator $\gamma_\Gamma^0 : H^s(\Omega) \to H^{s-1/2}(\Gamma)$ such that*

$$\left\|\gamma_\Gamma^0\, u\right\|_{H^{s-1/2}(\Gamma)} \le c_T \left\|u\right\|_{H^s(\Omega)}$$

holds for all $u \in H^s(\Omega)$ and a constant $c_T > 0$.

Proof. For a proof see [48, Theorem 3.38]. ■

The trace theorem stated in this work is a simplification of a more general trace theorem for $W_p^k(\Omega)$–spaces, see for example [1, Theorem 7.53] or [3, Theorem A6.6].

Using the trace theorem, the space $H_0^1(\Omega)$ can be characterized as the space of all $u \in H^1(\Omega)$ with $\gamma_\Gamma^0\, u = 0$.

Theorem 3.36 (Inverse Trace Theorem)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. There exists a continuous operator $\mathcal{E}_\Omega : H^{s-1/2}(\Gamma) \to H^s(\Omega)$ and a constant $c_E > 0$ such that*

$$\left\|\mathcal{E}_\Omega\, g\right\|_{H^s(\Omega)} \le c_E \left\|g\right\|_{H^{s-1/2}(\Gamma)}$$

holds for all $g \in H^{s-1/2}(\Gamma)$ and $1/2 < s < 3/2$. Furthermore, there holds $g = \gamma_\Gamma^0\, \mathcal{E}_\Omega\, g$ for all $g \in H^{s-1/2}(\Gamma)$.

Proof. For a proof see [48, Theorem 3.38]. ■

The inverse trace theorem is also true for higher order Sobolev spaces, see for example [1, p. 7.56]. If just a non–empty open subset $\Gamma_0 \subset \Gamma = \partial\Omega$ is considered, we define the space $H^s(\Gamma_0)$ as

$$H^s(\Gamma_0) := \left\{ g \in L_2(\Gamma_0) \mid \exists\, \widetilde{g} \in H^s(\Gamma) : \widetilde{g}_{|\Gamma_0} = g \right\}.$$

Equipped with the norm

$$\|g\|_{H^s(\Gamma_0)} = \inf_{\substack{\widetilde{g} \in H^s(\Gamma) \\ \widetilde{g}_{|\Gamma_0} = g}} \|\widetilde{g}\|_{H^s(\Gamma)}$$

the space $H^s(\Gamma_0)$ is again a Banach space.

Remark 3.37. *For a non–empty open subset $\Gamma_0 \subset \Gamma = \partial\Omega$, the Trace Theorem 3.35 remains true since*

$$\left\|\gamma^0_{\Gamma_0} u\right\|_{H^s(\Gamma_0)} \leq \left\|\gamma^0_{\Gamma} u\right\|_{H^s(\Gamma)} \leq c_T \|u\|_{H^s(\Omega)}$$

holds for all $u \in H^s(\Omega)$. A similar result holds for Theorem 3.36.

For a subset $\Gamma_0 \subset \partial\Omega$ and by using the trace operator we can define the space

$$H^1_{0,\Gamma_0}(\Omega) := \left\{ u \in H^1(\Omega) \mid \gamma^0_{\Gamma_0} u = 0 \right\}$$

which is a closed subspace of $H^1(\Omega)$ and there also holds the equivalence

$$c_N \|u\|_{H^1(\Omega)} \leq |u|_{H^1(\Omega)} \leq \|u\|_{H^1(\Omega)} \tag{3.8}$$

for some $0 < c_N < 1$ and all $u \in H^1_{0,\Gamma_0}(\Omega)$, we write $\|u\|_{H^1(\Omega)} \simeq |u|_{H^1(\Omega)}$.

The Trace Theorem 3.35 and the Inverse Trace Theorem 3.36 are one of the most crucial tools in the analysis of boundary value problems. One further important property of Sobolev spaces are imbeddings.

Theorem 3.38 (Imbedding Theorem)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and $1 \leq p, q < \infty$. For $l, k \in \mathbb{N}_0$ with $0 \leq l \leq k$ the following imbeddings hold.*

(1) If $(k-l)\,p \leq d$ and $k - d/p \geq l - d/q$, we have $W^k_p(\Omega) \subsetneq W^l_q(\Omega)$ and there exists a constant $c_I > 0$, such that $\|u\|_{W^l_q(\Omega)} \leq c_I \|u\|_{W^k_p(\Omega)}$ holds for all $u \in W^k_p(\Omega)$. The imbedding is compact if $k - d/p > l - d/q$ holds, we write $W^k_p(\Omega) \subsetneq_c W^l_q(\Omega)$.

(2) If $(k-l)\,p > d$ and $k - d/p \geq l + \gamma$ for $\gamma \in (0,1)$, we have $W^k_p(\Omega) \subsetneq \mathcal{C}^{l,\gamma}(\overline{\Omega})$ and there exists a constant $c_I > 0$, such that $\|u\|_{\mathcal{C}^{l,\gamma}(\overline{\Omega})} \leq c_I \|u\|_{W^k_p(\Omega)}$ holds for all $u \in W^k_p(\Omega)$. The imbedding is compact if $k - d/p > l + \gamma$ holds, we write $W^k_p(\Omega) \subsetneq_c \mathcal{C}^{l,\gamma}(\overline{\Omega})$.

Proof. For a proof see [2, Theorem 4.12, Remark 4.13, Theorem 6.3, Remark 6.4]. ■

Theorem 3.38 holds for integer order Sobolev spaces, but there is also a version of the imbedding theorem which holds for real order Sobolev spaces.

Theorem 3.39 (Imbedding Theorem). *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and $1 \leq p < \infty$ and let $s \in (0,1)$.*

(1) If $s\,p \leq d$ and $s - d/p \geq -d/q$, we have $W_p^s(\Omega) \subsetneq L_q(\Omega)$ and there exists a constant $c_I > 0$, such that $\|u\|_{L_q(\Omega)} \leq c_I \|u\|_{W_p^s(\Omega)}$ holds for all $u \in W_p^s(\Omega)$.

(2) If $s\,p > d$ and $s - d/p \geq \gamma$ for $\gamma \in (0,1)$, we have $W_p^s(\Omega) \subsetneq \mathcal{C}^{0,\gamma}(\overline{\Omega})$ and there exists a constant $c_I > 0$, such that $\|u\|_{\mathcal{C}^{0,\gamma}(\overline{\Omega})} \leq c_I \|u\|_{W_p^s(\Omega)}$ holds for all $u \in W_p^s(\Omega)$.

Proof. For a proof see [26, Theorem 6.7, Theorem 6.10, Theorem 8.2]. ■

Remark 3.40. *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with Lipschitz boundary $\Gamma = \partial\Omega$. We know, that the $H^{1/2}(\Gamma)$–norm is equivalent to the $\mathcal{H}^{1/2}(\Gamma)$–norm defined by*

$$\|g\|_{\mathcal{H}^{1/2}(\Gamma)} := \left(\sum_{i=1}^{N_C} \|g \circ T_i\|^2_{H^{1/2}(Q)} \right)^{1/2}$$

for $g \in H^{1/2}(\Gamma)$, see [40, Section 4.2]. In the above definition $N_C \in \mathbb{N}$ is the finite number of covers of Γ, T_i describes the transformation between the local and global coordinate system and $Q \subset \mathbb{R}^{d-1}$ is the parameter domain, see for example [40, Definition 3.3.1]. Therefore, the imbedding theorems are applicable with $d-1$. The statements remain true if just a part of the boundary, $\Gamma_0 \subset \Gamma = \partial\Omega$, is considered.

The final space we want to introduce is the space $H^{div}(\Omega)$, for further information see for example [62, Chapter 20].

Definition 3.41 ($H^{div}(\Omega)$–Space). *In view of Definition 3.31, we can define the space $H^{div}(\Omega)$ as*

$$H^{div}(\Omega) := \{\mathbf{q} \in \mathbf{L}_2(\Omega) \mid \nabla \cdot \mathbf{q} \in L_2(\Omega)\}.$$

This is the space of all functions in $\mathbf{L}_2(\Omega)$ such that the weak divergence is in $L_2(\Omega)$. Equipped with the norm

$$\|\mathbf{q}\|^2_{H^{div}(\Omega)} := \|\mathbf{q}\|^2_{\mathbf{L}_2(\Omega)} + \|\nabla \cdot \mathbf{q}\|^2_{L_2(\Omega)}$$

the space $H^{div}(\Omega)$ is a Banach space. By defining the inner product as

$$(\mathbf{q}, \mathbf{r})_{H^{div}(\Omega)} := (\mathbf{q}, \mathbf{r})_{\mathbf{L}_2(\Omega)} + (\nabla \cdot \mathbf{q}, \nabla \cdot \mathbf{r})_{L_2(\Omega)}$$

we see, that $H^{div}(\Omega)$ is a Hilbert space.

Consider for a moment the mapping $\mathbf{q} \mapsto \mathbf{q} \cdot \mathbf{n}$ defined from $[\mathcal{C}^\infty(\overline{\Omega})]^d$ into $L_\infty(\Gamma)$. As for the trace operator, this operator can be extended to the space $H^{div}(\Omega)$ in a continuous way.

Theorem 3.42. *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with Lipschitz boundary $\partial\Omega$. Then, the linear mapping $\mathbf{q} \mapsto \mathbf{q} \cdot \mathbf{n}$ can be extended to a continuous and surjective operator from $H^{div}(\Omega)$ into $H^{1/2}(\Gamma)'$. Furthermore, there exists a constant $c_{NT} > 0$, such that*

$$\|\mathbf{q} \cdot \mathbf{n}\|_{H^{1/2}(\Gamma)'} \leq c_{NT} \|\mathbf{q}\|_{H^{div}(\Omega)}$$

holds for all $\mathbf{q} \in H^{div}(\Omega)$.

Proof. For a proof see [62, Lemma 20.2]. ■

Due to the density of $\mathcal{C}^\infty(\overline{\Omega})$ in $H^1(\Omega)$ as well as the density of $[\mathcal{C}^\infty(\overline{\Omega})]^d$ in $H^{div}(\Omega)$, Green's formula can be extended to the following result, see [36, Theorem 2.4, Theorem 2.5].

Lemma 3.43 (Green's Formula)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Then Green's formula*

$$\int\limits_{\Omega} \Big[\mathbf{q} \cdot \nabla u + \nabla \cdot \mathbf{q}\, u\Big] \,\mathrm{d}\mathbf{x} = \langle \mathbf{q} \cdot \mathbf{n}, \gamma_\Gamma^0\, u\rangle_{H^{1/2}(\Gamma)' \times H^{1/2}(\Gamma)}$$

holds for all $u \in H^1(\Omega)$ and $\mathbf{q} \in H^{div}(\Omega)$.

We introduced the Sobolev spaces in Lipschitz domains and on boundaries of Lipschitz domains. Furthermore, we repeated the basic theorems and lemmata in Sobolev spaces which we need in this thesis. In the next section we will recall the mapping properties of superposition operators in Lebesgue spaces and Sobolev spaces.

3.3 Superposition Operators

Since we are dealing with nonlinear problems, superposition operators play an important role. This section is based on [5] and on [75]. In this section let Ω be a domain in the Euclidean space $\mathbb{R}^d$ and $m \in \mathbb{N}$ an arbitrary natural number.

Definition 3.44 (Superposition operator)**.** *Let $u_1, \dots, u_m$ be real valued functions defined almost everywhere on a domain $\Omega \subset \mathbb{R}^d$ and let $l : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be a real valued functional. Then we define the superposition operator (also known as Nemyckii operator) l as*

$$\Big(l(\mathbf{u})\Big)(\mathbf{x}) := l\Big(\mathbf{x}, u_1(\mathbf{x}), \dots, u_m(\mathbf{x})\Big)$$

by the pointwise application of l to $\mathbf{u} = (u_1, \ldots, u_m)^\top$ on Ω.

Let V be a normed space consisting of a subset of all measurable functions on the open domain Ω and let $m = 1$. If the superposition operator satisfies $l(u) \in V$ for all $u \in V$, we say that the operator l acts on the space V.

The next step is to consider the mapping properties of such an operator. Since we only deal with Sobolev spaces and Lebesgue spaces in this thesis, we restrict ourselves to mapping properties of superposition operators in such spaces. For further information about mapping properties in other spaces see [5].

Definition 3.45 (Carathéodory function)**.** *Let $l : \Omega \times \mathbb{R}^m \to \mathbb{R}$. If l satisfies*

(1) $l(\mathbf{x}, \cdot) : \mathbb{R}^m \to \mathbb{R}$ is continuous for almost all $\mathbf{x} \in \Omega$,
(2) $l(\cdot, \mathbf{s}) : \Omega \to \mathbb{R}$ is measurable for all $\mathbf{s} \in \mathbb{R}^m$,

then l is said to be a Carathéodory function.

The Carathéodory property is crucial for the following theorem about the mapping properties of superposition operators in Lebesgue spaces.

Theorem 3.46 (Acting conditions in Lebesgue spaces)**.** *Let $p_1, \ldots, p_m, q \in \mathbb{R}$ with $1 \le p_1, \ldots, p_m, q < \infty$ and let $l : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be a Carathéodory function satisfying*

$$|l(\mathbf{x}, \mathbf{s})| \le a(\mathbf{x}) + b \sum_{i=1}^{m} |s_i|^{p_i/q}$$

for almost all $\mathbf{x} \in \Omega$ and every $\mathbf{s} \in \mathbb{R}^m$ with $a \in L_p(\Omega)$ and a non negative b. Then the related superposition operator l maps $\prod_{i=1}^{m} L_{p_i}(\Omega)$ into $L_q(\Omega)$ continuously and it is furthermore bounded.

If there exists a function $a \in L_\infty(\Omega)$ such that

$$|l(\mathbf{x}, \mathbf{s})| \le a(\mathbf{x})$$

for almost all $\mathbf{x} \in \Omega$ and every $\mathbf{s} \in \mathbb{R}^m$, then l maps $\prod_{i=1}^{m} L_{p_i}(\Omega)$ into $L_\infty(\Omega)$.

Proof. See for example [5, Chapter 3] or [75, Section 26.3]. ■

We have already discussed some mapping properties of superposition operators. The following lemmata show some additional properties of superposition operators based on properties of the underlying function $l : \Omega \times \mathbb{R}^m \to \mathbb{R}$. Similar results to Lemma 3.47 and Lemma 3.48 can be found in [75, Section 26.3] where stronger assumptions on $l : \Omega \times \mathbb{R}^m \to \mathbb{R}$ are made.

Lemma 3.47. *Let $l : \Omega \times \mathbb{R} \to \mathbb{R}$ such that the related superposition operator l is a continuous mapping from $L_p(\Omega)$ to $L_q(\Omega)$ with $1 \le p, q < \infty$ and $1/p + 1/q = 1$. If $l(\mathbf{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ is monotonically increasing for almost all $\mathbf{x} \in \Omega$, then the operator $l : L_p(\Omega) \to L_q(\Omega)$ is monotone.*

Proof. In this proof we want to show that $l : L_p(\Omega) \to L_q(\Omega)$ is monotone, that is

$$\langle l(u) - l(v), u - v \rangle_{L_q(\Omega) \times L_p(\Omega)} = \int_\Omega \Big(l(u) - l(v) \Big) \Big(u - v \Big) \, \mathrm{d}\mathbf{x} \ge 0$$

for all $u, v \in L_p(\Omega)$. The idea is to show the pointwise monotonicity

$$\Big(l\Big(\mathbf{x}, u(\mathbf{x})\Big) - l\Big(\mathbf{x}, v(\mathbf{x})\Big) \Big) \Big(u(\mathbf{x}) - v(\mathbf{x}) \Big) \ge 0 \tag{3.9}$$

for almost all $\mathbf{x} \in \Omega$. For this, choose arbitrary $u, v \in L_p(\Omega)$. We have the decomposition $\Omega = \Omega^+ \cup \Omega^-$ with $\Omega^+, \Omega^- \subset \Omega$ such that $u \ge v$ almost everywhere in Ω^+ and $u < v$ almost everywhere in Ω^-. It is easy to verify that the equation (3.9) is pointwise satisfied almost everywhere in Ω^+ and almost everywhere in Ω^-. Therefore, inequality (3.9) is pointwise satisfied almost everywhere in Ω and we obtain

$$\int_\Omega \Big(l(u) - l(v) \Big) \Big(u - v \Big) \, \mathrm{d}\mathbf{x} \ge 0$$

for arbitrary $u, v \in L_p(\Omega)$. ∎

Lemma 3.48. *Let $l : \Omega \times \mathbb{R} \to \mathbb{R}$ be a mapping such that the related superposition operator l is a continuous mapping from $L_p(\Omega)$ to $L_q(\Omega)$ with $1 \le p, q < \infty$. If $l(\mathbf{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ satisfies $|l(\mathbf{x}, r) - l(\mathbf{x}, s)| \le c_L \, |r - s|^{p/q}$ with a constant $c_L > 0$ for almost all $\mathbf{x} \in \Omega$, then $\|l(u) - l(v)\|_{L_q(\Omega)} \le c_L \, \|u - v\|_{L_p(\Omega)}^{p/q}$ for all $u, v \in L_p(\Omega)$.*

Proof. As in the proof of Lemma 3.47 we choose arbitrary $u, v \in L_p(\Omega)$. Since $|l(\mathbf{x}, r) - l(\mathbf{x}, s)| \le c_L \, |r - s|^{p/q}$ holds for almost all $\mathbf{x} \in \Omega$ we have

$$\begin{aligned} \|l(u) - l(v)\|_{L_q(\Omega)}^q &= \int_\Omega |l(u) - l(v)|^q \, \mathrm{d}\mathbf{x} \le \int_\Omega {c_L}^q \, |u - v|^p \, \mathrm{d}\mathbf{x} \\ &= \left(c_L \, \|u - v\|_{L_p(\Omega)}^{p/q} \right)^q \end{aligned}$$

which shows the desired statement. ∎

Lemma 3.47 and Lemma 3.48 can be proven in the same way for L_p–spaces defined on a submanifold $\Gamma_0 \subset \partial\Omega$.

We have repeated the theorems and lemmata we need in this thesis concerning superposition operators in Lebesgue spaces. We will continue with superposition operator acting on Sobolev spaces.

Theorem 3.49. *Let $l : \mathbb{R} \to \mathbb{R}$ be a uniformly Lipschitz continuous function, let $1 \leq p < \infty$ and let $\Omega \subset \mathbb{R}^d$ be open. Then the related superposition operator l acts on $W^1_p(\Omega)$ continuously if either $l(0) = 0$ or $|\Omega|_d < \infty$. Furthermore, the representation*

$$\nabla l(u) = l'(u)\, \nabla u$$

holds in $\mathbf{L}_2(\Omega)$.

Proof. For a proof see [46, 47]. ■

In Theorem 3.49 we denote by l' the superposition operator related to the derivative of l, which exists almost everywhere due to Rademachers's theorem, see for example [31, Theorem 2 in Section 3.1.2].

Theorem 3.50. *Let $l : \mathbb{R} \to \mathbb{R}$ be a uniformly Lipschitz continuous function with Lipschitz constant $c_L > 0$, $p \in \mathbb{R}$ with $1 \leq p < \infty$ and $s \in (0, 1)$. Let $\Omega \subset \mathbb{R}^d$ be a domain with boundary $\Gamma = \partial\Omega$. Then the related superposition operator l acts on $W^s_p(\Gamma)$ continuously if either $l(0) = 0$ or $|\Gamma|_{d-1} < \infty$.*

Proof. First we want to show the mapping property of l, therefore let $u \in W^s_p(\Gamma)$. From

$$\begin{aligned}\|l(u)\|_{L_p(\Gamma)} &= \|l(u) - l(0) + l(0)\|_{L_p(\Gamma)} \leq \|l(u) - l(0)\|_{L_p(\Gamma)} + \|l(0)\|_{L_p(\Gamma)} \\ &\leq c_L \|u\|_{L_p(\Gamma)} + |l(0)|\, |\Gamma|^{1/p}_{d-1}\end{aligned}$$

and

$$\begin{aligned}\left\|[l(u)]_\gamma\right\|^p_{L_p(\Gamma\times\Gamma)} &= \int_\Gamma \int_\Gamma \frac{|l(u(\mathbf{x})) - l(u(\mathbf{y}))|^p}{|\mathbf{x} - \mathbf{y}|^{(d-1)+s\,p}}\, \mathrm{d}s_{\mathbf{y}}\, \mathrm{d}s_{\mathbf{x}} \\ &\leq c_L^p \int_\Gamma \int_\Gamma \frac{|u(\mathbf{x}) - u(\mathbf{y})|^p}{|\mathbf{x} - \mathbf{y}|^{(d-1)+s\,p}}\, \mathrm{d}s_{\mathbf{y}}\, \mathrm{d}s_{\mathbf{x}} = c_L^p \left\|[u]_\gamma\right\|^p_{L_p(\Gamma\times\Gamma)}\end{aligned}$$

we obtain that

$$\begin{aligned}\|l(u)\|^p_{W^s_p(\Gamma)} &= \|l(u)\|^p_{L_p(\Gamma)} + \left\|[l(u)]_\gamma\right\|^p_{L_p(\Gamma\times\Gamma)} \\ &\leq \left(c_L \|u\|_{L_p(\Gamma)} + |l(0)|\, |\Gamma|^{1/p}_{d-1}\right)^p + c_L^p \left\|[u]_\gamma\right\|^p_{L_p(\Gamma\times\Gamma)} < \infty\end{aligned}$$

and therefore $l : W^s_p(\Gamma) \to W^s_p(\Gamma)$.

From Theorem 3.46 we conclude that $l : L_p(\Gamma) \to L_p(\Gamma)$ is continuous. To show continuity in $W^s_p(\Gamma)$ we take a sequence $\{u_n\} \subset W^s_p(\Gamma) \subset L_p(\Gamma)$ such that $u_n \to u$ in $W^s_p(\Gamma)$. Since $\{u_n\}$ converges in $W^s_p(\Gamma)$ we have convergence in $L_p(\Gamma)$. This implies the existence of a subsequence $\{u_{n'}\} \subset \{u_n\}$ such that $u_{n'}(\mathbf{x}) \to u(\mathbf{x})$ for almost all $\mathbf{x} \in \Gamma$ as $n' \to \infty$. Next, define $f_{n'} : \Gamma \times \Gamma \to \mathbb{R}$ as

$$f_{n'}(\mathbf{x},\mathbf{y}) := \frac{l(u_{n'}(\mathbf{x})) - l(u_{n'}(\mathbf{y}))}{|\mathbf{x}-\mathbf{y}|^{\frac{(d-1)}{p}+s}} - \frac{l(u(\mathbf{x})) - l(u(\mathbf{y}))}{|\mathbf{x}-\mathbf{y}|^{\frac{(d-1)}{p}+s}}$$

for $n' \in \mathbb{N}$. Furthermore, define $\Sigma := \Gamma \times \Gamma$ as well as $\Sigma_D := \{(\mathbf{x},\mathbf{x}) \in \Sigma \mid \mathbf{x} \in \Gamma\}$. Since $u_{n'}$ converges to u for almost all $\mathbf{x} \in \Gamma$, we have

$$f_{n'}(\mathbf{x},\mathbf{y}) \to 0 \text{ as } n' \to \infty$$

for almost all $(\mathbf{x},\mathbf{y}) \in \Sigma \setminus \Sigma_D$. Due to the fact that $|\Sigma_D|_{(d-1)(d-1)} = 0$, we obtain convergence for almost all $(\mathbf{x},\mathbf{y}) \in \Sigma$.

With $g_{n'}(\mathbf{x},\mathbf{y}) := \frac{(2\,c_L)^p}{2}\,\Big([u_{n'}]_\gamma\,(\mathbf{x},\mathbf{y}) + [u]_\gamma\,(\mathbf{x},\mathbf{y})\Big)$ we have the estimate

$$\begin{aligned}
|f_{n'}(\mathbf{x},\mathbf{y})|^p &= \left| \frac{l(u_{n'}(\mathbf{x})) - l(u_{n'}(\mathbf{y}))}{|\mathbf{x}-\mathbf{y}|^{\frac{(d-1)}{p}+s}} - \frac{l(u(\mathbf{x})) - l(u(\mathbf{y}))}{|\mathbf{x}-\mathbf{y}|^{\frac{(d-1)}{p}+s}} \right|^p \\
&\le 2^{p-1}\, \frac{|l(u_{n'}(\mathbf{x})) - l(u_{n'}(\mathbf{y}))|^p + |l(u(\mathbf{x})) - l(u(\mathbf{y}))|^p}{|\mathbf{x}-\mathbf{y}|^{(d-1)+s\,p}} \\
&\le 2^{p-1}\, \frac{c_L^p\,\,|u_{n'}(\mathbf{x}) - u_{n'}(\mathbf{y})|^p + c_L^p\,\,|u(\mathbf{x}) - u(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{(d-1)+s\,p}} = g_{n'}(\mathbf{x},\mathbf{y})
\end{aligned}$$

for all $n' \in \mathbb{N}$. Since $u_{n'} \to u$ in $W_p^s(\Gamma)$, we obtain

$$\int_\Gamma \int_\Gamma g_{n'}(\mathbf{x},\mathbf{y})\,\mathrm{d}s_{\mathbf{y}}\,\mathrm{d}s_{\mathbf{x}} \to \int_\Gamma \int_\Gamma g(\mathbf{x},\mathbf{y})\,\mathrm{d}s_{\mathbf{y}}\,\mathrm{d}s_{\mathbf{x}}$$

with $g(\mathbf{x},\mathbf{y}) := (2\,c_L)^p\,[u]_\gamma\,(\mathbf{x},\mathbf{y})$. We have shown, that $f_{n'} \to 0$ almost everywhere in Σ, $|f_{n'}|^p \le g_{n'}$ almost everywhere in Σ and $g_{n'} \to g$ in $L_1(\Sigma)$. From Theorem 1.25 in [3] we obtain that $f_{n'} \to 0$ in $L_p(\Sigma)$. Thus we have

$$\begin{aligned}
&\left\| [l(u_{n'}) - l(u)]_\gamma \right\|^p_{L_p(\Gamma\times\Gamma)} = \\
&= \int_\Gamma \int_\Gamma \frac{\Big|\Big(l(u_{n'}(\mathbf{x})) - l(u((\mathbf{x}))\Big) - \Big(l(u_{n'}(\mathbf{y})) - l(u(\mathbf{y}))\Big)\Big|^p}{|\mathbf{x}-\mathbf{y}|^{(d-1)+s\,p}}\,\mathrm{d}s_{\mathbf{y}}\,\mathrm{d}s_{\mathbf{x}} \\
&= \int_\Gamma \int_\Gamma |f_{n'}(\mathbf{x},\mathbf{y})|^p\,\,\mathrm{d}s_{\mathbf{y}}\,\mathrm{d}s_{\mathbf{x}} \to 0
\end{aligned}$$

which implies

$$\|l(u_{n'}) - l(u)\|_{W_p^s(\Gamma)} \to 0.$$

To prove the convergence of $\|l(u_n) - l(u)\|_{W_p^s(\Gamma)} \to 0$, we show by contradiction that each subsequence $\{l(u_{\widehat{n}})\} \subset \{l(u_n)\}$ satisfies $l(u_{\widehat{n}}) \to l(u)$ in $W_p^s(\Gamma)$. From Lemma 4.34 in [6] we can then conclude that $\|l(u_n) - l(u)\|_{W_p^s(\Gamma)} \to 0$. Thus, let $\{l(u_{\widehat{n}})\} \subset \{l(u_n)\}$ be a subsequence with $\|l(u_{\widehat{n}}) - l(u)\|_{W_p^s(\Gamma)} > \varepsilon$ for all $\widehat{n} > \widehat{N}$. Since $u_n \to u$ in $W_p^s(\Gamma)$, the subsequence $\{u_{\widehat{n}}\} \subset \{u_n\}$ has a subsequence $\{u_{\widehat{n}'}\} \subset \{u_{\widehat{n}}\}$ which converges pointwise almost everywhere on Γ. In the same manner as before, we can can show that $\|l(u_{\widehat{n}'}) - l(u)\|_{W_p^s(\Gamma)} \to 0$ which contradicts our assumption on $\{l(u_{\widehat{n}})\}$ and finishes the proof. ∎

We have shown, that each uniformly Lipschitz continuous function $l : \mathbb{R} \to \mathbb{R}$ induces a continuous superposition operator acting on $H^1(\Omega)$ and on $H^{1/2}(\Gamma)$ with $\Gamma = \partial\Omega$. The following lemma shows, how these mappings are related to each other.

Lemma 3.51. *Let $l : \mathbb{R} \to \mathbb{R}$ be a uniformly Lipschitz continuous function. Furthermore, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with boundary $\Gamma = \partial\Omega$. Then there holds*

$$\gamma_\Gamma^0\, l(u) = l(\gamma_\Gamma^0\, u)$$

in $H^{1/2}(\Gamma)$. Here, the superposition operator l on the left hand side is a mapping $l : H^1(\Omega) \to H^1(\Omega)$, whereas the superposition operator l on the right hand side is a mapping $l : H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$.

Proof. Choose an arbitrary element $u \in H^1(\Omega)$. Since $\mathcal{C}^\infty(\overline{\Omega})$ is dense in $H^1(\Omega)$, there exists a sequence $u_n \subset \mathcal{C}^\infty(\overline{\Omega})$ such that $u_n \to u$ in $H^1(\Omega)$ as $n \to \infty$. For $u_n \in \mathcal{C}^\infty(\overline{\Omega})$ we have $\gamma_\Gamma^0\, l(u_n) = l(u_n)_{|\Gamma} = l(u_{n|\Gamma}) = l(\gamma_\Gamma^0\, u_n)$ due to the continuity of l and the regularity of u_n. From Theorem 3.49 and from Theorem 3.50 we get the continuity of the superposition operator on $H^1(\Omega)$ as well as on $H^{1/2}(\Gamma)$. Since $\gamma_\Gamma^0 : H^1(\Omega) \to H^{1/2}(\Gamma)$ is continuous, we obtain

$$\left\|\gamma_\Gamma^0\, l(u) - l(\gamma_\Gamma^0\, u)\right\|_{H^{1/2}(\Gamma)} = \lim_{n\to\infty} \left\|\gamma_\Gamma^0\, l(u_n) - l(\gamma_\Gamma^0\, u_n)\right\|_{H^{1/2}(\Gamma)} = 0$$

which proves the desired statement. ∎

We are done with the mathematical preliminaries. We have recalled some of the main tools from the functional analysis in Section 3.1 and we introduced function spaces as well as their properties in Section 3.2. In Section 3.3, we discussed superposition operators acting in Lebesgue spaces and in Sobolev spaces as well. In the next chapter we will derive a variational formulation for the Richards equation and we will discuss solvability and uniqueness of the derived formulation. Therefore, we will need the tools we repeated in this chapter.

4 VARIATIONAL FORMULATION

In this chapter we will derive a variational formulation for the Richards equation (2.3) which was discussed in Chapter 2. In the first section, Section 4.1, we will apply an implicit–explicit time discretization scheme and we obtain a series of stationary variational problems which depend on the previous time step. Next, in Section 4.2, we will discuss the solvability of the derived stationary variational problems as well as the uniqueness of the solution. Furthermore, a regularity result is given. After that, we consider in Section 4.3 the Richards equation in a homogeneous soil. To apply the solvability and uniqueness results from Section 4.2 we have to state certain assumptions on the nonlinear parameter functions θ and k introduced in Chapter 2. After clarifying this question, we apply the so called Kirchhoff transformation and we obtain, in a straight forward way, a partial differential equation which is now linear in its principal part. Finally, in Section 4.4, we apply the Kirchhoff transformation to the Richards equation considered in heterogeneous soil. This can not be done as straight forward as it was done in Section 4.3. We first have to rewrite the variational formulation using the primal hybrid formulation before we can apply local Kirchhoff transformations.

Let us begin with some preliminary assumptions. Assume, we have given a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, with boundary $\Gamma = \partial\Omega$ such that $\Gamma = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ and $|\Gamma_D|_{d-1} > 0$. We denote the outer unit normal on $\partial\Omega$ by $\mathbf{n}$.

The Richards equation (2.3) is a time dependent equation, hence we have to consider a space–time domain in which we want to solve the equation. Define the time interval $I := (0, T)$ for some $T \in \mathbb{R}_+$. By $Q := \Omega \times I$ we denote the corresponding space–time cylinder with surface $\Sigma := \Gamma \times I$ and base area $\Omega_0 := \Omega \times \{0\}$. In the same manner as for $\partial\Omega$ we can decompose the surface Σ into two disjoint parts Σ_D and Σ_N with $\Sigma_D := \Gamma_D \times I$ and $\Sigma_N := \Gamma_N \times I$, see Figure 4.1.

As already mentioned, we are interested in the solution of the initial boundary value problem (4.1) with initial and boundary conditions as discussed in Section 2.3.

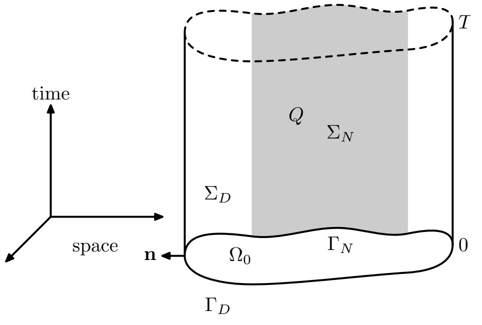


Figure 4.1: Sketch of a space–time cylinder.

Initial boundary value problem

Find $p : Q \to \mathbb{R}$ such that

$$\begin{aligned} n\,\frac{\partial \theta(p)}{\partial t} - \nabla \cdot \left(\frac{K}{\mu}\, k\Big(\theta(p)\Big)\, \nabla\Big(p - d\Big)\right) &= f \quad \text{in } Q,\\ \frac{K}{\mu}\, k\Big(\theta(p)\Big)\, \nabla\Big(p-d\Big)\cdot \mathbf{n} &= g_N \quad \text{on } \Sigma_N,\\ p &= g_D \quad \text{on } \Sigma_D,\\ p &= p_0 \quad \text{in } \Omega_0 \end{aligned} \tag{4.1}$$

holds for given f, g_N, g_D, and given initial datum p_0.

Note, that for the initial boundary value problem (4.1) we have to assume that the nonlinear parameter functions are of certain regularity. For the well posedness we have to assume $\theta \in \mathcal{C}^1(\Omega \times \mathbb{R})$ as well as $k \in \mathcal{C}^1(\Omega \times \mathbb{R})$. In the following section we want to derive a suitable variational formulation of the initial boundary value problem (4.1).

4.1 Time Discrete Variational Formulation

In this section we want to derive a time discrete variational formulation of the initial boundary value problem (4.1). We will apply a special implicit–explicit time discretization scheme, but first we have to discretize the time interval I.

Choose $M \in \mathbb{N}$ and fix corresponding discrete time steps $t_0, t_1, \ldots, t_M \in I$ such that $0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = T$ is a decomposition of the time interval I, that is

$$[0, T] = \bigcup_{m=1}^{M} [t_{m-1}, t_m],$$

see Figure 4.2. We denote the time step size by τ_m which is just defined by $\tau_m := t_m - t_{m-1}$ for each $m = 1, \ldots, M$.

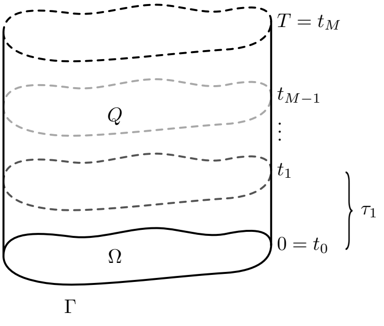


Figure 4.2: Time discretization of space–time cylinder Q.

For sufficient small time step size τ_m we can approximate the time derivative at t_m by the backward Euler method, that is

$$\frac{\partial \theta(\mathbf{x}, p(\mathbf{x}, t)))}{\partial t}\Big|_{t_m} \approx \frac{1}{\tau_m}\Big(\theta(\mathbf{x}, p(\mathbf{x}, t_m)) - \theta(\mathbf{x}, p(\mathbf{x}, t_{m-1}))\Big)$$

for $m = 1, \ldots, M$. Next, we approximate $p(\mathbf{x}, t_m)$ by functions $p_m(\mathbf{x})$ which satisfy the following boundary value problem for $f_m(\mathbf{x}) := f(\mathbf{x}, t_m)$ and boundary conditions $g_{N_m}(\mathbf{x}) := g_N(\mathbf{x}, t_m)$ as well as $g_{D_m}(\mathbf{x}) := g_D(\mathbf{x}, t_m)$.

Implicit time discrete boundary value problem

For each $m = 1, \ldots, M$ find $p_m : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} \frac{n}{\tau_m}\Big(\theta(p_m) - \theta(p_{m-1})\Big) - \nabla \cdot \Big(\frac{K}{\mu}\, k\big(\theta(p_m)\big)\, \nabla\big(p_m - d\big)\Big) &= f_m && \text{in } \Omega, \\ \frac{K}{\mu}\, k\big(\theta(p_m)\big)\, \nabla\big(p_m - d\big) \cdot \mathbf{n} &= g_{N_m} && \text{on } \Gamma_N, \\ p_m &= g_{D_m} && \text{on } \Gamma_D \end{aligned} \tag{4.2}$$

holds for given f_m, g_{N_m}, g_{D_m}, and given initial datum p_0.

Analogously to the initial boundary value problem (4.1), the existence of a classical solution $p_m \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^1(\overline{\Omega})$ is not guaranteed as long as the nonlinearity $k \circ \theta$ is not smooth enough. Hence we have to generalize the concept of solvability. Goal is to derive a weak formulation of the boundary value problem (4.2) where the solution is allowed to have lower regularity.

To derive such a formulation we define the space

$$V := H^1_{0,\Gamma_D}(\Omega)$$

which is a closed subspace of $H^1(\Omega)$ consisting of all functions in $H^1(\Omega)$ vanishing on the Dirichlet boundary Γ_D.

Next, we multiply the first line of the boundary value problem (4.2) with an arbitrary element $v \in V$ and integrate over Ω. Partial integration of the divergence term gives the identity

$$\begin{aligned}\int\limits_{\Omega} \frac{n}{\tau_m} \Big(\theta(p_m) - \theta(p_{m-1})\Big)\, v \,\mathrm{d}\mathbf{x} + \int\limits_{\Omega} \frac{K}{\mu}\, k\Big(\theta(p_m)\Big)\, \nabla\Big(p_m - d\Big) \cdot \nabla v \,\mathrm{d}\mathbf{x} = \\ = \int\limits_{\Omega} f_m\, v \,\mathrm{d}\mathbf{x} + \int\limits_{\partial\Omega} \frac{K}{\mu}\, k\Big(\theta(p_m)\Big)\, \nabla\Big(p_m - d\Big) \cdot \mathbf{n}\, \gamma^0_{\partial\Omega}\, v \,\mathrm{d}s_{\mathbf{x}}\end{aligned}$$

for all time steps $m = 1, \ldots, M$.

If we incorporate the Neumann boundary conditions, line two in problem (4.2), we obtain

$$\int\limits_{\partial\Omega} \frac{K}{\mu}\, k\Big(\theta(p_m)\Big)\, \nabla\Big(p_m - d\Big) \cdot \mathbf{n}\, \gamma^0_{\partial\Omega}\, v \,\mathrm{d}s_{\mathbf{x}} = \int\limits_{\Gamma_N} g_{N\,m}\, \gamma^0_{\Gamma_N}\, v \,\mathrm{d}s_{\mathbf{x}}$$

since $\gamma^0_{\Gamma_D}\, v = 0$ for all $v \in V$.

For the nonlinear diffusion term we apply a simple explicit time discretization of the form

$$k\Big(\theta(p_m)\Big)\, \nabla\Big(p_m - d\Big) \approx k\Big(\theta(p_m)\Big)\, \nabla p_m - k\Big(\theta(p_{m-1})\Big)\, \nabla d$$

where we keep the nonlinearity within the first term. Hence, we obtain the variational formulation to find $p_m \in H^1(\Omega)$, $\gamma^0_{\Gamma_D}\, p_m = g_{D_m}$, such that

$$\begin{aligned}\int\limits_{\Omega} \frac{n}{\tau_m}\, \theta(p_m)\, v \,\mathrm{d}\mathbf{x} + \int\limits_{\Omega} \frac{K}{\mu}\, k\Big(\theta(p_m)\Big)\, \nabla p_m \cdot \nabla v \,\mathrm{d}\mathbf{x} = \\ = \int\limits_{\Omega} \Big(f_m + \frac{n}{\tau_m}\, \theta(p_{m-1})\Big)\, v \,\mathrm{d}\mathbf{x} + \int\limits_{\Omega} \frac{K}{\mu}\, k\Big(\theta(p_{m-1})\Big)\, \nabla d \cdot \nabla v \,\mathrm{d}\mathbf{x} + \\ + \int\limits_{\Gamma_N} g_{N\,m}\, \gamma^0_{\Gamma_N}\, v \,\mathrm{d}s_{\mathbf{x}}\end{aligned} \tag{4.3}$$

for all $v \in V$.

In each time step we have to solve a stationary variational problem with the variational form (4.3) where the right hand side depends on the previous time step. Note, that the unknown $p_m \in H^1(\Omega)$ has to satisfy the Dirichlet boundary condition $\gamma^0_{\Gamma_D} p_m = g_{Dm}$. For this reason we consider an extension $p_{Dm} := \mathcal{E}_\Omega \, g_{Dm} \in H^1(\Omega)$ which satisfies the inhomogeneous Dirichlet boundary condition by construction. Thus we can write the unknown p_m as the sum $p_m = p_{0m} + p_{Dm}$ with $p_{0m} \in V$ and $p_{Dm} \in H^1(\Omega)$ defined as above. Due to this homogenization the new unknown now is $p_{0m} \in V$. For the rest of this thesis we will skip the subindex m which denotes the current time step and we will write p_{m-1} as q. Furthermore, will substitute p_0 by p for a better readability.

Thus, we obtain the following variational problem.

Implicit–explicit time discrete variational formulation

Find $p \in V$ such that

$$\begin{aligned} \int\limits_\Omega \frac{n}{\tau}\,\theta(p+p_D)\,v\,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\,k\Big(\theta(p+p_D)\Big)\,\nabla(p+p_D)\cdot\nabla v\,\mathrm{d}\mathbf{x} = \\ = \int\limits_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\,v\,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\,k\Big(\theta(q)\Big)\,\nabla d\cdot\nabla v\,\mathrm{d}\mathbf{x} + \\ + \int\limits_{\Gamma_N} g_N\,\gamma^0_{\Gamma_N}\,v\,\mathrm{d}s_{\mathbf{x}} \end{aligned} \tag{4.4}$$

for all $v \in V$.

Remark 4.1. *Due to the nonlinearity in the time derivative and the implicit–explicit time discretization scheme, it is hard to state convergence results as τ tends to zero. However, a good convergence behavior is observed in our numerical examples, see Chapter 6.*

We finally derived a variational problem which corresponds to the Richards equation (2.3) after applying a simple implicit–explicit time discretization scheme. The next step is to investigate solvability and uniqueness of solutions to the variational problem (4.4) which is done in the next section.

4.2 Solvability and Uniqueness

In this section we will discuss solvability of the variational problem (4.4) and uniqueness of the solution. To state results in a proper way, we assume that the coefficient functions

n and K of the variational problem (4.4) are functions in the space $L^+_\infty(\Omega)$ which is defined by

$$L^+_\infty(\Omega) := \left\{ u \in L_\infty(\Omega) \mid \operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} u(\mathbf{x}) > 0 \right\}.$$

This is not a restriction, since the porosity n and the permeability K are positive quantities, see Chapter 2.

For n, K in $L^+_\infty(\Omega)$ we define the constants

$$c_m := \operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} n(\mathbf{x}) \quad \text{and} \quad c_s := \frac{\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} K(\mathbf{x})}{\mu}$$

as well as

$$c_M := \|n\|_{L_\infty(\Omega)} \quad \text{and} \quad c_S := \frac{\|K\|_{L_\infty(\Omega)}}{\mu}$$

which are positive and bounded for a fixed $\mu \in \mathbb{R}_+$.

To obtain results for a more general class of parameter functions, we consider arbitrary nonlinearities θ and k satisfying the following assumption.

Assumption 4.2. *Let $\theta : \Omega \times \mathbb{R} \to \mathbb{R}$ and $k : \Omega \times \mathbb{R} \to \mathbb{R}$ be Carathéodory functions as in Definition 3.45 satisfying*

$$|\theta(\mathbf{x}, s)| \le a_\theta(\mathbf{x}) + b_\theta\,|s| \quad \textit{and} \quad |k(\mathbf{x}, s)| \le b_k$$

for an element $a_\theta \in L_2(\Omega)$ and non negative b_θ and b_k. Furthermore, assume that the following conditions hold for almost all $\mathbf{x} \in \Omega$.

(1) $\theta(\mathbf{x}, \cdot) \in \mathcal{C}^{0,1}(\mathbb{R})$ with Lipschitz constant $c_{L,\theta}$ and monotonically increasing.
(2) $k(\mathbf{x}, \cdot) \in \mathcal{C}(\mathbb{R}) \cap L_\infty(\mathbb{R})$.
(3) $k(\mathbf{x}, s) \ge c_{\alpha,k} > 0$ for all $s \in \mathbb{R}$.

We are now in the position to formulate the following theorem on the solvability of the variational problem (4.4).

Theorem 4.3. *Let $n, K \in L^+_\infty(\Omega)$, $\tau, \mu \in \mathbb{R}_+$, $f, q \in L_2(\Omega)$, $g_N \in L_2(\Gamma_N)$, $\nabla d \in \mathbf{L}_2(\Omega)$, $g_D \in H^{1/2}(\Gamma_D)$ and let Assumption 4.2 hold. Then the variational problem (4.4) has a solution $p \in V$. Furthermore, there holds*

$$\|p\|_{H^1(\Omega)} \le c \left(\|f\|_{L_2(\Omega)} + \frac{1}{\tau}\|q - p_D\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|g_D\|_{H^{1/2}(\Gamma_D)} + \|g_N\|_{L_2(\Gamma_N)} \right)$$

with $p_D := \mathcal{E}_\Omega\, g_D \in H^1(\Omega)$ and some positive constant $c(\Omega, \theta, k, K, n, \mu)$.

Proof. We split this proof into two parts. In the first part we show the solvability of the variational problem (4.4) using the theory of monotone operators we discussed in Section 3.1. In the second part we prove the boundedness of the solution.

Part A) Solvability: First, we define the operator $M : V \to V'$ by

$$\langle M(p), v \rangle_{V' \times V} := \int_\Omega \frac{n}{\tau}\, \theta(p + p_D)\, v \,\mathrm{d}\mathbf{x}$$

and the operator $\widetilde{S} : V \times V \to V'$ by

$$\langle \widetilde{S}(p,q), v \rangle_{V' \times V} := \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla(q + p_D) \cdot \nabla v \,\mathrm{d}\mathbf{x}$$

for $p, q, v \in V$. The operator $\widetilde{S}(p,p)$ will be denoted by $S(p)$, that is $S : V \to V'$ with

$$\langle S(p), v \rangle_{V' \times V} := \langle \widetilde{S}(p,p), v \rangle_{V' \times V}$$

for $p, v \in V$.

Goal is to prove that the operator $A := M + S$ is pseudomonotone, coercive and bounded. If we succeed, we can apply the main theorem on pseudomonotone operators, Theorem 3.21. We will start to show the pseudomonotonicity of A.

Pseudomonotonicity: By Proposition 27.7 in [75] we know that A is pseudomonotone if M is monotone and hemicontinuous and S is pseudomonotone. We will first show the properties of the operator M.

Consider the nonlinear parameter function θ. Due to the conditions in Assumption 4.2 and from Theorem 3.46, we conclude that θ induces a continuous and bounded superposition operator mapping $L_2(\Omega)$ to $L_2(\Omega)$.

Since $\theta(\mathbf{x}, s)$ is monotonically increasing in s, we obtain from Lemma 3.47 the monotonicity

$$\int_\Omega \Big(\theta(p) - \theta(q)\Big)\Big(p - q\Big) \,\mathrm{d}\mathbf{x} \geq 0$$

for arbitrary elements $p, q \in L_2(\Omega)$. Since the coefficient function $n \in L_\infty^+(\Omega)$ is bounded from below by $0 < c_m \leq n(\mathbf{x})$ for almost all $\mathbf{x} \in \Omega$, we have

$$\int_\Omega \frac{n}{\tau}\Big(\theta(p) - \theta(q)\Big)\Big(p - q\Big) \,\mathrm{d}\mathbf{x} \geq 0$$

for all $p, q \in L_2(\Omega)$. Since $V \subset H^1(\Omega) \subset L_2(\Omega)$ we have

$$\begin{aligned} \langle M(p) - M(q), p - q \rangle_{V' \times V} &= \int_\Omega \Big(\frac{n}{\tau}\, \theta(p + p_D) - \frac{n}{\tau}\, \theta(q + p_D)\Big)\Big(p - q\Big) \,\mathrm{d}\mathbf{x} = \\ &= \int_\Omega \frac{n}{\tau}\, \big(\theta(p + p_D) - \theta(q + p_D)\big)\Big((p + p_D) - (q + p_D)\Big) \,\mathrm{d}\mathbf{x} \geq 0 \end{aligned} \tag{4.5}$$

for all $p, q \in V$ and $p_D = \mathcal{E}_\Omega\, g_D \in H^1(\Omega)$ as in Theorem 4.3. This proves the desired monotonicity result of the operator $M : V \to V'$.

To show the hemicontinuity of M we prove that $M : V \to V'$ is continuous. Let $\{p_n\} \subset V$ be a sequence which converges to $p \in V$. Then there holds the estimate

$$\begin{aligned}\|M(p) - M(p_n)\|_{V'} &= \sup_{0 \neq v \in V} \frac{\langle M(p) - M(p_n), v\rangle_{V' \times V}}{\|v\|_{H^1(\Omega)}} \\ &= \sup_{0 \neq v \in V} \frac{1}{\|v\|_{H^1(\Omega)}} \int_\Omega \frac{n}{\tau} \Big(\theta(p + p_D) - \theta(p_n + p_D)\Big)\, v \,\mathrm{d}\mathbf{x} \\ &\leq \frac{c_M}{\tau} \|\theta(p + p_D) - \theta(p_n + p_D)\|_{L_2(\Omega)}\end{aligned}$$

with an upper bound which tends to zero since $\theta : L_2(\Omega) \to L_2(\Omega)$ is continuous. This proves the continuity of $M : V \to V'$ and due to the considerations made in Section 3.1, we conclude that $M : V \to V'$ is hemicontinuous.

So far, we have shown that the operator $M : V \to V'$ is monotone and hemicontinuous. The next step is to prove that the operator $S : V \to V'$ is pseudomonotone. We want to use Lemma 3.15 to show the desired property.

As done for the operator M, we want to define a suitable superposition operator and exploit its properties to show the desired pseudomonotonicity of S. Let $\mathbf{s} \in \mathbb{R}^2$ with $\mathbf{s} = (s_1, s_2)^\top$ and define $l(\mathbf{x}, \mathbf{s}) := k(\mathbf{x}, s_1)\, s_2$. Due to the assumptions on k, it is easy to verify that $l : \Omega \times \mathbb{R}^2 \to \mathbb{R}$ is again a Carathéodory function and there holds

$$|k(\mathbf{x}, s_1)\, s_2| \leq b_k\, |s_2| \leq b_k \Big(\, |s_1| + |s_2|\, \Big)$$

for arbitrary $s_1, s_2 \in \mathbb{R}$.

Theorem 3.46 implies that l induces a continuous and bounded superposition operator mapping the product space $L_2(\Omega) \times L_2(\Omega)$ to $L_2(\Omega)$.

To prove the desired hemicontinuity of $\widetilde{S}$ in each argument, we show that $\widetilde{S}$ is continuous in each argument. Therefore, let $\{p_n\} \subset L_2(\Omega)$ be a sequence which converges to $p \in L_2(\Omega)$ and let $\{q_n\} \subset V$ be a sequence which converges to $q \in V$. The convergence of $\{q_n\}$ in V implies $\partial_{x_i} q_n \to \partial_{x_i} q$ in $L_2(\Omega)$ for each $i = 1, \dots, d$. Consider now the norm of the difference of the images, that is

$$\Big\|\widetilde{S}(p, q) - \widetilde{S}(p_n, q_n)\Big\|_{V'} = \sup_{0 \neq v \in V} \frac{\langle \widetilde{S}(p, q) - \widetilde{S}(p_n, q_n), v\rangle_{V' \times V}}{\|v\|_{H^1(\Omega)}}.$$

The duality pairing can be written in terms of the superposition operator l as

$$\begin{aligned}
&\langle \widetilde{S}(p,q) - \widetilde{S}(p_n,q_n), v\rangle_{V'\times V} = \\
&= \sum_{i=1}^{d} \int_{\Omega} \frac{K}{\mu} \Big(k\big(\theta(p+p_D)\big)\, \partial_{x_i}(q+p_D) - k\big(\theta(p_n+p_D)\big)\, \partial_{x_i}(q_n+p_D)\Big) \partial_{x_i} v \,\mathrm{d}\mathbf{x} = \\
&\quad = \sum_{i=1}^{d} \int_{\Omega} \frac{K}{\mu} \Big(l\big(\theta(p+p_D), \partial_{x_i}(q+p_D)\big) - l\big(\theta(p_n+p_D), \partial_{x_i}(q_n+p_D)\big)\Big) \partial_{x_i} v \,\mathrm{d}\mathbf{x}.
\end{aligned}$$

The representation above can be estimated from above by

$$\begin{aligned}
\left\| \widetilde{S}(p,q) - \widetilde{S}(p_n,q_n) \right\|_{V'} &= \sup_{0\neq v\in V} \frac{\langle \widetilde{S}(p,q) - \widetilde{S}(p_n,q_n), v\rangle_{V'\times V}}{\|v\|_{H^1(\Omega)}} \leq \\
&\leq c_S \left(\sum_{i=1}^{d} \left\| l\big(\theta(p+p_D), \partial_{x_i}(q+p_D)\big) - l\big(\theta(p_n+p_D), \partial_{x_i}(q_n+p_D)\big) \right\|^2_{L_2(\Omega)} \right)^{1/2}
\end{aligned}$$

with a right hand side which converges to zero since $\theta : L_2(\Omega) \to L_2(\Omega)$ is continuous and the mapping $l : L_2(\Omega) \times L_2(\Omega) \to L_2(\Omega)$ is continuous as well. The continuity of $\widetilde{S} : L_2(\Omega) \times V \to L_2(\Omega)$ implies the hemicontinuity of $\widetilde{S}$ in each argument.

The next step is to show the boundedness of $\widetilde{S}(\cdot, q)$ for an arbitrary but fixed $q \in V$. From Theorem 3.46 and from the assumptions on $k : \Omega \times \mathbb{R} \to \mathbb{R}$, we conclude that $k(p + p_D) \in L_\infty(\Omega)$ for all $p \in V \subset L_2(\Omega)$. The estimate

$$\begin{aligned}
\left\| \widetilde{S}(p,q) \right\|_{V'} &= \sup_{0\neq v\in V} \frac{\langle \widetilde{S}(p,q), v\rangle_{V'\times V}}{\|v\|_{H^1(\Omega)}} \\
&= \sup_{0\neq v\in V} \frac{1}{\|v\|_{H^1(\Omega)}} \int_{\Omega} \frac{K}{\mu}\, k\big(\theta(p+p_D)\big)\, \nabla(q+p_D)\cdot \nabla v \,\mathrm{d}\mathbf{x} \\
&\leq c_S\, b_k\, \|q+p_D\|_{H^1(\Omega)}
\end{aligned}$$

implies the boundedness of $\widetilde{S}(\cdot, q) : V \to V'$ for all fixed $q \in V$. Hence, the conditions (1) and (2) in Lemma 3.15 are satisfied.

After that, we want to show the monotonicity condition (3) in Lemma 3.15. Due to (3) in Assumption 4.2, there holds for arbitrary elements $o, p, q \in V$ the estimate

$$\begin{aligned}
&\langle \widetilde{S}(o,p) - \widetilde{S}(p,q), p-q\rangle_{V'\times V} = \\
&= \int_{\Omega} \frac{K}{\mu} \Big(k\big(\theta(o+p_D)\big)\, \nabla(p+p_D) - k\big(\theta(o+p_D)\big)\, \nabla(q+p_D)\Big) \cdot \nabla(p-q) \,\mathrm{d}\mathbf{x} = \\
&\qquad = \int_{\Omega} \frac{K}{\mu}\, k\big(\theta(o+p_D)\big)\, |\nabla(p-q)|^2 \,\mathrm{d}\mathbf{x} \geq c_s\, c_{\alpha,k} |p-q|^2_{H^1(\Omega)} \geq 0 \quad (4.6)
\end{aligned}$$

which shows the desired result.

To show condition (4) in Lemma 3.15, assume we have a sequence $\{p_n\} \subset V$ which converges weakly to $p \in V$, that is $p_n \rightharpoonup p$ in V. Additionally assume that $\langle \widetilde{S}(p_n, p_n) - \widetilde{S}(p_n, p), p_n - p\rangle_{V' \times V} \to 0$ as n tends to infinity. The estimate (4.6) then implies

$$\langle \widetilde{S}(p_n, p_n) - \widetilde{S}(p_n, p), p_n - p\rangle_{V' \times V} \geq c_N \, c_S \, c_{\alpha,k} \, \|p_n - p\|^2_{H^1(\Omega)} \geq 0$$

which shows $p_n \to p$ in V. The continuity of $\widetilde{S} : L_2(\Omega) \times V \to V'$ gives $\widetilde{S}(p_n, q) \to \widetilde{S}(p, q)$ in V' which implies the weak convergence result.

For the last point, condition (5) in Lemma 3.15, we again assume that $\{p_n\} \subset V$ is a sequence which converges weakly to $p \in V$, that is $p_n \rightharpoonup p$ in V. Furthermore, we assume, that $\widetilde{S}(p_n, q) \rightharpoonup b$ in V' for some $q \in V$. Since $V \hookrightarrow_c L_2(\Omega)$ we have $p_n \to p$ in $L_2(\Omega)$ and due to the continuity of $\widetilde{S} : L_2(\Omega) \times V \to V'$ we see that $\langle \widetilde{S}(p_n, q), v\rangle_{V' \times V} \to \langle \widetilde{S}(p, q), v\rangle_{V' \times V}$. Since weak limits are unique, we have $b = \widetilde{S}(p, q)$ in V'. Consider now the estimate

$$\begin{aligned} \left| \langle \widetilde{S}(p_n, q), p_n\rangle_{V' \times V} - \langle b, p\rangle_{V' \times V} \right| &= \left| \langle \widetilde{S}(p_n, q), p_n\rangle_{V' \times V} - \langle \widetilde{S}(p, q), p\rangle_{V' \times V} \right| \leq \\ &\leq \left| \langle \widetilde{S}(p_n, q) - \widetilde{S}(p, q), p_n\rangle_{V' \times V} \right| + \left| \langle \widetilde{S}(p, q), p_n - p\rangle_{V' \times V} \right| \leq \\ &\leq \left\| \widetilde{S}(p_n, q) - \widetilde{S}(p, q) \right\|_{V'} \|p_n\|_{H^1(\Omega)} + \langle \widetilde{S}(p, q), p_n - p\rangle_{V' \times V}. \end{aligned}$$

Since weakly convergent sequences are bounded and $\widetilde{S} : L_2(\Omega) \times V \to V'$ is continuous, we see that the right hand side of the above estimate tends to zero as n tends to infinity.

The operator $\widetilde{S} : V \times V \to V'$ satisfies all assumptions of Lemma 3.15 which implies the pseudomonotonicity of the operator $S : V \to V'$ and further the pseudomonotonicity of $A = M + S : V \to V'$. The next step is to prove that the operator $A : V \to V'$ is coercive.

Coercivity: To show that the operator $A = M + S$ is coercive, we have to prove that

$$\frac{\langle A(p), p\rangle_{V' \times V}}{\|p\|_{H^1(\Omega)}} \to \infty$$

as $\|p\|_{H^1(\Omega)} \to \infty$. To show the desired result we consider each term separately.

We begin with the operator $M : V \to V'$. Let $p \in V$ be arbitrary but fixed, we have

$$\langle M(p), p\rangle_{V' \times V} = \langle M(p) - M(0), p\rangle_{V' \times V} + \langle M(0), p\rangle_{V' \times V}$$

and inequality (4.5) implies the estimate

$$\langle M(p) - M(0), p \rangle_{V' \times V} = \langle M(p) - M(0), p - 0 \rangle_{V' \times V} \geq 0. \tag{4.7}$$

The remaining duality pairing $\langle M(0), p \rangle_{V' \times V}$ can be estimated by using the Hölder inequality, see Theorem 3.28. Thus, we obtain

$$\langle M(0), p \rangle_{V' \times V} = \int_\Omega \frac{n}{\tau}\, \theta(p_D)\, p \,\mathrm{d}\mathbf{x} \leq \frac{c_M}{\tau} \left\| \theta(p_D) \right\|_{L_2(\Omega)} \left\| p \right\|_{H^1(\Omega)}$$

for all $p \in V$. Finally we can bound the duality pairing $\langle M(p), p \rangle_{V' \times V}$ from below by

$$\langle M(p), p \rangle_{V' \times V} \geq -\frac{c_M}{\tau} \left\| \theta(p_D) \right\|_{L_2(\Omega)} \left\| p \right\|_{H^1(\Omega)}$$

for all $p \in V$.

Next, we consider the operator S. We can write S as

$$\begin{aligned}
\langle S(p), p \rangle_{V' \times V} = \langle \widetilde{S}(p,p), p \rangle_{V' \times V} &= \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla (p + p_D) \cdot \nabla p \,\mathrm{d}\mathbf{x} = \\
= \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla p \cdot \nabla p \,\mathrm{d}\mathbf{x} &+ \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla p_D \cdot \nabla p \,\mathrm{d}\mathbf{x} = \\
&= \langle \widetilde{S}(p, p - p_D), p \rangle_{V' \times V} + \langle \widetilde{S}(p, 0), p \rangle_{V' \times V}
\end{aligned}$$

for all $p \in V$. The duality pairing $\langle \widetilde{S}(p, p - p_D), p \rangle_{V' \times V}$ can be bounded from below by

$$\langle \widetilde{S}(p, p - p_D), p \rangle_{V' \times V} = \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla p \cdot \nabla p \,\mathrm{d}\mathbf{x} \geq c_s\, c_{\alpha,k} \left| p \right|^2_{H^1(\Omega)} \tag{4.8}$$

for each $p \in V$. The second expression $\langle \widetilde{S}(p, 0), p \rangle_{V' \times V}$ can be bounded from above using Theorem 3.28 and we obtain

$$\langle \widetilde{S}(p, 0), p \rangle_{V' \times V} = \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla p_D \cdot \nabla p \,\mathrm{d}\mathbf{x} \leq c_S\, b_k \left\| p_D \right\|_{H^1(\Omega)} \left\| p \right\|_{H^1(\Omega)}.$$

If we use the norm equivalence in V, we have the following lower bound

$$\langle S(p), p \rangle_{V' \times V} \geq c_N^2\, c_s\, c_{\alpha,k} \left\| p \right\|^2_{H^1(\Omega)} - c_S\, b_k \left\| p_D \right\|_{H^1(\Omega)} \left\| p \right\|_{H^1(\Omega)}$$

for all $p \in V$.

Finally, the operator $A = M + S$ can be bounded from below by

$$\frac{\langle A(p), p\rangle_{V'\times V}}{\|p\|_{H^1(\Omega)}} \geq c_N^2\, c_s\, c_{\alpha,k}\, \|p\|_{H^1(\Omega)} - c_S\, b_k\, \|p_D\|_{H^1(\Omega)} - \frac{c_M}{\tau}\, \|\theta(p_D)\|_{L_2(\Omega)}$$

and the bound tends to infinity as $\|p\|_{H^1(\Omega)} \to \infty$. This proves the coercivity of the operator $A : V \to V'$.

So far we have shown that $A : V \to V'$ is a pseudomonotone and coercive operator. To be able to apply Theorem 3.21 we have to prove that A is a bounded operator. This is done in the next step.

Boundedness: To prove boundedness, we fix an arbitrary $p \in V$. We know, that $k(p + p_D) \in L_\infty(\Omega)$ for $p \in V \subset L_2(\Omega)$ with $\|k(p + p_D)\|_{L_\infty(\Omega)} \leq b_k$. Since θ maps $L_2(\Omega)$ to $L_2(\Omega)$, we obtain

$$\begin{aligned}\langle A(p), v\rangle_{V'\times V} &= \int\limits_\Omega \frac{n}{\tau}\, \theta(p + p_D)\, v\, \mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla(p + p_D) \cdot \nabla v\, \mathrm{d}\mathbf{x} \\ &\leq \left(\frac{c_M}{\tau}\, \|\theta(p + p_D)\|_{L_2(\Omega)} + c_S\, b_k\, \|p + p_D\|_{H^1(\Omega)}\right)\, \|v\|_{H^1(\Omega)}\end{aligned}$$

for all $v \in V$. The triangle inequality and Lemma 3.48 imply the estimate

$$\begin{aligned}\|\theta(p + p_D)\|_{L_2(\Omega)} &\leq \|\theta(p + p_D) - \theta(p_D)\|_{L_2(\Omega)} + \|\theta(p_D)\|_{L_2(\Omega)} \\ &\leq c_{L,\theta}\, \|p\|_{L_2(\Omega)} + \|\theta(p_D)\|_{L_2(\Omega)}.\end{aligned}$$

For the operator $A : V \to V'$ we obtain the upper bound

$$\begin{aligned}\|A(p)\|_{V'} &= \sup_{v\in V} \frac{\langle A(p), v\rangle_{V'\times V}}{\|v\|_{H^1(\Omega)}} \leq \\ &\leq \Big(\frac{c_M}{\tau}\, c_{L,\theta} + c_S\, b_k\Big)\, \|p\|_{H^1(\Omega)} + \frac{c_M}{\tau}\, \|\theta(p_D)\|_{L_2(\Omega)} + c_S\, b_k\, \|p_D\|_{H^1(\Omega)}\end{aligned}$$

which proves the boundedness.

We have proven that $A : V \to V'$ is a bounded, coercive and pseudomonotone operator. In the last step, we have to prove that F defined by

$$\langle F, v\rangle_{V'\times V} := \int\limits_\Omega \Big(f + \frac{n}{\tau}\, \theta(q)\Big)\, v\, \mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla v\, \mathrm{d}\mathbf{x} + \int\limits_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, v\, \mathrm{d}s_{\mathbf{x}}$$

is actually an element in V' to obtain a solution to the variational problem (4.4). It is easy to verify that F is a linear functional. Due to the assumption on the data and the considerations made so far, we obtain the estimate

$$\int_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, v \,\mathrm{d}\mathbf{x} \leq \Big(\|f\|_{L_2(\Omega)} + \frac{c_M}{\tau}\,\|\theta(q)\|_{L_2(\Omega)}\Big)\, \|v\|_{H^1(\Omega)}$$

as well as

$$\int_\Omega \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla v \,\mathrm{d}\mathbf{x} \leq c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_2(\Omega)}\, \|v\|_{H^1(\Omega)}.$$

Using the Trace Theorem 3.35 we have for the remaining surface term the upper bound

$$\int_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, v \,\mathrm{d}s_{\mathbf{x}} \leq c_T\, \|g_N\|_{L_2(\Gamma_N)}\, \|v\|_{H^1(\Omega)}$$

which implies

$$\|F\|_{V'} \leq \|f\|_{L_2(\Omega)} + \frac{c_M}{\tau}\,\|\theta(q)\|_{L_2(\Omega)} + c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_2(\Omega)} + c_T\, \|g_N\|_{L_2(\Gamma_N)}$$

and so $F \in V'$.

Since V is a real, reflexive and separable Banach space we can apply Theorem 3.21 and can conclude the existence of a $p \in V$ such that $A(p) = F$ in V' for $F \in V'$. Finally we obtain the existence of a solution to the variational problem (4.4). In the last part we want to prove the boundedness of the solution as stated in Theorem 4.3.

Part B) Boundedness of the solution: Let $p \in V$ be a solution to the variational problem (4.4). Since $p \in V$ is a valid test function we obtain the identity

$$\int_\Omega \frac{n}{\tau}\,\theta(p + p_D)\, p \,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla(p + p_D) \cdot \nabla p \,\mathrm{d}\mathbf{x} = \langle F, p\rangle_{V' \times V}$$

or in a more abstract way

$$\langle M(p) + \widetilde{S}(p,p), p\rangle_{V' \times V} = \langle F, p\rangle_{V' \times V}.$$

This can be written as

$$\langle M(p) - M(0) + \widetilde{S}(p, p - p_D), p\rangle_{V' \times V} = \langle F - M(0) - \widetilde{S}(p, 0), p\rangle_{V' \times V} \tag{4.9}$$

with the same notation as in the coercivity part.

Using the estimates (4.7) and (4.8) with equation (4.9), we obtain

$$\begin{aligned} c_N^2\, c_s\, c_{\alpha,k} \|p\|^2_{H^1(\Omega)} &\le \left\langle M(p) - M(0) + \widetilde{S}(p, p - p_D), p \right\rangle_{V' \times V} \\ &= \left\langle F - M(0) + \widetilde{S}(p, 0), p \right\rangle_{V' \times V} \end{aligned} \tag{4.10}$$

for a solution $p \in V$.

The right hand side in inequality (4.10) is given by the expression

$$\int_\Omega \left(f + \frac{n}{\tau} \Big(\theta(q) - \theta(p_D)\Big)\right) p \, \mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla p \, \mathrm{d}\mathbf{x} - \\ - \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p + p_D)\Big)\, \nabla p_D \cdot \nabla p \, \mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, p \, \mathrm{d}s_{\mathbf{x}}$$

which can be bounded from above by

$$\|f\|_{L_2(\Omega)} \|p\|_{H^1(\Omega)} + \frac{c_M}{\tau}\, c_{L,\theta}\, \|q - p_D\|_{L_2(\Omega)} \|p\|_{H^1(\Omega)} + c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_2(\Omega)} \|p\|_{H^1(\Omega)} + \\ c_S\, b_k\, c_E\, \|g_D\|_{H^{1/2}(\Gamma_D)} \|p\|_{H^1(\Omega)} + c_T\, \|g_N\|_{L_2(\Gamma_N)} \|p\|_{H^1(\Omega)} \tag{4.11}$$

using Lemma 3.48, the Trace Theorem 3.35 and the Inverse Trace Theorem 3.36.

Combining the estimate (4.10) and (4.11), we obtain

$$\begin{aligned} c_N^2\, c_s\, c_{\alpha,k}\, \|p\|_{H^1(\Omega)} \le{}& \|f\|_{L_2(\Omega)} + \frac{c_M}{\tau}\, c_{L,\theta}\, \|q - p_D\|_{L_2(\Omega)} \\ &+ c_S\, b_K\, \|\nabla d\|_{\mathbf{L}_2(\Omega)} + c_S\, b_K\, c_E\, \|g_D\|_{H^{1/2}(\Gamma_D)} \\ &+ c_T\, \|g_N\|_{L_2(\Gamma_N)}. \end{aligned}$$

If we set

$$c := \frac{\max\{1, c_S\, c_M\, c_{L,\theta}, b_K, c_S\, b_K\, c_E, c_T\}}{c_N^2\, c_s\, c_{\alpha,k}}$$

we get the desired bound for $p \in V$ solution to the variational problem (4.4). ∎

Theorem 4.3 states the existence of a solution to the variational problem (4.4) under certain assumptions on the nonlinear functions θ and k. The next theorem will provide a uniqueness result assuming a slightly stronger condition on the nonlinear function k.

Theorem 4.4. *Let $n, K \in L^+_\infty(\Omega)$, $\tau, \mu \in \mathbb{R}_+$, $f, q \in L_2(\Omega)$, $g_N \in L_2(\Gamma_N)$, $\nabla d \in \mathbf{L}_2(\Omega)$, $g_D \in H^{1/2}(\Gamma_D)$ and let Assumption 4.2 hold. In addition, assume that $k(\mathbf{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $c_{L,k}$ for almost all $\mathbf{x} \in \Omega$. Then, the solution $p \in V$ to the variational problem (4.4) is unique.*

Proof. To show uniqueness of the solution we follow the lines of the proof of Theorem 3.2 in [39]. In this work, the uniqueness result is proven for a right hand side in $L_2(\Omega)$. Since the right hand side of the variational problem (4.4) is in general not an element in $L_2(\Omega)$, we have to modify the proof.

First, we make the simple but essential observation, that the composite function $k(\mathbf{x}, \theta(\mathbf{x}, s))$ satisfies

$$\left|k\Big(\mathbf{x}, \theta(\mathbf{x}, s)\Big) - k\Big(\mathbf{x}, \theta(\mathbf{x}, r)\Big)\right| \leq c_{L,k}\ |\theta(\mathbf{x}, s) - \theta(\mathbf{x}, r)| \leq c_{L,k}\, c_{L,\theta}\ |s - r|$$

for all $s, r \in \mathbb{R}$ and almost all $\mathbf{x} \in \Omega$. We denote the Lipschitz constant of the composite function by c_L, that is $c_L := c_{L,k}\, c_{L,\theta}$.

Next, assume there are two solutions $p_1, p_2 \in V$ to the variational problem (4.4), that is

$$\langle M(p_i) + S(p_i), v\rangle_{V'\times V} = \langle M(p_i) + \widetilde{S}(p_i, p_i), v\rangle_{V'\times V} = \langle F, v\rangle_{V'\times V} \tag{4.12}$$

is satisfied for all $v \in V$ and for $i = 1, 2$. The operators $M, S, \widetilde{S}$ and the right hand side $F \in V'$ are defined as in the proof of Theorem 4.3.

For the two solutions we can define the domain

$$\Omega_1 := \{\mathbf{x} \in \Omega \mid p_2(\mathbf{x}) > p_1(\mathbf{x})\}$$

and we assume $|\Omega_1|_d > 0$. For an arbitrary $\varepsilon > 0$ define the subset

$$\Omega_\varepsilon := \{\mathbf{x} \in \Omega_1 \mid p_2(\mathbf{x}) - p_1(\mathbf{x}) > \varepsilon\}$$

and the function $v_\varepsilon := \min\{\varepsilon, (p_2 - p_1)^+\}$.

We know, that $|p|, p^+, p^- \in H^1(\Omega)$ for all $p \in H^1(\Omega)$, see for example [31, Theorem 4 in Section 4.2]. Since we can write the minimum as $\min\{p, q\} = \frac{1}{2}\Big((p+q) - |p - q|\Big) \in H^1(\Omega)$ we conclude that $v_\varepsilon \in H^1(\Omega)$.

For v_ε and ∇v_ε there holds the representation

$$v_\varepsilon = \begin{cases} \varepsilon & \text{in } \Omega_\varepsilon, \\ p_2 - p_1 & \text{in } \Omega_1 \setminus \Omega_\varepsilon, \\ 0 & \text{else} \end{cases} \qquad \text{and} \qquad \nabla v_\varepsilon = \begin{cases} 0 & \text{in } \Omega_\varepsilon, \\ \nabla(p_2 - p_1) & \text{in } \Omega_1 \setminus \Omega_\varepsilon, \\ 0 & \text{else} \end{cases}$$

and in addition $v_\varepsilon \geq 0$ almost everywhere in Ω. Furthermore, there holds $\gamma^0_{\Gamma_D}\, v_\varepsilon = 0$ and so $v_\varepsilon \in V$.

Due to the representation of v_ε we have

$$\begin{aligned}
c_s\, c_{\alpha,k}\, |v_\varepsilon|^2_{H^1(\Omega)} &\le \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p_1+p_D)\Big)\, \nabla v_\varepsilon \cdot \nabla v_\varepsilon \,\mathrm{d}\mathbf{x} \\
&= \int_{\Omega_1\setminus\Omega_\varepsilon} \frac{K}{\mu}\, k\Big(\theta(p_1+p_D)\Big)\, \nabla(p_2-p_1) \cdot \nabla v_\varepsilon \,\mathrm{d}\mathbf{x} \\
&= \int_\Omega \frac{K}{\mu}\, k\Big(\theta(p_1+p_D)\Big)\, \nabla\Big((p_2+p_D)-(p_1+p_D)\Big) \cdot \nabla v_\varepsilon \,\mathrm{d}\mathbf{x} \\
&= \langle \widetilde{S}(p_1,p_2) - S(p_1), v_\varepsilon\rangle_{V'\times V}.
\end{aligned}$$

Since p_1 and p_2 satisfy (4.12) for all $v \in V$, we obtain

$$c_s\, c_{\alpha,k}\, |v_\varepsilon|^2_{H^1(\Omega)} \le \langle M(p_1) - M(p_2) + \widetilde{S}(p_1,p_2) - \widetilde{S}(p_2,p_2), v_\varepsilon\rangle_{V'\times V}. \tag{4.13}$$

In a next step, we want to estimate the right hand side of (4.13). We consider first the duality pairing $\langle M(p_1) - M(p_2), v_\varepsilon\rangle_{V'\times V}$, which can be written as

$$\begin{aligned}
&\langle M(p_1) - M(p_2), v_\varepsilon\rangle_{V'\times V} = \int_\Omega \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, v_\varepsilon \,\mathrm{d}\mathbf{x} = \\
&= \int_{\Omega_1\setminus\Omega_\varepsilon} \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, v_\varepsilon \,\mathrm{d}\mathbf{x} + \int_{\Omega_\varepsilon} \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, v_\varepsilon \,\mathrm{d}\mathbf{x}
\end{aligned}$$

due to the representation of v_ε.

To estimate the two expressions in the right hand side we make the following observations. For the first expression we have $v_\varepsilon = (p_2 - p_1)$ in $\Omega_1 \setminus \Omega_\varepsilon$ and from Lemma 3.47 we conclude

$$\begin{aligned}
&\int_{\Omega_1\setminus\Omega_\varepsilon} \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, (p_2-p_1) \,\mathrm{d}\mathbf{x} = \\
&\qquad = \int_{\Omega_1\setminus\Omega_\varepsilon} \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\Big((p_2+p_D)-(p_1+p_D)\Big) \,\mathrm{d}\mathbf{x} \le 0.
\end{aligned}$$

Since $v_\varepsilon = \varepsilon$ and $p_2 > \varepsilon + p_1$ in Ω_ε, the second expression can be bounded by

$$\int_{\Omega_\varepsilon} \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, \varepsilon \,\mathrm{d}\mathbf{x} \le 0$$

since $\theta(\mathbf{x}, s)$ is monotonically increasing in $s \in \mathbb{R}$ for all $\mathbf{x} \in \Omega$. Combining these estimates we obtain

$$\langle M(p_1) - M(p_2), v_\varepsilon\rangle_{V'\times V} = \int_\Omega \frac{n}{\tau}\Big(\theta(p_1+p_D) - \theta(p_2+p_D)\Big)\, v_\varepsilon \,\mathrm{d}\mathbf{x} \le 0. \tag{4.14}$$

Next, we want to estimate the duality pairing $\langle \widetilde{S}(p_1,p_2) - \widetilde{S}(p_2,p_2), v_\varepsilon \rangle_{V' \times V}$. Due to the representation of v_ε we have

$$\begin{aligned}
&\langle \widetilde{S}(p_1,p_2) - \widetilde{S}(p_2,p_2), v_\varepsilon \rangle_{V' \times V} = \\
&\qquad = \int\limits_{\Omega_1 \setminus \Omega_\varepsilon} \frac{K}{\mu} \Big(k\Big(\theta(p_1+p_D)\Big) - k\Big(\theta(p_2+p_D)\Big)\Big) \, \nabla p_2 \cdot \nabla v_\varepsilon \, \mathrm{d}\mathbf{x} \le \\
&\qquad \le c_S \left\| k\Big(\theta(p_1+p_D)\Big) - k\Big(\theta(p_2+p_D)\Big) \right\|_{L_\infty(\Omega_1 \setminus \Omega_\varepsilon)} \|\nabla p_2\|_{\mathbf{L}_2(\Omega_1 \setminus \Omega_\varepsilon)} |v_\varepsilon|_{H^1(\Omega)}.
\end{aligned}$$

Since $0 < p_2 - p_1 \le \varepsilon$ in $\Omega_1 \setminus \Omega_\varepsilon$ and due to the Lipschitz continuity of θ and k we obtain

$$c_S \left\| k\Big(\theta(p_1+p_D)\Big) - k\Big(\theta(p_2+p_D)\Big) \right\|_{L_\infty(\Omega_1 \setminus \Omega_\varepsilon)} \le \varepsilon \, c_S \, c_L$$

which leads to the inequality

$$\langle \widetilde{S}(p_1,p_2) - \widetilde{S}(p_2,p_2), v_\varepsilon \rangle_{V' \times V} \le \varepsilon \, c_S \, c_L \, \|\nabla p_2\|_{\mathbf{L}_2(\Omega_1 \setminus \Omega_\varepsilon)} |v_\varepsilon|_{H^1(\Omega)}. \tag{4.15}$$

Combining the inequalities (4.13), (4.14) and (4.15), we have

$$\frac{1}{\varepsilon} |v_\varepsilon|_{H^1(\Omega)} \le \frac{c_S \, c_L}{c_s \, c_{\alpha,k}} \|\nabla p_2\|_{\mathbf{L}_2(\Omega_1 \setminus \Omega_\varepsilon)}. \tag{4.16}$$

The final step is to prove that $|\Omega_\varepsilon|_d$ tends to zero for $\varepsilon \searrow 0$. Since $\varepsilon < p_2 - p_1$ in Ω_ε, we obtain the upper bound

$$|\Omega_\varepsilon|_d = \frac{1}{\varepsilon^2} \int\limits_{\Omega_\varepsilon} \varepsilon^2 \, \mathrm{d}\mathbf{x} \le \frac{1}{\varepsilon^2} \int\limits_{\Omega_\varepsilon} |v_\varepsilon|^2 \, \mathrm{d}\mathbf{x} = \frac{1}{\varepsilon^2} \|v_\varepsilon\|^2_{L_2(\Omega_\varepsilon)} \le \frac{1}{\varepsilon^2} \|v_\varepsilon\|^2_{L_2(\Omega)}$$

which holds for all $\varepsilon > 0$.

The norm equivalence (3.8) in V with $0 < c_N < 1$ and the inequality (4.16) imply

$$|\Omega_\varepsilon|_d \le \frac{1}{\varepsilon^2} \left(\frac{1 - c_N^2}{c_N^2} \right) |v_\varepsilon|^2_{H^1(\Omega)} \le \left(\frac{1 - c_N^2}{c_N^2} \right) \left(\frac{c_S \, c_L}{c_s \, c_{\alpha,k}} \right)^2 \|\nabla p_2\|^2_{\mathbf{L}_2(\Omega_1 \setminus \Omega_\varepsilon)}$$

with a right hand side which tends to zero since $\Omega_\varepsilon \to \Omega_1$ by construction and hence $|\Omega_1 \setminus \Omega_\varepsilon|_d \to 0$ as $\varepsilon \searrow 0$. From

$$|\Omega_\varepsilon|_d \le |\Omega_1|_d = |\Omega_1|_d - |\Omega_\varepsilon|_d + |\Omega_\varepsilon|_d = |\Omega_1 \setminus \Omega_\varepsilon|_d + |\Omega_\varepsilon|_d$$

we conclude, that $|\Omega_1|_d$ tends to zero as $\varepsilon \searrow 0$ and so $p_1 \ge p_2$ almost everywhere in Ω.

By interchanging the role of p_1 and p_2 we obtain $p_2 \ge p_1$ almost everywhere in Ω which implies $p_1 = p_2$ almost everywhere in Ω and this proves the statement of Theorem 4.4. ∎

Next, we want to prove an $L_\infty(\Omega)$–bound for the solution to the variational problem (4.4). The proof is based on a special property of real valued functions, which is stated in the following lemma.

Lemma 4.5. *Let $r_0 \in \mathbb{R}$, $r_0 \geq 0$, and let $f : [r_0, \infty) \to \mathbb{R}$ be a non negative and non increasing function such that*

$$f(s) \leq c \frac{\big(f(r)\big)^\delta}{(s-r)^\sigma}$$

for $0 \leq r_0 \leq r < s$ with a constant $c > 0$. Then $\sigma > 0$ and $\delta > 1$ imply

$$f(r_0 + \varrho) = 0$$

for all $\varrho \geq C$, where $C := c^{\frac{1}{\sigma}} \big(f(r_0)\big)^{\frac{\delta-1}{\sigma}} 2^{\frac{\delta}{\delta-1}}$.

Proof. See [29, Lemma 1.4]. ∎

With this lemma we can prove the following statement about the regularity of solutions to the variational problem (4.4) in two and three space dimensions.

Lemma 4.6. *Let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ and let $q_2 = 2$ and $q_3 = {}^{12}\!/_5$. Let $n, K \in L^+_\infty(\Omega)$, $\tau, \mu \in \mathbb{R}_+$, $f, q \in L_2(\Omega)$, $g_N \in L_{q_d}(\Gamma_N)$, $\nabla d \in \mathbf{L}_6(\Omega)$, $g_D \in H^{1/2}(\Gamma_D) \cap L_\infty(\Gamma_D)$ and let Assumption 4.2 hold. In addition, assume that $\theta : \Omega \times \mathbb{R} \to \mathbb{R}$ satisfies one of the following conditions.*

(1) There exists a mapping $s : \Omega \to \mathbb{R}$ and constants $s_l, s_u \in \mathbb{R}$, such that $\theta(\mathbf{x}, s(\mathbf{x})) = 0$ and $s_l \leq s(\mathbf{x}) \leq s_u$ for all $\mathbf{x} \in \Omega$.

(2) There exists an element $a_\theta(\mathbf{x}) \in L_2(\Omega)$, such that $|\theta(\mathbf{x}, s)| \leq a_\theta(\mathbf{x})$ for all $\mathbf{x} \in \Omega$ and $s \in \mathbb{R}$.

Then $p + p_D \in H^1(\Omega) \cap L_\infty(\Omega)$ where $p \in V$ is a solution to the variational problem (4.4).

Proof. The proof is based on the ideas of the proof of Lemma 2.4 in [29]. We split the proof into three parts, in the first part we do some preliminary considerations. In the second part we show that $p + p_D \in H^1(\Omega)$ is bounded from above by some constant almost everywhere in Ω and in the last part we prove that $p + p_D \in H^1(\Omega)$ can also be bounded from below by a constant almost everywhere in Ω.

Part A) Preliminary Considerations: First, consider the following imbeddings, see Theorem 3.38, Theorem 3.39 and Remark 3.40. For $d = 2, 3$ we have

$$H^1(\Omega) \subsetneq L_r(\Omega) \qquad \text{and} \qquad H^{1/2}(\Gamma_N) \subsetneq L_s(\Gamma_N) \tag{4.17}$$

for all $r \in [1, \infty)$ and $s \in [1, \infty)$ if $d = 2$ and for all $r \in [1, 6]$ and $s \in [1, 4]$ if $d = 3$. The imbeddings imply the existence of two positive constants $c_{q,\Omega} > 0$ and $c_{r,\Gamma_N} > 0$ such that

$$\|p\|_{L_r(\Omega)} \leq c_{r,\Omega} \|p\|_{H^1(\Omega)} \qquad \text{and} \qquad \|g\|_{L_s(\Gamma_N)} \leq c_{s,\Gamma_N} \|g\|_{H^{1/2}(\Gamma_N)}$$

holds for all $p \in H^1(\Omega)$ and $g \in H^{1/2}(\Gamma_N)$. Next, set $r_d := 6$ and

$$s_d := \begin{cases} 6, & d = 2, \\ 4, & d = 3 \end{cases}$$

and thus we have

$$\frac{1}{r_d} + \frac{1}{3} + \frac{1}{2} = 1 \qquad \text{and} \qquad \frac{1}{s_d} + \frac{1}{3} + \frac{1}{q_d} = 1$$

for $d = 2, 3$.

Since g_D is assumed to be an element in $H^{1/2}(\Gamma_D) \cap L_\infty(\Gamma_D)$, we can define the constant $G_\infty := \|g_D\|_{L_\infty(\Gamma_D)}$. In the next two steps we prove the existence of an upper and lower bound for $p + p_D \in H^1(\Omega)$ where $p \in V$ is a solution to the variational problem (4.4). For the remaining part of the proof we write $\widetilde{p}$ for $p + p_D$.

Part B) Upper Bound: First, we choose an arbitrary but fixed constant $P_0 \geq G_\infty$. For $P_1 \geq P_0$ we define the sets

$$\Omega_1 := \{\mathbf{x} \in \Omega \mid \widetilde{p}(\mathbf{x}) > P_1\} \quad \text{and} \quad \Gamma_{N,1} := \left\{\mathbf{x} \in \Gamma_N \mid \gamma^0_{\Gamma_N} \widetilde{p}(\mathbf{x}) > P_1\right\}$$

as well as $p_1 \in H^1(\Omega)$ defined by $p_1 := \widetilde{p} - \min\{\widetilde{p}, P_1\}$. The function p_1 and its weak gradient ∇p_1 satisfy the representation

$$p_1 = \begin{cases} \widetilde{p} - P_1 & \text{in } \Omega_1, \\ 0 & \text{in } \Omega \setminus \Omega_1 \end{cases} \qquad \text{and} \qquad \nabla p_1 = \begin{cases} \nabla \widetilde{p} & \text{in } \Omega_1, \\ 0 & \text{in } \Omega \setminus \Omega_1 \end{cases}$$

in the domain Ω. By applying the trace operator we obtain $\gamma^0_{\Gamma_D} p_1 = 0$ and so $p_1 \in V$.

Since $p_1 \in V$ is a legal test function and $p \in V$ is a solution to the variational problem (4.4), we have the identity

$$\langle M(p) - M(P_1 - p_D) + S(p), p_1 \rangle_{V' \times V} = \langle F - M(P_1 - p_D), p_1 \rangle_{V' \times V}$$

using the same notation as in the proof of Theorem 4.3.

Since $p_1 \in V$ and due to its representation we have

$$
\begin{aligned}
c_s\, c_{\alpha,k}\, c_N^2\, \|p_1\|^2_{H^1(\Omega)} \le c_s\, c_{\alpha,k}\, |p_1|^2_{H^1(\Omega)} \le \\
\le \int_\Omega \frac{n}{\tau} \Big(\theta(p_1+P_1) - \theta(P_1)\Big) \Big((p_1+P_1) - P_1\Big)\,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\, k\Big(\theta(\widetilde{p})\Big)\, \nabla p_1 \cdot \nabla p_1 \,\mathrm{d}\mathbf{x} = \\
= \int_{\Omega_1} \frac{n}{\tau} \Big(\theta(p+p_D) - \theta(P_1)\Big)\, p_1\,\mathrm{d}\mathbf{x} + \int_{\Omega_1} \frac{K}{\mu}\, k\Big(\theta(p+p_D)\Big)\, \nabla(p+p_D) \cdot \nabla p_1 \,\mathrm{d}\mathbf{x} = \\
= \langle M(p) - M(P_1 - p_D) + S(p), p_1\rangle_{V'\times V} = \langle F - M(P_1 - p_D), p_1\rangle_{V'\times V}. \quad (4.18)
\end{aligned}
$$

Each term of the right hand side of the inequality (4.18) can be estimated by the Hölder inequality and the constants r_d, s_d chosen in the first part of this proof.

For the duality pairing $\langle F, p_1\rangle_{V'\times V}$, which is defined as

$$
\langle F, p_1\rangle_{V'\times V} = \int_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, p_1\,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla p_1\,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, p_1\,\mathrm{d}s_{\mathbf{x}},
$$

we consider each term explicitly. For the first expression we obtain

$$
\begin{aligned}
\int_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, p_1\,\mathrm{d}\mathbf{x} &= \int_{\Omega_1} \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, p_1\,\mathrm{d}\mathbf{x} \\
&\le \left\| f + \frac{n}{\tau}\,\theta(q) \right\|_{L_2(\Omega)} \|p_1\|_{L_{r_d}(\Omega)}\, |\Omega_1|_d^{1/3} \\
&\le c_{r_d,\Omega} \left\| f + \frac{n}{\tau}\,\theta(q) \right\|_{L_2(\Omega)} \|p_1\|_{H^1(\Omega)}\, |\Omega_1|_d^{1/3}
\end{aligned}
$$

while the second expression satisfies the upper bound

$$
\begin{aligned}
\int_\Omega \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla p_1\,\mathrm{d}\mathbf{x} &= \int_{\Omega_1} \frac{K}{\mu}\, k\Big(\theta(q)\Big)\, \nabla d \cdot \nabla p_1\,\mathrm{d}\mathbf{x} \\
&\le c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_6(\Omega)}\, \|p_1\|_{H^1(\Omega)}\, |\Omega_1|_d^{1/3}.
\end{aligned}
$$

The third expression can be estimated by

$$
\begin{aligned}
\int_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, p_1\,\mathrm{d}s_{\mathbf{x}} = \int_{\Gamma_{N,1}} g_N\, \gamma^0_{\Gamma_N}\, p_1\,\mathrm{d}s_{\mathbf{x}} &\le \|g_N\|_{L_{q_d}(\Gamma_N)} \left\|\gamma^0_{\Gamma_N}\, p_1\right\|_{L_{s_d}(\Gamma_N)} |\Gamma_{N,1}|_{d-1}^{1/3} \\
&\le c_{s_d,\Gamma_N}\, c_T\, \|g_N\|_{L_{q_d}(\Gamma_N)}\, \|p_1\|_{H^1(\Omega)}\, |\Gamma_{N,1}|_{d-1}^{1/3}.
\end{aligned}
$$

If we set

$$c_g := c_{r_d,\Omega} \left\| f + \frac{n}{\tau}\,\theta(q) \right\|_{L_2(\Omega)} + c_S\, b_k \, \|\nabla d\|_{\mathbf{L}_6(\Omega)} + c_{s_d,\Gamma_N}\, c_T \, \|g_N\|_{L_{q_d}(\Gamma_N)}$$

we obtain the upper bound

$$\langle F, p_1 \rangle_{V' \times V} \leq c_g \left(|\Omega_1|_d^{1/3} + |\Gamma_{N,1}|_{d-1}^{1/3} \right) \|p_1\|_{H^1(\Omega)}. \tag{4.19}$$

To estimate the duality pairing $\langle M(P_1 - p_D), p_1 \rangle_{V' \times V}$ we have to distinguish between the two cases (1) and (2).
(1) Since the choice of $P_0 \geq G_\infty$ is arbitrary, choose

$$P_0 := \max\{G_\infty, s_u\} \geq G_\infty.$$

Due to the choice of P_0 and since $P_1 \geq P_0$, we have $\theta(P_1) \geq 0$ and $p_1(\mathbf{x}) = \widetilde{p}(\mathbf{x}) - P_1 \geq 0$ almost everywhere in Ω_1. Thus, the estimate

$$-\int\limits_{\Omega} \frac{n}{\tau}\,\theta(P_1)\, p_1 \,\mathrm{d}\mathbf{x} = -\int\limits_{\Omega_1} \frac{n}{\tau}\,\theta(P_1)\, p_1 \,\mathrm{d}\mathbf{x} \leq 0$$

holds. Define $c_B := 0$ if (1) holds.
(2) In the second case choose $P_0 := G_\infty$. Then, there holds

$$-\int\limits_{\Omega} \frac{n}{\tau}\,\theta(P_1)\, p_1 \,\mathrm{d}\mathbf{x} = -\int\limits_{\Omega_1} \frac{n}{\tau}\,\theta(P_1)\, p_1 \,\mathrm{d}\mathbf{x} \leq \left(\frac{c_M}{\tau} \|a_\theta\|_{L_2(\Omega)} \right) \|p_1\|_{H^1(\Omega)} \, |\Omega_1|_d^{1/3}.$$

In this case define $c_B := \frac{c_M}{\tau} \|a_\theta\|_{L_2(\Omega)}$.

In both cases we get the upper bound

$$-\langle M(P_1 - p_D), p_1 \rangle_{V' \times V} \leq c_B \, \|p_1\|_{H^1(\Omega)} \, |\Omega_1|_d^{1/3}. \tag{4.20}$$

Using the estimates (4.19) and (4.20) in combination with the inequality (4.18) we get

$$\begin{aligned} \|p_1\|_{H^1(\Omega)} &\leq \frac{c_g + c_B}{c_s\, c_{\alpha,k}\, c_N^2} \left(|\Omega_1|_d^{1/3} + |\Gamma_{N,1}|_{d-1}^{1/3} \right) \\ &\leq \frac{c_g + c_B}{c_s\, c_{\alpha,k}\, c_N^2} \left(|\Omega_1|_d^{1/4} + |\Gamma_{N,1}|_{d-1}^{1/4} \right)^{4/3} \end{aligned} \tag{4.21}$$

Next, choose $P_2 > P_1$. In the same manner as for P_1 we can define Ω_2 and $\Gamma_{N,2}$. Since $\Omega_2 \subset \Omega_1$, we obtain

$$\begin{aligned} \|p_1\|_{L_4(\Omega_1)}^4 &= \int\limits_{\Omega_1} |p_1|^4 \,\mathrm{d}\mathbf{x} \geq \int\limits_{\Omega_2} |p_1|^4 \,\mathrm{d}\mathbf{x} = \int\limits_{\Omega_2} |\widetilde{p} - P_1|^4 \,\mathrm{d}\mathbf{x} \\ &\geq \int\limits_{\Omega_2} |P_2 - P_1|^4 \,\mathrm{d}\mathbf{x} = |P_2 - P_1|^4 \, |\Omega_2|_d. \end{aligned}$$

For $\Gamma_{N,2} \subset \Gamma_{N,1}$ the estimate

$$\begin{aligned}\left\| \gamma^0_{\Gamma_{N,1}} p_1 \right\|^4_{L_4(\Gamma_{N,1})} &= \int\limits_{\Gamma_{N,1}} \left| \gamma^0_{\Gamma_{N,1}} p_1 \right|^4 \, \mathrm{d}s_{\mathbf{x}} \geq \int\limits_{\Gamma_{N,2}} \left| \gamma^0_{\Gamma_{N,1}} p_1 \right|^4 \, \mathrm{d}s_{\mathbf{x}} \\ &= \int\limits_{\Gamma_{N,2}} \left| \gamma^0_{\Gamma_{N,1}} \widetilde{p} - P_1 \right|^4 \, \mathrm{d}s_{\mathbf{x}} \geq \int\limits_{\Gamma_{N,2}} \left| P_2 - P_1 \right|^4 \, \mathrm{d}s_{\mathbf{x}} \\ &= \left| P_2 - P_1 \right|^4 \, \left| \Gamma_{N,2} \right|_{d-1}\end{aligned}$$

holds. Combining the previous two estimates with the imbeddings (4.17), we have

$$|P_2 - P_1| \left(|\Gamma_{N,2}|^{1/4}_{d-1} + |\Omega_2|^{1/4}_{d} \right) \leq (c_{4,\Omega} + c_{4,\Gamma_N} \, c_T) \, \|p_1\|_{H^1(\Omega)} \tag{4.22}$$

If we define $c := \frac{(c_{4,\Omega}+c_{4,\Gamma_N}\, c_T)\,(c_g+c_B)}{c_s\, c_{\alpha,k}\, c_N^2}$ and $f(P_i) := |\Omega_i|^{1/4}_d + |\Gamma_{N,i}|^{1/4}_d$, inequality (4.21) and inequality (4.22) yield

$$f(P_2) \leq c \, \frac{f(P_1)^{4/3}}{(P_2 - P_1)}.$$

By applying Lemma 4.5 we get the existence of a constant $C_P \geq 0$, such that $f(P_0 + C) = 0$ for all $C \geq C_P$ which implies that $\widetilde{p}(\mathbf{x}) \leq P_0 + C_P$ for almost all $\mathbf{x} \in \overline{\Omega}$.

Part C) Lower Bound: To show the existence of a lower bound, we use the same technique as in the second part. Let $N_0 \geq G_\infty$ be arbitrary but fixed. For $N_1 \geq N_0$ define the sets

$$\Omega_1 := \{\mathbf{x} \in \Omega \mid \widetilde{p}(\mathbf{x}) < -N_1\} \quad \text{and} \quad \Gamma_{N,1} := \left\{\mathbf{x} \in \Gamma_N \mid \gamma^0_{\Gamma_N} \widetilde{p}(\mathbf{x}) < -N_1\right\}$$

as well as the function $p_1 \in H^1(\Omega)$ which is defined by $p_1 := \widetilde{p} - \max\{\widetilde{p}, -N_1\}$. As in the second part, p_1 and ∇p_1 satisfy the representation

$$p_1 := \begin{cases} \widetilde{p} + N_1 & \text{in } \Omega_1, \\ 0 & \text{in } \Omega \setminus \Omega_1 \end{cases} \quad \text{and} \quad \nabla p_1 := \begin{cases} \nabla \widetilde{p} & \text{in } \Omega_1, \\ 0 & \text{in } \Omega \setminus \Omega_1 \end{cases}$$

in the domain Ω. By applying the trace operator we obtain $\gamma^0_{\Gamma_D} p_1 = 0$ and so $p_1 \in V$. Since $p \in V$ is a solution to the variational problem (4.4) and $p_1 \in V$ is a test function, we have the identity

$$\langle M(p) - M(-N_1 - p_D) + S(p), p_1 \rangle_{V' \times V} = \langle F - M(-N_1 - p_D), p_1 \rangle_{V' \times V}.$$

We proceed as in the second part, we just have to take a closer look at the duality pairing $\langle M(-N_1 - p_D), p_1 \rangle_{V' \times V}$. Here we again distinguish between the two cases (1)

and (2) of Lemma 4.6.
(1) Since the choice of $N_0 \geq G_\infty$ is arbitrary, choose

$$N_0 := \max\{G_\infty, -s_l\} \geq G_\infty.$$

Due to the choice of N_0 and since $N_1 \geq N_0$, we have $\theta(-N_1) \leq 0$ and $p_1(\mathbf{x}) = \widetilde{p}(\mathbf{x}) + N_1 \leq 0$ almost everywhere in Ω_1. Thus the estimate

$$-\int\limits_\Omega \frac{n}{\tau}\,\theta(-N_1)\,p_1\,\mathrm{d}\mathbf{x} = -\int\limits_{\Omega_1} \frac{n}{\tau}\,\theta(-N_1)\,p_1\,\mathrm{d}\mathbf{x} \leq 0$$

holds. Define $c_B := 0$ if (1) holds.
(2) In the second case choose $N_0 := G_\infty$. Then, there holds

$$-\int\limits_\Omega \frac{n}{\tau}\,\theta(-N_1)\,p_1\,\mathrm{d}\mathbf{x} = -\int\limits_{\Omega_1} \frac{n}{\tau}\,\theta(-N_1)\,p_1\,\mathrm{d}\mathbf{x} \leq \left(\frac{c_M}{\tau}\|a_\theta\|_{L_2(\Omega)}\right)\|p_1\|_{H^1(\Omega)}\,|\Omega_1|_d^{1/3}.$$

In this case define $c_B := \frac{c_M}{\tau}\|a_\theta\|_{L_2(\Omega)}$.

Following the lines for the upper bound, we get the existence of a constant $C_N \geq 0$, such that $f(N_0 + C) = 0$ for all $C \geq C_N$ which implies that $\widetilde{p}(\mathbf{x}) \geq -(N_0 + C_N)$ for almost all $\mathbf{x} \in \overline{\Omega}$.

Since C_P and C_N depend on P_0 and N_0, they do not coincide in general. All in all we get the boundedness of $p + p_D$ almost everywhere in Ω, which proves the desired statement. ■

We have proven, that the variational problem (4.4) is uniquely solvable under suitable conditions on general nonlinear functions θ and k. In the next two sections we choose θ and k as introduced in Chapter 2 and we check if the conditions in Assumption 4.2 are satisfied. As mentioned in the introduction, we want to apply the Kirchhoff transformation to obtain a simplified equation with a linear principal part. In the next section, we will discuss the variational problem (4.4) in the context of a homogeneous soil whereas in the section after next, a heterogeneous soil is considered.

4.3 Kirchhoff Transformation in Homogeneous Soil

In this section we apply the Kirchhoff transformation, see [4, 12, 57], to the Richards equation in a homogeneous soil. As in Section 2.1, we assume to have a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, and soil parameter θ and k defined by (2.4) and (2.5), that is

$$\theta(\mathbf{x}, s) := \theta(s) \qquad \text{and} \qquad k(\mathbf{x}, s) := k(s)$$

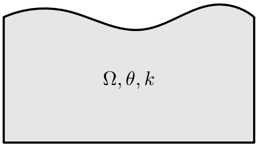


Figure 4.3: Homogeneous soil type.

for all $\mathbf{x} \in \Omega$ and $s \in \mathbb{R}$, see Figure 4.3.

Before we apply the Kirchhoff transformation, we want to see if the special choice of θ and k satisfy the conditions in Assumption 4.2. The saturation is prescribed by the monotonically increasing function θ mapping $\mathbb{R}$ to the interval $[\theta_{min}, \theta_{max}] \subset \mathbb{R}$. According to Chapter 2, the function θ is defined by

$$\theta(p) := \begin{cases} \left(\frac{p}{p_b}\right)^{-\lambda} \left(\theta_{max} - \theta_{min}\right) + \theta_{min} & \text{for } p \leq p_b, \\ \theta_{max} & \text{for } p > p_b, \end{cases}$$

for given constants $p_b < 0$ and $\lambda > 0$. The derivative of θ is

$$\theta'(p) = \begin{cases} \frac{\lambda}{(-p_b)} \left(\frac{p}{p_b}\right)^{-(\lambda+1)} \left(\theta_{max} - \theta_{min}\right) & \text{for } p < p_b, \\ 0 & \text{for } p > p_b, \end{cases}$$

which is defined everywhere except at $p = p_b < 0$, see Figure 4.4.

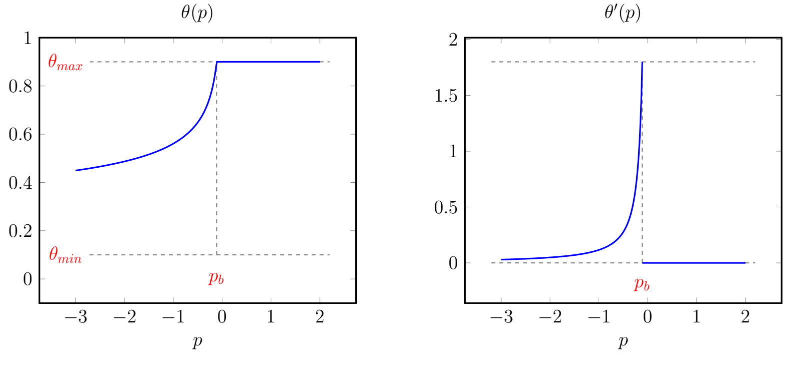


Figure 4.4: Saturation and its derivative.

It is easy to verify, that the saturation function θ is Lipschitz continuous. For $p_1, p_2 \leq p_b$ we have $|\theta(p_1) - \theta(p_2)| \leq c_{L,\theta}\, |p_1 - p_2|$ with the positive Lipschitz constant

$$c_{L,\theta} := \lim_{p \nearrow p_b} \theta'(p) = \frac{\lambda}{(-p_b)} \Big(\theta_{max} - \theta_{min}\Big).$$

For $p_1 \leq p_b \leq p_2$ we use the previous result and obtain

$$\begin{aligned}|\theta(p_1) - \theta(p_2)| = |\theta(p_1) - \theta(p_b)| \leq c_{L,\theta}\, |p_1 - p_b| = c_{L,\theta}(p_b - p_1)\\ \leq c_{L,\theta}(p_2 - p_1) = c_{L,\theta}\, |p_1 - p_2| \,.\end{aligned}$$

The Lipschitz continuity for $p_b \leq p_1, p_2$ is obvious since $\theta(p)$ is constant for all $p \geq p_b$. Hence, θ is Lipschitz continuous for all $p_1, p_2 \in \mathbb{R}$.

We conclude, that $\theta \in \mathcal{C}^{0,1}(\mathbb{R}) \cap L_\infty(\mathbb{R})$ and in addition, θ is monotonically increasing. Furthermore, we have the upper bound $\|\theta\|_{L_\infty(\mathbb{R})} = \theta_{max}$. Thus, the nonlinear function $\theta : \mathbb{R} \to \mathbb{R}$ fulfills condition (1) in Assumption 4.2 as well as condition (1) in Lemma 4.6. Furthermore, θ is a Carathéodory function and there holds

$$|\theta(s)| = |\theta(0) + \theta(s) - \theta(0)| \leq |\theta(0)| + |\theta(s) - \theta(0)| \leq \theta_{max} + c_{L,\theta}\ |s|$$

for all $s \in \mathbb{R}$. Hence, θ satisfies the conditions in Assumption 4.2.

Next, we consider the relative permeability k. It is given by the mapping $k : [\theta_{min}, \theta_{max}] \to [0,1] \subset \mathbb{R}$ with

$$k(\theta) := \left(\frac{\theta - \theta_{min}}{\theta_{max} - \theta_{min}}\right)^{3+2/\lambda}$$

which is monotonically increasing for a given $\lambda > 0$. We can extend k to a mapping $k : \mathbb{R} \to [0,1]$ by setting $k(\theta) = 0$ for $\theta < \theta_{min}$ and $k(\theta) = 1$ for $\theta > \theta_{max}$ which is still a monotonically increasing mapping. The derivative is

$$k'(\theta) = \begin{cases} 0 & \text{for } \theta < \theta_{min}, \\ \frac{3+2/\lambda}{\theta_{max}-\theta_{min}} \left(\frac{\theta-\theta_{min}}{\theta_{max}-\theta_{min}}\right)^{2+2/\lambda} & \text{for } \theta \in (\theta_{min}, \theta_{max}), \\ 0 & \text{for } \theta > \theta_{max}, \end{cases}$$

which is defined everywhere except at the points θ_{min} and θ_{max}, see Figure 4.5.

In the same manner as for θ, we can verify that k satisfies

$$|k(\theta_1) - k(\theta_2)| \leq c_{L,k}\, |\theta_1 - \theta_2|$$

for all $\theta_1, \theta_2 \in \mathbb{R}$ with a positive constant $c_{L,k}$ given by

$$c_{L,k} := \lim_{\theta \nearrow \theta_{max}} k'(\theta) = \frac{3 + 2/\lambda}{\theta_{max} - \theta_{min}}.$$

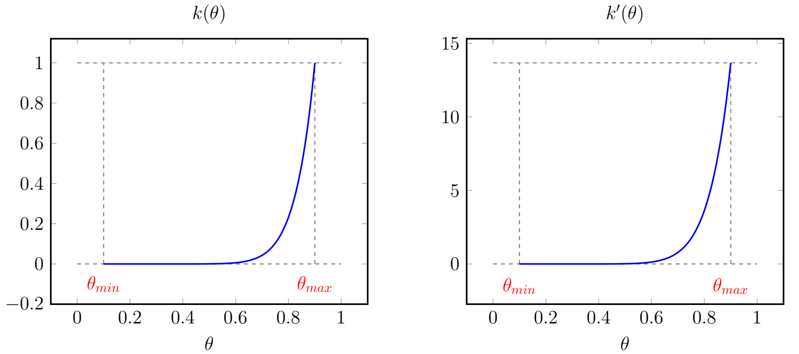


Figure 4.5: Relative permeability and its derivative in $[\theta_{min}, \theta_{max}]$.

This observation implies $k \in \mathcal{C}^{0,1}(\mathbb{R}) \cap L_\infty(\mathbb{R})$ with $\|k\|_{L_\infty(\mathbb{R})} = 1$. So far the function k fulfills condition (2) in Assumption 4.2 and the additional Lipschitz continuity condition of Theorem 4.4. In order to fulfill condition (3) in Assumption 4.2 we choose a fixed $\alpha \in (0, 1)$ arbitrarily small. We consider the modified permeability $k_\alpha : \mathbb{R} \to [\alpha, 1]$ which is defined by

$$k_\alpha(\theta) := \max\{\alpha, k(\theta)\} \tag{4.23}$$

for all $\theta \in \mathbb{R}$. The modified permeability k_α is still in $\mathcal{C}^{0,1}(\mathbb{R}) \cap L_\infty(\mathbb{R})$ with Lipschitz constant $c_{L,k}$ and in addition there holds $k_\alpha(\theta) \geq \alpha > 0$ for all $\theta \in \mathbb{R}$, see Figure 4.6. Thus, the saturation function θ and the modified permeability k_α fulfill the conditions

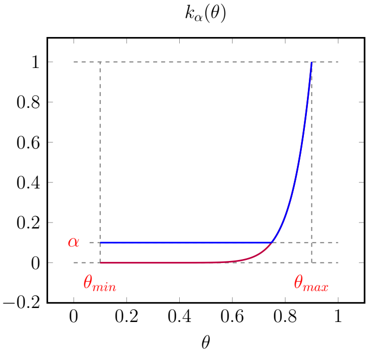


Figure 4.6: Modified relative permeability.

stated in Assumption 4.2.

In the case of a homogeneous soil we thus obtain the following variational problem.

Variational formulation for homogeneous soil

Find $p \in V$ such that

$$\int_\Omega \frac{n}{\tau}\,\theta(p+p_D)\,v\,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\,k_\alpha\Big(\theta(p+p_D)\Big)\,\nabla(p+p_D)\cdot\nabla v\,\mathrm{d}\mathbf{x} = \\ = \int_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\,v\,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\,k_\alpha\Big(\theta(q)\Big)\,\nabla d\cdot\nabla v\,\mathrm{d}\mathbf{x} + \\ + \int_{\Gamma_N} g_N\,\gamma^0_{\Gamma_N}\,v\,\mathrm{d}s_{\mathbf{x}} \tag{4.24}$$

for all $v \in V$ and $p_D := \mathcal{E}_\Omega\, g_D$.

Compared to the variational problem (4.4), the problem we obtain in the homogeneous case is formally identical. The major difference is that the nonlinear parameter functions do not depend on $\mathbf{x} \in \Omega$ explicitly. We can state the following corollary concerning solvability and uniqueness.

Corollary 4.7. *If $n, K \in L^+_\infty(\Omega)$, $\tau, \mu \in \mathbb{R}_+$, $f, q \in L_2(\Omega)$, $g_N \in L_2(\Gamma_N)$, $\nabla d \in \mathbf{L}_2(\Omega)$ and $g_D \in H^{1/2}(\Gamma_D)$, then the the variational problem (4.24) has a unique solution $p \in V$ and there holds*

$$\|p\|_{H^1(\Omega)} \le c\Big(\|f\|_{L_2(\Omega)} + \frac{1}{\tau}\|q-p_D\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|g_D\|_{H^{1/2}(\Gamma_D)} + \|g_N\|_{L_2(\Gamma_N)}\Big)$$

with some positive constant $c(\Omega, \theta, k, K, n, \mu)$. If, in addition, $g_N \in L_{q_d}(\Gamma_N)$ with q_d as in Lemma 4.6 and $g_D \in H^{1/2}(\Gamma_D) \cap L_\infty(\Gamma_D)$, then there holds $p + p_D \in H^1(\Omega) \cap L_\infty(\Omega)$.

Proof. From the previous considerations on the nonlinear parameter functions we have, that θ and k satisfy the conditions stated in Assumption 4.2. The unique solvability follows from Theorem 4.3 and Theorem 4.4. To show the regularity, we have to make sure that $\nabla d \in \mathbf{L}_6(\Omega)$. The function $d : \Omega \to \mathbb{R}$ was introduced in Chapter 2 as $d(x_1, \ldots, x_d) = \varrho\, g\, x_d$. The gradient of d is constant in each component and so in $\mathbf{L}_6(\Omega)$. Lemma 4.6 provides the desired regularity result. ∎

As already mentioned, we want to apply the Kirchhoff transformation to obtain a simplification of the variational problem (4.24). For this reason we define the mapping $\kappa : \mathbb{R} \to \mathbb{R}$ as

$$\kappa(s) := \int_0^s k_\alpha\Big(\theta(r)\Big)\,\mathrm{d}r \tag{4.25}$$

for all $s \in \mathbb{R}$. This mapping is known as Kirchhoff transformation and its properties are summarized in the following lemmata. The assertions we make are well known and are based on the fundamental theorem of calculus for Lebesgue points and on the theory of Lebesgue points, see [53, 67].

Lemma 4.8. *If $k_\alpha \circ \theta \in L_\infty(\mathbb{R})$ is non negative almost everywhere, then the Kirchhoff transformation $\kappa : \mathbb{R} \to \mathbb{R}$ defined by (4.25) is monotonically increasing and Lipschitz continuous with Lipschitz constant $\|\kappa'\|_{L_\infty(\mathbb{R})}$. In addition, we have $\kappa' = k_\alpha \circ \theta$ almost everywhere on $\mathbb{R}$.*

If $k_\alpha \circ \theta \in L_\infty(\mathbb{R})$ satisfies $k_\alpha \circ \theta \geq c > 0$ almost everywhere on $\mathbb{R}$, then the Kirchhoff transformation κ has a strictly monotonically increasing and Lipschitz continuous inverse $\kappa^{-1} : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant $\left\|\kappa^{-1\prime}\right\|_{L_\infty(\mathbb{R})}$. In addition, we have $\kappa^{-1\prime} = \frac{1}{\kappa' \circ \kappa^{-1}}$ almost everywhere in $\mathbb{R}$.

Proof. For a proof see [57, Lemma 4.2.1]. ∎

The composite function $k_\alpha \circ \theta$ defined by (4.23) and (2.5) with $\alpha > 0$ satisfies the assumptions in Lemma 4.8. In Figure 4.7 the Kirchhoff transformation as well as its inverse are depicted. The red line illustrates the limiting case $\alpha = 0$.

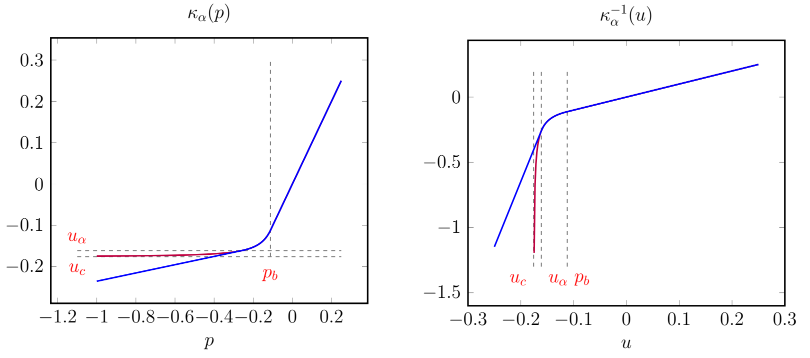


Figure 4.7: Kirchhoff transformation κ and its inverse κ^{-1}.

Since the Kirchhoff transformation κ satisfies the conditions in Theorem 3.49, we conclude that κ induces a superposition operator which acts on $H^1(\Omega)$ continuously. Furthermore, we have

$$\nabla \kappa(v) = \kappa'(v)\, \nabla v = k_\alpha\Big(\theta(v)\Big)\, \nabla v \tag{4.26}$$

in $\mathbf{L}_2(\Omega)$ for all $v \in H^1(\Omega)$. The representation (4.26) is crucial for later considerations. But we need further properties which are summarized in the following lemma.

Lemma 4.9. *Assume that $k_\alpha \circ \theta \in L_\infty(\mathbb{R})$ satisfies $k_\alpha \circ \theta \geq c > 0$ almost everywhere on $\mathbb{R}$. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and let $\kappa : \mathbb{R} \to \mathbb{R}$ be defined by (4.25). Then, the superposition operator $\kappa : H^1(\Omega) \to H^1(\Omega)$ is bounded.*

In addition, the inverse Kirchhoff transformation $\kappa^{-1} : \mathbb{R} \to \mathbb{R}$ induces a continuous superposition operator acting on $H^1(\Omega)$ which is also bounded.

Proof. From Theorem 3.49 we know that the superposition operator acts continuously on $H^1(\Omega)$. The $H^1(\Omega)$–norm of $\kappa(v)$ is

$$\|\kappa(v)\|^2_{H^1(\Omega)} = \|\kappa(v)\|^2_{L_2(\Omega)} + |\kappa(v)|^2_{H^1(\Omega)}$$

which we want to estimate separately. For the $L_2(\Omega)$–norm, we use Lemma 3.48 and Lemma 4.8 and thus we obtain

$$\|\kappa(v)\|^2_{L_2(\Omega)} = \|\kappa(v) - \kappa(0)\|^2_{L_2(\Omega)} \leq \|k_\alpha \circ \theta\|^2_{L_\infty(\mathbb{R})}\, \|v\|^2_{L_2(\Omega)}$$

for all $v \in H^1(\Omega)$. To estimate the $H^1(\Omega)$–seminorm, we use the representation formula (4.26) for the gradient of the Kirchhoff transformed. We obtain

$$\begin{aligned} |\kappa(v)|^2_{H^1(\Omega)} &= \int_\Omega |\nabla \kappa(v)|^2 \,\mathrm{d}\mathbf{x} = \int_\Omega \Big| k_\alpha\Big(\theta(v)\Big)\, \nabla v\Big|^2 \,\mathrm{d}\mathbf{x} \\ &\leq \|k_\alpha \circ \theta\|^2_{L_\infty(\mathbb{R})} \int_\Omega |\nabla v|^2 \,\mathrm{d}\mathbf{x} = \|k_\alpha \circ \theta\|^2_{L_\infty(\mathbb{R})}\, |v|^2_{H^1(\Omega)} \end{aligned}$$

and hence

$$\|\kappa(v)\|_{H^1(\Omega)} \leq \|k_\alpha \circ \theta\|_{L_\infty(\mathbb{R})}\, \|v\|_{H^1(\Omega)}$$

for all $v \in H^1(\Omega)$. Due to Lemma 4.8, the inverse transformation κ^{-1} exists and is Lipschitz continuous. Furthermore, we have

$$\kappa^{-1\prime}(s) = \frac{1}{\kappa'\Big(\kappa^{-1}(s)\Big)} = \frac{1}{k_\alpha\Big(\theta(\kappa^{-1}(s))\Big)}$$

for all $s \in \mathbb{R}$ and due to the boundedness of $k_\alpha \circ \theta$ we get

$$\frac{1}{\|k_\alpha \circ \theta\|_{L_\infty(\mathbb{R})}} \leq \frac{1}{k_\alpha\Big(\theta(\kappa^{-1}(s))\Big)} \leq \frac{1}{\alpha}$$

and therefore $\kappa^{-1\prime} \in L_\infty(\mathbb{R})$. Using the same technique as for $\kappa(v)$ we get

$$\left\|\kappa^{-1}(v)\right\|_{H^1(\Omega)} \leq \frac{1}{\alpha} \|v\|_{H^1(\Omega)}$$

for all $v \in H^1(\Omega)$. ■

Remark 4.10. *To guarantee the boundedness of the inverse superposition operator $\kappa^{-1} : H^1(\Omega) \to H^1(\Omega)$, we have to assume that the constant α in the definition of k_α, see (4.23), is strictly positive. The case $\alpha = 0$ is not covered by our framework but can be handled in the context of variational inequalities, see for example [12].*

We can now use the Kirchhoff transformation to shift one nonlinearity from the domain to the boundary. For this reason, we introduce a new quantity $u \in H^1(\Omega)$ which is defined by

$$u := \kappa(p + p_D) - u_D \tag{4.27}$$

with $u_D := \mathcal{E}_\Omega\, \kappa(g_D)$. The function $u \in H^1(\Omega)$ is known as generalized pressure. Theorem 3.50 implies $\kappa(g_D) \in H^{1/2}(\Gamma_D)$ which ensures that the definition of u_D is well posed. If we apply the trace operator, we obtain the identity

$$\gamma^0_{\Gamma_D} u = \gamma^0_{\Gamma_D} \kappa(p + p_D) - \gamma^0_{\Gamma_D} u_D = \gamma^0_{\Gamma_D} \kappa(p + p_D) - \kappa(g_D).$$

From Lemma 3.51 we conclude that

$$\gamma^0_{\Gamma_D} u = \kappa(\gamma^0_{\Gamma_D} p + \gamma^0_{\Gamma_D} p_D) - \kappa(g_D) = \kappa(g_D) - \kappa(g_D) = 0$$

and thus $u \in V$. If we use the representation formula (4.26) for the gradient, we obtain

$$\nabla u = \nabla \kappa(p + p_D) - \nabla u_D = k_\alpha\Big(\theta(p + p_D)\Big)\, \nabla(p + p_D) - \nabla u_D$$

in $\mathbf{L}_2(\Omega)$. If we plug this representation formula in the left hand side of the variational problem (4.24), we obtain

$$\begin{aligned}\int\limits_\Omega \frac{n}{\tau}\, \theta(p + p_D)\, v \,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, k_\alpha\Big(\theta(p + p_D)\Big)\, \nabla(p + p_D) \cdot \nabla v \,\mathrm{d}\mathbf{x} = \\ = \int\limits_\Omega \frac{n}{\tau}\, \theta\Big(\kappa^{-1}(u + u_D)\Big)\, v \,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, \nabla(u + u_D) \cdot \nabla v \,\mathrm{d}\mathbf{x}\end{aligned} \tag{4.28}$$

for all $v \in V$.

Some words to recap the latest lines. We have defined a new function $u \in V$ by (4.27) which we used to reformulate the variational formulation (4.24) and we obtained an equivalent representation, see (4.28). In contrast to the original variational formulation, the formulation of the Kirchhoff transformed u is now linear in its principal part. Thus, we were able to shift one nonlinearity from the domain to the boundary, hidden in the computation of $\kappa(g_D) \in H^{1/2}(\Gamma_D)$. This computation can be easy if the boundary condition g_D is sufficiently regular and an explicit representation of κ is known. For a general $k_\alpha \circ \theta$ the computation of $\kappa(g_D)$ can be a challenging task.

By setting $l := \theta \circ \kappa^{-1}$ we derive the following variational problem which is equivalent to the variational problem (4.24).

Transformed variational formulation for homogeneous soil

Find $u \in V$ such that

$$\begin{aligned}\int\limits_\Omega \frac{n}{\tau}\, l(u+u_D)\, v \,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, \nabla(u+u_D)\cdot\nabla v\,\mathrm{d}\mathbf{x} &= \\ = \int\limits_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, v\,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, k_\alpha\big(\theta(q)\big)\,\nabla d\cdot\nabla v\,\mathrm{d}\mathbf{x} &+ \\ + \int\limits_{\Gamma_N} g_N\,\gamma^0_{\Gamma_N}\, v\,\mathrm{d}s_{\mathbf{x}}&\end{aligned} \tag{4.29}$$

for each $v \in V$ and $u_D := \mathcal{E}_\Omega\, \kappa(g_D)$.

Instead of solving the fully nonlinear variational problem (4.24), we can equivalently solve the nonlinear variational problem (4.29) with a linear principal part. We just have to apply the Kirchhoff transformation to the given Dirichlet datum $g_D \in H^{1/2}(\Gamma_D)$. As already mentioned, we just have the implicit definition (4.25) of the Kirchhoff transformation. However, one big advantage of the model introduced by Brooks and Corey is, that we can compute an explicit representation of κ. Therefore, we first define $p_\alpha < 0$ as

$$p_\alpha := \frac{p_b}{\alpha^{1/(3\,\lambda+2)}}$$

which satisfies $k(p_\alpha) = \alpha$. We can rewrite $k_\alpha(\theta(p))$ defined by (4.23) as

$$k_\alpha\big(\theta(p)\big) := \begin{cases} \alpha & p \le p_\alpha, \\ \left(\frac{p}{p_b}\right)^{-3\,\lambda-2} & p_\alpha < p \le p_b, \\ 1 & p > p_b \end{cases}$$

for all $p \in \mathbb{R}$. By integration of $k_\alpha \circ \theta$, we obtain the explicit representation of the Kirchhoff transformation

$$\kappa(p) := \begin{cases} \alpha\,(p - p_\alpha) + \frac{p_b}{3\,\lambda+1}\left(\left(3\,\lambda+2\right) - \left(\frac{p_\alpha}{p_b}\right)^{-3\,\lambda-1}\right) & p \le p_\alpha, \\ \frac{p_b}{3\,\lambda+1}\left(\left(3\,\lambda+2\right) - \left(\frac{p}{p_b}\right)^{-3\,\lambda-1}\right) & p_\alpha < p \le p_b, \\ p & p > p_b \end{cases}$$

for all $p \in \mathbb{R}$. To give an explicit form of the inverse Kirchhoff transformation, we define

$$u_\alpha := \frac{p_b}{3\,\lambda+1}\left(\left(3\,\lambda+2\right) - \left(\frac{p_\alpha}{p_b}\right)^{-3\,\lambda-1}\right)$$

and we obtain

$$\kappa^{-1}(u) := \begin{cases} \frac{u-u_\alpha}{\alpha} + p_\alpha & u \le u_\alpha, \\ -\left(\frac{(-p_b)^{3\,\lambda+2}}{(3\,\lambda+2)\,(u-p_b)-u}\right)^{1/(3\,\lambda+1)} & u_\alpha < u \le p_b, \\ u & u > p_b \end{cases}$$

for all $u \in \mathbb{R}$, see Figure 4.7.

As we can see, the Kirchhoff transformation is a useful tool to transform a certain class of partial differential equations, which are quasilinear in the principal part, to partial differential equations, which are linear in the principle part. Especially for stationary problems the advantage is obvious. In this special case, quasilinear partial differential equations are transformed to linear partial differential equations. Next, we want to carry this idea over to the Richards equation in heterogeneous soil.

4.4 Kirchhoff Transformation in Heterogeneous Soil

In this section, we want to apply the Kirchhoff transformation to the Richards equation in a heterogeneous soil. But first we want to answer the question about solvability and uniqueness. As discussed in Section 2.2, we assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain with different layers. For each soil layer L_i we have corresponding parameter functions θ_i and k_i defined by (2.4) and (2.5). The global parameter functions, defined by (2.7) and (2.8), are given by

$$\theta(\mathbf{x}, s) = \theta_i(s) \quad \text{and} \quad k(\mathbf{x}, s) = k_i(s)$$

for $\mathbf{x} \in L_i$ and $s \in \mathbb{R}$, see Figure 4.8.

Due to the observations in Section 4.3, each θ_i is Lipschitz continuous and an element of the Banach space $L_\infty(\mathbb{R})$. For a constant $\alpha > 0$, the modified permeability $k_{\alpha,i}$,

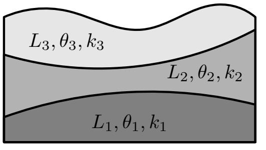


Figure 4.8: Heterogeneous soil type.

defined by (4.23), is Lipschitz continuous, in $L_\infty(\Omega)$ and $k_{\alpha,i} \geq \alpha > 0$. It is easy to verify, that the nonlinearities satisfy the conditions in Assumption 4.2.

If we write down the variational problem (4.4) in terms of the nonlinearities θ and k_α as discussed in this section, we obtain the following variational problem.

Variational formulation for heterogeneous soil

Find $p \in V$ such that

$$\begin{aligned}\int\limits_\Omega \frac{n}{\tau}\,\theta(p+p_D)\,v\,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\,k_\alpha\Big(\theta(p+p_D)\Big)\,\nabla(p+p_D)\cdot\nabla v\,\mathrm{d}\mathbf{x} =& \\ = \int\limits_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\,v\,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\,k_\alpha\Big(\theta(q)\Big)\,\nabla d\cdot\nabla v\,\mathrm{d}\mathbf{x} +& \\ + \int\limits_{\Gamma_N} g_N\,\gamma^0_{\Gamma_N}\,v\,\mathrm{d}s_{\mathbf{x}}&\end{aligned} \tag{4.30}$$

for all $v \in V$ and $p_D := \mathcal{E}_\Omega\, g_D$.

The variational problem (4.30) looks formally like the variational problem (4.24), but the nonlinearities in the variational problem (4.30) depend explicitly on $\mathbf{x} \in \Omega$. Furthermore, it is rather easy to prove the following corollary using the considerations made in Section 4.3 and the results of Section 4.2.

Corollary 4.11. *If $n, K \in L^+_\infty(\Omega)$, $\tau, \mu \in \mathbb{R}_+$, $f, q \in L_2(\Omega)$, $g_N \in L_2(\Gamma_N)$, $\nabla d \in \mathbf{L}_2(\Omega)$ and $g_D \in H^{1/2}(\Gamma_D)$, then the variational problem (4.30) has a unique solution $p \in V$ and there holds*

$$\|p\|_{H^1(\Omega)} \leq c\left(\|f\|_{L_2(\Omega)} + \frac{1}{\tau}\|q - p_D\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|g_D\|_{H^{1/2}(\Gamma_D)} + \|g_N\|_{L_2(\Gamma_N)}\right)$$

with some positive constant $c(\Omega, \theta, k, K, n, \mu)$. If, in addition, $g_N \in L_{q_d}(\Gamma_N)$ with q_d as in Lemma 4.6 and $g_D \in H^{1/2}(\Gamma_D) \cap L_\infty(\Gamma_D)$, then there holds $p + p_D \in H^1(\Omega) \cap L_\infty(\Omega)$.

Proof. Follow the lines of the proof to Corollary 4.7. ■

Next, we want to apply the Kirchhoff transformation like in Section 4.3 to obtain a variational problem with a linear principal part. In contrast to the Kirchhoff transformation defined by (4.25), we obtain in this setting a Kirchhoff transformation κ which depends on $\mathbf{x} \in \Omega$, that is

$$\kappa(\mathbf{x}, s) := \int_0^s k_\alpha(\mathbf{x}, \theta(\mathbf{x}, r))\,\mathrm{d}r$$

for $\mathbf{x} \in \Omega$ and $s \in \mathbb{R}$. Unfortunately we can not apply Theorem 3.49, since the composite function $k_\alpha \circ \theta$ depends on $\mathbf{x} \in \Omega$.

The idea is to exploit the structure of the nonlinearities. Within each soil layer, the nonlinear parameter functions θ and k are independent of $\mathbf{x} \in \Omega$. Goal is to find an equivalent formulation, such that the nonlinear form

$$\int_\Omega \frac{n}{\tau}\,\theta(p + p_D)\,v\,\mathrm{d}\mathbf{x} + \int_\Omega \frac{K}{\mu}\,k_\alpha\Big(\theta(p + p_D)\Big)\,\nabla(p + p_D)\cdot\nabla v\,\mathrm{d}\mathbf{x}$$

can be written as the sum of local forms, that is

$$\sum_{i=1}^{N_L}\left(\int_{L_i} \frac{n}{\tau}\,\theta(p + p_D)\,v\,\mathrm{d}\mathbf{x} + \int_{L_i} \frac{K}{\mu}\,k_\alpha\Big(\theta(p + p_D)\Big)\,\nabla(p + p_D)\cdot\nabla v\,\mathrm{d}\mathbf{x}\right),$$

plus additional coupling conditions.

To introduce a suitable formulation we need the following definition.

Definition 4.12 (Admissible decomposition)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. The family $\mathcal{D}_\Omega^N := \{\Omega_i\}_{i=1}^N$ is called admissible decomposition of Ω, if the following conditions hold.*

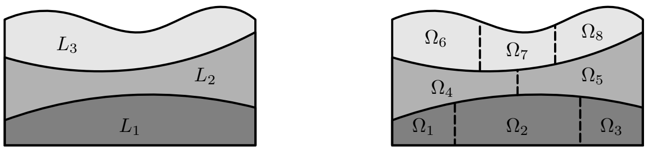
(1) $\Omega_i \subset \Omega$ is an open and bounded Lipschitz domain for $i = 1, \ldots, N$.
(2) $\Omega_i \cap \Omega_j = \emptyset$ for all $1 \le i, j \le N$ with $i \neq j$.
(3) $\overline{\Omega} = \bigcup_{\Omega_i \in \mathcal{D}_\Omega^N} \overline{\Omega_i}$.

By $\mathbf{n}_i$ we denote the outer unit normal of $\Omega_i \in \mathcal{D}_\Omega^N$ and let $\mathcal{S}$ denote the skeleton of $\mathcal{D}_\Omega^N$ defined by

$$\mathcal{S} := \bigcup_{\Omega_i \in \mathcal{D}_\Omega^N} \partial\Omega_i.$$

Furthermore, we define the interface Γ_{ij} by $\Gamma_{ij} := \partial\Omega_i \cap \partial\Omega_j$ for each pair of indices i, j with $1 \le i, j \le N$, $i \neq j$.

Since we want to consider the Richards equation in a heterogeneous soil, the decomposition of Ω can not be chosen arbitrarily. In Section 2.2 we assumed, that Ω consists of several soil layers which we denoted by $L_i \subset \Omega$ for $i = 1, \ldots, N_L$, see Figure 4.9a. This decomposition is a natural decomposition of our domain induced by the soil parameter and therefore from the problem itself. From now on, we assume that $\mathcal{D}_\Omega^N$, $N \geq N_L$, is an admissible decomposition of Ω which resolve the given natural decomposition, see Figure 4.9b.



(a) Natural decomposition. (b) General decomposition.

Figure 4.9: Sketch of natural and general decomposition.

After we clarified the question concerning the decomposition of our domain, we can continue with the decomposition of the nonlinear form. This is done in the following subsection.

4.4.1 The Primal Hybrid Framework

In this subsection we want to discuss the primal hybrid method as well as the tools and statements we need to realize this method. This subsection is based on the work of [13, 19, 51, 64]. The idea of the primal hybrid formulation is to consider a larger trial space by removing the constraint of continuity at the interfaces. These spaces are known as broken Sobolev spaces and are the fundamentals of discontinuous Galerkin methods, see [27, Section 1.2.5]. For our purpose we need an extension of the Sobolev space $H^1(\Omega)$ which is defined as follows.

Definition 4.13 (Broken Sobolev space)**.** *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}_\Omega^N$. The broken Sobolev space $X^1(\Omega)$ is defined by*

$$X^1(\Omega) := \left\{ u \in L_2(\Omega) \mid u_{|\Omega_i} \in H^1(\Omega_i) \; \forall \, \Omega_i \in \mathcal{D}_\Omega^N \right\}$$

where

$$\|u\|_{X^1(\Omega)} := \left(\sum_{i=1}^{N} \left\| u_{|\Omega_i} \right\|^2_{H^1(\Omega_i)} \right)^{1/2}$$

defines the corresponding norm.

It is easy to verify, that $\|u\|_{H^1(\Omega)} = \|u\|_{X^1(\Omega)}$ for all $u \in H^1(\Omega)$ and thus $H^1(\Omega) \subsetneq X^1(\Omega)$, that is $H^1(\Omega)$ is continuously imbedded in $X^1(\Omega)$. Since one can not neglect the constraint of continuity completely, we introduce Lagrange multiplier to obtain continuity. For the introduction of the Lagrange multiplier in a proper way, we need additional spaces. First, we introduce the following product trace space

$$X^{1/2}(\mathcal{S}) := \prod_{i=1}^{N} H^{1/2}(\partial\Omega_i)$$

and its dual space

$$X^{1/2}(\mathcal{S})' := \prod_{i=1}^{N} H^{1/2}(\partial\Omega_i)'.$$

We can define the trace operator $\gamma^0_{\mathcal{S}} : X^1(\Omega) \to X^{1/2}(\mathcal{S})$ as

$$\gamma^0_{\mathcal{S}}\, u := \left(\gamma^0_{\partial\Omega_1}\, u_{|\Omega_1}, \ldots, \gamma^0_{\partial\Omega_N}\, u_{|\Omega_N}\right)^\top$$

which satisfies the stability estimate

$$\left\|\gamma^0_{\mathcal{S}}\, u\right\|_{X^{1/2}(\mathcal{S})} := \left(\sum_{i=1}^{N} \left\|\gamma^0_{\partial\Omega_i}\, u_{|\Omega_i}\right\|^2_{H^{1/2}(\partial\Omega_i)}\right)^{1/2} \leq c_{T,\mathcal{S}}\, \|u\|_{X^1(\Omega_i)}$$

for all $u \in X^1(\Omega)$. The constant $c_{T,\mathcal{S}}$ is defined by $c_{T,\mathcal{S}} := \max_{i=1,\ldots,N} c_{T,i}$ with $c_{T,i}$ from the Trace Theorem 3.35 for $H^1(\Omega_i)$. In the same manner, we define the extension operator $\mathcal{E}_\Omega : X^{1/2}(\mathcal{S}) \to X^1(\Omega)$ as

$$\mathcal{E}_\Omega\, g := \left(\mathcal{E}_{\Omega_1}\, g_1, \ldots, \mathcal{E}_{\Omega_N}\, g_N\right)^\top$$

for all $(g_1, \ldots, g_N)^\top = g \in X^{1/2}(\mathcal{S})$. The extension operator satisfies

$$\|\mathcal{E}_\Omega\, g\|_{X^1(\Omega)} := \left(\sum_{i=1}^{N} \left\|\mathcal{E}_{\Omega_i}\, g_i\right\|^2_{H^1(\Omega_i)}\right)^{1/2} \leq c_{E,\mathcal{S}}\, \|g\|_{X^{1/2}(\mathcal{S})}$$

with $c_{E,\mathcal{S}} := \max_{i=1,\ldots,N} c_{E,i}$ and $c_{E,i}$ from the Inverse Trace Theorem 3.36 for $H^{1/2}(\partial\Omega_i)$.

Furthermore, we need the space $H^{div}_{0,\Gamma_N}(\Omega)$ which is defined by

$$H^{div}_{0,\Gamma_N}(\Omega) := \left\{\mathbf{q} \in H^{div}(\Omega) \mid 0 = \langle \mathbf{q}\cdot\mathbf{n}, \gamma^0_{\partial\Omega}\, v\rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} \quad \forall\, v \in V\right\}. \tag{4.31}$$

The space $H^{div}_{0,\Gamma_N}(\Omega)$ is a closed subspace of $H^{div}(\Omega)$ introduced in Section 3.2. Loosely speaking, the elements $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$ satisfy "$\mathbf{q}\cdot\mathbf{n} \equiv 0$" on Γ_N. We are now in the position to define a proper trial space for the Lagrange multiplier which is needed to obtain continuity of elements $u \in X^1(\Omega)$ on the skeleton.

Definition 4.14. *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}_\Omega^N$. Furthermore, let $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ with $|\Gamma_D|_{d-1} > 0$. The Lagrange multiplier space $Q_{0,\Gamma_N}(\mathcal{S})$ is then defined by*

$$Q_{0,\Gamma_N}(\mathcal{S}) := \left\{ \nu \in X^{1/2}(\mathcal{S})' \;\middle|\; \exists\, \mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega) : \mathbf{q} \cdot \mathbf{n}_i = \nu \text{ on } \partial\Omega_i \;\forall\, \Omega_i \in \mathcal{D}_\Omega^N \right\}$$

which is a closed subspace of the space $X^{1/2}(\mathcal{S})'$. The corresponding norm is

$$\|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} := \inf_{\substack{\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega) \\ \mathbf{q}\cdot\mathbf{n}_i = \nu \text{ on } \partial\Omega_i \\ \forall\, \Omega_i \in \mathcal{D}_\Omega^N}} \|\mathbf{q}\|_{H(div,\Omega)}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$.

The natural norm $\|\cdot\|_{Q_{0,\Gamma_N}(\mathcal{S})}$ defined on the space $Q_{0,\Gamma_N}(\mathcal{S})$ is due to its definition hard to handle. The following lemma provides an equivalence result with a more accessible norm.

Lemma 4.15. *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}_\Omega^N$. Furthermore, let $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ with $|\Gamma_D|_{d-1} > 0$. Then, there holds*

$$\frac{1}{c_{T,\mathcal{S}}} \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} \le \|\nu\|_{X^{1/2}(\mathcal{S})'} \le c_{E,\mathcal{S}} \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$, we write $\|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} \simeq \|\nu\|_{X^{1/2}(\mathcal{S})'}$.

Proof. Let $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$ and $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, such that $\mathbf{q} \cdot \mathbf{n}_i = \nu$ on $\partial\Omega_i$ for all $\Omega_i \in \mathcal{D}_\Omega^N$.

Using Trace Theorem 3.35 and Inverse Trace Theorem 3.36 for $H^1(\Omega_i)$, we obtain

$$\begin{aligned} \|\nu\|_{X^{1/2}(\mathcal{S})'} &= \sup_{0 \ne g \in X^{1/2}(\mathcal{S})} \frac{\sum_{i=1}^N \langle \nu, g_i \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|g\|_{X^{1/2}(\mathcal{S})}} \\ &= \sup_{0 \ne g \in X^{1/2}(\mathcal{S})} \frac{\sum_{i=1}^N \langle \mathbf{q}\cdot\mathbf{n}_i, \gamma^0_{\partial\Omega_i} \mathcal{E}_{\Omega_i} g \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|g\|_{X^{1/2}(\mathcal{S})}}. \end{aligned} \tag{4.32}$$

If we apply Green's formula, Lemma 3.43, for each duality pairing, we get

$$\begin{aligned} \langle \mathbf{q}\cdot\mathbf{n}_i, \gamma^0_{\partial\Omega_i} \mathcal{E}_{\Omega_i} g \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)} &= \int_{\Omega_i} \left(\mathbf{q} \cdot \nabla \mathcal{E}_{\Omega_i} g + \nabla \cdot \mathbf{q}\, \mathcal{E}_{\Omega_i} g \right) d\mathbf{x} \\ &\le \|\mathbf{q}\|_{H^{div}(\Omega_i)} \left\| \mathcal{E}_{\Omega_i} g \right\|_{H^1(\Omega_i)} \\ &\le c_{E,i} \|\mathbf{q}\|_{H^{div}(\Omega_i)} \|g\|_{H^{1/2}(\partial\Omega_i)} \end{aligned} \tag{4.33}$$

with $c_{E,i}$ from the Inverse Trace Theorem 3.36 for $H^{1/2}(\partial\Omega_i)$. If we plug in estimate (4.33) in the representation (4.32), we obtain

$$\|\nu\|_{X^{1/2}(\mathcal{S})'} \leq \sup_{0\neq g\in X^{1/2}(\mathcal{S})} \frac{\sum\limits_{i=1}^{N} c_{E,i}\,\|\mathbf{q}\|_{H^{div}(\Omega_i)}\,\|g\|_{H^{1/2}(\partial\Omega_i)}}{\|g\|_{X^{1/2}(\mathcal{S})}} \leq c_{E,\mathcal{S}}\|\mathbf{q}\|_{H^{div}(\Omega)}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$ and $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, such that $\mathbf{q}\cdot\mathbf{n}_i = \nu$ on $\partial\Omega_i$ for all $\Omega_i \in \mathcal{D}^N_\Omega$. If we take the infimum over all such $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, we get the inequality

$$\frac{1}{c_{E,\mathcal{S}}}\,\|\nu\|_{X^{1/2}(\mathcal{S})'} \leq \inf_{\substack{\mathbf{q}\in H^{div}_{0,\Gamma_N}(\Omega)\\ \mathbf{q}\cdot\mathbf{n}_i=\nu \text{ on } \partial\Omega_i\\ \forall\,\Omega_i\in\mathcal{D}^N_\Omega}} \|\mathbf{q}\|_{H^{div}(\Omega)} = \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})}$$

which proves the first inequality we have to show for the norm equivalence.

To prove the reverse inequality, we consider elements $u_i \in H^1(\Omega_i)$, satisfying

$$\int_{\Omega_i} \nabla u_i\cdot\nabla v\,\mathrm{d}\mathbf{x} + \int_{\Omega_i} u_i\,v\,\mathrm{d}\mathbf{x} = \langle \nu, \gamma^0_{\partial\Omega_i}\,v\rangle_{H^{1/2}(\partial\Omega_i)'\times H^{1/2}(\partial\Omega_i)}$$

for all $v \in H^1(\Omega_i)$. For each u_i there holds the estimate

$$\|u_i\|_{H^1(\Omega_i)} \leq c_{T,i}\,\|\nu\|_{H^{1/2}(\partial\Omega_i)'} \tag{4.34}$$

with $c_{T,i}$ from the Trace Theorem 3.35 for $H^1(\Omega_i)$.

Choose $\mathbf{w}_i = \nabla u_i$ in each Ω_i, therefore we obtain $\nabla\cdot\mathbf{w}_i = u_i$ in Ω_i, see Definition 3.31, and we have $\mathbf{w}_i\cdot\mathbf{n}_i = \nu$ on $\partial\Omega_i$. In addition, there holds

$$\|\mathbf{w}_i\|_{H^{div}(\Omega_i)} = \|u_i\|_{H^1(\Omega_i)} \leq c_{T,i}\,\|\nu\|_{H^{1/2}(\partial\Omega_i)'} \tag{4.35}$$

for all $\Omega_i \in \mathcal{D}^N_\Omega$. Define a global $\mathbf{w} \in H^{div}_{0,\Gamma_N}(\Omega)$ by $\mathbf{w}_{|\Omega_i} = \mathbf{w}_i$. If we incorporate estimate (4.35), we obtain

$$\begin{aligned}
\|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} &= \inf_{\substack{\mathbf{q}\in H^{div}_{0,\Gamma_N}(\Omega)\\ \mathbf{q}\cdot\mathbf{n}_i=\nu \text{ on } \partial\Omega_i\\ \forall\,\Omega_i\in\mathcal{D}^N_\Omega}} \|\mathbf{q}\|_{H^{div}(\Omega)} \leq \|\mathbf{w}\|_{H^{div}(\Omega)} = \left(\sum_{i=1}^{N} \|\mathbf{w}_i\|^2_{H^{div}(\Omega_i)}\right)^{1/2} \\
&\leq \left(\sum_{i=1}^{N} c^2_{T,i}\,\|\nu\|^2_{H^{1/2}(\partial\Omega_i)'}\right)^{1/2} \leq c_{T,\mathcal{S}}\,\|\nu\|_{X^{1/2}(\mathcal{S})'}
\end{aligned}$$

which completes the proof. ∎

The equivalence of the norms will play an important role in the primal hybrid framework. The following proposition is as well important but rather obvious compared to Lemma 4.15.

Proposition 4.16. *Let $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. Then $\nu \in H^{1/2}(\partial\Omega)'$ and there holds*

$$\|\nu\|_{H^{1/2}(\partial\Omega)'} \leq c_E \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})}$$

with c_E from the Inverse Trace Theorem 3.36.

Proof. Let $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$ and $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, such that $\mathbf{q} \cdot \mathbf{n}_i = \nu$ on $\partial\Omega_i$ for all $\Omega_i \in \mathcal{D}^N_\Omega$. Since $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, the normal trace of $\mathbf{q}$ on $\partial\Omega$ is well defined, that is $\mathbf{q} \cdot \mathbf{n} =: \nu \in H^{1/2}(\partial\Omega)'$. In addition, we have

$$\begin{aligned}\langle \nu, g\rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} &= \langle \mathbf{q}\cdot\mathbf{n}, g\rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} = \int_\Omega \Big(\mathbf{q} \cdot \nabla \mathcal{E}_\Omega\, g + \nabla \cdot \mathbf{q}\, \mathcal{E}_\Omega\, g\Big)\,\mathrm{d}\mathbf{x} \\ &\leq \|\mathbf{q}\|_{H^{div}(\Omega)} \|\mathcal{E}_\Omega\, g\|_{H^1(\Omega)} \leq c_E \|\mathbf{q}\|_{H^{div}(\Omega)} \|g\|_{H^{1/2}(\partial\Omega)}\end{aligned}$$

for all such $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$. Taking the infimum over all $\mathbf{q} \in H^{div}_{0,\Gamma_N}(\Omega)$, we obtain

$$\|\nu\|_{H^{1/2}(\partial\Omega)'} \leq c_E \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. ■

We now have the right spaces to work with, but we still need to realize the correct side condition for the Lagrange multiplier. In other words, we still need a condition which allows for a characterization of V in the lager space $X^1(\Omega)$. To realize such a condition, consider the bilinear form $b : X^1(\Omega) \times Q_{0,\Gamma_N}(\mathcal{S}) \to \mathbb{R}$ defined by

$$b(u, \nu) := -\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, u_{|\Omega_i}\rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}$$

for all $u \in X^1(\Omega)$ and $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. The bilinear form $b(\cdot,\cdot)$ induces an operator $B : X^1(\Omega) \to Q_{0,\Gamma_N}(\mathcal{S})'$ given by

$$\langle Bu, \nu\rangle_{Q_{0,\Gamma_N}(\mathcal{S})' \times Q_{0,\Gamma_N}(\mathcal{S})} := b(u, \nu)$$

and its adjoint operator $B' : Q_{0,\Gamma_N}(\mathcal{S}) \to X^1(\Omega)'$ defined by

$$\langle B'\nu, u\rangle_{X^1(\Omega)' \times X^1(\Omega)} := b(u, \nu)$$

for all $u \in X^1(\Omega)$ and $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. Furthermore, there holds the estimate

$$b(u,\nu) = -\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, u_{|\Omega_i} \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)} \leq c_{T,\mathcal{S}}\, \|\nu\|_{X^{1/2}(\mathcal{S})'}\, \|u\|_{X^1(\Omega)}$$

for arbitrary $u \in X^1(\Omega)$ and $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. Lemma 4.15 then implies

$$\|Bu\|_{Q_{0,\Gamma_N}(\mathcal{S})'} \leq c_{T,\mathcal{S}}\, c_{E,\mathcal{S}}\, \|u\|_{X^1(\Omega)} \quad \text{and} \quad \|B'\nu\|_{X^1(\Omega)'} \leq c_{T,\mathcal{S}}\, c_{E,\mathcal{S}}\, \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})}$$

and from Proposition 3.10 we can conclude the continuity of the operators B and B'.

Next, we want to use the operator $B : X^1(\Omega) \to Q_{0,\Gamma_N}(\mathcal{S})'$ to characterize the space V in the larger space $X^1(\Omega)$. The following lemma provides the desired condition.

Lemma 4.17. *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}^N_\Omega$ with $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ where $|\Gamma_D|_{d-1} > 0$. Then there holds*

$$V = \operatorname{Ker} B$$

with $\operatorname{Ker} B := \{u \in X^1(\Omega) \mid b(u,\nu) = 0 \;\forall\, \nu \in Q_{0,\Gamma_N}(\mathcal{S})\}$.

Proof. See for example [19, Proposition 2.1.1]. ■

Lemma 4.17 is the principal item in the primal hybrid framework. It allows the characterization of functions $u \in H^1(\Omega)$ satisfying homogeneous Dirichlet boundary conditions in the space $X^1(\Omega)$. But, we even can characterize functions $u \in H^1(\Omega)$ satisfying inhomogeneous Dirichlet boundary conditions in the space $X^1(\Omega)$.

Lemma 4.18. *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}^N_\Omega$ with $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ where $|\Gamma_D|_{d-1} > 0$. Furthermore, let $g \in H^{1/2}(\Gamma_D)$. The affine space*

$$V_g := \left\{ v \in H^1(\Omega) \mid \gamma^0_{\Gamma_D}\, v = g \text{ in } H^{1/2}(\Gamma_D) \right\}$$

can then be characterized in $X^1(\Omega)$ as the subset $\operatorname{Ker}_g B \subset X^1(\Omega)$ defined by

$$\operatorname{Ker}_g B := \left\{ u \in X^1(\Omega) \,\middle|\, b(u,\nu) = -\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g \rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} \;\forall\, \nu \in Q_{0,\Gamma_N}(\mathcal{S}) \right\}$$

which is again an affine subspace in $X^1(\Omega)$.

Proof. To prove the desired statement, we show that the affine subspaces are included in each other.

$\mathbf{V_g} \subset \mathrm{Ker}_{\mathbf{g}}\,\mathbf{B}$: First, we want to show that $V_g \subset \mathrm{Ker}_g\, B$. Let $v \in H^1(\Omega)$ such that $\gamma^0_{\Gamma_D}\, v = g$ in $H^{1/2}(\Gamma_D)$. We have $v - \mathcal{E}_\Omega\, g \in V$ and from Lemma 4.17 we obtain $b(v - \mathcal{E}_\Omega\, g, \nu) = 0$ for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. This is equivalent to

$$
\begin{aligned}
b(v,\nu) = b(\mathcal{E}_\Omega\, g,\nu) &= -\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, (\mathcal{E}_\Omega\, g)_{|\Omega_i} \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)} \\
&= -\sum_{i=1}^{N} \int_{\Omega_i} \Big(\mathbf{q}\cdot\nabla(\mathcal{E}_\Omega\, g)_{|\Omega_i} + \nabla\cdot\mathbf{q}\,(\mathcal{E}_\Omega\, g)_{|\Omega_i} \Big)\, \mathrm{d}\mathbf{x} \\
&= -\int_{\Omega} \Big(\mathbf{q}\cdot\nabla\mathcal{E}_\Omega\, g + \nabla\cdot\mathbf{q}\,\mathcal{E}_\Omega\, g \Big)\, \mathrm{d}\mathbf{x} \\
&= -\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g \rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} .
\end{aligned}
\tag{4.36}
$$

Since $H^1(\Omega) \subset X^1(\Omega)$ we see that $v \in \mathrm{Ker}_g\, B$ and therefore $V_g \subset \mathrm{Ker}_g\, B$.

$\mathrm{Ker}_{\mathbf{g}}\,\mathbf{B} \subset \mathbf{V_g}$: Next, we want to prove $\mathrm{Ker}_g\, B \subset V_g$. Therefore, let $v \in X^1(\Omega)$, such that v satisfies

$$
b(u,\nu) = -\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g \rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)}
$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. From (4.36) we obtain, that the right hand side satisfies the representation

$$
-\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g \rangle_{H^{1/2}(\partial\Omega)' \times H^{1/2}(\partial\Omega)} = b(\mathcal{E}_\Omega\, g, \nu)
$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$, thus $b(u - \mathcal{E}_\Omega\, g, \nu) = 0$. Lemma 4.17 then implies $u - \mathcal{E}_\Omega\, g =: w \in V$ and therefore $u = w + \mathcal{E}_\Omega\, g \in H^1(\Omega)$. If we apply the trace operator, we obtain $\gamma^0_{\Gamma_D}\, u = \gamma^0_{\Gamma_D}\, w + \gamma^0_{\Gamma_D}\, \mathcal{E}_\Omega\, g = 0 + g = g$. This proves the inclusion $\mathrm{Ker}_g\, B \subset V_g$ which finishes the proof. ■

The last important property we need is the inf–sup–stability of the operator $B : X^1(\Omega) \to Q_{0,\Gamma_N}(\mathcal{S})$. This stability ensures the surjectivity of B, see Theorem 3.16.

Lemma 4.19. *Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz domain with an admissible decomposition $\mathcal{D}^N_\Omega$ with $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ where $|\Gamma_D|_{d-1} > 0$. Then*

$$
\inf_{0\neq\nu\in Q_{0,\Gamma_N}(\mathcal{S})}\ \sup_{0\neq u\in X^1(\Omega)} \frac{\langle Bu,\nu\rangle_{Q_{0,\Gamma_N}(\mathcal{S})'\times Q_{0,\Gamma_N}(\mathcal{S})}}{\|\nu\|_{X^{1/2}(\mathcal{S})'}\ \|u\|_{X^1(\Omega)}} \geq \frac{1}{c_{T,\mathcal{S}}\, c_{E,\mathcal{S}}} > 0
$$

holds.

Proof. Let $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$ be an arbitrary element. Due to Lemma 4.15, we have

$$\frac{1}{c_{T,\mathcal{S}}} \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} \leq \|\nu\|_{X^{1/2}(\mathcal{S})'} = \sup_{0 \neq g \in X^{1/2}(\mathcal{S})} \frac{\sum_{i=1}^{N} \langle \nu, g \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|g\|_{X^{1/2}(\mathcal{S})}}. \tag{4.37}$$

Since $\|\mathcal{E}_\Omega\, g\|_{X^1(\Omega)} \leq c_{E,\mathcal{S}} \|g\|_{X^{1/2}(\mathcal{S})}$ for all $g \in X^{1/2}(\mathcal{S})$, we have

$$\begin{aligned}
\frac{1}{c_{E,\mathcal{S}}\, c_{T,\mathcal{S}}} \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} &\leq \sup_{0 \neq g \in X^{1/2}(\mathcal{S})} \frac{\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, \mathcal{E}_{\Omega_i}\, g \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|\mathcal{E}_\Omega\, g\|_{X^1(\Omega)}} \\
&\leq \sup_{0 \neq u \in X^1(\Omega)} \frac{\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, u_{|\Omega_i} \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|u\|_{X^1(\Omega)}}
\end{aligned}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. Since the supremum of the duality pairing taken over all $u \in X^1(\Omega)$ equals the supremum of the duality pairing taken over $-u \in X^1(\Omega)$, we obtain

$$\begin{aligned}
\frac{1}{c_{E,\mathcal{S}}\, c_{T,\mathcal{S}}} \|\nu\|_{Q_{0,\Gamma_N}(\mathcal{S})} &\leq \sup_{0 \neq u \in X^1(\Omega)} \frac{-\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, u_{|\Omega_i} \rangle_{H^{1/2}(\partial\Omega_i)' \times H^{1/2}(\partial\Omega_i)}}{\|u\|_{X^1(\Omega)}} \\
&= \sup_{0 \neq u \in X^1(\Omega)} \frac{\langle Bu, \nu \rangle_{Q_{0,\Gamma_N}(\mathcal{S})' \times Q_{0,\Gamma_N}(\mathcal{S})}}{\|u\|_{X^1(\Omega)}}
\end{aligned}$$

for all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$. Taking the infimum over all $\nu \in Q_{0,\Gamma_N}(\mathcal{S})$, we obtain the desired statement. ∎

We have derived the tools we need to split the nonlinear operator in the variational problem (4.30) into the sum of local operators to apply local Kirchhoff transformations. This is discussed in the next subsection.

4.4.2 The Primal Hybrid Formulation

In this subsection, we introduce for the rest of this thesis X and Q as

$$X := X^1(\Omega) \quad \text{and} \quad Q := Q_{0,\Gamma_N}(\mathcal{S}).$$

Furthermore, we use the notation $v_i := v_{|\Omega_i}$ for functions $v \in X$.

Using the same notation as in Section 4.2, we have the nonlinear operators $M : V \to V'$ and $S : V \to V'$ from the variational problem (4.30). The operators M and S satisfy the integral representation

$$\langle M(p), v \rangle_{V' \times V} := \int_\Omega \frac{n}{\tau}\, \theta(p + p_D)\, v \,\mathrm{d}\mathbf{x}$$

and

$$\langle S(p), v \rangle_{V' \times V} := \int_\Omega \frac{K}{\mu}\, k_\alpha\Big(\theta(p + p_D)\Big)\, \nabla(p + p_D) \cdot \nabla v \,\mathrm{d}\mathbf{x}$$

for all $p, g \in V$ and $\mathcal{E}_\Omega\, g_D =: p_D \in V_{g_D}$.

We can rewrite the operator $M : V \to V'$ by replacing the integral over Ω with the sum of integrals over Ω_i with $\Omega_i \in \mathcal{D}_\Omega^N$. That is

$$\langle M(p), v \rangle_{V' \times V} := \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau}\, \theta(p + p_D)_{|\Omega_i}\, v_{|\Omega_i} \,\mathrm{d}\mathbf{x} \tag{4.38}$$

for $p, v \in V$. Note, that the right hand side in equation (4.38) is still well defined, if we choose $p + p_D \in X$ and $v \in X$. Thus, we can introduce an extension of M which will be denoted as $\widehat{M}$ defined by

$$\langle \widehat{M}(p), v \rangle_{X' \times X} := \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau}\, \theta(p)_{|\Omega_i}\, v_{|\Omega_i} \,\mathrm{d}\mathbf{x}$$

for all $p, v \in X$. There holds $\langle \widehat{M}(p + p_D), v \rangle_{X' \times X} = \langle M(p), v \rangle_{V' \times V}$ for all $p, v \in V$. In the definition of $\widehat{M}$ we do not consider the Dirichlet extension which satisfies the inhomogeneous Dirichlet boundary condition. This condition is incorporated in a weak sense via an additional condition. Due to the special structure of the nonlinear function θ, see (2.7), and due to the choice of the domain decomposition $\mathcal{D}_\Omega^N$, we can write $\theta(p)_{|\Omega_i}$ as $\theta_i(p_i)$ within each subdomain $\Omega_i \in \mathcal{D}_\Omega^N$. Thus, $\widehat{M}$ can be written as

$$\langle \widehat{M}(p), v \rangle_{X' \times X} = \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau}\, \theta_i(p_i)\, v_i \,\mathrm{d}\mathbf{x}$$

for $p, v \in X^1(\Omega)$. In the same manner, we can extend $S : V \to V'$ to an operator $\widehat{S} : X \to X'$ defined by

$$\langle \widehat{S}(p), v \rangle_{X' \times X} := \sum_{i=1}^{N} \int_{\Omega_i} \frac{K}{\mu}\, k_\alpha\Big(\theta(p)\Big)_{|\Omega_i}\, \nabla p_{|\Omega_i} \cdot \nabla v_{|\Omega_i} \,\mathrm{d}\mathbf{x}$$

for $p, v \in X$. Since $\mathcal{D}_\Omega^N$ resolves the natural decomposition, see Figure 4.9b, and due to the structure of θ and k_α, we obtain $k_\alpha\big(\theta(p)\big)_{|\Omega_i} = k_{\alpha,i}\big(\theta_i(p_i)\big)$ within each $\Omega_i \in \mathcal{D}_\Omega^N$. Hence, $\widehat{S}$ can be written as

$$\langle \widehat{S}(p), v\rangle_{X'\times X} = \sum_{i=1}^{N} \int\limits_{\Omega_i} \frac{K}{\mu}\, k_{\alpha,i}\big(\theta_i(p_i)\big)\, \nabla p_i \cdot \nabla v_i \,\mathrm{d}\mathbf{x}$$

for $p, v \in X$. Furthermore, there holds $\langle \widehat{S}(p+p_D), v\rangle_{X'\times X} = \langle S(p), v\rangle_{V'\times V}$ for all $p, v \in V$. We dropped the Dirichlet extension in the definition of $\widehat{M}$ and $\widehat{S}$. As already mentioned, we will incorporate the inhomogeneous Dirichlet boundary condition in a weak sense as an additional condition. For this reason define $G \in Q'$ by the duality pairing

$$\langle G, \nu\rangle_{Q'\times Q} := -\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g_D\rangle_{H^{1/2}(\partial\Omega)'\times H^{1/2}(\partial\Omega)} \tag{4.39}$$

for all $\nu \in Q$. Proposition 4.16 implies the estimate

$$\begin{aligned} \|G\|_{Q'} &= \sup_{0\neq\nu\in Q} \frac{\langle G, \nu\rangle_{Q'\times Q}}{\|\nu\|_Q} = \sup_{0\neq\nu\in Q} \frac{-\langle \nu, \gamma^0_{\partial\Omega}\, \mathcal{E}_\Omega\, g\rangle_{H^{1/2}(\partial\Omega)'\times H^{1/2}(\partial\Omega)}}{\|\nu\|_Q} \\ &\leq c_T\, c_E \sup_{0\neq\nu\in Q} \frac{\|\nu\|_{H^{1/2}(\partial\Omega)'}\, \|g_D\|_{H^{1/2}(\Gamma_D)}}{\|\nu\|_Q} \leq c_T\, c_E^2\, \|g_D\|_{H^{1/2}(\Gamma_D)} \end{aligned}$$

which shows, that G is indeed an element in Q'. Next, take a closer look at the functional $F \in V'$ which is defined by

$$\langle F, v\rangle_{V'\times V} := \int\limits_\Omega \Big(f + \frac{n}{\tau}\,\theta(q)\Big)\, v \,\mathrm{d}\mathbf{x} + \int\limits_\Omega \frac{K}{\mu}\, k_\alpha\big(\theta(q)\big)\, \nabla d \cdot \nabla v \,\mathrm{d}\mathbf{x} + \int\limits_{\Gamma_N} g_N\, \gamma^0_{\Gamma_N}\, v \,\mathrm{d}s_{\mathbf{x}}$$

for all $v \in V$. Since f and g_N are assumed to be square integrable, that is $f \in L_2(\Omega)$ and $g_N \in L_2(\Gamma_N)$, we see, that F is even an element in the smaller space $X' \subset V'$. As in the proof of Theorem 4.3 we obtain the estimate

$$\|F\|_{X'} \leq \|f\|_{L_2(\Omega)} + \frac{c_M}{\tau}\, \|\theta(q)\|_{L_2(\Omega)} + c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_2(\Omega)} + c_T\, \|g_N\|_{L_2(\Gamma_N)}.$$

We can rewrite the variational problem (4.30) using the previous defined operators and we obtain the following abstract variational problem.

Primal hybrid formulation for heterogeneous soil

Find $p \in X$ and $\lambda \in Q$ such that

$$\begin{aligned} \langle \widehat{M}(p) + \widehat{S}(p), v \rangle_{X' \times X} + \langle B'\lambda, v \rangle_{X' \times X} &= \langle F, v \rangle_{X' \times X} \\ \langle Bp, \nu \rangle_{Q' \times Q} &= \langle G, \nu \rangle_{Q' \times Q} \end{aligned} \tag{4.40}$$

for each $v \in X$ and $\nu \in Q$.

Using the results from Section 4.4.1 we can prove the following statement, which shows the equivalence of the variational problem (4.30) and the primal hybrid formulation (4.40).

Lemma 4.20. *The variational problem (4.30) is uniquely solvable, iff the variational problem (4.40) is uniquely solvable with $G \in Q'$ defined by (4.39) and $F \in X'$.*

For the solution $(p, \lambda) \in X \times Q$ to the variational problem (4.40) there holds

$$\|p\|_X \leq \widehat{c}_1 \left(\|f\|_{L_2(\Omega)} + \frac{1}{\tau} \|q - p_D\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|g_D\|_{H^{1/2}(\Gamma_D)} + \|g_N\|_{L_2(\Gamma_N)} \right)$$

and

$$\|\lambda\|_Q \leq \widehat{c}_2 \left(\|f\|_{L_2(\Omega)} + \frac{1}{\tau} \|q - p\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|p\|_{H^1(\Omega)} + \|g_N\|_{L_2(\Gamma_N)} \right)$$

with some positive constants $\widehat{c}_1(\Omega, \theta, k, K, n, \mu)$ and $\widehat{c}_2(\Omega, \theta, k, K, n, \mu)$.

Proof. To show the equivalence, we show that the unique solvability of the variational problem (4.30) implies unique solvability of the primal hybrid formulation (4.40) and vice versa.

(4.30) ⇒ (4.40): First, assume $p \in V$ is the unique solution to the variational problem (4.30), that is

$$\langle M(p) + S(p), v \rangle_{V' \times V} = \langle F, v \rangle_{V' \times V} \tag{4.41}$$

for all $v \in V$. We want to prove, that $p + p_D =: \widetilde{p} \in X$ is the unique solution to the variational problem (4.40).

From Lemma 4.19 and Theorem 3.16, we conclude the existence of an operator $B^\dagger \in \mathcal{L}(Q', X)$ such that $B \circ B^\dagger = I$ on Q'. This implies $\operatorname{Im} B = Q'$. The Closed Range Theorem, see [72, Section VII.5], implies $\operatorname{Im} B' = (\operatorname{Ker} B)^\circ$ with

$$(\operatorname{Ker} B)^\circ := \left\{ F \in X' \mid 0 = \langle F, v \rangle_{X' \times X} \ \forall v \in \operatorname{Ker} B \right\}.$$

That is, for each $F \in (\operatorname{Ker} B)^\circ$ there exists a $\lambda \in Q$ such that $B'\lambda = F$ in X'.

Next, define $\overline{F} \in X'$ by

$$\langle \overline{F}, v\rangle_{X'\times X} := \langle F - \widehat{M}(\widetilde{p}) - \widehat{M}(\widetilde{p}), v\rangle_{X'\times X}$$

for all $v \in X$. From equation (4.41) and Lemma 4.17 we have $\overline{F} \in (\operatorname{Ker} B)^\circ$ and due to our previous considerations there exists a $\lambda \in Q$, such that

$$\langle B'\lambda, v\rangle_{X'\times X} = \langle \overline{F}, v\rangle_{X'\times X} = \langle F - \widehat{M}(\widetilde{p}) - \widehat{M}(\widetilde{p}), v\rangle_{X'\times X} \tag{4.42}$$

holds for all $v \in X$.

From Lemma 4.18 we conclude that

$$\langle B\widetilde{p}, \nu\rangle_{Q'\times Q} = \langle G, \nu\rangle_{Q'\times Q}$$

holds for all $\nu \in Q$. Thus $\widetilde{p} \in X$ and $\lambda \in Q$ is a solution to the primal hybrid formulation (4.40).

It still remains to prove the uniqueness of the solution. The uniqueness of $\widetilde{p} \in X$ follows from the uniqueness of $p \in V$. Next, assume we have two solutions $\lambda_1, \lambda_2 \in Q$. From Lemma 4.19 we obtain

$$\begin{aligned}\frac{1}{c_{T,\mathcal{S}}\, c_{E,\mathcal{S}}} \|\lambda_1 - \lambda_2\|_Q &\le \sup_{0\neq v\in X} \frac{\langle Bv, \lambda_1 - \lambda_2\rangle_{Q'\times Q}}{\|v\|_X} \\ &= \sup_{0\neq v\in X} \frac{\langle v, B'\lambda_1 - B'\lambda_2\rangle_{X'\times X}}{\|v\|_X}\end{aligned}$$

which is zero due to (4.42) and so $\lambda_1 = \lambda_2$. Therefore, $\widetilde{p} \in X$ and $\lambda \in Q$ are the unique solution to the primal hybrid formulation(4.40).

(4.40) $\Rightarrow$ (4.30): To show the reverse implication, let us assume we have a unique solution $\widetilde{p} \in X$ and $\lambda \in Q$ to the primal hybrid formulation (4.40). Since $\widetilde{p} \in X$ is a solution, $\widetilde{p}$ satisfies

$$\langle B\widetilde{p}, \nu\rangle_{Q'\times Q} = \langle G, \nu\rangle_{Q'\times Q}$$

for all $\nu \in Q$. From Lemma 4.18 we conclude $\widetilde{p} \in V_{g_D}$. Next, we define $p \in V$ by $p := \widetilde{p} - p_D$. From the first equation in the primal hybrid formulation (4.40), we obtain

$$\langle \widehat{M}(p + p_D) + \widehat{S}(p + p_D), v\rangle_{X'\times X} + \langle B'\lambda, v\rangle_{X'\times X} = \langle F, v\rangle_{X'\times X}$$

for all $v \in X$. If we choose an arbitrary test function $v \in V \subset X$, we obtain

$$\langle B'\lambda, v\rangle_{X'\times X} = \langle \lambda, Bv\rangle_{X'\times X} = 0$$

and therefore

$$\langle M(p) + S(p), v\rangle_{V'\times V} = \langle F, v\rangle_{V'\times V}$$

for each $v \in V$. The uniqueness of $p \in V$ follows from the uniqueness of $\widetilde{p} \in X$.

Boundedness of the solution: It remains to show the desired estimates for the solution to the primal hybrid formulation (4.40). First, we consider $\widetilde{p} \in X$. From Lemma 4.18 we conclude, that $\widetilde{p} \in H^1(\Omega)$ and there holds $\|\widetilde{p}\|_X = \|\widetilde{p}\|_{H^1(\Omega)}$. The triangle inequality for $\widetilde{p} = p + p_D$ and the Inverse Trace Theorem 3.36 yield

$$\|\widetilde{p}\|_{H^1(\Omega)} = \|p + p_D\|_{H^1(\Omega)} \leq \|p\|_{H^1(\Omega)} + c_E \|g_D\|_{H^{1/2}(\Gamma_D)}.$$

Since $p \in H^1(\Omega)$ is a solution to the variational problem (4.30), p satisfies the estimate in Corollary 4.11 and furthermore the sharper estimate in the proof of Theorem 4.3. Therefore, we obtain the upper bound

$$\begin{aligned} c_N^2\, c_s\, c_{\alpha,k} \|\widetilde{p}\|_X \leq \|f\|_{L_2(\Omega)} &+ \frac{c_M}{\tau}\, c_{L,\theta} \|q - p_D\|_{L_2(\Omega)} \\ &+ c_S\, b_k \|\nabla d\|_{\mathbf{L}_2(\Omega)} + c_T \|g_N\|_{L_2(\Gamma_N)} \\ &+ c_E \left(c_S\, b_k + c_N^2\, c_s\, c_{\alpha,k}\right) \|g_D\|_{H^{1/2}(\Gamma_D)} \end{aligned}$$

and by setting $\widehat{c}_1 := \frac{1}{c_N^2\, c_s\, c_{\alpha,k}} \max\{1, c_M\, c_{L,\theta}, c_S\, b_k, c_T, c_E\, (c_S\, b_k + c_N^2\, c_s\, c_{\alpha,k})\}$, we get the desired estimate for the norm of $\widetilde{p}$.

Next, we want to estimate $\lambda \in Q$. The stability condition in Lemma 4.19 and the identity (4.42) gives

$$\begin{aligned} \frac{1}{c_{T,\mathcal{S}}\, c_{E,\mathcal{S}}} \|\lambda\|_Q \leq \sup_{0 \neq v \in X} \frac{\langle Bv, \lambda\rangle_{Q' \times Q}}{\|v\|_X} &= \sup_{0 \neq v \in X} \frac{\langle B'\lambda, v\rangle_{X' \times X}}{\|v\|_X} \\ &= \sup_{0 \neq v \in X} \frac{\langle F - \widehat{M}(\widetilde{p}) - \widehat{S}(\widetilde{p}), v\rangle_{X' \times X}}{\|v\|_X}. \end{aligned}$$

The numerator in the right hand side of the above estimate satisfies the following representation

$$\begin{aligned} \langle F - \widehat{M}(\widetilde{p}) - \widehat{S}(\widetilde{p}), v\rangle_{X' \times X} &= \sum_{i=1}^N \int_{\Omega_i} \left\{ f_i + \frac{n}{\tau}\Big(\theta_i(q_i) - \theta_i(\widetilde{p}_i)\Big)\right\} v_i \,\mathrm{d}\mathbf{x} + \\ &+ \sum_{i=1}^N \int_{\Omega_i} \frac{K}{\mu} \Big\{ k_{\alpha,i}\Big(\theta_i(q_i)\Big) \nabla d_i - k_{\alpha,i}\Big(\theta_i(\widetilde{p}_i)\Big) \nabla \widetilde{p}_i \Big\} \cdot \nabla v_i \,\mathrm{d}\mathbf{x} + \\ &+ \sum_{i=1}^N \int_{\Gamma_{N,i}} g_{N,i}\, \gamma^0_{\Gamma_{N,i}}\, v_i \,\mathrm{d}s_{\mathbf{x}} \end{aligned}$$

for all $v \in X$. In the same manner as in the proof of Theorem 4.3, we obtain the estimate

$$\frac{\langle F - \widehat{M}(\widetilde{p}) - \widehat{S}(\widetilde{p}), v\rangle_{X' \times X}}{\|v\|_X} \leq \|f\|_{L_2(\Omega)} + \frac{c_M}{\tau}\, c_{L,\theta}\, \|q - \widetilde{p}\|_{L_2(\Omega)} + \\ + c_S\, b_k\, \|\nabla d\|_{\mathbf{L}_2(\Omega)} + c_S\, b_k\, \|\widetilde{p}\|_{H^1(\Omega)} + c_{T,\mathcal{S}}\, \|g_N\|_{L_2(\Gamma_N)}$$

since $p + p_D = \widetilde{p}$ is an element in $H^1(\Omega)$. If we define the constant $\widehat{c}_2$ by $\widehat{c}_2 := c_{T,\mathcal{S}}\, c_{E,\mathcal{S}} \max\{1, c_M\, c_{L,\theta}, c_S\, b_k, c_S\, b_k, c_{T,\mathcal{S}}\}$, we get

$$\|\lambda\|_Q \leq \widehat{c}_2 \Big(\|f\|_{L_2(\Omega)} + \frac{1}{\tau}\|q - \widetilde{p}\|_{L_2(\Omega)} + \|\nabla d\|_{\mathbf{L}_2(\Omega)} + \|\widetilde{p}\|_{H^1(\Omega)} + \|g_N\|_{L_2(\Gamma_N)} \Big)$$

which proves the desired bound. ■

As desired, the operators $\widehat{M} : X \to X'$ and $\widehat{S} : X \to X'$ can be written as the sum of local acting operators and the corresponding variational problem (4.40) is equivalent to the variational problem (4.30) which was our starting point. The next step is to apply the Kirchhoff transformation to the local acting operators. For this reason we recall the definition of the operator $\widehat{S} : X \to X'$, which is defined by

$$\langle \widehat{S}(p), v\rangle_{X' \times X} = \sum_{i=1}^{N} \int_{\Omega_i} \frac{K}{\mu}\, k_{\alpha,i}\Big(\theta_i(p_i)\Big)\, \nabla p_i \cdot \nabla v_i \, \mathrm{d}\mathbf{x}$$

for all $v \in X$. Due to the special choice of the decomposition $\mathcal{D}_\Omega^N$, the nonlinearities just depend on the unknown p within each subdomain Ω_i, that is

$$k_{\alpha,i}\Big(\theta_i(s)\Big)(\mathbf{x}) = k_{\alpha,i}\Big(\theta_i(s)\Big)$$

independent of $\mathbf{x} \in \Omega_i$. As done in the previous Section 4.3, we can define the mapping $\kappa_i : \mathbb{R} \to \mathbb{R}$ for each $i = 1, \dots, N$ as

$$\kappa_i(s) := \int_0^s \kappa_{\alpha,i}\Big(\theta_i(r)\Big)\, \mathrm{d}r$$

for all $s \in \mathbb{R}$. Each κ_i induces a superposition operator which acts on $H^1(\Omega_i)$ continuously, see Theorem 3.49. Furthermore, Lemma 4.9 remains true for κ_i as an operator from $H^1(\Omega_i)$ to $H^1(\Omega_i)$.

Thus, we can introduce new local functions $u_i \in H^1(\Omega_i)$ as the Kirchhoff transformations of p_i, that is $u_i := \kappa_i(p_i)$. As in the homogeneous case, the gradient of u_i satisfies

$$\nabla u_i = \nabla \kappa_i(p_i) = k_{\alpha,i}\Big(\theta_i(p_i)\Big)\, \nabla p_i \tag{4.43}$$

in $\mathbf{L}_2(\Omega_i)$, see Theorem 3.49.

From Lemma 4.9 we conclude that

$$\|u_i\|_{H^1(\Omega_i)} \leq c_{B,k_{\alpha,i}} \|p_i\|_{H^1(\Omega_i)}$$

with $c_{B,k_{\alpha,i}} := \|k_{\alpha,i}\|_{L_\infty(\mathbb{R})}$ for all $i = 1, \ldots, N$. If we set $c_{B,k_\alpha} := \max\limits_{i=1,\ldots,N} c_{B,k_{\alpha,i}}$ we obtain

$$\|u\|_X = \left(\sum_{i=1}^{N} \|u_i\|^2_{H^1(\Omega_i)} \right)^{1/2} \leq \left(\sum_{i=1}^{N} c^2_{B,k_{\alpha,i}} \|p_i\|^2_{H^1(\Omega_i)} \right)^{1/2} \leq c_{B,k_\alpha} \|p\|_X$$

and therefore $u \in X$.

Analogously, we conclude from Lemma 4.9 the existence of the continuous and bounded inverse operators $\kappa_i^{-1} : H^1(\Omega_i) \to H^1(\Omega_i)$ and there holds

$$\|p\|_X = \left(\sum_{i=1}^{N} \|p_i\|^2_{H^1(\Omega_i)} \right)^{1/2} \leq \left(\sum_{i=1}^{N} \frac{1}{\alpha^2} \|u_i\|^2_{H^1(\Omega_i)} \right)^{1/2} \leq \frac{1}{\alpha} \|u\|_X$$

for $u \in X$ defined as above.

As done in the homogeneous case, we want to rewrite the variational problem (4.40) in terms of $u \in X$. For this reason we consider each term separately.

First consider the nonlinear operator $\widehat{M} : X \to X'$. Using the representation $p_i := \kappa_i^{-1}(u_i)$ within each Ω_i, we obtain

$$\begin{aligned} \langle \widehat{M}(p), v \rangle_{X' \times X} &= \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau} \, \theta_i(p_i) \, v_i \, \mathrm{d}\mathbf{x} = \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau} \, \theta_i\Big(\kappa_i^{-1}(u_i)\Big) \, v_i \, \mathrm{d}\mathbf{x} \\ &= \sum_{i=1}^{N} \int_{\Omega_i} \frac{n}{\tau} \, l_i(u_i) \, v_i \, \mathrm{d}\mathbf{x} =: \langle L(u), v \rangle_{X' \times X} \end{aligned}$$

with $l_i := \theta_i \circ \kappa_i^{-1}$ for $i = 1, \ldots, N$. Next, consider the operator $\widehat{S} : X \to X'$. We use the representation (4.43) for the gradient and obtain

$$\begin{aligned} \langle \widehat{S}(p), v \rangle_{X' \times X} &= \sum_{i=1}^{N} \int_{\Omega_i} \frac{K}{\mu} \, k_{\alpha,i}\Big(\theta_i(p_i)\Big) \, \nabla p_i \cdot \nabla v_i \, \mathrm{d}\mathbf{x} \\ &= \sum_{i=1}^{N} \int_{\Omega_i} \frac{K}{\mu} \, \nabla u_i \cdot \nabla v_i \, \mathrm{d}\mathbf{x} =: \langle Au, v \rangle_{X' \times X} \end{aligned}$$

which is now a linear operator $A : X \to X'$. Unfortunately we have to rewrite the linear coupling condition $\langle Bp, \nu \rangle_{Q' \times Q}$. Since $\kappa_i : \mathbb{R} \to \mathbb{R}$ satisfies the conditions of Lemma 3.51, we have

$$\begin{aligned} \langle Bp, \nu \rangle_{Q' \times Q} &= -\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, p_i \rangle_{Q' \times Q} = -\sum_{i=1}^{N} \langle \nu, \gamma^0_{\partial\Omega_i}\, \kappa_i^{-1}(u_i) \rangle_{Q' \times Q} \\ &= -\sum_{i=1}^{N} \langle \nu, \kappa_i^{-1}(\gamma^0_{\partial\Omega_i}\, u_i) \rangle_{Q' \times Q} =: \langle C(u), \nu \rangle_{Q' \times Q} \end{aligned}$$

for all $\nu \in Q$.

Thus we were able to transform the nonlinear variational problem (4.40) into the variational problem (4.44) with a linear principal part, but with nonlinear coupling conditions.

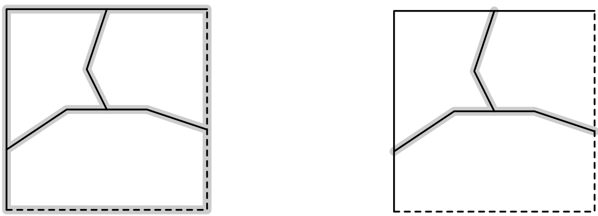
Transformed primal hybrid formulation for heterogeneous soil
Find $u \in X$ and $\lambda \in Q$ such that $$\begin{aligned} \langle L(u) + Au, v \rangle_{X' \times X} + \langle B'\lambda, v \rangle_{X' \times X} &= \langle F, v \rangle_{X' \times X} \\ \langle C(u), \nu \rangle_{Q' \times Q} &= \langle G, \nu \rangle_{Q' \times Q} \end{aligned} \tag{4.44}$$ for all $v \in X$ and $\nu \in Q$.

Since the operators induced by the local Kirchhoff transformation are isomorphisms, we obtain unique solvability of the variational problem (4.44) from the unique solvability of the variational problem (4.40).

Some words to recap the work done so far. In Section 4.1 we applied an implicit–explicit time discretization scheme and derived the corresponding variational problem (4.4). The next step was to investigate the unique solvability of the variational problem (4.4). The results were presented in Section 4.2 for general nonlinearities satisfying certain conditions, see Assumption 4.2. Next, in Section 4.3, we considered the case of a homogeneous soil with nonlinearities introduced in Section 2.1. We made sure that this special choice of nonlinearities fulfills the conditions to ensure unique solvability. In addition, we applied the Kirchhoff transformation and obtained the transformed variational problem (4.29) with a linear principal part. In the last section, Section 4.4, we considered the case of a heterogeneous soil with nonlinearities discussed in Section 2.2. As in Section 4.3, we were able to show unique solvability. To apply the Kirchhoff transformation we had to do some additional work. With the help of the primal hybrid formulation it was possible to apply local Kirchhoff transformations and we obtained the transformed variational problem (4.44). In the next chapter we want to discuss discretization and linearization techniques since the problem is still a nonlinear problem. We will also analyze the discrete linearized variational problem in view of solvability and uniqueness.

5 LINEARIZATION, DISCRETIZATION AND IMPLEMENTATION

In this chapter we want to discuss discretiaztion and linearization strategies for the solution of the variational problem (4.44). We restrict ourselves to the two and three dimensional case, that is $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$. As already mentioned in the introduction, see Chapter 1, we want to apply the mortar finite element method to discretize the variational problem (4.44). In contrast to the derived primal hybrid formulation, for which the Lagrange multiplier is defined on the entire skeleton $\mathcal{S}$, the discrete Lagrange multiplier of the mortar finite element method is just defined on the inner skeleton $\mathcal{S} \backslash \partial\Omega$, see Figure 5.1 for $d = 2$. To fit the primal hybrid formulation into



(a) Primal hybrid formulation. (b) Mortar finite element method.

Figure 5.1: Domain of definition of the Lagrange multiplier.

the context of the mortar finite element method, we make the following modification. Assume, we have given discrete trial spaces $\widetilde{X}_h$ and $\widetilde{Q}_h$ which satisfy the inclusion

$$\widetilde{X}_h \subset X \cap \prod_{i=1}^{N} \mathcal{C}^0(\overline{\Omega}_i) \quad \text{and} \quad \widetilde{Q}_h \subset Q \cap L_2(\mathcal{S})$$

for some discretization parameter h. For elements $u_h \in \widetilde{X}_h$ we write again $u_{h,i}$ for $u_{h|_{\Omega_i}}$ and since $u_{h,i} \in \mathcal{C}^0(\overline{\Omega}_i)$ we write $\gamma^0_{\partial\Omega_i} u_{h,i}$ as $u_{h,i|_{\partial\Omega_i}}$. Due to the definition of Q, see (4.31), we conclude that every $\nu_h \in \widetilde{Q}_h$ satisfies $\nu_{h|_{\Gamma_N}} \equiv 0$. Next, suppose there exist elements $u_{D,h} \in \widetilde{X}_h$, such that

$$\left(\kappa_i^{-1}(u_{D,h,i|_{\Gamma_{D,i}}}), \nu_{h|_{\Gamma_{D,i}}}\right)_{L_2(\Gamma_{D,i})} = \left(\gamma^0_{\Gamma_{D,i}} (\mathcal{E}_\Omega\, g_D)_{|\Omega_i}, \nu_{h|_{\Gamma_{D,i}}}\right)_{L_2(\Gamma_{D,i})} \tag{5.1}$$

for all $\nu_h \in \widetilde{Q}_h$ and for all Dirichlet boundary parts $\Gamma_{D,i} := \Gamma_D \cap \partial\Omega_i$ with $\Gamma_{D,i} \neq \emptyset$. Since the discrete trial and test space for the Lagrange multiplier is assumed to be a subspace of $L_2(\mathcal{S})$, we can rewrite the linear coupling condition in the variational problem (4.44) as

$$\begin{aligned}\left\langle B'\nu_h, v_h\right\rangle_{X'\times X} &= -\sum_{i=1}^{N} \left(\nu_h, v_{h,i|_{\partial\Omega_i}}\right)_{L_2(\partial\Omega_i)} = \\ &= -\sum_{\substack{\Gamma_{ij}\neq\emptyset \\ i<j}} \left(\nu_h, v_{h,i|_{\Gamma_{ij}}} - v_{h,j|_{\Gamma_{ij}}}\right)_{L_2(\Gamma_{ij})} - \sum_{\Gamma_{D,i}\neq\emptyset} \left(\nu_h, v_{h,i|_{\Gamma_{D,i}}}\right)_{L_2(\Gamma_{D,i})}\end{aligned}$$

for $v_h \in \widetilde{X}_h$ and $\nu_h \in \widetilde{Q}_h$. Here, according to Definition 4.12, the interfaces $\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j$ are defined for $1 \le i, j \le N$, $i \neq j$. In the above splitting we neglect the $L_2(\Gamma_{N,i})$–inner product on the Neumann boundary parts since $\nu_h \in \widetilde{Q}_h$ is assumed to vanish on Γ_N. It is of course possible to sum over all interfaces Γ_{ij} with $i > j$, but since $\Gamma_{ij} = \Gamma_{ji}$ we just have to ensure that each interface is counted once.

We can further rewrite the linear coupling condition in terms of jumps across the interfaces Γ_{ij} as

$$\left\langle B'\nu_h, v_h\right\rangle_{X'\times X} = -\sum_{\substack{\Gamma_{ij}\neq\emptyset \\ i<j}} \left(\nu_h, [\![v_h]\!]_{\Gamma_{ij}}\right)_{L_2(\Gamma_{ij})} - \sum_{\Gamma_{D,i}\neq\emptyset} \left(\nu_h, v_{h,i|_{\Gamma_{D,i}}}\right)_{L_2(\Gamma_{D,i})}$$

with the jump

$$[\![v_h]\!]_{\Gamma_{ij}} := v_{h,i|_{\Gamma_{ij}}} - v_{h,j|_{\Gamma_{ij}}}$$

for all $v_h \in \widetilde{X}_h$ and $\nu_h \in \widetilde{Q}_h$. In the same manner we can rewrite the nonlinear coupling term as

$$\begin{aligned}\left\langle C(v_h), \nu_h\right\rangle_{Q'\times Q} &= -\sum_{i=1}^{N} \left(\nu_h, \kappa_i^{-1}(v_{h,i|_{\partial\Omega_i}})\right)_{L_2(\partial\Omega_i)} = \\ &= -\sum_{\substack{\Gamma_{ij}\neq\emptyset \\ i<j}} \left(\nu_h, [\![\kappa^{-1}(v_h)]\!]_{\Gamma_{ij}}\right)_{L_2(\Gamma_{ij})} - \sum_{\Gamma_{D,i}\neq\emptyset} \left(\nu_h, \kappa_i^{-1}(v_{h,i|_{\Gamma_{D,i}}})\right)_{L_2(\Gamma_{D,i})}\end{aligned}$$

for all $v_h \in \widetilde{X}_h$ and $\nu_h \in \widetilde{Q}_h$. The nonlinear jump across the interfaces Γ_{ij} is defined by

$$[\![\kappa^{-1}(v_h)]\!]_{\Gamma_{ij}} := \kappa_i^{-1}(v_{h,i|_{\Gamma_{ij}}}) - \kappa_j^{-1}(v_{h,j|_{\Gamma_{ij}}})$$

for all $v_h \in \widetilde{X}_h$.

Next, we introduce the subspace $X_h \subset \widetilde{X}_h$ as the space of all $v_h \in \widetilde{X}_h$ satisfying homogeneous Dirichlet boundary conditions, that is

$$X_h := \left\{u_h \in \widetilde{X}_h \mid u_{h|_{\Gamma_D}} = 0\right\}.$$

Using the space X_h, we can consider elements $\tilde{u}_h \in \widetilde{X}_h$ defined by $\tilde{u}_h := u_h + u_{D,h}$ with $u_h \in X_h$. Thus, we obtain

$$\begin{aligned}\left\langle C(u_h + u_{D,h}), \nu_h \right\rangle_{Q' \times Q} = &- \sum_{\substack{\Gamma_{ij} \neq \emptyset \\ i<j}} \left(\nu_h, [\![\kappa^{-1}(u_h + u_{D,h})]\!]_{\Gamma_{ij}}\right)_{L_2(\Gamma_{ij})} \\ &- \sum_{\Gamma_{D,i} \neq \emptyset} \left(\nu_h, \kappa_i^{-1}(u_{h|_{\Gamma_{D,i}}} + u_{D,h|_{\Gamma_{D,i}}})\right)_{L_2(\Gamma_{D,i})}\end{aligned}$$

and since $u_h \in X_h$ vanishes on Γ_D, we have

$$\begin{aligned}\left\langle C(u_h + u_{D,h}), \nu_h \right\rangle_{Q' \times Q} = &- \sum_{\substack{\Gamma_{ij} \neq \emptyset \\ i<j}} \left(\nu_h, [\![\kappa^{-1}(u_h + u_{D,h})]\!]_{\Gamma_{ij}}\right)_{L_2(\Gamma_{ij})} \\ &- \sum_{\Gamma_{D,i} \neq \emptyset} \left(\nu_h, \kappa_i^{-1}(u_{D,h|_{\Gamma_{D,i}}})\right)_{L_2(\Gamma_{D,i})}\end{aligned}$$

for all $\nu_h \in \widetilde{Q}_h$. Since the function $u_{D,h} \in \widetilde{X}_h$ satisfies (5.1), we can rewrite the second sum in the right hand side as

$$\begin{aligned}\sum_{\Gamma_{D,i} \neq \emptyset} \left(\nu_h, \kappa_i^{-1}(u_{D,h,i|_{\Gamma_{D,i}}})\right)_{L_2(\Gamma_{D,i})} &= \sum_{\Gamma_{D,i} \neq \emptyset} \left(\nu_h, \gamma^0_{\Gamma_{D,i}}\left(\mathcal{E}_\Omega\, g_D\right)_{|\Omega_i}\right)_{L_2(\Gamma_{D,i})} = \\ = \left(\nu_h, \gamma^0_{\Gamma_D}\left(\mathcal{E}_\Omega\, g_D\right)\right)_{L_2(\Gamma_D)} &= \left(\nu_h, \gamma^0_{\partial\Omega}\left(\mathcal{E}_\Omega\, g_D\right)\right)_{L_2(\partial\Omega)} = \left\langle G, \nu_h \right\rangle_{Q' \times Q}\end{aligned}$$

for all $\nu_h \in \widetilde{Q}_h$ and $G \in Q'$ defined by (4.39). We see, that $u_h + u_{D,h}$ with $u_h \in X_h$ satisfies the inhomogeneous Dirichlet boundary conditions, we just have to ensure that $u_h + u_{D,h}$ satisfies the nonlinear jump across the interface, that is

$$- \sum_{\substack{\Gamma_{ij} \neq \emptyset \\ i<j}} \left(\nu_h, [\![\kappa^{-1}(u_h + u_{D,h})]\!]_{\Gamma_{ij}}\right)_{L_2(\Gamma_{ij})} = 0$$

for all $\nu_h \in \widetilde{Q}_h$. From the continuous variational problem we know, that testing with functions $v \in V$ which vanish on Γ_D is sufficient to ensure that the partial differential equation is satisfied in the domain. For this reason, we just test with functions $v_h \in X_h$ in the discrete variational problem. Thus, the values of λ_h on Γ_D are negligible.

If we restrict ourselves to the spaces X_h and Q_h, which is defined by

$$Q_h := \left\{ v_{h|_{\mathcal{S} \setminus \partial\Omega}} \mid v_h \in \widetilde{Q}_h \right\},$$

we obtain the following modified variational problem.

Modified discrete variational problem

Find $u_h \in X_h$ and $\lambda_h \in Q_h$ such that

$$\begin{aligned} m(u_h + u_{D,h}, v_h) + a(u_h, v_h) + b(v_h, \lambda_h) &= f(v_h) - a(u_{D,h}, v_h) \\ c(u_h + u_{D,h}, \nu_h) &= 0 \end{aligned} \tag{5.2}$$

for each $v_h \in X_h$ and $\nu_h \in Q_h$.

The coupling forms in the above variational problem are given by

$$b(v_h, \lambda_h) := - \sum_{\substack{\Gamma_{ij} \neq \emptyset \\ i<j}} \int_{\Gamma_{ij}} [\![v_h]\!]_{\Gamma_{ij}} \lambda_h \, \mathrm{d}s_{\mathbf{x}},$$

$$c(u_h + u_{D,h}, \nu_h) := - \sum_{\substack{\Gamma_{ij} \neq \emptyset \\ i<j}} \int_{\Gamma_{ij}} [\![\kappa^{-1}(u_h + u_{D,h})]\!]_{\Gamma_{ij}} \nu_h \, \mathrm{d}s_{\mathbf{x}}$$

and for the linear and bilinear forms we have

$$\begin{aligned} m(u_h + u_{D,h}, v_h) &:= \sum_{i=1}^{N} \int_{\Omega_i} l_i(u_{h,i} + u_{D,h,i}) \, v_{h,i} \, \mathrm{d}\mathbf{x}, \\ a(u_h, v_h) &:= \sum_{i=1}^{N} \int_{\Omega_i} \nabla u_{h,i} \cdot \nabla v_{h,i} \, \mathrm{d}\mathbf{x}, \\ f(v_h) &:= \sum_{i=1}^{N} \int_{\Omega_i} \left(\Big(f + \frac{n}{\tau} \, \theta(q) \Big) \, v_{h,i} + \frac{K}{\mu} \, k_\alpha\Big(\theta(q)\Big) \, \nabla d \cdot \nabla v_{h,i} \right) \mathrm{d}\mathbf{x} \\ &\quad + \sum_{i=1}^{N} \int_{\Gamma_{N,i}} g_N \, \gamma^0_{\Gamma_{N,i}} \, v_{h,i} \, \mathrm{d}s_{\mathbf{x}} \end{aligned}$$

for all $u_h, v_h \in X_h$ and $\lambda_h, \nu_h \in Q_h$.

In the following sections, we discuss linearization and discretization techniques and we briefly describe some implementation details. We start with the Newton's method to obtain a linear variational problem in Section 5.1. In Section 5.2 we introduce suitable trial spaces known from the mortar finite element method. Next, in Section 5.3 we discuss unique solvability of the discrete problem. Finally, in Section 5.4, we take a closer look at the implementation.

5.1 Linearization

The modified variational problem (5.2) is still a nonlinear problem which we have to solve. We want to apply the Newton method to obtain a linear variational problem, see [25] for further information. In the next section we recall some of the basics concerning Newtons's method.

5.1.1 The Newton Method

Let V, W be two real Banach spaces and let $A : V \to W$ be a nonlinear operator. Consider the nonlinear equation to find $u \in V$, such that

$$A(u) = 0 \tag{5.3}$$

in W. We want to solve equation (5.3) by applying Newton's method. Formally, we can write the method in the following way.

Choose an initial guess $u^0 \in V$. For $k \in \mathbb{N}$ compute $u^k \in V$ as

$$u^k = u^{k-1} + \delta u$$

with the update $\delta u \in V$ which is a solution to

$$DA(u^{k-1})\,\delta u = -A(u^{k-1}) \tag{5.4}$$

in W.

To state the above linear problem (5.4) in a proper way, we have to define the derivative of an operator in Banach spaces. For further information see [9, Section 2.1C] or [73, Section 4.2].

Definition 5.1 (Gâteaux derivative)**.** *Let V, W be two real Banach spaces and let $A : V \to W$ be a nonlinear operator. A is said to be Gâteaux differentiable in $v \in V$, iff*

$$\lim_{t\to 0} \frac{A(v + t\,h) - A(v)}{t} = DA(v)h, \quad \forall\, h \in V$$

exists in W with $DA(v) \in \mathcal{L}(V, W)$. In this case, $DA(v)$ is called Gâteaux derivative of A in v and we have

$$\frac{\mathrm{d}}{\mathrm{d}t} A(v + t\,h)_{|t=0} = DA(v).$$

The operator A is said to be Gâteaux differentiable on $U \subset V$ if A is Gâteaux differentiable for all $v \in U$.

An extension of the Gâteaux derivative is the Fréchet derivative which is defined as follows.

Definition 5.2 (Fréchet derivative)**.** *Let V, W be two real Banach spaces and let $A : V \to W$ be a nonlinear operator. Then, A is said to be Fréchet differentiable in $v \in V$, iff there exists a linear operator $DA(v) \in \mathcal{L}(V, W)$, such that*

$$\lim_{\substack{h \in V \\ \|h\|_V \to 0}} \frac{\|A(v+h) - A(v) - DA(v)h\|_W}{\|h\|_V} = 0$$

holds. In this case $DA(v)$ is called Fréchet derivative of A in v. The operator A is said to be Fréchet differentiable on $U \subset V$ if A is Fréchet differentiable for all $v \in U$. If $DA : V \to \mathcal{L}(V, W)$ is continuous, then A is called continuously Fréchet differentiable.

It is easy to see, that Fréchet differentiability in $v \in V$ implies Gâteaux differentiability in $v \in V$. This follows immediately from the definition. The reverse implication holds, if the Gâteaux derivative is continuous in $v \in V$, see [9, Theorem 2.1.13]. The following ensures the convergence of the Newton method to solve a nonlinear problem of the form (5.3).

Theorem 5.3 (Kantorovich Theorem)**.** *Let V, W be two real Banach spaces and let $U \subset V$ be an open and convex subset. Furthermore, let $A : V \to W$ be continuously Fréchet differentiable on U. For an initial guess $u^0 \in U$ let $DA(u^0) \in \mathcal{L}(V, W)$ be invertible and assume, that*

$$\begin{aligned} \left\| DA(u^0)^{-1} A(u^0) \right\|_V &\leq \alpha, \\ \left\| DA(u^0)^{-1} \left(DA(v) - DA(w) \right) \right\|_{V \to V} &\leq \gamma \|v - w\|_V \end{aligned}$$

for all $u, v \in U$. If $h_0 := \alpha\gamma \leq 1/2$ holds, then the sequence $\{u^k\}$ obtained by the Newton iteration is well defined and converges to u with $A(u) = 0$. If $h_0 < 1/2$, the convergence is of second order.

Proof. For a proof see [25, Theorem 2.1]. ∎

In Theorem 5.3 Fréchet differetiability is required to show solvability of the nonlinear equation (5.3). It is also possible to prove convergence results of Newton like methods for operators which are just Gâteaux differentiable, see for example [54, 66]. But then additional assumptions are made on the nonlinear operator.

In the next subsection we want to apply the Newton method to the modified variational problem (5.2). Due to the representation theorem of Riesz, the considerations made in this subsection are also applicable to forms $a(\cdot, \cdot)$ which are nonlinear in the first argument.

5.1.2 Newton Linearization

To apply the Newton method to the modified variational problem (5.2), we have to compute the Fréchet derivative of the nonlinear forms $m(\cdot,\cdot)$ and $c(\cdot,\cdot)$ of the modified variational problem (5.2). As already mentioned, the Fréchet derivative and the Gâteaux derivative of an operator $A : V \to W$ coincide if $DA : V \to \mathcal{L}(V,W)$ is continuous. For this reason we use the representation

$$DA(v) = \frac{\mathrm{d}}{\mathrm{d}t} A(v + t\,h)_{|t=0}$$

to determine the Fréchet derivative, see Definition 5.1.

For the nonlinear form $m(\cdot,\cdot)$ we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} m(w_h + t\,u_h + u_{D,h}, v_h)_{|t=0} &= \sum_{i=1}^{N} \int_{\Omega_i} l_i{}'(w_{h,i} + u_{D,h,i})\, u_{h,i}\, v_{h,i}\,\mathrm{d}\mathbf{x} \\ &=: m'(w_h + u_{D,h}, u_h, v_h) \end{aligned} \tag{5.5}$$

for $w_h, u_h, v_h \in X_h$. The derivative of the nonlinear form $m(\cdot,\cdot)$ in $w_h + u_{D,h}$ is given by the bilinear form $m'(w_h + u_{D,h}, \cdot,\cdot)$. We answer the question whether the operator is well defined or not in Section 5.2 where we introduce a concrete discrete trial space X_h and a suitable norm. For the nonlinear form $c(\cdot,\cdot)$ we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} c(w_h + t\,u_h + u_{D,h}, \nu_h)_{|t=0} &= -\sum_{\substack{\Gamma_{ij}\neq\emptyset \\ i<j}} \int_{\Gamma_{ij}} [\![\kappa^{-1\prime}(w_h + u_{D,h})\, u_h]\!]_{\Gamma_{ij}} \nu_h\,\mathrm{d}s_{\mathbf{x}} \\ &=: c'(w_h + u_{D,h}, u_h, \nu_h) \end{aligned} \tag{5.6}$$

for $w_h, u_h \in X_h$ and $\nu_h \in Q_h$. The derivative in $w_h + u_{D,h}$ is given by the bilinear form $c'(w_h + u_{D,h}, \cdot,\cdot)$ which will be discussed in more detail in Section 5.2. The Newton iteration for the modified variational problem (5.2) reads as follows.

Choose an initial guess $u_h^0 \in X_h$ and $\lambda_h^0 \in Q_h$. Compute $u_h^k \in X_h$ and $\lambda_h^k \in Q_h$ as

$$u_h^k := u_h^{k-1} + \delta u_h \qquad \text{and} \qquad \lambda_h^k := \lambda_h^{k-1} + \delta\lambda_h \tag{5.7}$$

with the updates $\delta u_h \in X_h$ and $\delta\lambda_h \in Q_h$ which are solutions to

$$\begin{aligned} m'(u_h^{k-1} + u_{h,D}, \delta u_h, v_h) + a(\delta u_h, v_h) + b(v_h, \delta\lambda_h) &= \widetilde{f}(v_h) \\ c'(u_h^{k-1} + u_{h,D}, \delta u_h, \nu_h) &= \widetilde{g}(\nu_h) \end{aligned}$$

for all $v_h \in X_h$ and $\nu_h \in Q_h$. The right hand sides are

$$\widetilde{f}(v_h) := f(v_h) - a(u_h^{k-1} + u_{D,h}, v_h) - b(v_h, \lambda_h^{k-1}) - m(u_h^{k-1} + u_{D,h}, v_h)$$

for all $v_h \in X_h$ and

$$\widetilde{g}(v_h) := -c(u_h^{k-1} + u_{h,D}, \nu_h)$$

for all $\nu_h \in Q_h$.

If we plug in the representation (5.7) for δu_h and $\delta\lambda_h$, we can compute $u_h^k \in X_h$ and $\lambda_h^k \in Q_h$ directly as solution to the following variational problem.

Newton iteration for the modified discrete variational problem

For $w_h \in X_h$ find $u_h \in X_h$ and $\lambda_h \in Q_h$ such that

$$\begin{aligned} m'(w_h + u_{h,D}, u_h, v_h) + a(u_h, v_h) + b(v_h, \lambda_h) &= f(w_h, v_h) \\ c'(w_h + u_{h,D}, u_h, \nu_h) &= g(w_h, \nu_h) \end{aligned} \tag{5.8}$$

for all $v_h \in X_h$ and $\nu_h \in Q_h$.

In the variational problem (5.8) the given $w_h \in X_h$ corresponds to the previous Newton iteration and the right hand sides are

$$f(w_h, v_h) := f(v_h) - a(u_{D,h}, v_h) + m'(w_h + u_{D,h}, w_h, v_h) - m(w_h + u_{D,h}, v_h)$$

for all $v_h \in X_h$ and

$$g(w_h, \nu_h) := c'(w_h + u_{D,h}, w_h, \nu_h) - c(w_h + u_{D,h}, v_h)$$

for all $\nu_h \in Q_h$.

We applied the Newton method to the modified nonlinear variational problem (5.2) and obtained the linear variational problem (5.8). All the considerations are done with respect to general discrete trial spaces X_h and Q_h. The next step is to define concrete realizations of the trial spaces, which is done in Section 5.2 and to verify the existence of the derivatives defined by (5.5) and (5.6).

5.2 Space Discretization

In this section we discuss the discretization of trial and test spaces fitting to the variational problem (5.8), see for example [8, 11, 70, 71]. We want to start with the trial space X_h. To define appropriate spaces, we assume that $\Omega \subset \mathbb{R}^2$ is a bounded polygonal domain or $\Omega \subset \mathbb{R}^3$ is a bounded polyhedral domain. We suppose that the subdomains of the decomposition $\mathcal{D}_\Omega^N$ are again polygonal and polyhedral domains, respectively. Furthermore, we expect that the decomposition is a geometrical conforming and

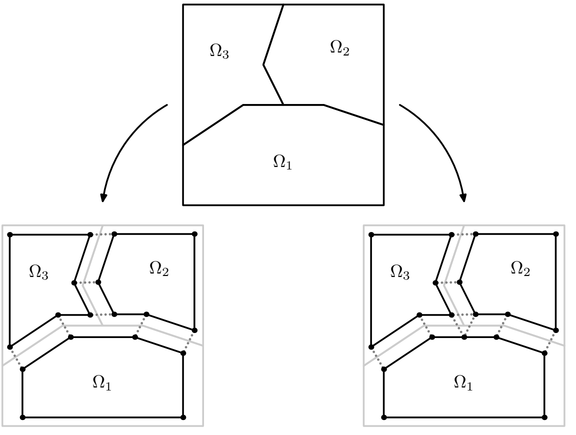


Figure 5.2: Geometrical non–conforming (left) and conforming (right) situation.

admissible decomposition where the intersection of the boundaries $\partial\Omega_i \cap \partial\Omega_j$ is either empty, a vertex, a common edge or in the three dimensional case a common face, see Figure 5.2. Due to the primal hybrid formulation, the trial and test functions just act locally in each $\Omega_i \in \mathcal{D}_\Omega^N$ and are coupled via suitable conditions. This allows the independent definition of discrete trial spaces in each subdomain. In this thesis we restrict ourselves to piecewise linear trial functions. Nevertheless, it is also possible to define trial spaces of higher order, see [45]. But before we introduce the trial spaces, we have to discretize the underlying computational domain in a proper way.

Definition 5.4 (Domain discretization)**.** *Let $h \in \mathbb{R}$, $h > 0$, and let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a polygonal and polyhedral domain, respectively. Furthermore, let $\mathcal{D}_\Omega^N$ be an admissible decomposition of Ω. By $\mathcal{T}_{h,i} := \{T_k\}_{k=1}^{N_{T,i}}$ we denote the discretization of the subdomain $\Omega_i \in \mathcal{D}_\Omega^N$, that is*

$$\overline{\Omega}_i = \bigcup_{i=1}^{N_{T,i}} \overline{T}_i$$

with nonoverlapping triangles T_k for $d = 2$ or tetrahedra T_k for $d = 3$.

For each finite element $T \in \mathcal{T}_{h,i}$ we define its volume $|T| := \int_T \mathrm{d}\mathbf{x}$, its local mesh size $h_T := |T|^{1/d}$ and its diameter $d_T := \sup_{\mathbf{x},\mathbf{y} \in T} |\mathbf{x} - \mathbf{y}|$. For each $\mathcal{T}_{h,i}$ we define the global mesh size h_i by

$$h_i = h_{max,i} := \max_{T \in \mathcal{T}_{h,i}} h_T$$

and the minimal local mesh size is given by

$$h_{min,i} := \min_{T \in \mathcal{T}_{h,i}} h_T.$$

We assume, that each discretization $\mathcal{T}_{h,i}$ is globally quasi–uniform, that is

$$\frac{h_{max,i}}{h_{min,i}} \leq c_{G,i} \leq c_G$$

with $c_{G,i}, c_G \in \mathbb{R}^+$.

The global domain discretization will be denoted by $\mathcal{T}_h$ and is defined by

$$\mathcal{T}_h := \prod_{i=1}^{N} \mathcal{T}_{h,i},$$

see for example Figure 5.3 for $\Omega \subset \mathbb{R}^2$.

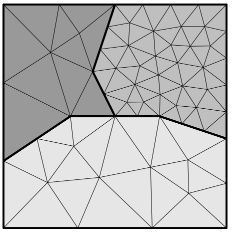


Figure 5.3: Discretization of the subdomains with different mesh sizes.

We have clarified the question concerning the domain discretization, now we can use the triangulation $\mathcal{T}_h$ to introduce an appropriate trial space X_h.

Definition 5.5 (Discrete trial space X_h)**.** *Let $\mathcal{T}_h$ be an admissible triangulation of the domain Ω with respect to the decomposition $\mathcal{D}_\Omega^N$. For each $\mathcal{T}_{h,i}$ consider*

$$S^1(\mathcal{T}_{h,i}) := \left\{ v \in \mathcal{C}(\overline{\Omega}_i) \mid v_{|T} \in P^1(T)\ \forall\, T \in \mathcal{T}_{h,i} \right\}$$

the space of continuous and piecewise linear functions defined on Ω_i. Next, define

$$X_{h,i} := \begin{cases} S^1(\mathcal{T}_{h,i}) \cap H^1_{0,\Gamma_{D,i}}(\Omega_i) & \text{if } \Gamma_{D,i} \neq \emptyset, \\ S^1(\mathcal{T}_{h,i}) & \text{if } \Gamma_{D,i} = \emptyset \end{cases}$$

which can be written as the span of a family of linear basis functions, that is $X_{h,i} = \operatorname{span}\{\varphi_{k,i}\}_{k=1}^{\mathring{N}_{V,i}}$. The global trial space is the product of the local spaces, that is

$$X_h := \prod_{i=1}^{N} X_{h,i} = \prod_{i=1}^{N} \operatorname{span}\{\varphi_{k,i}\}_{k=1}^{\mathring{N}_{V,i}}.$$

By $\mathring{N}_{V,i}$ *we denote the number of vertices which are not on the Dirichlet boundary* $\Gamma_{D,i}$.

The next step is to define a suitable discrete trial space for the Lagrange multiplier. Due to the fact, that the Lagrange multiplier act on the skeleton $\mathcal{S} \setminus \partial\Omega$, we have to find a suitable discretization of the interfaces Γ_{ij} first.

Definition 5.6 (Interface discretization)**.** *Let* $\mathcal{T}_h$ *be an admissible triangulation of the domain* Ω *with respect to the decomposition* $\mathcal{D}_\Omega^N$. *For each non–empty interface* $\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j$, $i \neq j$, *we define the interface discretization by*

$$\mathcal{I}_{h,ij} := \left\{ E \subset \mathbb{R}^{d-1} \mid \exists T \in \mathcal{T}_{h,i} : \overline{E} = \overline{T} \cap \overline{\Gamma}_{ij} \text{ and } |\overline{E}|_{d-1} > 0 \right\}.$$

By $N_{E,ij}$ *we denote the number of elements, that is* $N_{E,ij} := |\mathcal{I}_{h,ij}|$. *For* $d = 2$ *the elements are edges of triangles and for* $d = 3$ *the elements are faces of tetrahedra. In general* $\mathcal{I}_{h,ij} \neq \mathcal{I}_{h,ji}$, *see Figure 5.4. Since* $\mathcal{T}_{h,i}$ *is assumed to be globally quasi–uniform, each interface discretization* $\mathcal{I}_{h,ij}$ *is globally quasi–uniform as well.*

Analogously to the domain discretization, we define for each interface element $E \in \mathcal{I}_{h,ij}$ *its volume* $|E| := \int_E \mathrm{d}s_{\mathbf{x}}$, *its local mesh size* $h_E := |E|^{1/(d-1)}$ *and its diameter* $d_E := \sup_{\mathbf{x},\mathbf{y}\in E} |\mathbf{x} - \mathbf{y}|$.

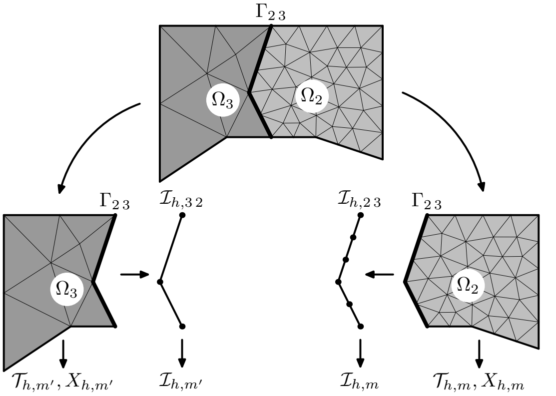


Figure 5.4: Different discretizations of the interface Γ_{ij}.

Since every interface Γ_{ij} can be discretized in two different ways, we have to fix a unique discretization for each interface to define the discrete Lagrange multiplier in a proper way. In view of a better approximation property, we want to choose the finer discretization. The precise definition reads as follows.

Definition 5.7 (Mortar and non–mortar side)**.** *Let $\mathcal{T}_h$ be an admissible triangulation of the domain Ω with respect to the decomposition $\mathcal{D}_\Omega^N$. For each non–empty interface Γ_{ij} fix the ordered pair m in the following way*

$$m := \begin{cases} (i,j) & \textit{if } (N_{E,ij} > N_{E,ji}) \textit{ or } (N_{E,ij} = N_{E,ji} \textit{ and } i < j)\,, \\ (j,i) & \textit{if } (N_{E,ji} > N_{E,ij}) \textit{ or } (N_{E,ji} = N_{E,ij} \textit{ and } j < i)\,. \end{cases}$$

For each non–empty interface, we denote the non–mortar side by $\mathcal{I}_{h,m} := \mathcal{I}_{h,kl}$ as well as the corresponding domain discretization $\mathcal{T}_{h,m} := \mathcal{T}_{h,k}$. The mortar side is then given by $\mathcal{I}_{h,m'} := \mathcal{I}_{h,lk}$ and the corresponding domain discretization is $\mathcal{T}_{h,m'} := \mathcal{T}_{h,l}$, see Figure 5.4. Analogously we define $X_{h,m} := X_{h,k}$ and $X_{h,m'} := X_{h,l}$.

We use the above definition to ensure that each interface is only counted once in the coupling terms, see the introduction of Chapter 5.

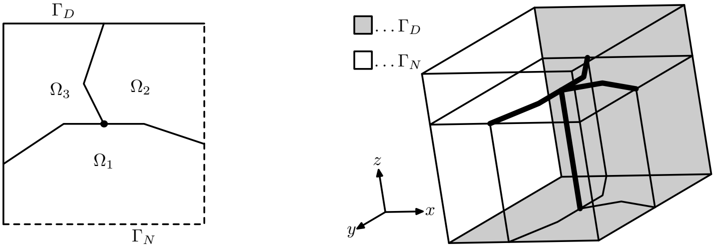
Of course, the choice of the interface discretization is arbitrary but has to be fixed. It is even possible to choose an interface discretization $\mathcal{I}_{h,m}$ which is inherited neither from $\mathcal{T}_{h,m}$ nor from $\mathcal{T}_{h,m'}$. Then, suitable conditions has to be imposed on the mesh or on the trial space to ensure stability of the coupling condition, see for example [20].

Next, we want to construct the trial space for the Lagrange multiplier which is defined on the interface. There are several possibilities to choose this space, see [71, Section 1.2.4] for $d = 2$ or [41] for $d = 3$. In this thesis we want to consider piecewise constant functions defined on a modified dual interface discretization, whose construction is discussed in the following.

To define the modified dual interface mesh, we have to introduce Γ_C, which is the set of all cross elements. For $d = 2$ these elements are points on the boundary of interfaces and for $d = 3$ these elements are edges on the boundary of interfaces. A cross element is an element, which belongs to the boundaries of more than two subdomains, see Figure 5.5.

For a fixed interface Γ_m, we denote the number of vertices on Γ_m by $N_{V,m}$ and $\mathring{N}_{V,m}$ denotes the number of vertices on Γ_m which are neither on Γ_D nor on Γ_C. We distinguish between the two and three dimensional case to construct the modified dual mesh of the interface discretization $\mathcal{I}_{h,m}$. Consider the case $d = 2$ first.

Let Γ_m be an arbitrary interface and consider its discretization $\mathcal{I}_{h,m}$. By $\mathcal{V}_m$ we denote the set of all vertices belonging to $\overline{\Gamma}_m$. First, we split each edge element $E \in \mathcal{I}_{h,m}$ into two sub elements of equal area by connecting the end points with the midpoint. Next, for each vertex $\mathbf{x}_{k,m} \in \mathcal{V}_m$ we unite all sub elements having $\mathbf{x}_{k,m}$ as a vertex. We denote this new partition by $\mathcal{I}_{h,m}{}'$ and there holds $|\mathcal{I}_{h,m}{}'| = N_{V,m}$ by construction. We now modify the dual interface mesh by the following instruction.



(a) Cross points for $d = 2$. (b) Cross edges for $d = 3$.

Figure 5.5: Cross points and cross edges.

(1) If the edge element $E \in \mathcal{I}_{h,m}$ has one vertex on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we unite it with the finite element $E' \in \mathcal{I}_{h,m}{}'$ which corresponds to the second vertex.

In Figure 5.6 the construction of a modified dual mesh is depicted for the two dimensional case. Due to the construction, the number of finite volumes in the

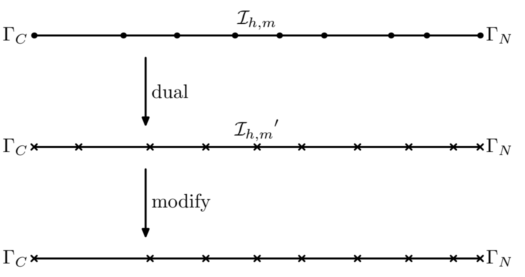
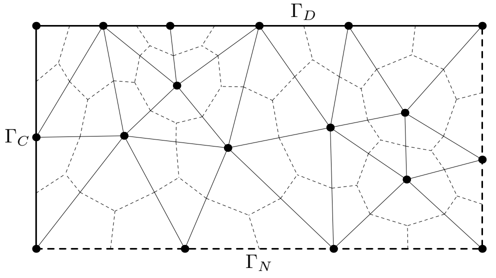


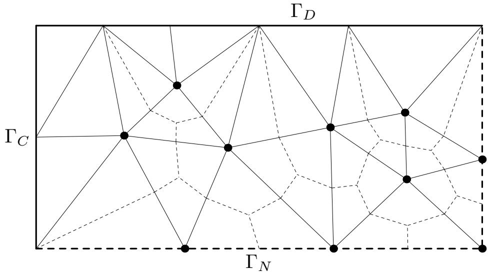
Figure 5.6: Construction of the modified dual interface mesh.

modified dual interface mesh is $\mathring{N}_{V,m}$.

Next, we want to discuss the three dimensional case. Let Γ_m be an arbitrary interface with discretization $\mathcal{I}_{h,m}$. We again denote by $\mathcal{V}_m$ the set of all vertices on $\overline{\Gamma}_m$. Analogously to the two dimensional case, we split each face element $E \in \mathcal{I}_{h,m}$ into three sub elements of equal size by connecting its midcenter with the midpoints of each edge. Next, for each vertex $\mathbf{x}_{k,m} \in \mathcal{V}_m$ we unite all sub elements having $\mathbf{x}_{k,m}$ as a vertex. This partition is denoted by $\mathcal{I}_{h,m}{}'$ and there holds $|\mathcal{I}_{h,m}{}'| = N_{V,m}$ by construction, see Figure 5.7a. In the same manner as in the two dimensional case, we want to modify the dual mesh to reduce its number of finite elements to $\mathring{N}_{V,m}$. This can be achieved by the following instructions.



(a) Construction of the dual interface mesh.



(b) Construction of the modified dual interface mesh.

(1) If the face element $E \in \mathcal{I}_{h,m}$ has all of its three vertices on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we attach this triangle to a adjacent one sharing one common internal edge.

(2) If the face element $E \in \mathcal{I}_{h,m}$ has two vertices on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we unite E and all attached triangles with the element $E' \in \mathcal{I}_{h,m}{}'$ which corresponds to the inner vertex.

(3) If the face element $E \in \mathcal{I}_{h,m}$ has one vertex on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, then we split the triangle by connecting this vertex with the midpoint of the opposite edge. The new triangles are united with the finite elements in $\mathcal{I}_{h,m}{}'$ which corresponds to the two inner vertices.

Due to the construction, the number of the finite elements in the modified dual interface mesh is $\mathring{N}_{V,m}$, see for example Figure 5.7b.

In a next step we use the constructed modified dual interface discretization to define the basis functions for the discrete Lagrange multiplier trial space. So for each element E_k in the modified dual interface discretization we define the corresponding basis function $\psi_{k,m}$ to be one on E_k and zero else, see Figure 5.8 for $d = 2$. We denote the

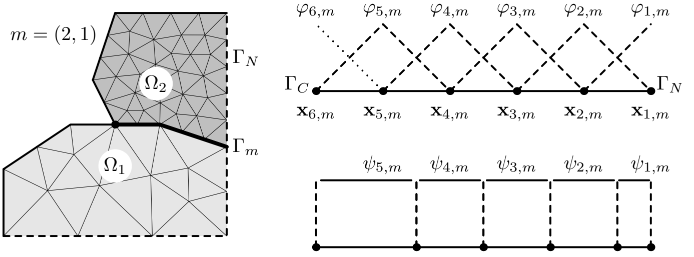


Figure 5.8: Construction of the basis functions for the discrete Lagrange multiplier for $d = 2$.

trial space on the interface Γ_m as

$$Q_{h,m} := \operatorname{span}\{\psi_{k,m}\}_{k=1}^{\mathring{N}_{V,m}} \tag{5.9}$$

and there holds $\dim Q_{h,m} = \mathring{N}_{V,m}$ by construction.

Definition 5.8 (Discrete trial space Q_h)**.** *Let $\mathcal{T}_h$ be an admissible triangulation of the domain Ω with respect to the decomposition $\mathcal{D}_\Omega^N$. For each interface Γ_{ij} we fix the mortar side and the non–mortar side according to Definition 5.7 and construct the piecewise constant trial space $Q_{h,m}$ as described for (5.9). The global trial space is defined as the product space, that is*

$$Q_h := \prod_{\Gamma_m} Q_{h,m} = \prod_{\Gamma_m} \operatorname{span}\{\psi_{k,m}\}_{k=1}^{\mathring{N}_{V,m}}.$$

Since the coupling conditions are defined on the interface, we introduce for completeness the trace space on the interfaces Γ_m. For each $m = (k, l)$, the trace space $W_{h,m}$ is given by

$$W_{h,m} := \left\{ t_h \in L_2(\Gamma_{kl}) \mid t_h = \gamma^0_{\Gamma_{kl}} u_h : u_h \in X_{h,m} \right\} \tag{5.10}$$

and the modified trace space

$$\mathring{W}_{h,m} := \left\{ t_h \in W_{h,m} \mid t_h|_{\overline{\Gamma}_C \cup \overline{\Gamma}_D} = 0 \right\} \tag{5.11}$$

which satisfies homogeneous Dirichlet conditions on cross points in $\mathbb{R}^2$ and on cross edges in $\mathbb{R}^3$ respectively as well as on the Dirichlet boundary Γ_D. The trace spaces $W_{h,m}$ and $\mathring{W}_{h,m}$ are spanned by the traces of the basis functions of $X_{h,m}$ which does not vanish on Γ_m. Thus, we can write

$$W_{h,m} := \operatorname{span}\{\varphi_{k,m}\}_{k=1}^{N_{V,m}} \qquad \text{and} \qquad \mathring{W}_{h,m} := \operatorname{span}\{\varphi_{k,m}\}_{k=1}^{\mathring{N}_{V,m}}$$

with $\mathring{N}_{V,m} \leq N_{V,m}$.

We have discussed how to construct proper trial spaces for the two and three dimensional case respectively. In the next section we want to answer whether the discrete problem is solvable or not. To answer this question we first have to introduce the right norms for the discrete trial spaces.

5.3 Solvability and Uniqueness

We have constructed basis functions for the discrete trial spaces but to state solvability results, we have to equip the spaces with the right norm. As already mentioned we want to use the tools and results achieved in the mortar finite element framework. For the trial space X_h we choose $\|\cdot\|_X$ as defined in Definition 4.13, that is we consider $(X_h, \|\cdot\|_X)$. The question about the right norm for the space Q_h is much harder to answer, see [16]. The most natural choice would be the dual $H^{1/2}$–norm, but since we are dealing with a saddle point problem we have to ensure the inf–sup–stability of the side condition, see [18]. Unfortunately no stability estimates for this norm has been established so far. However, there are estimates for the dual $H^{1/2}_{00}$–norm, see [8], but the estimates only hold for domains with no cross elements. One further common approach is to consider the space $Q_{h,m}$ as a subspace of $Q_{2,m} \subset L_2(\Gamma_m)$ defined by

$$Q_{2,m} := \left\{ \nu \in L_2(\Gamma_m) \mid \|\nu\|_{Q_{2,m}} < \infty \right\} \tag{5.12}$$

equipped with the h–weighted L_2–norm

$$\|\nu\|^2_{Q_{2,m}} := \sum_{E \in \mathcal{I}_{h,m}} h_E \|\nu\|^2_{L_2(E)}.$$

The dual space is therefore given as

$$Q_{2,m}{}' := \left\{ u \in L_2(\Gamma_m) \mid \|u\|_{Q_{2,m}{}'} < \infty \right\}$$

equipped with

$$\|u\|^2_{Q_{2,m}{}'} := \sum_{E \in \mathcal{I}_{h,m}} \frac{1}{h_E} \|u\|^2_{L_2(E)}.$$

The global spaces are then defined by

$$Q_2 := \prod_{\Gamma_m} Q_{2,m} \qquad \text{with} \qquad \|u\|^2_{Q_2} := \sum_{\Gamma_m} \|u\|^2_{Q_{2,m}}$$

and $Q_2{}'$ analogously. For our considerations we choose for the discrete Lagrange multiplier the h–weighted L_2–norm, that is we consider $(Q_{h,m}, \|\cdot\|_{Q_{2,m}})$ as well as $(W_{h,m}, \|\cdot\|_{Q_{2,m}{}'})$ for each interface Γ_m.

Furthermore, we have to assume that for each element in the non–mortar discretization $E \in \mathcal{I}_{h,m}$ the corresponding domain element $T \in X_{h,m}$ has at least one vertex not on Γ_m. Under this assumption, the discrete extension with zero $\mathcal{E}_{h,\Omega_m} : W_{h,m} \to X_{h,m}$ is bounded, that is

$$\left\| \mathcal{E}_{h,\Omega_m} w_h \right\|_{H^1(\Omega_i)} \leq c_{E,h} \left\| w_h \right\|_{Q_{2,m}'} \tag{5.13}$$

for some positive constant $c_{E,h} > 0$ independent of the mesh size h.

In Section 5.1 we briefly discussed the Newton method and we have seen, that Fréchet differentiability is essential for the convergence of the Newton iteration. The following two lemmata show under which conditions the forms $m(\cdot,\cdot)$ and $c(\cdot,\cdot)$ in the variational problem (5.2) are Fréchet differentiable.

Lemma 5.9. *Suppose the nonlinear function $l_i := \theta_i \circ \kappa_i^{-1} : \mathbb{R} \to \mathbb{R}$ is a function in $\mathcal{C}^1(\mathbb{R})$ and has in addition a Lipschitz continuous and bounded derivative $l_i' : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant $c_{L,l_i'} > 0$ and upper bound $c_{B,l_i'} := \|l_i'\|_{L_\infty(\mathbb{R})} < \infty$. Then the nonlinear form $m(\cdot,\cdot)$ is Fréchet differentiable.*

Proof. To show Fréchet differentiability we have to estimate the expression

$$\begin{aligned}\sup_{0 \neq w \in X} \frac{m(v+h,w) - m(v,w) - m'(v,h,w)}{\|v\|_X} &= \\ &= \sup_{0 \neq w \in X} \left[\frac{1}{\|v\|_X} \sum_{i=1}^{N} \int_{\Omega_i} \Big(l_i(v_i + h_i) - l_i(v_i) - l_i'(v_i)\, h_i \Big) w_i \,\mathrm{d}\mathbf{x} \right]\end{aligned}$$

which is bounded by the $L_2(\Omega)$–norm

$$\|l(v+h) - l(v) - l'(v)\, h\|_{L_2(\Omega)} \tag{5.14}$$

with $v, h \in X$. The functions $l(v)$ are due to previous considerations elements in $L_2(\Omega)$. The boundedness of $l' : \mathbb{R} \to \mathbb{R}$ ensures that the expression $l'(v + u_D)\, h$ is an element in $L_2(\Omega)$ and thus the norm in (5.14) is well defined. Next, we want to estimate the expression (5.14). Since $l_i : \mathbb{R} \to \mathbb{R}$ is assumed to be differentiable, there holds

$$l_i(r+s) - l_i(r) = \int_0^s l_i'(r+t)\,\mathrm{d}t$$

and the Lipschitz continuity of the derivative implies

$$|l_i(r+s) - l_i(r) - l_i'(r)\, s| = \left| \int_0^s \Big[l_i'(r+t) - l_i'(r) \Big] \mathrm{d}t \right| \leq \frac{c_{L,l_i'}}{2}\, |s|^2$$

for all $r, s \in \mathbb{R}$. Thus, for each $\Omega_i \in \mathcal{D}_\Omega^N$ and $v, h \in X$ we have

$$\|l_i(v+h) - l_i(v) - l_i'(v)\,h\|^2_{L_2(\Omega_i)} \le \left(\frac{c_{L,l_i'}}{2}\right)^2 \int\limits_{\Omega_i} |v|^4 \,\mathrm{d}\mathbf{x} = \left(\frac{c_{L,l_i'}}{2}\right)^2 \|h\|^4_{L_4(\Omega_i)}$$

and the Imbedding Theorem 3.38 implies

$$\|l_i(v+h) - l_i(v) - l_i'(v)\,h\|_{L_2(\Omega_i)} \le c_I \,\frac{c_{L,l_i'}}{2}\, \|h\|^2_{H^1(\Omega_i)}$$

for each subdomain and with a positive constant c_I. Summing over all subdomains, we get

$$\|l(v+h) - l(v) - l'(v)\,h\|_{L_2(\Omega)} \le c_I \,\frac{c_{L,l'}}{2}\, \|h\|^2_X$$

for $v, h \in X$ and $c_{L,l'} := \max\limits_{i=1,\dots,N_L} c_{L,l_i'}$. If we combine all these observations, we obtain

$$\lim_{\|h\|_X \to 0} \frac{1}{\|h\|_X} \sup_{0 \ne w \in X} \frac{m(v+h,w) - m(v,w) - m'(v,h,w)}{\|v\|_X} \le$$

$$\le \lim_{\|h\|_X \to 0} \frac{\|l(v+h) - l(v) - l'(v)\,h\|_{L_2(\Omega)}}{\|h\|_X} \le c_I \,\frac{c_{L,l'}}{2} \lim_{\|h\|_X \to 0} \frac{\|h\|^2_X}{\|h\|_X} = 0$$

which proves the Fréchet differentiability. ■

The next step is to prove the Fréchet differentiability of the nonlinear form $c(\cdot,\cdot)$ in the variational problem (5.2). Since we consider the discrete trial space Q_h as a subspace of the space Q_2, we consider $c : X \times Q_2 \to \mathbb{R}$.

Lemma 5.10. *The nonlinear form $c : X \times Q_2 \to \mathbb{R}$ is Fréchet differentiable.*

Proof. The form $c(\cdot,\cdot)$ as defined in the variational problem (5.2) is given by the sum of the nonlinear jumps, that is

$$\begin{aligned} c(v,\nu) &= -\sum_{\Gamma_m} \int\limits_{\Gamma_m} \left(\kappa_k^{-1}(\gamma^0_{\Gamma_m} v_k) - \kappa_l^{-1}(\gamma^0_{\Gamma_m} v_l)\right) \nu \,\mathrm{d}s_{\mathbf{x}} \\ &= \sum_{\Gamma_m} \left(\int\limits_{\Gamma_m} \kappa_l^{-1}(\gamma^0_{\Gamma_m} v_l)\, \nu \,\mathrm{d}s_{\mathbf{x}} - \int\limits_{\Gamma_m} \kappa_k^{-1}(\gamma^0_{\Gamma_m} v_k)\, \nu \,\mathrm{d}s_{\mathbf{x}} \right) \end{aligned}$$

for $v \in X$ and $\nu \in Q_2$. For the following we define $s := \gamma^0_{\Gamma_m} v_k$ and $t := \gamma^0_{\Gamma_m} h_k$ with $s, t \in H^{1/2}(\Gamma_m)$ For an arbitrary element $E \in \mathcal{I}_{h,m}$ we have

$$\int\limits_E \left(\kappa_i^{-1}(s+t) - \kappa_i^{-1}(s) - {\kappa_i^{-1}}'(s)\,t\right) \nu \,\mathrm{d}s_{\mathbf{x}} \le$$

$$\le \int\limits_E \left|\kappa_i^{-1}(s+t) - \kappa_i^{-1}(s) - {\kappa_i^{-1}}'(s)\,t\right| \,|\nu| \,\mathrm{d}s_{\mathbf{x}}$$

and since each $\kappa_i^{-1} : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, see Lemma 4.8, we obtain

$$\int_E \left(\kappa_i^{-1}(s+t) - \kappa_i^{-1}(s) - \kappa_i^{-1\prime}(s)\,t\right) \nu \,\mathrm{d}s_{\mathbf{x}} \le \frac{1}{\alpha} \int_E |t|^2 \; |\nu| \; \mathrm{d}s_{\mathbf{x}}$$
$$\le \frac{1}{\alpha} \left(\frac{1}{h_E^{1/2}} \left\|t\right\|^2_{L_4(E)} h_E^{1/2} \left\|\nu\right\|_{L_2(E)} \right)$$

with $\alpha > 0$. Summing over all elements $E \in \mathcal{I}_{h,m}$ yields

$$\int_{\Gamma_m} \left(\kappa_i^{-1}(s+t) - \kappa_i^{-1}(s) - \kappa_i^{-1\prime}(s)\,t\right) \nu \,\mathrm{d}s_{\mathbf{x}} \le$$
$$\le \sum_{E \in \mathcal{I}_{h,m}} \frac{1}{\alpha} \left(\frac{1}{h_E^{1/2}} \left\|t\right\|^2_{L_4(E)} h_E^{1/2} \left\|\nu\right\|_{L_2(E)} \right)$$

and by applying the Cauchy–Schwarz–inequality we get

$$\int_{\Gamma_m} \left(\kappa_i^{-1}(s+t) - \kappa_i^{-1}(s) - \kappa_i^{-1\prime}(s)\,t\right) \nu \,\mathrm{d}s_{\mathbf{x}} \le \frac{1}{\alpha\, h_{min}^{1/2}} \left\|t\right\|^2_{L_4(\Gamma_m)} \left\|\nu\right\|_{Q_{2,m}}$$

for all $s, t \in H^{1/2}(\Gamma_m)$ and $\nu \in Q_{2,m}$. If we set $s := \gamma^0_{\Gamma_m}\, v_l$ and $t := \gamma^0_{\Gamma_m}\, h_l$, we obtain exactly the same estimate. Thus, for $c : X \times Q_2 \to \mathbb{R}$ we obtain

$$c(v+h,\nu) - c(v,\nu) - c'(v,h,\nu) \le$$
$$\le \frac{1}{\alpha\, h_{min}^{1/2}} \sum_{\Gamma_m} \left(\left(\left\|\gamma^0_{\Gamma_m}\, h_l\right\|^2_{L_4(\Gamma_m)} + \left\|\gamma^0_{\Gamma_m}\, h_k\right\|^2_{L_4(\Gamma_m)} \right) \left\|\nu\right\|_{Q_{2,m}} \right)$$

which leads to the estimate

$$c(v+h,\nu) - c(v,\nu) - c'(v,h,\nu) \le$$
$$\le \frac{2^{1/2}}{\alpha\, h_{min}^{1/2}} \left(\sum_{\Gamma_m} \left(\left\|\gamma^0_{\Gamma_m}\, h_l\right\|^4_{L_4(\Gamma_m)} + \left\|\gamma^0_{\Gamma_m}\, h_k\right\|^4_{L_4(\Gamma_m)} \right) \right)^{1/2} \left\|\nu\right\|_{Q_2}.$$

The sum of the right hand sides, can be rewritten as the sum over all boundaries $\partial\Omega_i \setminus \partial\Omega$ and thus there holds

$$\sum_{\Gamma_m} \left(\left\|\gamma^0_{\Gamma_m}\, h_l\right\|^4_{L_4(\Gamma_m)} + \left\|\gamma^0_{\Gamma_m}\, h_k\right\|^4_{L_4(\Gamma_m)} \right) \le \left(\sum_i \left\|\gamma^0_{\partial\Omega_i}\, h_i\right\|^4_{L_4(\partial\Omega_i)} \right)^{1/2}$$
$$\le \sum_i \left\|\gamma^0_{\partial\Omega_i}\, h_i\right\|^2_{L_4(\partial\Omega_i)}$$

for all $h \in X$. Due to the Imbedding Theorem 3.38 and the Trace Theorem 3.35 we have

$$\sum_i \left\|\gamma^0_{\partial\Omega_i} h_i\right\|^2_{L_4(\partial\Omega_i)} \leq c_I \sum_i \left\|\gamma^0_{\partial\Omega_i} h_i\right\|^2_{H^{1/2}(\partial\Omega_i)} \leq c_I\, c_{T,\mathcal{S}} \sum_i \|h_i\|^2_{H^1(\Omega_i)}$$

for all $h \in X$. If we combine all these estimates we obtain

$$c(v+h,\nu) - c(v,\nu) - c'(v,h,\nu) \leq c_I\, c_{T,\mathcal{S}} \frac{8^{1/2}}{\alpha\, h^{1/2}_{min}} \|h\|^2_X \|\nu\|_{Q_2}$$

and in the same manner as in the proof of Lemma 5.9 we can conclude the Fréchet differentiability. ■

In order to apply the Kantorovich Theorem 5.3 we have to ensure that the linearized operator is invertible. Thus, we have to verify, that for each right hand side there exists a unique solution to the linearized variational problem (5.8). In contrast to saddle point problems which arise for example from the Stokes equation, see [36, 63], the saddle point problem (5.8) is neither symmetric nor skew symmetric. For this reason, the standard theory of mixed formulations discussed in [13, 19] is not applicable. In order to show solvability of such generalized saddle point problems, we refer to [10, 23] or [50] where the following theorem is proven.

Theorem 5.11. *Let V, W be two Hilbert spaces and let $A : V \to V'$, $B : V \to W'$ and $C : V \to W'$ be linear and bounded operators. Furthermore, assume that*

$$\sup_{0\neq v\in \operatorname{Ker} B} \frac{\langle Au, v\rangle_{V'\times V}}{\|v\|_V} \geq \alpha_A \|u\|_V \qquad \forall\, u \in \operatorname{Ker} C, \tag{5.15a}$$

$$\sup_{0\neq u\in \operatorname{Ker} C} \langle Au, v\rangle_{V'\times V} > 0 \qquad \forall\, v \in \operatorname{Ker} B, v \not\equiv 0, \tag{5.15b}$$

$$\sup_{0\neq v\in V} \frac{\langle Bv, q\rangle_{W'\times W}}{\|v\|_V} \geq \alpha_B \|q\|_W \qquad \forall\, q \in W, \tag{5.15c}$$

$$\sup_{0\neq v\in V} \frac{\langle Cv, q\rangle_{W'\times W}}{\|v\|_V} \geq \alpha_C \|q\|_W \qquad \forall\, q \in W \tag{5.15d}$$

hold for positive constants α_A, α_B and α_C. Then for any $f \in V'$ and $g \in W'$, there exists a unique pair $(u,p) \in V \times W$, such that

$$\begin{aligned} \langle Au, v\rangle_{V'\times V} + \langle B'p, v\rangle_{V'\times V} &= \langle f, v\rangle_{V'\times V} \\ \langle Cu, q\rangle_{W'\times W} &= \langle g, q\rangle_{W'\times W} \end{aligned}$$

for all $v \in V$ and $q \in W$. The unique solution satisfies the following bound

$$\begin{aligned} \|u\|_V &\leq \frac{1}{\alpha_C}\left(1 + \frac{\beta_a}{\alpha_a}\right) \|g\|_{W'} + \frac{1}{\alpha_A}\|f\|_{V'} \\ \|p\|_W &\leq \frac{1}{\alpha_B}\Big(\|f\|_{V'} + \beta_a \|u\|_V\Big) \end{aligned}$$

with $\beta_a = \|A\|_{\mathcal{L}(V,V')}$.

The main idea behind the proof of this theorem is to apply Theorem 3.17 to the operator A on the kernels of the operators B and C and to apply Theorem 3.16 to B and C separately. We want to apply Theorem 5.11 to the linearized variational problem (5.8). Thus, we have to show the stability estimates (5.15a)–(5.15d).

Stability of the linear coupling: To show the inf–sup–stability of the linear coupling condition, the mortar projection plays a major role. For each interface Γ_m the mortar projection $\Pi_m : L_2(\Gamma_m) \to \mathring{W}_{h,m}$ is defined by

$$\int\limits_{\Gamma_m} (\Pi_m v - v)\nu_h \, \mathrm{d}s_{\mathbf{x}} = 0 \tag{5.16}$$

for all $\nu_h \in Q_{h,m}$.

Furthermore, we need the following auxiliary result.

Lemma 5.12. *Let* $w_h \in \mathring{W}_{h,m}$ *and* $\omega_h \in Q_{h,m}$ *with*

$$w_h = \sum_{k=1}^{\mathring{N}_{V,m}} w_k \, \varphi_{k,m} \qquad \textit{and} \qquad \omega_h = \sum_{k=1}^{\mathring{N}_{V,m}} w_k \, \psi_{k,m},$$

then there holds $\|w_h\|_{L_2(\Gamma_m)} \simeq \|\omega_h\|_{L_2(\Gamma_m)}$.

Proof. To prove the above statement, we first consider the element mass matrix $M_E \in \mathbb{R}^{d\times d}$ for an arbitrary $E \in \mathcal{I}_{h,m}$ with entries

$$(M_E)_{kl} := \int\limits_E \varphi_{e_k} \, \varphi_{e_l} \, \mathrm{d}s_{\mathbf{x}}$$

for $0 \le k, l \le d-1$. The global vertex index e_l corresponds to the local vertex index l belonging to E. The eigenvalues of the matrix M_E are

$$\lambda_0 = \frac{|E|}{d} \quad \text{and} \quad \lambda_1 = \ldots \lambda_{d-1} = \frac{|E|}{d\,(d+1)},$$

see [60, Lemma 9.4]. For each element E with global vertex indices $e_0, \ldots, e_{d-1}$, we denote by $\mathbf{w}_E = (w_{e_0}, \ldots, w_{e_{d-1}})^\top \in \mathbb{R}^d$ its local coefficient vector, with $w_{e_l} = 0$ if the corresponding vertex $\mathbf{x}_{e_l,m} \in \partial\Gamma_m \cup \overline{\Gamma}_D$. Thus, we obtain

$$\begin{aligned} \|w_h\|^2_{L_2(E)} &= \int\limits_E |w_h|^2 \, \mathrm{d}s_{\mathbf{x}} = \sum_{k,l=0}^{d-1} w_{e_k} \, w_{e_l} \int\limits_E \varphi_{e_k} \, \varphi_{e_l} \, \mathrm{d}s_{\mathbf{x}} \\ &= \mathbf{w}_E^\top \, M_E \, \mathbf{w}_E \simeq |E| \; \mathbf{w}_E^\top \, \mathbf{w}_E \end{aligned} \tag{5.17}$$

for each $E \in \mathcal{I}_{h,m}$. Since each vertex belongs to a fixed number of elements E, we obtain

$$\begin{aligned}\|w_h\|^2_{L_2(\Gamma_m)} &= \sum_{E\in\mathcal{I}_{h,m}} \|w_h\|^2_{L_2(E)} \simeq \sum_{E\in\mathcal{I}_{h,m}} h_E^{d-1} \sum_{k=0}^{d-1} w_{e_k}^2 \\ &\simeq \sum_{k=1}^{\mathring{N}_{V,m}} h_k^{d-1}\, w_k^2 \simeq \sum_{k=1}^{\mathring{N}_{V,m}} \int\limits_{\operatorname{supp}\psi_{k,m}} |w_k\,\psi_{k,m}|^2 \,\mathrm{d}s_{\mathbf{x}} = \|\omega_h\|^2_{L_2(\Gamma_m)}\end{aligned}$$

which proves the desired result. By h_k we denote the maximum of the h_E with $E \in \mathcal{I}_{h,m}$ attached to the vertex $\mathbf{x}_{k,m}$. ■

This auxiliary result is essential to prove the following equivalence.

Lemma 5.13. *Let $w_h \in \mathring{W}_{h,m}$ and $\omega_h \in Q_{h,m}$ be as in Lemma 5.12, then there holds*

$$1 \simeq \frac{(w_h,\omega_h)_{L_2(\Gamma_m)}}{\|w_h\|_{Q_{2,m}'}\,\|\omega_h\|_{Q_{2,m}}}$$

with constants independent of the mesh size h.

Proof. The idea of this proof is to estimate the eigenvalues of the element matrices to show the desired result, see for example [61] for further information.

For $w_h \in \mathring{W}_{h,m}$ and $\omega_h \in Q_{h,m}$ the $L_2(\Gamma_m)$–inner product is given as

$$\begin{aligned}(w_h,\omega_h)_{L_2(\Gamma_m)} &= \int\limits_{\Gamma_m} w_h\,\omega_h \,\mathrm{d}s_{\mathbf{x}} = \sum_{E\in\mathcal{I}_{h,m}} \int\limits_E w_h\,\omega_h\,\mathrm{d}s_{\mathbf{x}} \\ &= \sum_{E\in\mathcal{I}_{h,m}} \sum_{k,l=1}^{\mathring{N}_{V,m}} w_k\,w_l \int\limits_E \psi_{k,m}\,\varphi_{l,m}\,\mathrm{d}s_{\mathbf{x}} = \sum_{E\in\mathcal{I}_{h,m}} \mathbf{w}_E^\top M_{E,1}\,\mathbf{w}_E\end{aligned}$$

with the local coefficient vector $\mathbf{w}_E$ and the mixed mass matrix $M_{E,1}$ with entries

$$(M_{E,1})_{kl} := \int\limits_E \psi_{e_k}\,\varphi_{e_l}\,\mathrm{d}s_{\mathbf{x}}.$$

Next, we want to estimate $\mathbf{w}_E^\top M_{E,1}\,\mathbf{w}_E$, therefore we distinguish between the two and three space dimensional case. We consider first the two dimensional case.

d = 2: Let $E \in \mathcal{I}_{h,m}$ be an arbitrary edge element. For E, either no vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ or one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$. If one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we have

$$M_{E,1} = \frac{|E|}{2}$$

with eigenvalue $\lambda_0 = |E| \, \frac{1}{2}$. If no vertex of E is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ we obtain

$$M_{E,1} = \frac{|E|}{8} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

with eigenvalues $\lambda_0 = |E| \, \frac{1}{2}$ and $\lambda_1 = |E| \, \frac{1}{4}$.

d = 3: In the three dimensional case, we have to consider four cases for a element E. Either no vertex of E is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ or two vertices are on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$. The fourth case is that all three vertices are on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ which is due the construction of the spaces not possible for $d = 2$. The latter case can be neglected, since the integral $\int_E \psi_{k,m} \, \varphi_{l,m} \, \mathrm{d}s_{\mathbf{x}} = \int_E \varphi_{l,m} \, \mathrm{d}s_{\mathbf{x}}$ vanishes due to $\operatorname{supp} \varphi_{l,m} \cap E = \emptyset$ for all $1 \le l \le \mathring{N}_{V,m}$. If two vertices belong to $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we obtain

$$M_{E,1} = \frac{|E|}{3}$$

with eigenvalue $\lambda_0 = |E| \, \frac{1}{3}$. If one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we obtain

$$M_{E,1} = \frac{|E|}{12} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

with eigenvalues $\lambda_0 = |E| \, \frac{1}{3}$ and $\lambda_1 = |E| \, \frac{1}{6}$. Otherwise we have

$$M_{E,1} = \frac{|E|}{108} \begin{pmatrix} 22 & 7 & 7 \\ 7 & 22 & 7 \\ 7 & 7 & 22 \end{pmatrix}$$

with eigenvalues $\lambda_0 = |E| \, \frac{1}{3}$ and $\lambda_1 = \lambda_2 = |E| \, \frac{5}{36}$ if no vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$.

The case–by–case analysis implies $\mathbf{w}_E^\top M_{E,1} \, \mathbf{w}_E \simeq |E| \, \mathbf{w}_E^\top \mathbf{w}_E$ and (5.17) implies

$$(w_h, \omega_h)_{L_2(\Gamma_m)} \simeq \sum_{E \in \mathcal{I}_{h,m}} |E| \, \mathbf{w}_E^\top \mathbf{w}_E \simeq \sum_{E \in \mathcal{I}_{h,m}} \mathbf{w}_E^\top M_E \, \mathbf{w}_E = \|w_h\|^2_{L_2(\Gamma_m)}.$$

If we apply Lemma 5.12 we obtain

$$(w_h, \omega_h)_{L_2(\Gamma_m)} \simeq \|w_h\|_{L_2(\Gamma_m)} \, \|\omega_h\|_{L_2(\Gamma_m)}$$

and the global quasi–uniformity of the interface discretization finishes the proof. ∎

With Lemma 5.13 it is easy to see that $\Pi_m : {Q_{2,m}}' \to \mathring{W}_{h,m}$ given by (5.16) is well defined and satisfies

$$c \left\| \Pi_m v \right\|_{{Q_{2,m}}'} \leq \left\| v \right\|_{{Q_{2,m}}'}$$

with $c > 0$ independent of the mesh size h. The stability of the linear mortar projection is essential in the proof of the following lemma.

Lemma 5.14. *Let the discrete trial spaces X_h and Q_h be as in Definition 5.5 and Definition 5.8. Furthermore, assume that the discrete extension operator satisfyies the stability estimate (5.13). Then, there exists a positive constant $\alpha_B > 0$, such that*

$$\sup_{v_h \in X_h} \frac{b(v_h, \nu_h)}{\left\| v_h \right\|_X} \geq \alpha_B \left\| \nu_h \right\|_{Q_2}$$

holds for all $\nu_h \in Q_h$.

Proof. For a proof of Lemma 5.14 see [71, Lemma 1.9]. ∎

Lemma 5.14 provides the stability estimate (5.15c) in Theorem 5.11) which was rather easy to show and follows directly from the mortar finite element framework. The next step is to prove the stability of the nonlinear coupling condition of the linearized variational problem (5.8).

Stability of the nonlinear coupling: To show the inf–sup–stability of the nonlinear coupling condition, we define a weighted mortar projection. For each interface Γ_m the weighted mortar projection $\Psi_m : L_2(\Gamma_m) \to \mathring{W}_{h,m}$ is defined by

$$\int\limits_{\Gamma_m} (\varrho_m \Psi_m v - v) \nu_h \, \mathrm{d}s_{\mathbf{x}} = 0 \tag{5.18}$$

for all $\nu_h \in Q_{h,m}$ and some weight function $\varrho_m \in L_\infty(\Gamma_m) \cap L_\infty^+(\Gamma_m)$. The goal is to prove an equivalence similar to the one in Lemma 5.13 which was essential to show the stability of the linear coupling. To show such an equivalence we define

$$r_{E,m} := \min_{\mathbf{x} \in E} \varrho_m(\mathbf{x}) > 0 \qquad \text{and} \qquad R_{E,m} := \max_{\mathbf{x} \in E} \varrho_m(\mathbf{x}) < \infty$$

for each element $E \in \mathcal{I}_{h,m}$.

Lemma 5.15. *Let $w_h \in \mathring{W}_{h,m}$ and $\omega_h \in Q_{h,m}$ be as in Lemma 5.12. If there exists a $\rho > 0$ such that*

$$\rho \leq \begin{cases} (3\, r_{E,m} - R_{E,m}), & d = 2, \\ (\frac{11}{7}\, r_{E,m} - R_{E,m}), & d = 3 \end{cases} \tag{5.19}$$

holds, then

$$1 \simeq \frac{(\varrho_m\, w_h, \omega_h)_{L_2(\Gamma_m)}}{\|w_h\|_{Q_{2,m}{}'}\, \|\omega_h\|_{Q_{2,m}}}.$$

The constants in the above equivalence depend on the weight function ϱ_m and ρ but not on the mesh size h.

Proof. The idea of the proof is similar to the idea in the proof of Lemma 5.13. We first consider the $L_2(\Gamma_m)$–inner product, that is

$$\begin{aligned}(\varrho_m\, w_h, \omega_h)_{L_2(\Gamma_m)} &= \int\limits_{\Gamma_m} \varrho_m\, w_h\, \omega_h \,\mathrm{d}s_{\mathbf{x}} = \sum_{E\in\mathcal{I}_{h,m}} \int\limits_{E} \varrho_m\, w_h\, \omega_h \,\mathrm{d}s_{\mathbf{x}} \\ &= \sum_{E\in\mathcal{I}_{h,m}} \sum_{k,l=1}^{\mathring{N}_{V,m}} w_k\, w_l \int\limits_{E} \varrho_m\, \psi_{k,m}\, \varphi_{l,m} \,\mathrm{d}s_{\mathbf{x}} = \sum_{E\in\mathcal{I}_{h,m}} \mathbf{w}_E^\top M_{E,2}\, \mathbf{w}_E\end{aligned}$$

with the weighted mixed mass matrix $M_{E,2}$. The entries of $M_{E,2}$ are given by

$$(M_{E,2})_{kl} := \int\limits_{E} \varrho_m\, \psi_{e_k}\, \varphi_{e_l} \,\mathrm{d}s_{\mathbf{x}}$$

and it is easy to verify that $M_{E,2}$ is not symmetric anymore. Thus, we use the representation

$$\mathbf{w}_E^\top M_{E,2}\, \mathbf{w}_E = \mathbf{w}_E^\top M_{E,2}^{sym}\, \mathbf{w}_E.$$

with $M_{E,2}^{sym} := \frac{1}{2}\,(M_{E,2} + M_{E,2}^\top)$. The essential step in the proof of Lemma 5.13 was the computation of the eigenvalues of the mixed matrix $M_{E,1}$ which was done in a straight forward way. This can not be done for the matrix $M_{E,2}^{sym}$ explicitly, thus we just estimate the eigenvalues. The estimates are based on the following property of the entries $(M_{E,2})_{kl}$, namely

$$0 \le r_{E,m}\,(M_{E,1})_{kl} \le (M_{E,2})_{kl} \le R_{E,m}\,(M_{E,1})_{kl}.$$

If we exploit the symmetry of $M_{E,1}$ we obtain

$$0 \le r_{E,m}\,(M_{E,1})_{kl} \le (M_{E,2}^{sym})_{kl} \le R_{E,m}\,(M_{E,1})_{kl}.$$

Next, we will estimate the eigenvalues of $M_{E,2}^{sym}$ using the theorem of Gerschgorin, see [34, Satz II]. The symmetry of $M_{E,2}^{sym}$ ensures that all eigenvalues are real numbers in the union of all discs D_i with center $(M_{E,2}^{sym})_{ii}$ and radius $R_i = \sum_{j\neq i}(M_{E,2}^{sym})_{ij}$. For this reason we will estimate the expressions $D_i - R_i$ and $D_i + R_i$. We will again distinguish between the two and three space dimensional case.

d = 2: Let $E \in \mathcal{I}_{h,m}$ be an arbitrary edge element. For E, either no vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ or one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$. If one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we obtain

$$M_{E,2} = M_{E,2}^{sym} = \int_E \varrho_m \, \varphi_{e_0,m} \, \mathrm{d}s_{\mathbf{x}}$$

and thus $r_{E,m} \, |E| \, \frac{1}{2} \leq \lambda_0 \leq R_{E,m} \, |E| \, \frac{1}{2}$. If no vertex of E is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$ we obtain $M_{E,2} \in \mathbb{R}^{2\times 2}$. The first Gerschgorin circle can be bounded from below by

$$\begin{aligned} 0 < \frac{\rho}{8} \, |E| &\leq \frac{3\, r_{E,m} - R_{E,m}}{8} \, |E| = r_{E,m} \, \frac{3}{8} \, |E| - \frac{1}{2} \left(R_{E,m} \, \frac{1}{8} \, |E| + R_{E,m} \, \frac{1}{8} \, |E| \right) \\ &\leq r_{E,m} \, (M_{E,1})_{11} - \frac{1}{2} \Big(R_{E,m} \, (M_{E,1})_{12} + R_{E,m} \, (M_{E,1})_{21} \Big) \\ &\leq (M_{E,2})_{11} - \frac{1}{2} \Big((M_{E,2})_{12} + (M_{E,2})_{21} \Big) = D_1 - R_1 \end{aligned}$$

and for the upper bound we have

$$\begin{aligned} D_1 + R_1 &= (M_{E,2})_{11} + \frac{1}{2} \Big((M_{E,2})_{12} + (M_{E,2})_{21} \Big) \\ &\leq R_{E,m} \, (M_{E,1})_{11} + \frac{1}{2} \Big(R_{E,m} \, (M_{E,1})_{12} + R_{E,m} \, (M_{E,1})_{21} \Big) \leq R_{E,m} \frac{1}{2} \, |E| \, . \end{aligned}$$

The lower bound $D_2 - R_2$ and the upper bound $D_2 + R_2$ of the second Gerschgorin circle can be estimated in the same way and we get the bounds $\frac{\varrho}{8} \, |E| \leq \lambda_0, \lambda_1 \leq R_{E,m} \frac{1}{2} \, |E|$ for the eigenvalues of $M_{E,2}^{sym}$.

d = 3: According to the proof of Lemma 5.13, we just have to consider the following three cases. If two vertices belong to $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we obtain

$$M_{E,2} = M_{E,2}^{sym} = \int_E \varrho_m \, \varphi_{e_0,m} \, \mathrm{d}s_{\mathbf{x}}$$

and thus $r_{E,m} \, |E| \, \frac{1}{3} \leq \lambda_0 \leq R_{E,m} \, |E| \, \frac{1}{3}$. If one vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we have $M_{E,2} \in \mathbb{R}^{2\times 2}$ and as in the two dimensional case we get from the first row

$$\begin{aligned} 0 < \rho \left(\frac{1}{12} + \frac{(10/7)}{12} \right) |E| &\leq \frac{(11/7) \; r_{E,m} - R_{E,m}}{12} \, |E| + \frac{(10/7) \; r_{E,m}}{12} \, |E| \\ &\leq \frac{3\, r_{E,m} - R_{E,m}}{12} \, |E| \leq D_1 - R_1 \end{aligned}$$

and

$$D_1 + R_1 \leq R_{E,m} \frac{1}{3} \, |E| \, .$$

The bounds of the second Gerschgorin circle can be estimated in the same way. In this case we can bound the eigenvalues of $M_{E,2}^{sym}$ by $\frac{17\rho}{84}\,|E| \leq \lambda_0, \lambda_1 \leq R_{E,m}\frac{1}{3}\,|E|$. If no vertex is on $\overline{\Gamma}_C \cup \overline{\Gamma}_D$, we have to bound three Gerschgorin circles. The procedure is similar to the previous estimates. The lower bound for the first circle is

$$\begin{aligned} 0 < \frac{7\,\rho}{54}\,|E| \leq \frac{11\,r_{E,m} - 7\,R_{E,m}}{54}\,|E| &\leq r_{E,m}\,\frac{22}{108}\,|E| - R_{E,m}\frac{14}{108}\,|E| \\ &= r_{E,m}\,\frac{22}{108}\,|E| - \frac{1}{2}\,R_{E,m}\frac{28}{108}\,|E| \leq D_1 - R_1 \end{aligned}$$

and the upper bound is given by

$$D_1 + R_1 \leq R_{E,m}\frac{1}{3}\,|E|\,.$$

For the remaining Gerschgorin circles we obtain the same bounds, therefore we can bound the eigenvalues by $\frac{7\rho}{54}\,|E| \leq \lambda_0, \lambda_1, \lambda_2 \leq R_{E,m}\frac{1}{3}\,|E|$.

The case–by–case analysis implies $\mathbf{w}_E^\top\,M_{E,2}\,\mathbf{w}_E = \mathbf{w}_E^\top\,M_{E,2}^{sym}\,\mathbf{w}_E \simeq |E|\;\mathbf{w}_E^\top\,\mathbf{w}_E$ and (5.17) implies

$$\left(\varrho_m\,w_h, \omega_h\right)_{L_2(\Gamma_m)} \simeq \sum_{E\in\mathcal{I}_{h,m}} |E|\;\mathbf{w}_E^\top\,\mathbf{w}_E \simeq \sum_{E\in\mathcal{I}_{h,m}} \mathbf{w}_E^\top\,M_E\,\mathbf{w}_E = \left\|w_h\right\|^2_{L_2(\Gamma_m)}.$$

If we apply Lemma 5.12 we obtain

$$\left\|w_h\right\|_{L_2(\Gamma_m)} \simeq \frac{\left(\varrho_m\,w_h, \omega_h\right)_{L_2(\Gamma_m)}}{\left\|\omega_h\right\|_{L_2(\Gamma_m)}}$$

and if we incorporate the global quasi–uniformity of the interface discretization, we have

$$\left\|w_h\right\|_{Q_{2,m}'} \simeq \frac{\left(\varrho_m\,w_h, \omega_h\right)_{L_2(\Gamma_m)}}{\left\|\omega_h\right\|_{Q_{2,m}}}$$

which yields the desired result. ■

Lemma 5.15 seems to be quite restrictive since the weight function ϱ_m has to satisfy a certain smoothness on Γ_m. It is rather hard to state a general stability result for general weight functions, since the shape of the weight function has a massive influence on the eigenvalues of the matrix $M_{E,2}^{sym}$. We want to show the influence in the following example in which we compute the smallest eigenvalue of $M_{E,2}^{sym}$ for a set of different weight functions.

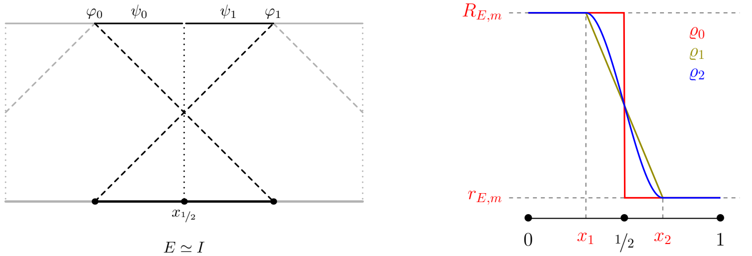
Example 5.16. *Consider an arbitrary edge E in $\mathbb{R}^2$ which can be transformed to the reference element $I = [0,1]$. On I we have*

$$\varphi_0(x) = 1 - x \qquad \textit{and} \qquad \varphi_1(x) = x$$

as well as

$$\psi_0(x) = \begin{cases} 1 & x < 1/2 \\ 0 & \textit{else} \end{cases} \qquad \textit{and} \qquad \psi_1(x) = \begin{cases} 1 & x > 1/2 \\ 0 & \textit{else} \end{cases}$$

for $x \in I$, see Figure 5.9a. Next, we consider three different nonlinear weights ϱ_0, ϱ_1 and ϱ_2. We choose ϱ_0 to be piecewise constant, ϱ_1 is continuous and piecewise linear and ϱ_2 is assumed to be differentiable and piecewise cubic, see Figure 5.9b. For all



(a) Basis functions on E. (b) Different weights ϱ_m.

experiments we fix $r_{E,m} = 1$ and vary the upper bound $R_{E,m}$ of the weight function as well as the interval $G := (x_1, x_2)$ in which the weight function is not constant. We denote by $x_{1/2} = 1/2$ the center point of the reference element I. This point is of special interest since we switch between ψ_0 and ψ_1 at $x_{1/2}$.

We compute the smallest eigenvalue of the symmetric and weighted mixed mass matrix $M_{E,2}^{sym}$ given by

$$M_{E,2}^{sym} := \frac{1}{2} \begin{pmatrix} 2 \int_0^1 \varrho_i \, \psi_0 \, \varphi_0 \, \mathrm{d}s_{\mathbf{x}} & \int_0^1 \varrho_i \left(\psi_0 \, \varphi_1 + \psi_1 \, \varphi_0 \right) \varrho_i \, \mathrm{d}s_{\mathbf{x}} \\ \int_0^1 \varrho_i \left(\psi_1 \, \varphi_0 + \psi_0 \, \varphi_1 \right) \mathrm{d}s_{\mathbf{x}} & 2 \int_0^1 \varrho_i \, \psi_1 \, \varphi_1 \, \mathrm{d}s_{\mathbf{x}} \end{pmatrix}$$

as it was defined in the proof of Lemma 5.15.

For the first experiment we set $x_1 = 0.3$ and $x_2 = 0.7$. The non–constant part of the weight function is therefore located around the center point $x_{1/2}$ the of I. For an increasing upper bound $R_{E,m}$ we can compute the smallest eigenvalue of $M_{E,2}^{sym}$ with respect to the different weight functions, see Table 5.1. In the case of a piecewise

$R_{E,m}$	2	3	30	60	120	180	240
ϱ_0	0.2973	0.2993	0.0401	-0.2635	-0.8718	-1.4802	-2.0887
ϱ_1	0.3166	0.3452	0.8065	1.3023	2.2927	3.2830	4.2732
ϱ_2	0.3116	0.3332	0.6030	0.8864	1.4520	2.0174	2.5828

Table 5.1: Smallest eigenvalue of $M_{E,2}^{sym}$ with centered $G = (0.3, 0.7)$.

constant weight, the smallest eigenvalue becomes smaller and smaller if the upper bound of ϱ_0 becomes larger. In the case of a continuous and differentiable function, the smallest eigenvalue stays away from zero and seems to increase.

For the second experiment we shrink the interval G to $x_1 = 0.45$ and $x_2 = 0.55$. The interval G is still located around the center point $x_{1/2}$ but the length of G was decreased. The continuity of ϱ_1 and the differentiability of ϱ_2 are obviously preserved, but the gradient within G becomes steeper. The smallest eigenvalue of $M_{E,2}^{sym}$ is listed in Table 5.2. Obviously we obtain the same result as for the piecewise constant weight

$R_{E,m}$	2	3	30	60	120	180	240
ϱ_0	0.2973	0.2993	0.0401	-0.2635	-0.8718	-1.4802	-2.0887
ϱ_1	0.3019	0.3100	0.2162	0.0960	-0.1452	-0.3867	-0.6282
ϱ_2	0.3007	0.3073	0.1714	0.0045	-0.3301	-0.6650	-0.9999

Table 5.2: Smallest eigenvalue of $M_{E,2}^{sym}$ with centered $G = (0.45, 0.55)$.

function. In the other cases we see, that the eigenvalues become negative if we increase the upper bound of the corresponding weight function. Thus, the smallest eigenvalue depends on the gradient of the weight function.

For the third experiment shift G to the right by setting $x_1 = 0.55$ and $x_2 = 0.65$. The length of G is still the same, but the point $x_{1/2}$ is located outside of G. In Table 5.3 the computed smallest eigenvalues using this setup are listed. As we can see, the smallest

$R_{E,m}$	2	3	30	60	120	180	240
ϱ_0	0.3260	0.3698	1.2961	2.3114	4.3410	6.3705	8.4000
ϱ_1	0.3265	0.3708	1.3103	2.3403	4.3994	6.4583	8.5172
ϱ_2	0.3263	0.3704	1.3046	2.3287	4.3761	6.4232	8.4703

Table 5.3: Smallest eigenvalue of $M_{E,2}^{sym}$ with shifted $G = (0.55, 0.65)$.

eigenvalue of $M_{E,2}^{sym}$ stays away from zero and the values become larger for increasing upper bounds $R_{E,m}$. Thus, the location of G plays an important role for the smallest eigenvalues.

For the last experiment we shift G to the left by setting $x_1 = 0.35$ and $x_2 = 0.45$. The center point $x_{1/2}$ of I is again located outside of G. If we compute the smallest we obtain Table 5.4. Within this setting, we see that the eigenvalues become negative for

$R_{E,m}$	2	3	30	60	120	180	240
ϱ_0	0.3052	0.3150	0.2004	0.0534	-0.2417	-0.5370	-0.8324
ϱ_1	0.3050	0.3147	0.1985	0.0499	-0.2486	-0.5474	-0.8462
ϱ_2	0.3050	0.3148	0.1993	0.0513	-0.2458	-0.5432	-0.8407

Table 5.4: Smallest eigenvalue of $M_{E,2}^{sym}$ with shifted $G = (0.35, 0.45)$.

all three weight functions, even though the center point $x_{1/2}$ is located outside of G. Thus, even the location of the bulk influences the smallest eigenvalue of $M_{E,2}^{sym}$.

We see, that without the explicit knowledge of the shape of the weight function ϱ_m, it is hard check whether the equivalence in Lemma 5.15 is satisfied or not if ϱ_m does not satisfy condition (5.18).

With Lemma 5.15 it is again easy to see that $\Psi_m : {Q_{2,m}}' \to \mathring{W}_{h,m}$ given by (5.18) is well defined and satisfies

$$c \left\| \Psi_m v \right\|_{{Q_{2,m}}'} \leq \left\| v \right\|_{{Q_{2,m}}'}$$

with a $c > 0$ depending on ϱ_m and ρ but independent of the mesh size h.

Lemma 5.17. *Let the discrete trial spaces X_h and Q_h be as in Definition 5.5 and Definition 5.8. Furthermore, assume that the discrete extension operator satisfies the stability estimate (5.13). If on each interface Γ_m the weight function $\varrho_m := {\kappa_k^{-1}}'(w_{h,k|_{\Gamma_m}} + u_{D,h,k|_{\Gamma_m}})$ satisfies the condition (5.19), then there exists a positive constant $\alpha_C > 0$, such that*

$$\sup_{v_h \in X_h} \frac{c'(w_h + u_{D,h}, v_h, \nu_h)}{\left\| v_h \right\|_X} \geq \alpha_C \left\| \nu_h \right\|_{Q_2}$$

holds for all $\nu_h \in Q_h$.

Proof. The proof of Lemma 5.17 follows again the lines of the proof of [71, Lemma 1.9] using the weighted mortar projection defined by (5.18). ∎

Thus, Lemma 5.17 provides the stability estimate (5.15d) in Theorem 5.11. Unfortunately we were not able to prove the conditions (5.15a) and (5.15b) of Theorem 5.11. Thus, we have to assume that the bilinear form given by $m'(w_h + u_{D,h}, \cdot, \cdot) + a(\cdot, \cdot)$ induces a bijective operator between the two discrete kernels. In the numerical examples stability issues were only observed in cases of a mesh which was too coarse.

Remark 5.18. *Due to the discrete Newton linearization and the nonconforming space discretization, a convergence result with respect to the space discretization parameter h is still an open problem. In our numerical examples the convergenve of the Newton method is of second order and we observe a good convergence behavior as h tends to zero.*

This leads to the last section in this chapter where we briefly describe the implementation of the described method.

5.4 Implementation

In this section, we briefly discuss the implementation which was used to compute the approximate solution. Thus, we have to solve a sequence of the following linearized problem.

Newton iteration for the modified discrete variational problem

For $w_h \in X_h$ find $u_h \in X_h$ and $\lambda_h \in Q_h$ such that

$$\begin{aligned} m'(w_h + u_{D,h}, u_h, v_h) + a(u_h, v_h) + b(v_h, \lambda_h) &= f(w_h, v_h) \\ c'(w_h + u_{D,h}, u_h, \nu_h) &= g(w_h, \nu_h) \end{aligned}$$

for all $v_h \in X_h$ and $\nu_h \in Q_h$.

The bilinear forms in the variational problem are

$$\begin{aligned} m'(w_h + u_{D,h}, u_h, v_h) &= \sum_{i=1}^{N} \int_{\Omega_i} l_i{}'(w_{h,i} + u_{D,h,i})\, u_{h,i}\, v_{h,i}\, \mathrm{d}\mathbf{x}, \\ a(u_h, v_h) &= \sum_{i=1}^{N} \int_{\Omega_i} \nabla u_{h,i} \cdot \nabla v_{h,i}\, \mathrm{d}\mathbf{x}, \\ b(v_h, \lambda_h) &= -\sum_{\Gamma_m} \int_{\Gamma_m} [\![v_h]\!]_{\Gamma_m}\, \lambda_h\, \mathrm{d}s_{\mathbf{x}}, \\ c'(w_h + u_{D,h}, u_h, \nu_h) &= -\sum_{\Gamma_m} \int_{\Gamma_m} [\![\kappa^{-1\prime}(w_h + u_{D,h})\, u_h]\!]_{\Gamma_m} \nu_h\, \mathrm{d}s_{\mathbf{x}} \end{aligned} \tag{5.20}$$

for $w_h, u_h, v_h \in X_h$ and $\lambda_h, \nu_h \in Q_h$. The linear form in the right hand side are as in Subsection 5.1.2 given by

$$\begin{aligned} &f(w_h, v_h) = f(v_h) - a(u_{D,h}, v_h) + m'(w_h + u_{D,h}, w_h, v_h) - m(w_h + u_{D,h}, v_h), \\ &g(w_h, \nu_h) = c'(w_h + u_{D,h}, w_h, \nu_h) - c(w_h + u_{D,h}, v_h) \end{aligned} \tag{5.21}$$

for $w_h, v_h \in X_h$ and $\nu_h \in Q_h$. Due to Definition 5.5 and Definition 5.8, the trial spaces X_h as well as Q_h can be written as

$$X_h = \prod_{i=1}^{N} \operatorname{span}\{\varphi_{k,i}\}_{k=1}^{\mathring{N}_{V,i}} = \operatorname{span}\{\varphi_k\}_{k=1}^{N_X}$$

with $N_X := \dim X_h = \sum_{i=1}^{N} \mathring{N}_{V,i}$ and

$$Q_h = \prod_{\Gamma_m} \operatorname{span}\{\psi_{k,m}\}_{k=1}^{\mathring{N}_{V,m}} = \operatorname{span}\{\psi_k\}_{k=1}^{N_M}$$

with $N_M := \dim Q_h = \sum_{\Gamma_m} \mathring{N}_{V,m}$. Thus, we can express the approximate solution $(u_h, \lambda_h) \in X_h \times Q_h$ as

$$u_h = \sum_{k=1}^{N_X} u_k\, \varphi_k \quad \text{and} \quad \lambda_h = \sum_{k=1}^{N_M} \lambda_k\, \psi_k$$

with $\mathbf{u} \in \mathbb{R}^{N_X}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{N_M}$. By plugging the representation of the approximate solution into the linear and bilinear forms (5.21) and (5.20) we obtain the corresponding equivalent discrete linear system by testing with each basis function of X_h and Q_h,

$$\left(\begin{array}{cccc|c} A_1 & 0 & \cdots & 0 & B_1^\top \\ 0 & A_2 & \cdots & 0 & B_2^\top \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_N & B_N^\top \\ \hline C_1 & C_2 & \cdots & C_N & 0 \end{array}\right) \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_N \\ \mathbf{g} \end{pmatrix}$$

for the unknown coefficient vectors $\mathbf{u} \in \mathbb{R}^{N_X}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{N_M}$. The block matrices A_i are square matrices that is, $A_i \in \mathbb{R}^{\mathring{N}_{V,i} \times \mathring{N}_{V,i}}$ and they are independent of each other. The coupling is realized by the matrices $B_i^\top \in \mathbb{R}^{\mathring{N}_{V,i} \times N_M}$ and $C_i \in \mathbb{R}^{N_M \times \mathring{N}_{V,i}}$. The independency of the blocks A_i allows the following Schur complement reformulation of the above system. That is, instead of solving the entire $(N_X + N_M) \times (N_X + N_M)$ system, we solve

$$S\,\boldsymbol{\lambda} = \widetilde{\mathbf{g}}$$

with the Schur complement $S \in \mathbb{R}^{N_M \times N_M}$ and $\widetilde{\mathbf{g}} \in \mathbb{R}^{N_M}$ given by

$$S = \sum_{i=1}^{N} C_i A_i^\dagger B_i^\top \qquad \text{and} \qquad \widetilde{\mathbf{g}} = \sum_{i=1}^{N} C_i A_i^\dagger \mathbf{f}_i - \mathbf{g}.$$

By $A_i^\dagger$ we denote the pseudo inverse of A_i which can be computed for each subdomain in parallel. In view of efficiency, we implemented the solving algorithm into our in–house C++ software using a distributed memory model which was realized by using the OpenMPI implementation of the Message–Passing–Interface, see [33]. To obtain a decomposition of our domain mesh, we used the external GMSH tool, see [35], which allows to create two and three dimensional meshes and to partition them. Each partition was saved separately to provide one single mesh file for each parallel MPI–thread. To speed up the local assembling routines we used a shared memory parallelization using OpenMP, see [24]. To compute the pseudo inverse we use the PARDISO solver, see [44, 56]. A pseudo code of the solving routine is depicted in Algorithm 1, which we want to discuss briefly.

Algorithm 1 Solving algorithm for given $M \in \mathbb{N}$, $\varepsilon > 0$, initial condition $\mathbf{u}_i^0$ and initial guess $\mathbf{l}_i^0$

1: loadMesh()
2: ***exchangeInterfaceData***()
3: $t = 0.0$
4: $\widetilde{\mathbf{u}}_i = \mathbf{u}_i^0$
5: $\widetilde{\mathbf{l}}_i = \mathbf{l}_i^0$
6: **for** $m = 1$ **to** M **do**
7:     $t = t + \tau_m$
8:     **repeat**
9:         $A_i,\, B_i,\, C_i,\, \mathbf{f}_i^m,\, \mathbf{g}_i^m = \mathit{assembleSystem}(\widetilde{\mathbf{u}}_i,\, \mathbf{u}_i^{m-1})$
10:         $A_i^{-1} = \mathit{pseudoInverse}(A_i)$
11:         $S_i = C_i\, A_i^{-1}\, B_i^\top$
12:         $\widetilde{\mathbf{g}}_i = C_i\, A_i^{-1}\, \mathbf{f}_i^m - \mathbf{g}_i^m$
13:         $\mathbf{l}_i^m = \boldsymbol{parallelGMRES}(S_i,\, \widetilde{\mathbf{g}}_i,\, \widetilde{\mathbf{l}}_i)$
14:         $\mathbf{u}_i^m = A_i^{-1}\, \mathbf{f}_i^m - A_i^{-1}\, B_i^\top\, \mathbf{l}_i^m$
15:         $\mathit{tol} = \boldsymbol{terminationCriteria}(S_i,\, \widetilde{\mathbf{g}}_i,\, \widetilde{\mathbf{l}}_i,\, \mathbf{l}_i^m)$
16:         $\widetilde{\mathbf{u}}_i = \mathbf{u}_i^m$
17:         $\widetilde{\mathbf{l}}_i = \mathbf{l}_i^m$
18:     **until** $\mathit{tol} < \varepsilon$
19:     writeSolution()
20: **end for**

Before starting the Algorithm 1, we have to create the mesh partitions using the GMSH tool. In **Line 1** each thread loads the corresponding mesh partition which was created by GMSH in advance. Then, the interface information of the local partitions is exchanged, see **Line 2**. Here we have global communication. In **Line 3–5** we set the current time $t = 0.0$ and initialize the local previous Newton iteration $\widetilde{\mathbf{u}}_i$ and $\widetilde{\mathbf{l}}_i$. Then, we start the time stepping scheme in **Lines 6** and update the current time in **Lines 7**. In **Line 8** we start the Newton iteration. We can assemble the local system matrices and vectors independent of the other MPI–threads in **Line 9** and compute the corresponding local Schur complement system in **Line 10–12**. To solve the global Schur complement system, **Line 13**, we use a parallel iterative solver which needs to perform global communication. Since the Schur complement system is non symmetric, we use a parallel GMRES solver, see for example [38, Section 6.2]. As an initial guess for the solver, we use the previous Newton iteration $\widetilde{\mathbf{l}}_i$. From $\mathbf{l}_i^m$ we can compute the local vectors $\mathbf{u}_i^m$, see **Line 14**. To check the termination criteria for the Newton method, **Line 15**, global communication has to be performed. In **Line 16, 17** we update the previous Newton iterations $\widetilde{\mathbf{u}}_i$ and $\widetilde{\mathbf{l}}_i$. If the termination criteria is satisfied, we stop the Newton iteration and step into the next time step, **Line 6**. If the criteria is not satisfied, we have to perform a further Newton iteration, **Line 8**. The parallel work flow is depicted in Figure 5.10.

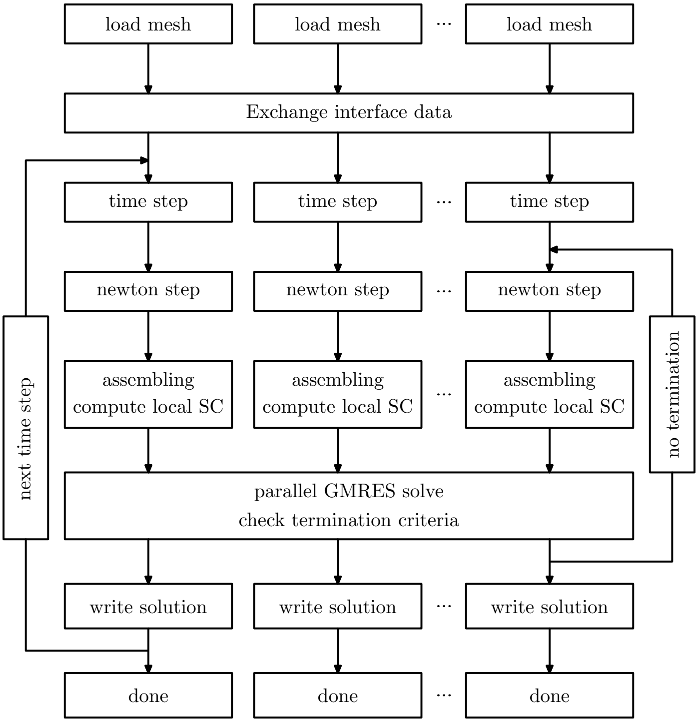


Figure 5.10: Parallel program flow chart.

Some word to summarize this chapter. In the beginning of this chapter we did some modifications of the variational problem (4.44) we derived in Section 4.4. These modifications were necessary to fit the primal hybrid formulation into the mortar finite element context. In Section 5.1 we discussed the linearization of the modified variational problem using Newton's method. Next, in Section 5.2, the construction of proper trial spaces for $d = 2, 3$ was discussed. After that, we wanted to answer the question whether the discrete problem if solvable or not, see Section 5.3. In the last section, Section 5.4, we briefly described the algorithm we have implemented to solve the derived discrete problem. In the next chapter we present numerical experiments in two and three space dimensions.

6 NUMERICAL EXPERIMENTS

In this last chapter we present numerical experiments which were computed using the method described in Chapter 4 and Chapter 5. That is, we approximate the generalized pressure $u \in X$ and $\lambda \in Q$ by solving a sequence of discrete and linearized variational problems of the form (5.8). Since we are interested in the physical pressure we have to apply the inverse Kirchhoff transformation as discussed in Section 4.3 and Section 4.4.

The soil parameter we used for the computations are listed in Table 6.1. For the modification of the relative permeability k we choose $\alpha = 0.025$, see Definition 4.23. The gravitational constant g as well as the viscosity μ are normalized to one. In both

parameter	n	K	θ_{min}	θ_{max}	p_b	λ
sand	0.437	6.54e-5	0.046	0.94	-0.073	0.694
sandy loam	0.453	6.06e-6	0.091	0.94	-0.147	0.378
loam	0.463	3.67e-6	0.058	0.87	-0.112	0.252

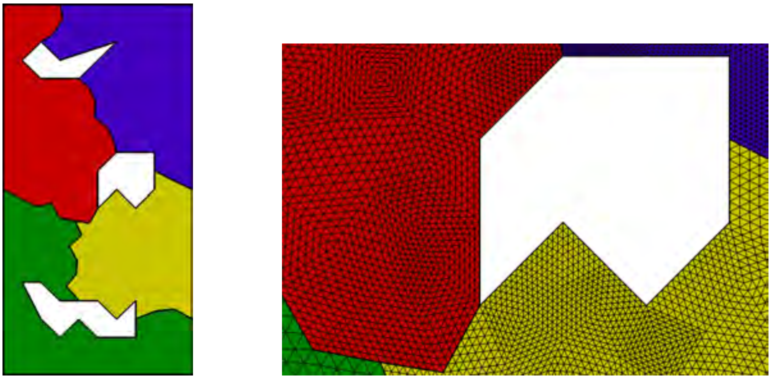
Table 6.1: Soil parameter.

experiments we choose an equidistant decomposition of the time interval $I = (0, T)$ with time step size $\tau = 1.0$. We further assume, that we do not have sinks or sources within our computational domain Ω, that is $f \equiv 0$.

In the following we want to present a two and three dimensional experiment respectively. First we consider the experiment in two space dimensions.

6.1 Experiment in 2D

For the experiment in two space dimensions, we consider the computational domain $\Omega \subset \mathbb{R}^2$ depicted in Figure 6.1. The soil parameter for each subdomain are chosen in the following way. The red (top left) and the yellow (bottom right) region behaves like sand, the blue (top right) region behaves like sandy loam and the green (bottom left) region behaves like loam. We have a Dirichlet boundary on the top, that is $\Gamma_D = (0, 1) \times \{2\}$ and the remaining boundary is considered as a Neumann boundary, that is $\Gamma_N = \partial\Omega \setminus \Gamma_D$. The method discussed in Chapter 5 allows different discretiza-



(a) Decomposition. (b) Triangulation of the subdomains.

Figure 6.1: Computational domain.

tion of each subdomain Ω_i. For this experiment the discretization is depicted in Figure 6.1b.

According to Section 2.3, we want to apply a pressure on the Dirichlet boundary, that is

$$g_D(\mathbf{x}, t) := \begin{cases} -5\left(1 - \frac{t}{10}\right), & t < 10, \\ 0, & t \geq 10 \end{cases}$$

for all $\mathbf{x} \in \Gamma_D$ and $t > 0$. On Γ_N we consider a no outflow condition, that is $g_N(\mathbf{x}, t) \equiv 0$. Since we solve the Kirchhoff transformed variational problem (5.8), we have to transform the Dirichlet boundary condition as discussed in Section 4.3, that is

$$h_{D,i}(\mathbf{x}, t) := \kappa_i(g_D(\mathbf{x}, t))$$

for $\mathbf{x} \in \Gamma_{D,i}$ and $t > 0$. Due to our considerations made in Section 4.3 we can compute the Kirchhoff transformed Dirichlet datum h_D explicitly. The Neumann boundary condition remains unchanged when applying the Kirchhoff transformation. The initial datum is given by the constant $p_{0,h} \equiv -5$ and thus $u_{0,h,i} = \kappa_i^{-1}(p_{0,h,i})$ for the subdomains, see Figure 6.2. The end time T is set to $T = 62000$ for this computation. We compute the generalized pressure $u_h + u_{D,h}$ of the variational problem (5.8), and transform it to the physical pressure by applying the inverse Kirchhoff transformation, that is $p_{h,i} = \kappa_i^{-1}(u_{h,i} + u_{D,h,i})$ in each subdomain Ω_i. The solutions depicted for different time steps on the following pages show the physical pressure p_h on the left hand side and the computed discontinuous generalized pressure u_h on the right hand side. For a better identification of the pressure profile we plotted contour lines at several levels.

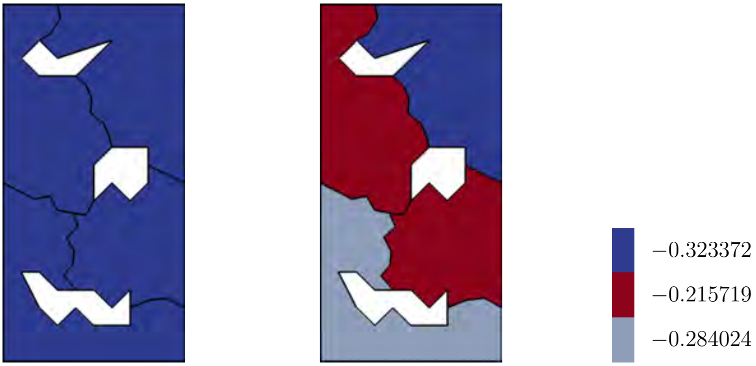
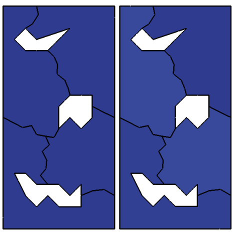


Figure 6.2: Initial datum $p_{0,h}$ and $u_{0,h}$.



Solution at $t = 0$.

Solution at $t = 20$.

Solution at $t = 60$.

Solution at $t = 100$.

Solution at $t = 140$.

Solution at $t = 200$.

Solution at $t = 270$.

Solution at $t = 340$.

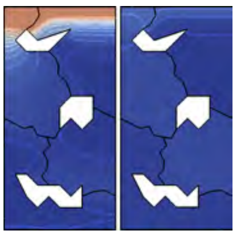
Solution at $t = 440$.

Solution at $t = 55$.

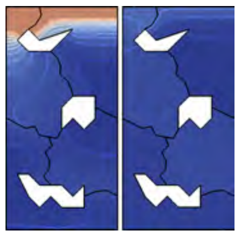
Solution at $t = 680$.

Solution at $t = 840$.

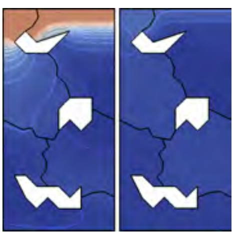
Solution at $t = 1020$.



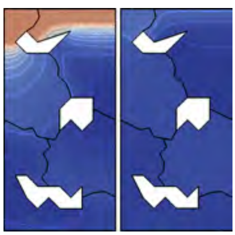
Solution at $t = 1250$.



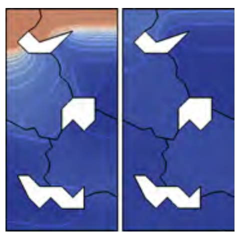
Solution at $t = 1510$.



Solution at $t = 1820$.

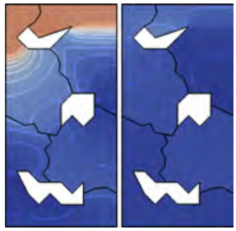


Solution at $t = 2190$.

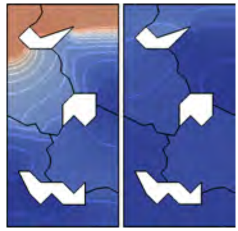


Solution at $t = 2630$.

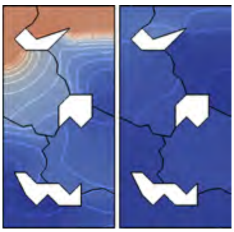
Solution at $t = 3150$.



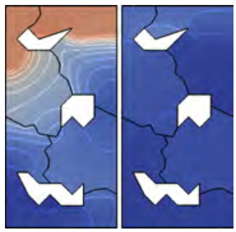
Solution at $t = 3770$.



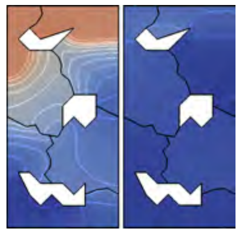
Solution at $t = 4510$.



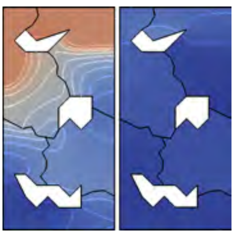
Solution at $t = 5380$.



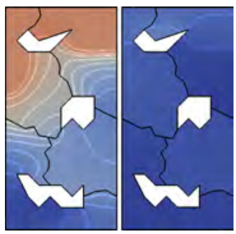
Solution at $t = 6420$.



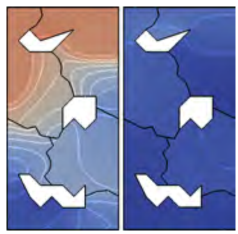
Solution at $t = 7650$.



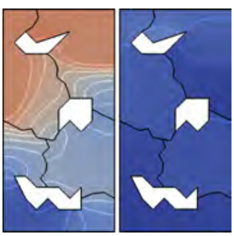
Solution at $t = 9110$.



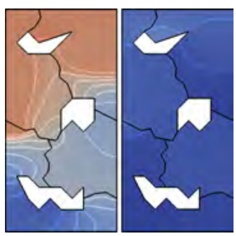
Solution at $t = 10850$.



Solution at $t = 12920$.



Solution at $t = 15370$.

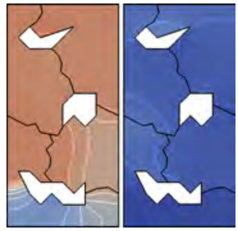


Solution at $t = 18280$.

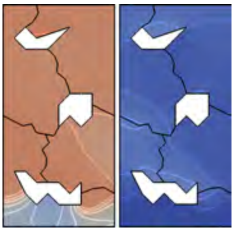
Solution at $t = 21730$.

Solution at $t = 25830$.

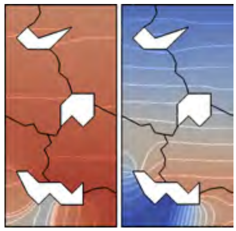
Solution at $t = 30710$.



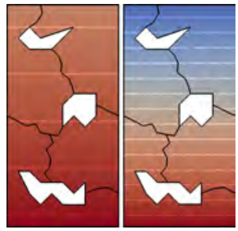
Solution at $t = 36490$.



Solution at $t = 43360$.

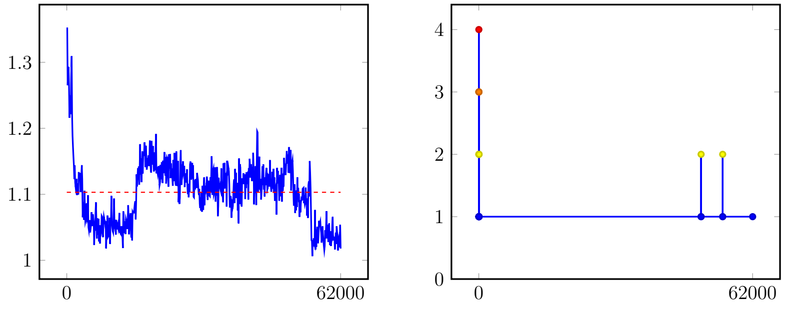


Solution at $t = 51520$.



Solution at $t = 60000$.

In the following we plotted the average runtime per time step on the left hand side and on the right hand side we plotted the number of Newton iterations per time step. We see, that the number of Newton iterations is for almost every time step equal



(a) Average runtime per time step. (b) Newton iteration per time step.

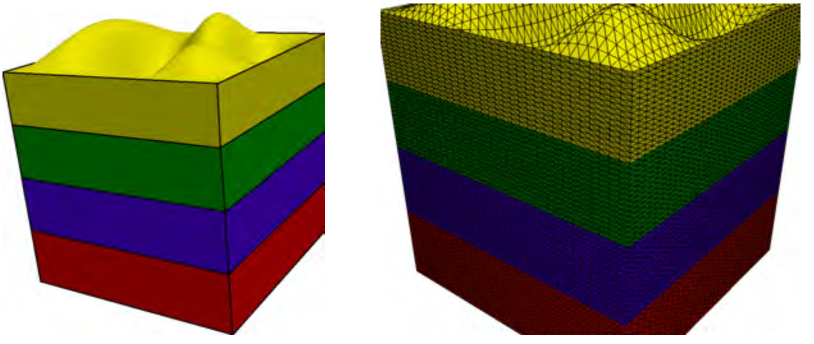
to one. We observe that the number of iterations is a little higher at the beginning which can be explained with the sudden increase of the pressure at Γ_D.

6.2 Experiment in 3D

For the experiment in three space dimensions, we consider the computational domain $\Omega \subset \mathbb{R}^3$ depicted in Figure 6.16. The computational domain is a transformation of the unit cube $Q = (0,1)^3 \subset \mathbb{R}^3$ which was decomposed in four layers of equal height, that is $\Omega = F(Q)$ for some prescribed $F : \mathbb{R}^3 \to \mathbb{R}^3$. In the three dimensional case we use a global uniform discretization of Ω, see Figure 6.16b. As in the two dimensional case, we have to specify the soil parameter of each subdomain. The red (first from the bottom) and yellow (first from the top) region behaves like sand, the blue (second from the bottom) region behaves like sandy loam and the green (second from the top) region behaves like loam. The corresponding parameters are listed in Table 6.1. We have a Dirichlet boundary Γ_D with $\Gamma_D = F(x_1, x_2, 1)$, $x_1, x_2 \in (0,1)$, and a Neumann boundary $\Gamma_N = \partial\Omega \setminus \Gamma_D$. Analogously to the two dimensional experiment we apply a pressure at Γ_D to simulate a surface water. The associated boundary condition is defined by

$$g_D(\mathbf{x}, t) := \begin{cases} \frac{t-10}{2}, & t < 10, \\ 0, & t \geq 10 \end{cases}$$

for all $\mathbf{x} \in \Gamma_D$ and $t > 0$. On Γ_N we consider a no outflow condition, that is $g_N(\mathbf{x}, t) \equiv 0$.



(a) Computational domain. (b) Triangulation of the subdomains.

Figure 6.16: Computational domain.

As done in Section 6.1, we have to transform the Dirichlet boundary condition since we solve the Kirchhoff transformed variational problem (5.8), that is

$$h_{D,i}(\mathbf{x},t) := \kappa_i(g_D(\mathbf{x},t))$$

for $\mathbf{x} \in \Gamma_{D,i}$ and $t > 0$. The initial datum is given by $p_{0,h} \equiv -5$ and for the generalize pressure we obtain $u_{0,h,i} = \kappa_i(p_{0,h})$, see Figure 6.17. The end time T for this experiment is $T = 20000$. In contrast to the two dimensional experiment we just depict the the

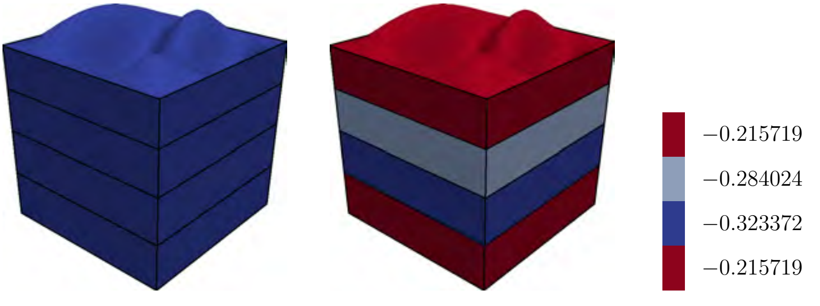
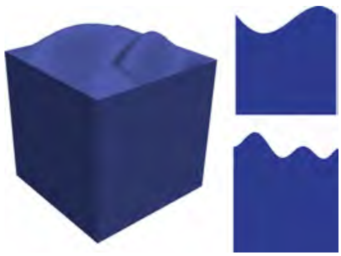
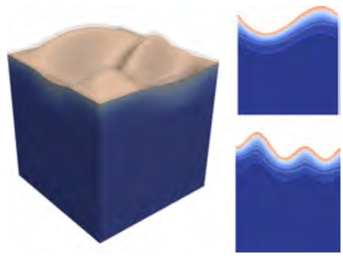


Figure 6.17: Initial datum $p_{0,h}$ and $u_{0,h}$.

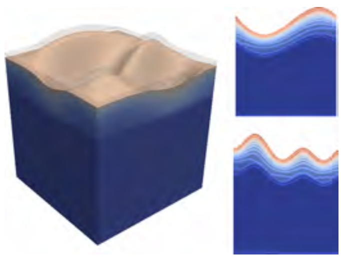
physical pressure p_h which is of interest. To illustrate the evolution of the pressure, we track the pressure at level -0.84. We also plot a vertical cut through the computational domain Ω in x–direction and y–direction respectively.



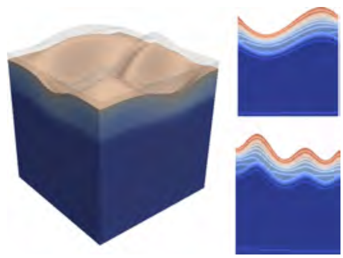
Solution at $t = 0$.



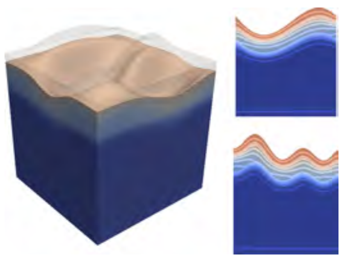
Solution at $t = 30$.



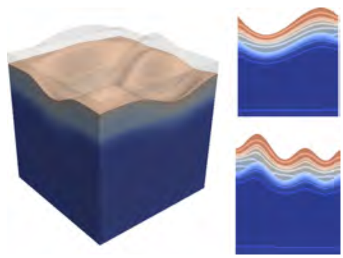
Solution at $t = 60$.



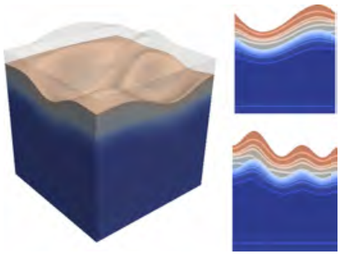
Solution at $t = 110$.



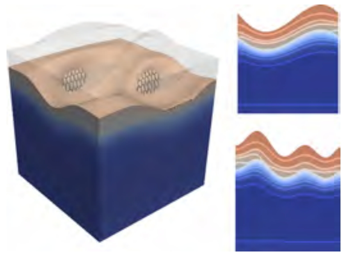
Solution at $t = 170$.



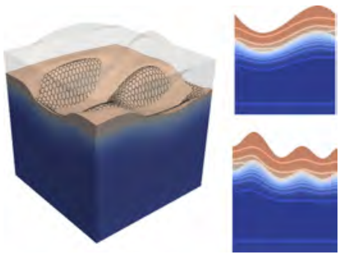
Solution at $t = 240$.



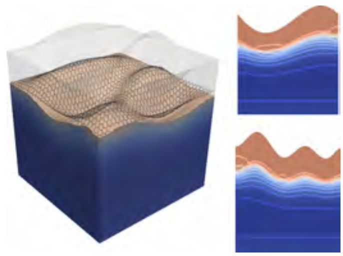
Solution at $t = 350$.



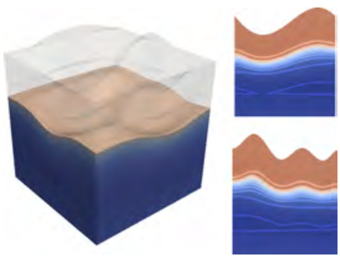
Solution at $t = 500$.



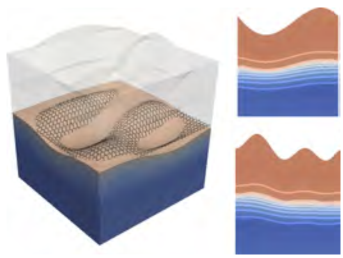
Solution at $t = 730$.



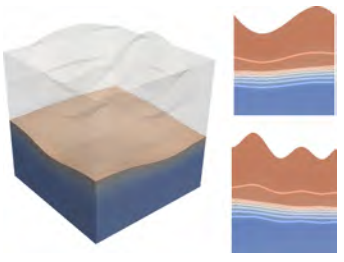
Solution at $t = 1150$.



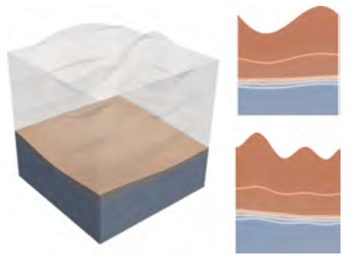
Solution at $t = 2060$.



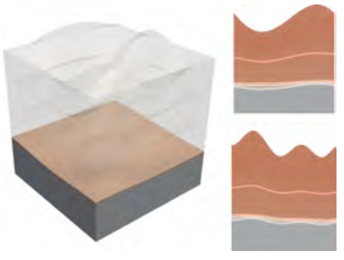
Solution at $t = 5280$.



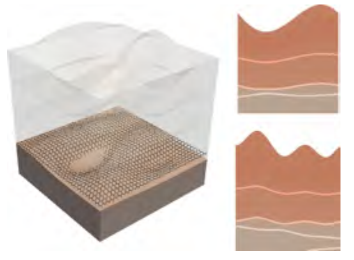
Solution at $t = 8080$.



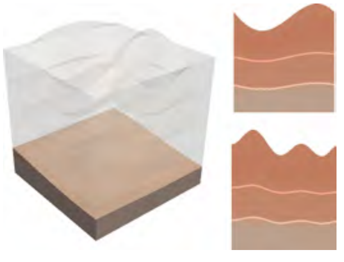
Solution at $t = 11820$.



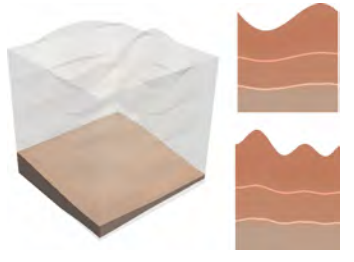
Solution at $t = 14620$.



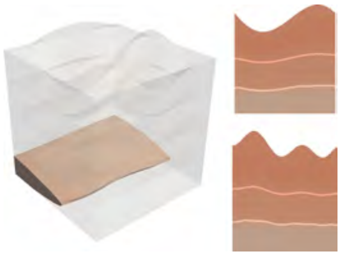
Solution at $t = 17780$.



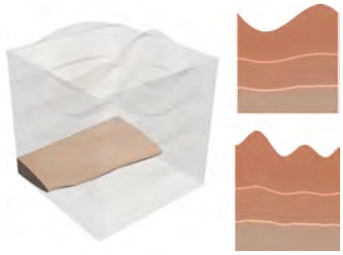
Solution at $t = 18690$.



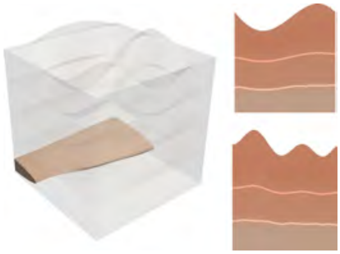
Solution at $t = 19100$.



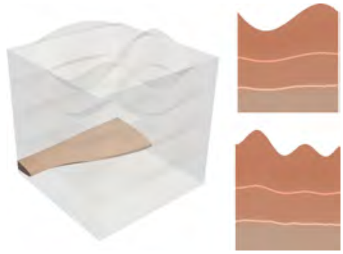
Solution at $t = 19330$.



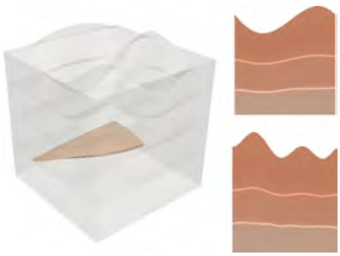
Solution at $t = 19480$.



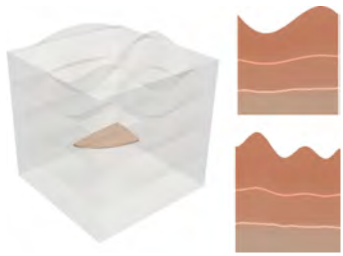
Solution at $t = 19580$.



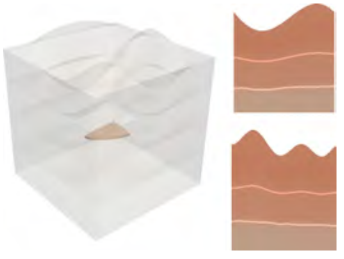
Solution at $t = 19660$.



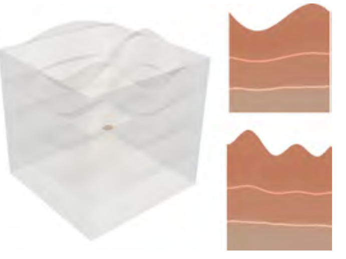
Solution at $t = 19740$.



Solution at $t = 19850$.

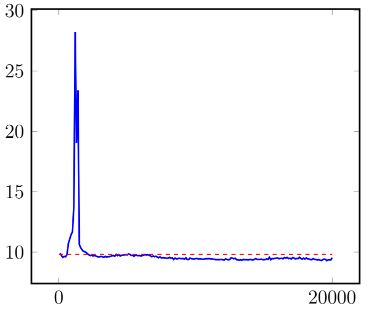


Solution at $t = 19920$.

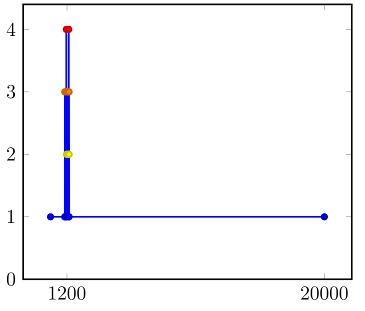


Solution at $t = 20000$.

As for the two dimensional experiment, we plotted the average runtime per time step and the number of Newton iterations per time step. As we can see in Figure 6.31a,



(a) Average runtime per time step.



(b) Newton iteration per time step.

the number of Newton iterations and thus the runtime of the corresponding time steps increases for $t \in [1100, 1300]$. In this period, the change from a low pressure level to a high pressure level leaves the yellow layer and enters the green layer which seems to have an impact on the Newton iteration.

7 CONCLUSION

In this thesis we considered a novel approach to solve the pressure formulation of the Richards equation to simulate the flow of water in a heterogeneous porous medium using the so called Kirchhoff transformation. In Chapter 2 we derived the Richards equation from the principle of mass balance and we discussed the equation in a homogeneous soil and in a heterogeneous soil. After we recalled functional analytical tools in Chapter 3, we analyzed the Richards equation. After we applied an implicit–explicit time stepping scheme, we derived a series of stationary variational problems. In Section 3.2 we were able to prove unique solvability of the stationary variational problem with general nonlinearities. In Section 3.2 we considered the equation in a homogeneous soil using the model derived in Section 2.1 and showed unique solvability under certain assumptions. We applied the Kirchhoff transformation and obtained a partial differential equation with a linear principal part. In Section 3.3 we extended the results from the homogeneous soil case to the heterogeneous soil case. We had to do some additional work but finally we derived a system of partial differential equations with linear principal part coupled via a nonlinear transmission condition. The similarity to the discrete mortar finite element method was crucial for its application to compute the approximate solution. Since the problem is still a nonlinear problem we applied the Newton method to obtain a linearized problem, see Section 5.1. In Section 5.2 and Section 5.3 we discussed the discretization of the computational domain to obtain suitable trial spaces as well as the stability of the linearized system. In the last section we briefly discussed some implementation details since a lot of work was invested to obtain a parallel code. Finally, in Chapter 6, numerical experiments in two and three space dimensions were presented.

Outlook and Open Problems

With this thesis we just did a first step into the field of nonlinear transmission problems. We just discussed the Richards equation in this thesis but the results achieved in Chapter 3 are also applicable to the nonlinear heat equation of the form

$$\frac{\partial u}{\partial t} - \nabla \cdot \Big(c(u) \, \nabla u \Big) = f$$

where the thermal conductivity depends on the unknown temperature u.

Nevertheless, there are still open problems concerning the numerical treatment of the variational problem (4.40). First of all, one has to think about a more conform way from the continuous variational problem (4.40) to the discrete counterpart (5.2). The question about the right norms for the Lagrange multiplier is still an open question and consequently inf–sup–stability of the coupling conditions. Since the linearization of the nonlinear coupling condition contains an additional weight, one may introduce a different test space for the Lagrange multiplier to obtain better stability results, thus a Galerkin-Petrov method has to be considered. As we could see, the solvability of the discrete saddle point problem is in general still an open problem. The next open question is the convergence of the approximation method. Since the problem is highly nonlinear, convergence results are often hard to show and in most cases only under restrictive assumptions on the nonlinearities. But the question concerning convergence of the approximate solution already starts with the time stepping scheme. Furthermore, the construction of a suitable preconditioner remains an untouched problem.

BIBLIOGRAPHY

[1] R. Adams. *Sobolev spaces*. Pure and Applied Mathematics, Vol. 65. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975, pp. xviii+268.

[2] R. Adams and J. Fournier. *Sobolev spaces*. Second. Vol. 140. Pure and Applied Mathematics (Amsterdam). Elsevier/Academic Press, Amsterdam, 2003, pp. xiv+305.

[3] H. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer-Lehrbuch Masterclass. 6., überarbeitete Auflage. Springer Berlin Heidelberg, 2012.

[4] H. Alt and S. Luckhaus. "Quasilinear elliptic-parabolic differential equations". In: *Math. Z.* 183.3 (1983), pp. 311–341.

[5] J. Appell and P. Zabrejko. *Nonlinear superposition operators*. Vol. 95. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1990, pp. viii+311.

[6] W. Arendt and K. Urban. *Partielle Differenzialgleichungen*. Vol. 1. Spektrum Akademischer Verlag. Springer, Heidelberg, 2010, pp. xii+353.

[7] J. Bear. *Dynamics of Fluids In Porous Media*. Dynamics of Fluids in Porous Media Bd. 1. American Elsevier Publishing Company, 1972.

[8] F. Belgacem. "The Mortar finite element method with Lagrange multipliers". In: *Numer. Math.* 84 (1999), pp. 173–197.

[9] M. Berger. *Nonlinearity and functional analysis*. Lectures on nonlinear problems in mathematical analysis, Pure and Applied Mathematics. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1977, pp. xix+417.

[10] C. Bernardi, C. Canuto, and Y. Maday. "Generalized inf-sup conditions for Chebyshev spectral approximation of the Stokes problem". In: *SIAM J. Numer. Anal.* 25.6 (1988), pp. 1237–1271.

[11] C. Bernardi, Y. Maday, and A. Patera. "A new nonconforming approach to domain decomposition: the mortar element method". In: *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XI (Paris, 1989–1991)*. Vol. 299. Pitman Res. Notes Math. Ser. Longman Sci. Tech., Harlow, 1994, pp. 13–51.

[12] H. Berninger. “Domain Decomposition Methods for Elliptic Problems with Jumping Nonlinearities and Application to the Richards Equation”. PhD thesis. Freie Universität Berlin, 2008.

[13] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*. Vol. 44. Springer Series in Computational Mathematics. Heidelberg: Springer, 2013, pp. xiv+685.

[14] D. Braess. *Finite elements*. Third. Theory, fast solvers, and applications in elasticity theory, Translated from the German by Larry L. Schumaker. Cambridge: Cambridge University Press, 2007, pp. xviii+365.

[15] D. Braess and W. Dahmen. “Stability estimates of the mortar finite element method for 3-dimensional problems”. In: *East-West J. Numer. Math.* 6.4 (1998), pp. 249–263.

[16] D. Braess and W. Dahmen. “The mortar element method revisited—what are the right norms?” In: *Domain decomposition methods in science and engineering (Lyon, 2000)*. Theory Eng. Appl. Comput. Methods. Internat. Center Numer. Methods Eng. (CIMNE), Barcelona, 2002, pp. 27–40.

[17] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011, pp. xiv+599.

[18] F. Brezzi. “On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers”. In: *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge* 8.R-2 (1974), pp. 129–151.

[19] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. Vol. 15. Springer Series in Computational Mathematics. New York: Springer-Verlag, 1991, pp. x+350.

[20] F. Brezzi and D. Marini. “Error estimates for the three-field formulation with bubble stabilization”. In: *Math. Comp.* 70.235 (2001), pp. 911–934.

[21] R. Brooks and A. Corey. *Hydraulic Properties of Porous Media*. Colorado State University Hydrology Papers. Colorado State University, 1964.

[22] R. Brooks and A. Corey. “Properties of porous media affecting fluid flow”. In: *J. Irrig. Drain. Div., Am. Soc. Civil Eng.* (1966), pp. 61–88.

[23] P. Ciarlet Jr., J. Huang, and J. Zou. “Some observations on generalized saddle-point problems”. In: *SIAM J. Matrix Anal. Appl.* 25.1 (2003), pp. 224–236.

[24] L. Dagum and R. Menon. “OpenMP: an industry standard API for shared-memory programming”. In: *Computational Science & Engineering, IEEE* 5.1 (1998), pp. 46–55.

[25] P. Deuflhard. *Newton methods for nonlinear problems*. Vol. 35. Springer Series in Computational Mathematics. Affine invariance and adaptive algorithms. Springer-Verlag, Berlin, 2004, pp. xii+424.

[26] E. Di Nezza, G. Palatucci, and E. Valdinoci. “Hitchhiker’s guide to the fractional Sobolev spaces”. In: *Bull. Sci. Math.* 136.5 (2012), pp. 521–573.

[27] D. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*. Vol. 69. Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer, Heidelberg, 2012, pp. xviii+384.

[28] M. Dobrowolski. *Angewandte Funktionalanalysis: Funktionalanalysis, Sobolev-Räume und elliptische Differentialgleichungen*. Springer-Lehrbuch Masterclass. Springer Berlin Heidelberg, 2010.

[29] P. Drábek, A. Kufner, and F. Nicolosi. *Quasilinear elliptic equations with degenerations and singularities*. Vol. 5. de Gruyter Series in Nonlinear Analysis and Applications. Walter de Gruyter & Co., Berlin, 1997, pp. xii+219.

[30] C. Eck, H. Garcke, and P. Knabner. *Mathematische Modellierung*. Berlin: Springer, 2008, pp. xiv+503.

[31] L. Evans and R. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992, pp. viii+268.

[32] J. Franců. “Monotone operators. A survey directed to applications to differential equations”. In: *Apl. Mat.* 35.4 (1990), pp. 257–301.

[33] E. Gabriel et al. “Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation”. In: *Proceedings, 11th European PVM/MPI Users’ Group Meeting*. Budapest, Hungary, Sept. 2004, pp. 97–104.

[34] S. Geršgorin. “Über die Abgrenzung der Eigenwerte einer Matrix”. In: *Bulletin de l’Académie des Sciences de l’URSS. Classe des sciences mathématiques et na* 6 (1931), pp. 749–754.

[35] C. Geuzaine and J.-F. Remacle. “Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities”. In: *International Journal for Numerical Methods in Engineering* 79.11 (2009), pp. 1309–1331.

[36] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*. Vol. 5. Springer Series in Computational Mathematics. Theory and algorithms. Berlin: Springer-Verlag, 1986, pp. x+374.

[37] P. Grisvard. *Elliptic problems in nonsmooth domains*. Vol. 24. Monographs and Studies in Mathematics. Pitman (Advanced Publishing Program), Boston, MA, 1985, pp. xiv+410.

[38] G. Haase. *Parallelisierung numerischer Algorithmen für partielle Differentialgleichungen*. B. G. Teubner, Stuttgart, 1999, p. 175.

[39] I. Hlaváček, M. Křížek, and J. Malý. “On Galerkin approximations of a quasilinear nonpotential elliptic problem of a nonmonotone type”. In: *J. Math. Anal. Appl.* 184.1 (1994), pp. 168–189.

[40] G. Hsiao and W. Wendland. *Boundary integral equations*. Vol. 164. Applied Mathematical Sciences. Springer-Verlag, Berlin, 2008, pp. xx+618.

[41] C. Kim et al. “Multiplier spaces for the mortar finite element method in three dimensions”. In: *SIAM J. Numer. Anal.* 39.2 (2001), pp. 519–538.

[42] R. Krause. “Monotone Multigrid Methods for Signorini’s Problem with Friction”. PhD thesis. Freie Universität Berlin, 2001.

[43] E. Kreyszig. *Introductory functional analysis with applications*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1989, pp. xvi+688.

[44] A. Kuzmin, M. Luisier, and O. Schenk. “Euro-Par 2013 Parallel Processing: 19th International Conference, Aachen, Germany, August 26-30, 2013. Proceedings”. In: ed. by F. Wolf, B. Mohr, and D. Mey. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. Chap. Fast Methods for Computing Selected Elements of the Green’s Function in Massively Parallel Nanoelectronic Device Simulations, pp. 533–544.

[45] B. Lamichhane. “Higher Order Mortar Finite Elements with Dual Lagrange Multiplier Spaces and Applications”. PhD thesis. Universität Stuttgart, 2006.

[46] M. Marcus and V. Mizel. “Absolute continuity on tracks and mappings of Sobolev spaces”. In: *Arch. Rational Mech. Anal.* 45 (1972), pp. 294–320.

[47] M. Marcus and V. Mizel. “Every superposition operator mapping one Sobolev space into another is continuous”. In: *J. Funct. Anal.* 33.2 (1979), pp. 217–229.

[48] W. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge: Cambridge University Press, 2000, pp. xiv+357.

[49] N. Meyers and J. Serrin. “$H = W$”. In: *Proc. Nat. Acad. Sci. U.S.A.* 51 (1964), pp. 1055–1056.

[50] R. Nicolaides. “Existence, uniqueness and approximation for generalized saddle point problems”. In: *SIAM J. Numer. Anal.* 19.2 (1982), pp. 349–357.

[51] P.-A. Raviart and J. Thomas. “Primal hybrid finite element methods for 2nd order elliptic equations”. In: *Math. Comp.* 31.138 (1977), pp. 391–413.

[52] L. A. Richards. “Capillary Conduction of Liquids Through Porous Mediums”. In: *Physics* 1 (Nov. 1931), pp. 318–333.

[53] W. Rudin. *Real and complex analysis*. Third. McGraw-Hill Book Co., New York, 1987, pp. xiv+416.

[54] D. Sahu and K. Singh. “On Newton-like method for solving generalized nonlinear operator equations in Banach spaces”. In: *Bull. Malays. Math. Sci. Soc. (2)* 36.3 (2013), pp. 723–731.

[55] S. Sauter and C. Schwab. *Boundary element methods*. Vol. 39. Springer Series in Computational Mathematics. Translated and expanded from the 2004 German original. Springer-Verlag, Berlin, 2011, pp. xviii+561.

[56] O. Schenk and K. Gärtner. "Solving unsymmetric sparse systems of linear equations with PARDISO". In: *Future Generation Computer Systems* 20.3 (2004), pp. 475–487.

[57] S. Schreiber. "Nichtüberlappende Gebietszerlegungsmethoden für lineare und quasilineare (monotone und nichtmonotone) Probleme". PhD thesis. Universität Kassel, 2009.

[58] R. Showalter. *Monotone operators in Banach space and nonlinear partial differential equations*. Vol. 49. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 1997, pp. xiv+278.

[59] D. Stefanica. "Domain Decomposition Methods for Mortar Finite Elements". PhD thesis. New York University, 1999.

[60] O. Steinbach. *Numerical approximation methods for elliptic boundary value problems*. Finite and boundary elements, Translated from the 2003 German original. New York: Springer, 2008, pp. xii+386.

[61] O. Steinbach. *Stability estimates for hybrid coupled domain decomposition methods*. Vol. 1809. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2003, pp. vi+120.

[62] L. Tartar. *An introduction to Sobolev spaces and interpolation spaces*. Vol. 3. Lecture Notes of the Unione Matematica Italiana. Springer, Berlin; UMI, Bologna, 2007, pp. xxvi+218.

[63] R. Temam. *Navier-Stokes equations*. Third. Vol. 2. Studies in Mathematics and its Applications. Theory and numerical analysis, With an appendix by F. Thomasset. Amsterdam: North-Holland Publishing Co., 1984, pp. xii+526.

[64] J.-M. Thomas. "Méthode des éléments finis hybrides duaux pour les problèmes elliptiques du second ordre". fre. In: *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* 10.R3 (1976), pp. 51–79.

[65] M. Van Genuchten. "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils". In: *Soil science society of America journal* 44.5 (1980), pp. 892–898.

[66] V. Vijesh and P. Subrahmanyam. "A variant of Newton's method and its application". In: *J. Math. Anal. Appl.* 342.1 (2008), pp. 361–370.

[67] W. Walter. *Analysis 2*. Analysis. Springer Berlin Heidelberg, 2002.

[68] D. Werner. *Funktionalanalysis*. 7., korrigierte und erweiterte Auflage. Berlin Heidelberg: Springer-Verlag, 2011, pp. xiii+552.

[69] J. Wloka. *Partial differential equations*. Translated from the German by C. B. Thomas and M. J. Thomas. Cambridge: Cambridge University Press, 1987, pp. xii+518.

[70] B. Wohlmuth. "A mortar finite element method using dual spaces for the Lagrange multiplier". In: *SIAM J. Numer. Anal.* 38.3 (2000), pp. 989–1012.

[71] B. Wohlmuth. *Discretization methods and iterative solvers based on domain decomposition*. Vol. 17. Lecture Notes in Computational Science and Engineering. Berlin: Springer-Verlag, 2001, pp. x+197.

[72] K. Yosida. *Functional analysis*. Fourth. Die Grundlehren der mathematischen Wissenschaften, Band 123. Springer-Verlag, New York-Heidelberg, 1974, pp. xii+496.

[73] E. Zeidler. *Nonlinear functional analysis and its applications. I*. Fixed-point theorems, Translated from the German by Peter R. Wadsack. Springer-Verlag, New York, 1986, pp. xxi+897.

[74] E. Zeidler. *Nonlinear functional analysis and its applications. II/A*. Linear monotone operators, Translated from the German by the author and Leo F. Boron. New York: Springer-Verlag, 1990, pp. xviii+467.

[75] E. Zeidler. *Nonlinear functional analysis and its applications. II/B*. Nonlinear monotone operators, Translated from the German by the author and Leo F. Boron. Springer-Verlag, New York, 1990, i–xvi and 469–1202.

[76] E. Ziemer. *Weakly differentiable functions*. Vol. 120. Graduate Texts in Mathematics. Sobolev spaces and functions of bounded variation. Springer-Verlag, New York, 1989, pp. xvi+308.

Monographic Series TU Graz

Computation in Engineering and Science

Vol. 1 Steffen Alvermann
Effective Viscoelastic Behavior of Cellular Auxetic Materials
2008
ISBN 978-3-902465-92-4

Vol. 2 Sendy Franciscus Tantono
The Mechanical Behaviour of a Soilbag under Vertical Compression
2008
ISBN 978-3-902465-97-9

Vol. 3 Thomas Rüberg
Non-conforming FEM/BEM Coupling in Time Domain
2008
ISBN 978-3-902465-98-6

Vol. 4 Dimitrios E. Kiousis
Biomechanical and Computational Modeling of Atherosclerotic Arteries
2008
ISBN 978-3-85125-023-7

Vol. 5 Lars Kielhorn
A Time-Domain Symmetric Galerkin BEM for Viscoelastodynamics
2009
ISBN 978-3-85125-042-8

Vol. 6 Gerhard Unger
Analysis of Boundary Element Methods for Laplacian Eigenvalue Problems
2009
ISBN 978-3-85125-081-7

Monographic Series TU Graz

Computation in Engineering and Science

Vol. 7 Gerhard Sommer
Mechanical Properties of Healthy and Diseased Human Arteries
2010
ISBN 978-3-85125-111-1

Vol. 8 Mathias Nenning
Infinite Elements for Elasto- and Poroelastodynamics
2010
ISBN 978-3-85125-130-2

Vol. 9 Thanh Xuan Phan
Boundary Element Methods for Boundary Control Problems
2011
ISBN 978-3-85125-149-4

Vol. 10 Loris Nagler
Simulation of Sound Transmission through Poroelastic Plate-like Structures
2011
ISBN 978-3-85125-153-1

Vol. 11 Markus Windisch
Boundary Element Tearing and Interconnecting Methods for Acoustic and Electromagnetic Scattering
2011
ISBN 978-3-85125-152-4

Vol. 12 Christian Walchshofer
Analysis of the Dynamics at the Base of a Lifted Strongly Buoyant Jet Flame Using Direct Numerical Simulation
2011
ISBN 978-3-85125-185-2

Monographic Series TU Graz

Computation in Engineering and Science

Vol. 13 Matthias Messner
Fast Boundary Element Methods in Acoustics
2012
ISBN 978-3-85125-202-6

Vol. 14 Peter Urthaler
Analysis of Boundary Element Methods for Wave Propagation in Porous Media
2012
ISBN 978-3-85125-216-3

Vol. 15 Peng Li
Boundary Element Method for Wave Propagation in Partially Saturated Poroelastic Continua
2012
ISBN 978-3-85125-236-1

Vol. 16 Andreas Jörg Schriefl
Quantification of Collagen Fiber Morphologies in Human Arterial Walls
2013
ISBN 978-3-85125-238-5

Vol. 17 Thomas S. E. Eriksson
Cardiovascular Mechanics
2013
ISBN 978-3-85125-277-4

Vol. 18 Jianhua Tong
Biomechanics of Abdominal Aortic Aneurysms
2013
ISBN 978-3-85125-279-8

Monographic Series TU Graz

Computation in Engineering and Science

Vol. 19 Jonathan Rohleder
Titchmarsh–Weyl Theory and Inverse Problems for Elliptic Differential Operators
2013
ISBN 978-3-85125-283-5

Vol. 20 Martin Neumüller
Space-Time Methods
2013
ISBN 978-3-85125-290-3

Vol. 21 Michael J. Unterberger
Microstructurally-Motivated Constitutive Modeling of Cross-Linked Filamentous Actin Networks
2013
ISBN 978-3-85125-303-0

Vol. 22 Vladimir Lotoreichik
Singular Values and Trace Formulae for Resolvent Power Differences of Self-Adjoint Elliptic Operators
2013
ISBN 978-3-85125-304-7

Vol. 23 Michael Meßner
A Fast Multipole Galerkin Boundary Element Method for the Transient Heat Equation
2014
ISBN 978-3-85125-350-4

Vol. 24 Lorenz Johannes John
Optimal Boundary Control in Energy Spaces
2014
ISBN 978-3-85125-373-3

Monographic Series TU Graz

Computation in Engineering and Science

Vol. 25 Hannah Weisbecker
Softening and Damage Behavior of Human Arteries
2014
ISBN 978-3-85125-370-2

Vol. 26 Bernhard Kager
Efficient Convolution Quadrature based Boundary Element Formulation for Time-Domain Elastodynamics
2015
ISBN 978-3-85125-382-5

Vol. 27 Christoph M. Augustin
Classical and All-floating FETI Methods with Applications to Biomechanical Models
2015
ISBN 978-3-85125-418-1

Vol. 28 Elias Karabelas
Space-Time Discontinuous Galerkin Methods for Cardiac Electromechanics
2016
ISBN 978-3-85125-461-7

Vol. 29 Thomas Traub
A Kernel Interpolation Based Fast Multipole Method for Elastodynamic Problems
2016
ISBN 978-3-85125-465-5

Vol. 30 Matthias Gsell
Mortar Domain Decomposition Methods for Quasilinear Problems and Applications
2017
ISBN 978-3-85125-522-5