# The Navier–Stokes–Darcy problem

Moritz Hoffmann

September 29, 2013

# Contents

# 1 Introduction

The objective of this thesis is to discuss the Navier–Stokes–Darcy problem for incompressible fluids. The Navier–Stokes–Darcy problem is a coupled problem of free flow together with flow through porous media. Applications are for example the filtration process of blood through vessel walls in the field of medicine or the flow of a river and its riverbed in the field of geosciences.

In order to study this problem, several steps are made. As the goal is the coupled Navier–Stokes–Darcy problem, the Darcy problem that models flow through porous media, the linear Stokes problem that models free viscous flow and the nonlinear Navier–Stokes problem that models turbulent flow are first considered separately. The Stokes problem can be derived from the Navier–Stokes problem and does not possess the difficulty of being nonlinear. Nevertheless, some of the theory can be reused which is why it is considered before introducing the Navier–Stokes problem.

After having considered these separate problems, the coupled Stokes–Darcy and the coupled Navier–Stokes–Darcy problems are discussed. Again, some of the theory of the Stokes–Darcy problem can be reused for the Navier–Stokes–Darcy problem. The focus is on how exactly to couple these systems and how one can solve the resulting problems. In particular, it turns out that the standard method to couple the systems does not work well for small values of hydraulic permeability. Therefore, an alternative approach is discussed as well. Basically, the discussed methods decompose the coupled problem into separated problems that have a part of the boundary of their domain in common. This part isn called interface. Then the idea is to solve the problems iteratively, i.e., solve one problem, make an interface update, solve the other problem, make another interface update and so on. For the Navier–Stokes–Darcy problem, one additionally needs to deal with the nonlinearity of the Navier–Stokes equations. This is done by using a fixed-point iteration which approximates the Navier–Stokes equations by linear Oseen-type equations. In the context of the coupled problem the question arises, how one should restrict the number of fixed-point iterations per interface iteration to get optimal performance.

This question and others are discussed with the help of two numerical examples. One example compares the numerical results of a Navier–Stokes–Darcy problem with an analytical solution, the other example is a model of a riverbed and is compared to results of the literature.

# 2 Models for flow problems

This chapter is about two models for free flow and one model for flow through porous media. First, there is a general part which deals with saddle-point problems, which arise in the Stokes model and in the linearization of the Navier–Stokes model. Then the specific models are presented.

## 2.1 Saddle point problems

Saddle point type problems are variational problems which can be, under certain hypotheses, formulated in terms of a function where one searches for the infimum of one parameter and the supremum of the other (hence "Saddle point"). This type of problem arises for instance considering a weak formulation of the Stokes equations or the linearized Navier–Stokes equations.

This section follows §4 of the first chapter of [9]. For simplicity some parts are omitted. Let $X$ and $M$ be two real Hilbert spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_M$ respectively and their dual spaces $X'$ and $M'$. For two given continuous bilinear forms

$$\begin{aligned} a(\cdot,\cdot) &: X \times X \to \mathbb{R}, \\ b(\cdot,\cdot) &: M \times M \to \mathbb{R} \end{aligned}$$

consider the following problem: For given $\ell \in X', \chi \in M'$, find $(u,\lambda) \in X \times M$ such that

$$\left\{ \begin{array}{rll} a(u,v) + b(v,\lambda) & = \langle \ell, v\rangle & \forall v \in X, \\ b(u,\mu) & = \langle \chi, \mu\rangle & \forall \mu \in M. \end{array} \right. \tag{1}$$

In the following we will discuss existence and uniqueness of a solution of (1). Therefore define two continuous linear operators associated with $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$

$$\begin{array}{ll} A : X \to X',\ \langle Au, v\rangle = a(u,v) & \forall v \in X, \\ B : X \to M',\ \langle Bv, \mu\rangle = b(v,\mu) & \forall \mu \in M, \end{array}$$

and the dual operator of B

$$B' : M \to X',\ \langle B'\mu, v\rangle = \langle Bv, \mu\rangle = b(v,\mu) \qquad \forall v \in X.$$

Then, one can reformulate problem (1): Find $(u,\lambda) \in X \times M$ satisfying

$$\left\{ \begin{array}{rl} Au + B'\lambda & = \ell \text{ in } X', \\ Bu & = \chi \text{ in } M'. \end{array} \right.$$

Next, set $V = \ker B \subset X$ and define for each $\chi \in M'$

$$V(\chi) = \{v \in X : Bv = \chi\} = \{v \in X : b(v,\mu) = \langle \chi, \mu\rangle \quad \forall \mu \in M\},\ V(0) = V.$$

Then problem (1) can be associated with another problem, which reads as follows: Find $u \in V(\chi)$ such that

$$a(u,v) = \langle \ell, v\rangle \qquad \forall v \in V. \tag{2}$$

**2.1.1 Remark.** If $(u,\lambda) \in X \times M$ is a solution of (1), then $u$ is also a solution of (2), since $u \in V(\chi)$ because of the second equation of (1) and

$$\begin{aligned} a(u,v) + b(v,\lambda) &= a(u,v) && \text{, since } v \in V = \ker B, \\ &= \langle \ell, v\rangle. \end{aligned}$$

### 2.1.1 Existence and uniqueness of a solution

The following theorem is Theorem 4.1 of §4 of Chapter I of [9].

**2.1.2 Theorem**[Existence and uniqueness]**.** Let us assume the following hypotheses:

(i) The bilinear form $a(\cdot,\cdot)$ is coercive on $V$: There exists a constant $\alpha > 0$ such that

$$a(v,v) \geq \alpha \|v\|_X^2 \quad \forall v \in V = \ker B.$$

(ii) The bilinear form $b(\cdot,\cdot)$ satisfies the inf-sup condition: There exists a constant $\beta > 0$ such that

$$\inf_{\mu \in M} \sup_{v \in X} \frac{b(v,\mu)}{\|v\|_X \|\mu\|_M} \geq \beta.$$

Then problem (2) has a unique solution $u \in V(\chi)$ and there exists a unique $\lambda \in M$ such that the pair $(u, \lambda)$ is the unique solution of problem (1). Moreover, the mapping

$$X' \times M' \to X \times M : (\ell, \chi) \mapsto (u, \lambda)$$

is an isomorphism.

*Proof.* See Theorem 4.1 of §4 of Chapter I of [9].

**2.1.3 Remark.** The isomorphism property shows that for each right-hand side it is possible to find a solution and vice versa, for each solution there exists a corresponding right-hand side. It also implies continuity, hence small changes to the right-hand side have only a small impact on the solution.

**2.1.4 Remark.** The problem stated in this section is also known as a saddle point type problem. There, one searches in a special form of this problem for a saddle point (i.e., one parameter gets minimized, one gets maximized). If $a(\cdot, \cdot)$ fulfills the assumptions of Theorem 2.1.2 and it is semi positive definite and symmetric, the solution of the saddle point problem coincides with the solution of problem (1).

## 2.2 The Stokes equations

The Stokes equations describe the motion of fluids that have a high viscosity. They are in fact a limit case of the more general Navier–Stokes equations, where the nonlinear term becomes negligible due to large viscosity.
Let $\Omega \subset \mathbb{R}^d$ be open and bounded with Lipschitz boundary, $d \in \{2, 3\}$ and $\mathbf{n}$ the outer normal vector of $\partial\Omega$. Then the Stokes problem reads as follows:
Find $(\mathbf{u}, p) : \Omega \times \Omega \to \mathbb{R}^d \times \mathbb{R}$, such that, in the Laplace form,

$$\begin{cases} -\nu\Delta\mathbf{u} + \nabla p & = \mathbf{f} \text{ in } \Omega, \\ \nabla \cdot \mathbf{u} & = 0 \text{ in } \Omega, \end{cases} \tag{3}$$

or in the Cauchy-Stress form,

$$\begin{cases} -\nabla \cdot \mathbb{T}(\mathbf{u}, p) & = \mathbf{f} \text{ in } \Omega, \\ \nabla \cdot \mathbf{u} & = 0 \text{ in } \Omega, \end{cases} \tag{4}$$

with $\mathbb{T}(\mathbf{u}, p) := 2\nu\, \mathbb{D}(\mathbf{u}) - p\,\mathbb{I}$ being the Cauchy stress tensor, $\nu > 0$ the kinematic viscosity and

$$\mathbb{D}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T)$$

the symmetric part of $\nabla\mathbf{u}$, also called deformation tensor. The function $\mathbf{f} \in L^2(\Omega)$ is called the source term.

**2.2.1 Remark.** For (4) one needs $\mathbf{u}$ to be just two times differentiable, but not two times continuously differentiable. In order to be able to apply the theorem of Schwarz on $\mathbf{u}$, it is required to chose it in $(C^2(\Omega) \cap C^1(\bar{\Omega}))^d$. Under these circumstances the systems are equivalent. Indeed:

$$\begin{aligned}
-\nabla\cdot\mathbb{T}(\mathbf{u},p) &= -\nabla\cdot(2\nu\,\mathbb{D}(\mathbf{u})-p\,\mathbb{I}) = -\nabla\cdot(2\nu\,\mathbb{D}(\mathbf{u}))+\nabla p \\
&= -\nu\nabla\cdot\left(\tfrac{\partial u_i}{\partial x_j}+\tfrac{\partial u_j}{\partial x_i}\right)_{i,j=1}^{d}+\nabla p \\
&= -\nu\left(\textstyle\sum_{j=1}^{d}\tfrac{\partial^2 u_j}{\partial x_i\partial x_j}+\sum_{j=1}^{d}\tfrac{\partial^2 u_i}{\partial x_j^2}\right)_{i=1}^{d}+\nabla p \\
&= -\nu\left(\textstyle\sum_{j=1}^{d}\tfrac{\partial^2 u_j}{\partial x_j\partial x_i}+\Delta u_i\right)_{i=1}^{d}+\nabla p && \text{, Schwarz' theorem,} \\
&= -\nu\left(\tfrac{\partial}{\partial x_i}(\nabla\cdot\mathbf{u})+\Delta u_i\right)_{i=1}^{d}+\nabla p \\
&= -\nu\Delta\mathbf{u}+\nabla p=\mathbf{f} && \text{, with } \nabla\cdot\mathbf{u}=0.
\end{aligned}$$

Even though these two systems are equivalent under the preceding regularity assumptions, the finite element discretizations yield different results, which is the reason why they are considered separately.

### 2.2.1 Stokes boundary conditions

On the boundary we consider Dirichlet and Neumann-type conditions. Therefore split $\partial\Omega$ into two disjoint (relatively open) parts $\Gamma_D, \Gamma_N$ such that

- $\bar{\Gamma}_D \cup \bar{\Gamma}_N = \partial\Omega$,
- $\Gamma_N \cap \Gamma_D = \emptyset$,
- $\mathrm{meas}(\Gamma_D) > 0$ (for simplicity).

Furthermore we assume the Dirichlet boundary conditions to be homogeneous, i.e., for $T_N \in H^{-1/2}(\Gamma_N)$ we have

$$\mathbf{u}=\mathbf{0} \qquad \text{on } \Gamma_D, \tag{5}$$

$$(\nu\nabla\mathbf{u}-p\,\mathbb{I})\cdot\mathbf{n}=T_N \qquad \text{on } \Gamma_N \text{ for (3)}, \tag{6}$$

$$\mathbb{T}(\mathbf{u},p)\cdot\mathbf{n}=T_N \qquad \text{on } \Gamma_N \text{ for (4)}. \tag{7}$$

**2.2.2 Remark.** For a given non-homogeneous Dirichlet boundary $\mathbf{u}=\mathbf{u}_D\in(H^{1/2}(\Gamma_D))^d$, one can just consider $\mathbf{U}:=\mathbf{u}-\mathbf{u}_D$ for (3) or (4). Then the Dirichlet boundary turns out to be homogeneous for $\mathbf{U}$. Note, that $\mathbf{u}-\mathbf{u}_D$ at first makes no sense, because $\mathbf{u}$ is defined on $\Omega$ and $\mathbf{u}_D$ on the boundary. Hence one needs to extend $\mathbf{u}_D$ into $\Omega$. The existence of such an extension is guaranteed as the trace operator is surjective. Also the right-hand side of the original problem transforms:

$$\begin{cases}-\nu\Delta\mathbf{U}+\nabla p &= \mathbf{f}\\ -\nabla\cdot\mathbb{T}(\mathbf{U},p) &= \mathbf{f}\end{cases} \Rightarrow \begin{cases}-\nu\Delta\mathbf{u}+\nabla p &= \mathbf{f}+\nu\Delta\mathbf{u}_D =: \tilde{\mathbf{f}},\\ -\nabla\cdot\mathbb{T}(\mathbf{u},p) &= \mathbf{f}+\nabla\cdot(2\nu\,\mathbb{D}(\mathbf{u}_D)) =: \tilde{\mathbf{f}},\end{cases}$$

$$\nabla\cdot\mathbf{U}=-\nabla\cdot\mathbf{u}_D.$$

The choice of spaces for the boundary conditions is reasonable, since it will turn out that in the weak formulation one searches for a solution $\mathbf{u}$ in $H^1(\Omega)$. Therefore, the corresponding space of traces is $H^{1-1/2}(\Gamma_N)=H^{1/2}(\Gamma_N)$. As $T_N$ acts as a functional on $\Gamma_N$, it has to be an element of the dual space $H^{-1/2}(\Gamma_N)$.

### 2.2.2 Weak formulation

Assume $(\mathbf{u},p)$ is a classical solution with $\mathbf{u}\in(C^2(\Omega)\cap C^1(\bar{\Omega}))^d$ and $p\in C^1(\Omega)\cap C(\bar{\Omega})$. Take test functions $\mathbf{v}$ for the first equation and $q$ for the second equation of (3) and (4) with $\mathbf{v}$ vanishing close to the Dirichlet boundary, that is $\mathbf{v}\in(C^\infty_{\Gamma_D}(\Omega))^d$ with

$$C^\infty_{\Gamma_D}(\Omega):=\left\{v\in C^\infty(\Omega)\cap H^1(\Omega):\substack{\exists U\subset\mathbb{R}^d\text{ open neighborhood of }\Gamma_D\\ \text{s.t. }v(x)=0\ \forall x\in U\cap\Omega}\right\}$$

and $q\in C^\infty(\Omega)$. The intersection of $C^\infty(\Omega)$ and $H^1(\Omega)$ is needed as $C^\infty(\Omega)$ is not a subset of $H^1(\Omega)$ (but the intersection is dense in $H^1(\Omega)$, see [12]) and the space is going to be completed with the $H^1$ norm, as it does not have enough structure to make Hilbert-/Banach-space theory applicable.

After having multiplied with these test functions, one needs to integrate over $\Omega$ and apply integration by parts to obtain the weak formulation. This is done separately for the two different Stokes formulations (3) and (4).

(i) For (3): One has

$$\int\limits_\Omega -\nu\Delta\mathbf{u}\cdot\mathbf{v}+\int\limits_\Omega(\nabla p)\cdot\mathbf{v}=\int\limits_\Omega\mathbf{f}\cdot\mathbf{v}.$$

The second integral can be transformed as follows:

$$\begin{aligned}\int\limits_\Omega(\nabla p)\cdot\mathbf{v}&=\int\limits_\Omega\sum_{i=1}^d\frac{\partial p}{\partial x_i}v_i\\ &=\int\limits_\Omega\sum_{i=1}^d\frac{\partial}{\partial x_i}(pv_i)-p\frac{\partial v_i}{\partial x_i}\\ &=\int\limits_\Omega\nabla\cdot(p\mathbf{v})-p\nabla\cdot\mathbf{v} &&\text{, using product rule,}\\ &=\int\limits_{\partial\Omega}(p\mathbf{v})\cdot\mathbf{n}-\int\limits_\Omega p\nabla\cdot\mathbf{v} &&\text{, using the Gaussian theorem.}\end{aligned}$$

Considering the first term:

$$\begin{aligned}-\int\limits_\Omega\nu\Delta\mathbf{u}\cdot\mathbf{v}&=-\int\limits_\Omega\nu\sum_{i=1}^d\Delta u_iv_i\\ &=-\int\limits_{\partial\Omega}\sum_{i=1}^d\nu(\nabla u_i\cdot\mathbf{n})v_i+\int\limits_\Omega\nu\sum_{i=1}^d\nabla u_i\nabla v_i &&\text{, using integration by parts,}\\ &=-\int\limits_{\partial\Omega}\nu(\nabla\mathbf{u}\cdot\mathbf{v})\cdot\mathbf{n}+\int\limits_\Omega(\nu\nabla\mathbf{u}):(\nabla\mathbf{v}),\end{aligned}$$

with $A : B := \sum_{i=1}^d \sum_{j=1}^d A_{ij} \cdot B_{ij}$. Adding the two terms yields

$$\begin{aligned}\int\limits_\Omega -\nu\Delta\mathbf{u}\cdot\mathbf{v} + \int\limits_\Omega (\nabla p)\cdot\mathbf{v} &= \int\limits_\Omega (\nu\nabla\mathbf{u}) : (\nabla\mathbf{v}) - \int\limits_\Omega p\nabla\cdot\mathbf{v} - \int\limits_{\partial\Omega} \nu(\nabla\mathbf{u}\cdot\mathbf{v})\cdot\mathbf{n} - (p\mathbf{v})\cdot\mathbf{n} \\ &= (\nu\nabla\mathbf{u}, \nabla\mathbf{v})_0 - (p, \nabla\cdot\mathbf{v})_0 - \int\limits_{\Gamma_N} (\nu\nabla\mathbf{u} - p\,\mathbb{I})\mathbf{n}\cdot\mathbf{v} \\ &= (\nu\nabla\mathbf{u}, \nabla\mathbf{v})_0 - (p, \nabla\cdot\mathbf{v})_0 - \langle T_N, \mathbf{v}\rangle_{\Gamma_N}.\end{aligned}$$

The test functions are zero at the Dirichlet boundary, hence the integral disappears there and one only has to consider the Neumann boundary parts. All together one gets

$$(\nu\nabla\mathbf{u}, \nabla\mathbf{v})_0 - (p, \nabla\cdot\mathbf{v})_0 = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v}\rangle_{\Gamma_N}.$$

(ii) For (4): One has

$$\int\limits_\Omega -\nabla\cdot\mathbb{T}(\mathbf{u},p)\cdot\mathbf{v} = -\int\limits_\Omega (\nabla\cdot(2\nu\,\mathbb{D}(\mathbf{u})))\cdot\mathbf{v} + \int\limits_\Omega (\nabla p)\cdot\mathbf{v} = \int\limits_\Omega \mathbf{f}\cdot\mathbf{v}$$

The second integral and the integral on the right hand side are the same as in the first case, so focus on the first integral: Define $M := 2\nu\,\mathbb{D}(\mathbf{u})$ and note, that this matrix (tensor) is symmetric. Then one has:

$$\begin{aligned}-(\nabla\cdot(2\nu\,\mathbb{D}(\mathbf{u})))\cdot\mathbf{v} &= -(\nabla\cdot M)\cdot\mathbf{v} \\ &= -\left(\textstyle\sum_{j=1}^d \frac{\partial M_{ij}}{\partial x_j} \quad \cdots \quad \sum_{j=1}^d \frac{\partial M_{dj}}{\partial x_j}\right)_{i=1}^d \cdot\mathbf{v} \\ &= -\sum_{i=1}^d\sum_{j=1}^d \frac{\partial M_{ij}}{\partial x_j} v_i \\ &= \sum_{i=1}^d\sum_{j=1}^d M_{ij}\frac{\partial v_i}{\partial x_j} - \frac{\partial}{\partial x_j}(M_{ij}v_i) && \text{, product rule,} \\ &= \sum_{i=1}^d\sum_{j=1}^d M_{ij}\frac{\partial v_i}{\partial x_j} - \frac{\partial}{\partial x_j}(M_{ji}v_j) && \text{, because } M = M^T\text{,} \\ &= M : \nabla\mathbf{v} - \nabla\cdot(M\mathbf{v}).\end{aligned}$$

This gives us with the Gaussian Theorem

$$\begin{aligned}\int\limits_\Omega -\nabla\cdot\mathbb{T}(\mathbf{u},p)\cdot\mathbf{v} &= \int\limits_\Omega M : \nabla\mathbf{v} - \int\limits_\Omega p\nabla\cdot\mathbf{v} + \int\limits_{\partial\Omega} (p\mathbf{v})\cdot\mathbf{n} - (M\mathbf{v})\cdot\mathbf{n} \\ &= \int\limits_\Omega M : \nabla\mathbf{v} - \int\limits_\Omega p\nabla\cdot\mathbf{v} - \int\limits_{\partial\Omega} (M\mathbf{v} - p\mathbf{v})\cdot\mathbf{n} \\ &= \int\limits_\Omega M : \nabla\mathbf{v} - \int\limits_\Omega p\nabla\cdot\mathbf{v} - \langle T_N, \mathbf{v}\rangle_{\Gamma_N}.\end{aligned}$$

The last step works, because the test functions are zero at the boundary. The integrand can be reformulated as follows:

$$\begin{aligned}
M : \nabla \mathbf{v} &= \sum_{i=1}^{d} \sum_{j=1}^{d} M_{ij} \frac{\partial v_i}{\partial x_j} \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{2}(M_{ij} + M_{ji}) \frac{\partial v_i}{\partial x_j} && \text{, as } M = M^T, \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{2} M_{ij} \left( \frac{\partial v_j}{\partial x_i} + \frac{\partial v_i}{\partial x_j} \right) \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} 2\nu (\mathbb{D}(\mathbf{u}))_{ij} \cdot \frac{1}{2} \left( \frac{\partial v_j}{\partial x_i} + \frac{\partial v_i}{\partial x_j} \right) \\
&= (2\nu\, \mathbb{D}(\mathbf{u})) : \mathbb{D}(\mathbf{v}),
\end{aligned}$$

which then gives

$$M : \nabla \mathbf{v} - p \nabla \cdot \mathbf{v} = M : \nabla \mathbf{v} - p \sum_{i=1}^{d} \frac{\partial v_i}{\partial x_i} = (2\nu\, \mathbb{D}(\mathbf{u}) - p\, \mathbb{I}) : \mathbb{D}(\mathbf{v}) = \mathbb{T}(\mathbf{u}, p) : \mathbb{D}(\mathbf{v}).$$

So we finally get

$$(\mathbb{T}(\mathbf{u}, p), \mathbb{D}(\mathbf{v}))_0 = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v} \rangle_{\Gamma_N}.$$

The second equations of (3) and (4) are the identical. Multiplication of a test function $q \in C^\infty(\Omega)$ and integration over $\Omega$ yields

$$\int_\Omega (\nabla \cdot \mathbf{u}) q = (\nabla \cdot \mathbf{u}, q)_0 = (-\nabla \cdot \mathbf{u}_D, q),$$

which gives the set of equations

$$\begin{cases}
(\nu \nabla \mathbf{u}, \nabla \mathbf{v})_0 - (p, \nabla \cdot \mathbf{v})_0 & = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v} \rangle_{\Gamma_N} \quad \text{for (3)}, \\
(\mathbb{T}(\mathbf{u}, p), \mathbb{D}(\mathbf{v}))_0 & = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v} \rangle_{\Gamma_N} \quad \text{for (4)}, \\
(\nabla \cdot \mathbf{u}, q)_0 & = (-\nabla \cdot \mathbf{u}_D, q)_0.
\end{cases}$$

These equations can be written in the form

$$\begin{cases}
a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) & = \langle \tilde{\mathbf{f}}, \mathbf{v} \rangle, \\
b(\mathbf{u}, q) & = (\nabla \cdot \mathbf{u}_D, q).
\end{cases}$$

For the Laplace form (3) almost nothing has to be done to apply the theory of Section 2.1.1, it follows

$$\begin{cases}
a(\mathbf{u}, \mathbf{v}) & = (\nu \nabla \mathbf{u}, \nabla \mathbf{v})_0, \\
b(\mathbf{v}, p) & = -(\nabla \cdot \mathbf{v}, p)_0, \\
\langle \tilde{\mathbf{f}}, \mathbf{v} \rangle & = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v} \rangle_{\Gamma_N}.
\end{cases} \tag{8}$$

In the second case (4) one can see, that

$$\mathbb{T}(\mathbf{u}, p) : \mathbb{D}(\mathbf{v}) = (2\nu\, \mathbb{D}(\mathbf{u}) - p\, \mathbb{I}) : \mathbb{D}(\mathbf{v}) = (2\nu\, \mathbb{D}(\mathbf{u})) : \mathbb{D}(\mathbf{v}) - p \nabla \cdot \mathbf{v},$$

so we get

$$\begin{cases} a(\mathbf{u},\mathbf{v}) & = (2\nu\, \mathbb{D}(\mathbf{u}), \mathbb{D}(\mathbf{v}))_0, \\ b(\mathbf{v},p) & = -(\nabla\cdot\mathbf{v}, p)_0, \\ \langle\tilde{\mathbf{f}},\mathbf{v}\rangle & = (\mathbf{f},\mathbf{v})_0 + \langle T_N, \mathbf{v}\rangle_{\Gamma_N}. \end{cases} \tag{9}$$

**2.2.3 Remark.** The test functions $q$ are from $C^\infty(\Omega)$ which is not complete in the $\|\cdot\|_0$ or $\|\cdot\|_{H^1(\Omega)}$ norm, so one cannot apply the Ritz/Galerkin method in the end as they rely on Hilbert and Banach space theory. But $C^\infty(\Omega)$ is dense in the $L^2(\Omega)$ so one can just take the completion in the $L^2$-norm as test space and obtain $L^2(\Omega)$.
A similar approach can be applied for the test functions $\mathbf{v} \in (C^\infty_{\Gamma_D}(\Omega))^d$. There one can take the completion with respect to the $H^1$-Norm:

$$V := \left(\overline{C^\infty_{\Gamma_D}(\Omega)}^{H^1(\Omega)}\right)^d \tag{10}$$

where

$$\|u\|_{H^1(\Omega)} := \left(\|\nabla u\|_0^2 + \|u\|_0^2\right)^{\frac{1}{2}},$$

and obtain a space which has functions that vanish on the Dirichlet boundary and therefore is in between $(H^1(\Omega))^d$ and $(H^1_0(\Omega))^d$:

$$V = \left\{\mathbf{v} \in \left(H^1(\Omega)\right)^d : \mathbf{v}\big|_{\Gamma_D} = \mathbf{0}\right\},\ (H^1_0(\Omega))^d \subset V \subset (H^1(\Omega))^d,$$

where the restriction of $\mathbf{v}$ is meant in the sense of traces. Even though we completed the test space with respect to the $H^1$-norm we are going to equip the space $V$ with the $|\cdot|_{H^1(\Omega)}$ semi-norm

$$|u|_{H^1(\Omega)} := \|\nabla u\|_0,$$

which will turn out to be advantageous in Theorem 2.2.5 about existence and uniqueness of a weak solution. This semi norm is in fact a norm of $V$ because of the Poincaré inequality

$$\int\limits_\Omega u\cdot u \le C \int\limits_\Omega \nabla u \cdot \nabla u \Rightarrow \|u\|_0 \le C|u|_{H^1(\Omega)} \le C\|u\|_{H^1(\Omega)} \tag{11}$$

and the property that $H^1(\Omega) \subset H^0(\Omega) = L^2(\Omega)$, so

$$|v|_{H^1(\Omega)} =: \|v\|_V = 0 \Rightarrow \|v\|_0 \le 0 \Rightarrow v = 0.$$

Finally the weak problem reads as follows:
Find $(\mathbf{u}, p) \in V \times L^2(\Omega)$ such that for all $\mathbf{v} \in V$ and for all $q \in L^2(\Omega)$ it holds

$$\begin{cases} a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) & = \langle\tilde{\mathbf{f}}, v\rangle, \\ b(\mathbf{u},q) & = (\nabla\cdot\mathbf{u}_D, q)_0. \end{cases} \tag{12}$$

**2.2.4 Remark.** If $\Gamma_N = \emptyset$, the pressure $p$ does not appear in any boundary conditions. Looking at (3),(4) motivates, that it is just fixed up to a constant because it only appears as gradient. Indeed, taking a solution $(\mathbf{u}_0, p_0)$ of the weak Stokes problem (12) and shifting $p_0$

by a constant $c \in \mathbb{R}$ yields

$$\begin{aligned} a(\mathbf{u}_0, \mathbf{v}) + b(\mathbf{v}, p_0 + c) &= \langle \tilde{\mathbf{f}}, v \rangle \\ &\Leftrightarrow b(\mathbf{v}, c) = 0 && \text{, due to linearity,} \\ &\Leftrightarrow c \int_\Omega \nabla \cdot \mathbf{v} = 0 && \text{, per definition,} \\ &\Leftrightarrow c \int_{\Gamma_D} \mathbf{v} \cdot \mathbf{n} = 0 && \text{, using the Gaussian theorem,} \\ &\Leftrightarrow 0 = 0 && \text{, as } \mathbf{v}\big|_{\Gamma_D} = 0. \end{aligned}$$

Therefore, $L^2(\Omega)$ as test space would give a pressure that is defined up to a constant. In this case, one fixes this constant for example by changing the ansatz and test space to

$$L_0^2(\Omega) := \left\{ v \in L^2(\Omega) : \int_\Omega v = 0 \right\}.$$

**2.2.5 Theorem**[Existence and uniqueness of the Stokes problem]**.** The Stokes problem (12) has one and only one solution.

*Proof.* Problem (12) has the form of the problem discussed in Section 2.1.1. In this case, (1) transforms to (12) with $X = V$, $M = L^2(\Omega)$, $\ell = \langle \tilde{\mathbf{f}}, \cdot \rangle$, $\chi = (\nabla \cdot \mathbf{u}_D, \cdot)_0$. For existence and uniqueness from Theorem 2.1.2, one needs to check continuity and coercivity for $a(\cdot,\cdot)$ and the inf-sup condition and continuity for $b(\cdot,\cdot)$.

(i) Continuity of $a(\cdot,\cdot)$, for simplicity just shown for (8):

$$\begin{aligned} |a(\mathbf{v}, \mathbf{w})| &= |(\nu \nabla \mathbf{v}, \nabla \mathbf{w})_0| \\ &\leq \nu \|\mathbf{v}\|_V \|\mathbf{w}\|_V && \text{, using the Cauchy–Schwarz inequality.} \end{aligned}$$

(ii) Coercivity of $a(\cdot,\cdot)$:

- For (8):

$$a(\mathbf{v}, \mathbf{v}) = (\nu \nabla \mathbf{v}, \nabla \mathbf{v})_0 = \nu \|\mathbf{v}\|_V^2,$$

  so $a(\cdot,\cdot)$ is coercive.

- For (9): In this case, the first Korn inequality is needed. It states (see Section 2.2, Lemma 6 of [13]), that there is for all $\mathbf{v} \in V$ a constant $\kappa \geq 0$, such that

$$\|\mathbf{v}\|_{H^1(\Omega)}^2 \leq \kappa \int_\Omega \sum_{i,j=1}^d \frac{1}{4} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right)^2.$$

  So we get

$$\begin{aligned} a(\mathbf{v}, \mathbf{v}) &= (2\nu \, \mathbb{D}(\mathbf{v}), \mathbb{D}(\mathbf{v}))_0 \\ &= \nu \int_\Omega \sum_{i,j=1}^d \frac{1}{2} \left( \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right)^2 \\ &\geq \frac{\nu}{\kappa} \|\mathbf{v}\|_{H^1(\Omega)}^2 && \text{, using the first Korn inequality,} \\ &\geq C \frac{\nu}{\kappa} \|\mathbf{v}\|_V^2 && \text{, using the Poincaré inequality (11).} \end{aligned}$$

(iii) Continuity of $b(\cdot,\cdot)$: First of all it is $\|\mathbf{x}\|_{\ell^1} \leq \sqrt{d}\|\mathbf{x}\|_{\ell^2}$ for all $\mathbf{x} \in \mathbb{R}^d$, see Section 8.4.1 of [2]. Now let $A = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d\times d}$ be a $d \times d$-matrix, then

$$\begin{aligned}\left|\sum_{i=1}^d a_{ii}\right| &\leq \sum_{i=1}^d |a_{ii}| \\ &\leq \sqrt{d}\left(\sum_{i=1}^d a_{ii}^2\right)^{1/2} && \text{, using the above inequality,} \\ &\leq \sqrt{d}\left(\sum_{i,j=1}^d a_{ij}^2\right)^{1/2} .\end{aligned}$$

This yields for $b(\cdot,\cdot)$ :

$$\begin{aligned}|b(\mathbf{v},q)| = |(\nabla\cdot\mathbf{v},q)_0| &\leq \|\nabla\cdot\mathbf{v}\|_0\|q\|_0 && \text{, using the Cauchy–Schwarz inequality,} \\ &= \left|\int_\Omega \sum_{i=1}^d \frac{\partial v_i}{\partial x_i}\right| \|q\|_0 \\ &\leq \sqrt{d}\left(\int_\Omega \sum_{i,j=1}^d (\frac{\partial v_i}{\partial x_j})^2\right)^{1/2} \|q\|_0 \\ &= \sqrt{d}\|\nabla\mathbf{v}\|_0\|q\|_0 = \sqrt{d}\|\mathbf{v}\|_V\|q\|_0 && \text{, per definition of } \|\cdot\|_V.\end{aligned}$$

(iv) Inf-sup condition: The inf-sup condition is for both problems the same. It is proven in Theorem 3.7 of Chapter I in [9].

Hence there exists one and only one solution. □

## 2.3 The Navier–Stokes equations

The Navier–Stokes equations describe, similarly to the Stokes equations, the motion of fluids. The difference is an additional term which describes convective acceleration. Convective acceleration is the change of velocity with respect to the position (and not of a fixed position with respect to time, which would be "local acceleration"). This kind of acceleration is the main contributor to turbulences. The term however is nonlinear, which increases the difficulties in analyzing and solving these equations compared with the Stokes equations.

The Navier–Stokes problem of an incompressible flow in the steady case (i.e., without time dependence) reads as follows:

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $\mathbf{f} \in L^2(\Omega)$, then find $(\mathbf{u},p) : \Omega \to \mathbb{R}^d \times \mathbb{R}$, such that

$$\begin{cases} -\nabla\cdot\mathbb{T}(\mathbf{u},p) + (\mathbf{u}\cdot\nabla)\cdot\mathbf{u} &= \mathbf{f} \quad \text{in } \Omega, \\ \nabla\cdot\mathbf{u} &= 0 \quad \text{in } \Omega, \end{cases} \tag{13}$$

where $\mathbf{u}$ denotes the velocity, $p$ the pressure and $\mathbb{T}(\mathbf{u},p) := 2\nu\,\mathbb{D}(\mathbf{u}) - p\,\mathbb{I}$ the Cauchy stress tensor, as already seen in (4). The boundary conditions are the same as in the Stokes problem, assuming a homogeneous Dirichlet boundary for simplicity: Let $\partial\Omega$ be split into two disjoint, relatively open parts $\Gamma_D$ and $\Gamma_N$ for the Dirichlet boundary and Neumann boundary part, respectively, with $\bar{\Gamma}_N \cup \bar{\Gamma}_D = \partial\Omega$ and $\Gamma_D \cap \Gamma_N = \emptyset$. Then for $T_N \in H^{-1/2}(\Gamma_N)$ it holds, that

$$\begin{cases} \mathbf{u} = 0 & \text{on } \Gamma_D, \quad (14) \\ \mathbb{T}(\mathbf{u}, p) \cdot \mathbf{n} = T_N & \text{on } \Gamma_N. \quad (15) \end{cases}$$

**Remark.** One searches for $\mathbf{u} \in (H^1(\Omega))^d$, which means the corresponding space of traces is $(H^{1/2}(\partial\Omega))^d$ and therefore the space for the Neumann conditions is the dual space $H^{-1/2}(\Gamma_N)$ such that (15) makes sense.
The homogeneous Dirichlet boundary conditions can be assumed, as inhomogeneous conditions would matter in the analytical, but not that much in the numerical approach. There they are incorporated by introducing an identity block in the matrix and writing the boundary values to the right-hand side.

### 2.3.1 Weak formulation

Introducing a trilinear form

$$c(\mathbf{w}, \mathbf{u}, \mathbf{v}) = \int_\Omega ((\mathbf{w} \cdot \nabla) \cdot \mathbf{u}) \cdot \mathbf{v}, \tag{16}$$

multiplication with test functions $\mathbf{v} \in V$ for the first equation, $q \in L^2(\Omega)$ for the second equation (see Remark 2.2.3 for choice of spaces and definition of $V$) and integration by parts yields as in the Stokes case the set of equations

$$\begin{cases} a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) + c(\mathbf{u}, \mathbf{u}, \mathbf{v}) & = \langle \tilde{\mathbf{f}}, \mathbf{v} \rangle, \\ b(\mathbf{u}, q) & = 0 \end{cases} \tag{17}$$

with

$$\begin{cases} a(\mathbf{u}, \mathbf{v}) & = (2\nu\, \mathbb{D}(\mathbf{u}), \mathbb{D}(\mathbf{v}))_0, \\ b(\mathbf{v}, p) & = -(\nabla \cdot \mathbf{v}, p)_0, \\ c(\mathbf{w}, \mathbf{u}, \mathbf{v}) & = ((\mathbf{w} \cdot \nabla) \cdot \mathbf{u}, \mathbf{v})_0, \\ \langle \tilde{\mathbf{f}}, \mathbf{v} \rangle & = (\mathbf{f}, \mathbf{v})_0 + \langle T_N, \mathbf{v} \rangle_{\Gamma_N}. \end{cases} \tag{18}$$

The trilinearity of $c(\cdot, \cdot, \cdot)$ makes the whole problem nonlinear. The Navier–Stokes problem then reads as follows: Find $(\mathbf{u}, p) \in V \times L^2(\Omega)$ such that for all $\mathbf{v} \in V$ and $q \in L^2(\Omega)$ the equations (17) hold.

### 2.3.2 Fixed-point iteration

As the weak formulation is not linear anymore, one cannot simply assemble a big system of linear equations and solve it - one needs to iteratively approximate the solution. There are several approaches, but the most commonly used method is performing a fixed-point iteration which yields Oseen-type equations.

**Algorithm**[See Chapter 11 of [8]]. Given a $\mathbf{u}^0 \in V$, find $\mathbf{u}^m \in V$ and $p^m \in L^2(\Omega)$ for $m > 0$ with

$$\begin{cases} a(\mathbf{u}^m, \mathbf{v}) + b(\mathbf{v}, p^m) + c(\mathbf{u}^{m-1}, \mathbf{u}^m, \mathbf{v}) & = \langle \tilde{\mathbf{f}}, \mathbf{v} \rangle \\ b(\mathbf{u}^m, q) & = 0 \end{cases}$$

for all $\mathbf{v} \in V$, $q \in L^2(\Omega)$.

In this way one obtains in each step a linear system, as $\mathbf{u}^{m-1}$ is known and therefore $c(\mathbf{u}^{m-1},\cdot,\cdot)$ transforms into a bilinear form. This kind of equations is known as Oseen-type equations. A common choice for $\mathbf{u}^0$ is the solution of the according Stokes problem, i.e., without nonlinear term.
As solving the nonlinear Navier–Stokes problem involves solving linear Oseen problems iteratively, existence and uniqueness of a solution of these problems are of interest.

**2.3.1 Theorem**[Existence and uniqueness of an Oseen type problem]**.** Let $\mathbf{w} \in (L^\infty(\Omega))^d$ with $\nabla \cdot \mathbf{w} = 0$ and $\mathbf{w} \cdot \mathbf{n} > 0$ on $\Gamma_N$ be fixed. The Oseen problem to find $(\mathbf{u}, p) \in V \times L^2(\Omega)$ such that

$$\begin{cases} a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) + c(\mathbf{w},\mathbf{u},\mathbf{v}) & = \langle \tilde{\mathbf{f}}, \mathbf{v}\rangle, \\ \hfill b(\mathbf{u},q) & = 0 \end{cases}$$

for all $\mathbf{v} \in V$, $q \in L^2(\Omega)$ has one and only one solution.

*Proof.* As the Oseen problem is of the form of problem (1), we have to check the assumptions of Theorem 2.1.2 in order to get existence and uniqueness.

(i) Continuity of $a(\cdot,\cdot) + c(\mathbf{w},\cdot,\cdot)$: Continuity of $a(\cdot,\cdot)$ was already proven in Theorem 2.2.5. Continuity of $c(\mathbf{w},\cdot,\cdot)$:

$$\begin{aligned} c(\mathbf{w},\mathbf{v}_1,\mathbf{v}_2) &= ((\mathbf{w}\cdot\nabla)\cdot\mathbf{v}_1,\mathbf{v}_2)_0 \\ &\le \|(\mathbf{w}\cdot\nabla)\cdot\mathbf{v}_1\|_0\|\mathbf{v}_2\|_0 && \text{, using the Cauchy–Schwarz inequality,} \\ &\le \|\mathbf{w}\|_{L^\infty(\Omega)}\|(\mathbf{1}\cdot\nabla)\cdot\mathbf{v}_1\|_0\|\mathbf{v}_2\|_0 \\ &\le \|\mathbf{w}\|_{L^\infty(\Omega)}\|\mathbf{v}_1\|_V\|\mathbf{v}_2\|_0 \\ &\le \|\mathbf{w}\|_{L^\infty(\Omega)}\|\mathbf{v}_1\|_V\|\mathbf{v}_2\|_V && \text{, using the Poincaré inequality (11).} \end{aligned}$$

(ii) Coercivity of $a(\cdot,\cdot) + c(\mathbf{w},\cdot,\cdot)$ on $\ker B = \{\mathbf{v} \in V : b(\mathbf{v},q) = 0\ \forall q \in Q\}$: Let $\mathbf{v} \in \ker B$, then

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) + c(\mathbf{w},\mathbf{v},\mathbf{v}) &= (2\nu\,\mathbb{D}(\mathbf{v}),\mathbb{D}(\mathbf{v}))_0 + ((\mathbf{w}\cdot\nabla)\cdot\mathbf{v},\mathbf{v})_0 \\ &\ge C\|\mathbf{v}\|_V + ((\mathbf{w}\cdot\nabla)\cdot\mathbf{v},\mathbf{v})_0 && \text{, using coercivity of } a(\cdot,\cdot), \\ &\ge C\|\mathbf{v}\|_V, \end{aligned}$$

because having a closer look at the second term yields

$$\begin{aligned} ((\mathbf{w}\cdot\nabla)\mathbf{v},\mathbf{v})_0 &= \sum_{i,j=1}^d \int_\Omega w_j \frac{\partial v_i}{\partial x_j} v_i \\ &= \sum_{i,j=1}^d \int_\Omega \frac{\partial}{\partial x_j}(w_j v_i) v_i - \sum_{i=1}^d \int_\Omega (\nabla\cdot\mathbf{w}) v_i v_i && \text{, using product rule,} \\ &= \sum_{i,j=1}^d \int_\Omega \frac{\partial}{\partial x_j}(w_j v_i) v_i && \text{, as } \nabla\cdot\mathbf{w} = 0, \\ &= \sum_{i,j=1}^d \int_{\Gamma_N} w_j v_i v_i \cdot \mathbf{n} - ((\mathbf{w}\cdot\nabla)\mathbf{v},\mathbf{v})_0 \\ &\ge 0 && \text{, with } \mathbf{w}\cdot\mathbf{n} > 0 \text{ on } \Gamma_N, \\ \Rightarrow & ((\mathbf{w}\cdot\nabla)\mathbf{v},\mathbf{v})_0 \ge 0. \end{aligned}$$

(iii) Inf-sup condition: The inf-sup condition is fulfilled, as $b(\cdot,\cdot)$ is the same as in the Stokes problem, see proof of Theorem 2.2.5.

□

## 2.4 The Darcy equations

Darcy's law describes the flow of a fluid through a porous medium. It first was formulated on empirical results, later one found that it could also be deduced from the Navier–Stokes equations. It usually reads as follows (called "mixed form"):
For $\Omega \subset \mathbb{R}^d$ a bounded domain with Lipschitz boundary, $f \in L^2(\Omega)$, find $\mathbf{u} : \Omega \to \mathbb{R}^d$ and $\varphi : \Omega \to \mathbb{R}$ such that

$$\begin{cases} \mathbf{u} + \mathbb{K}\,\nabla\varphi &= 0 \quad \text{in } \Omega, \\ \nabla \cdot \mathbf{u} &= f \quad \text{in } \Omega, \end{cases}$$

with $\varphi$ being the so called piezometric head which basically represents the fluid pressure in $\Omega$, $\mathbf{u}$ describing the velocity and $\mathbb{K}$ being the hydraulic conductivity tensor, describing the characteristics of the porous medium and properties of the fluid. Here, $f$ is called the source term (of fluid).
However, for simplicity we are going to consider the so called "primal form" which can be deduced by taking the divergence of the first equation and substituting the second one:

$$-\nabla \cdot \mathbb{K}\,\nabla\varphi = f \quad \text{in } \Omega. \tag{19}$$

For a complete description of the problem, one needs boundary conditions as well. As in the case of Stokes, the focus is on Dirichlet and Neumann-type boundaries. Decompose $\partial\Omega$ into two relatively open, disjoint parts $\Gamma_{\text{nat}}, \Gamma_{\text{ess}} \subset \partial\Omega$ such that

- $\Gamma_{\text{nat}} \cap \Gamma_{\text{ess}} = \emptyset$,
- $\bar{\Gamma}_{\text{nat}} \cup \bar{\Gamma}_{\text{ess}} = \partial\Omega$.

Then a solution of (19) should fulfill

$$\begin{cases} (-\,\mathbb{K}\,\nabla\varphi)\cdot\mathbf{n} = u_{\text{nat}} & \text{on } \Gamma_{\text{nat}}, \\ \varphi = \varphi_{\text{ess}} & \text{on } \Gamma_{\text{ess}}, \end{cases} \tag{20}$$

with $u_{\text{nat}}$ being a prescribed velocity and $\varphi_{\text{ess}}$ being a prescribed pressure.

**Remark.** In this case, $\Gamma_{\text{nat}}$ corresponds to the Neumann part of the boundary (prescribed derivative) and $\Gamma_{\text{ess}}$ to the Dirichlet part (prescribed value). If one considers the mixed form of the Darcy problem, the boundary conditions would switch, i.e., the mixed form Dirichlet condition coincides with the primal form Neumann condition and vice versa. To avoid this confusion, it makes sense to talk about natural and essential boundary conditions. Natural boundary conditions are boundary conditions that can be incorporated into the weak formulation by substitution, essential boundary conditions are the ones that have impact on the ansatz and test space.

### 2.4.1 Weak formulation

Even though it can be generally assumed that $\mathbb{K}$ is symmetric and even positive definite, we will only consider $\mathbb{K} = K \cdot \mathbb{I}$, $K > 0$ for simplicity.
Consider a test function $\psi$ which vanishes close to the essential boundary, i.e.,

$$\psi \in C^\infty_{\Gamma_\text{ess}}(\Omega) = \left\{ v \in C^\infty(\Omega) : v\big|_{\Gamma_\text{ess}} = 0 \right\}.$$

Multiplication (19) with these test functions and integration by parts yields:

$$\begin{aligned}
\int\limits_\Omega -(\nabla \cdot \mathbb{K}\,\nabla\varphi)\psi &= -\int\limits_\Omega \nabla \cdot \begin{pmatrix} K\frac{\partial\varphi}{\partial x_1} \\ \vdots \\ K\frac{\partial\varphi}{\partial x_d} \end{pmatrix} \psi = -K \int\limits_\Omega \sum_{i=1}^d \frac{\partial^2\varphi}{\partial x_i^2}\psi \\
&= -K \int\limits_\Omega \sum_{i=1}^d \frac{\partial}{\partial x_i}\left(\frac{\partial\varphi}{\partial x_i}\psi\right) - \frac{\partial\varphi}{\partial x_i}\frac{\partial\psi}{\partial x_i} && \text{, product rule,} \\
&= -K \int\limits_{\partial\Omega} \sum_{i=1}^d \frac{\partial\varphi}{\partial x_i}\psi \cdot n + K \int\limits_\Omega \nabla\varphi \cdot \nabla\psi && \text{, Gaussian theorem,} \\
&= \int\limits_{\Gamma_\text{nat}} -K\nabla\varphi \cdot \psi \cdot \mathbf{n} + (\mathbb{K}\,\nabla\varphi, \nabla\psi)_0 \\
&= (\mathbb{K}\,\nabla\varphi, \nabla\psi)_0 + \langle u_\text{nat}, \psi\rangle_{\Gamma_\text{nat}}.
\end{aligned}$$

So one ends up with the form

$$(\mathbb{K}\,\nabla\varphi, \nabla\psi)_0 = (f, \psi)_0 - \langle u_\text{nat}, \psi\rangle_{\Gamma_\text{nat}}, \tag{21}$$

which still can be simplified as we assume $f \equiv 0$ in our case. This gives:

$$(\mathbb{K}\,\nabla\varphi, \nabla\psi)_0 = -\langle u_\text{nat}, \psi\rangle_{\Gamma_\text{nat}}. \tag{22}$$

In order to be able to apply Hilbert space theory, we need to complete our test space as it is not complete in the norm of $H^1(\Omega)$. One then gets a new test space

$$V = \overline{C^\infty_{\Gamma_\text{ess}}(\Omega)}^{H^1(\Omega)},$$

which contains the functions which are zero close to the essential boundary in the sense of traces. Hence it is between $H^1_0(\Omega)$ and $H^1(\Omega)$. Analogously as in the Stokes problem, we can equip the space $V$ with the semi norm $|\cdot|_{H^1(\Omega)} =: \|\cdot\|_V$, which in fact is a norm on $V$ due to the Poincaré inequality (11).
The weak formulation reads as follows: Find $\varphi \in H^1(\Omega)$ with $\varphi - \varphi_\text{ess} \in V$ such that

$$a(\varphi, \psi) = (\mathbb{K}\,\nabla\varphi_\text{ess}, \nabla\psi)_0 - \langle u_\text{nat}, \psi\rangle_{\Gamma_\text{nat}} =: \langle \tilde{f}, \psi\rangle \qquad \forall\psi \in V, \tag{23}$$

with

$$a : V \times V \to \mathbb{R},\ a(\varphi, \psi) = (\mathbb{K}\,\nabla\varphi, \nabla\psi)_0.$$

Note that $\varphi - \varphi_\text{ess}$ at first is not reasonable because $\varphi$ is defined on $\Omega$ and $\varphi_\text{ess}$ on the boundary. So one needs to extend $\varphi_\text{ess}$ into the interior. The existence of such an extension is guaranteed because the trace operator is surjective.

**2.4.1 Theorem**[Existence and uniqueness of the Darcy problem]**.** Problem (23) with $\tilde{f} \in L^2(\Omega)$, $u_{\text{nat}} \in H^{-1/2}(\Gamma_{\text{nat}})$ and $\varphi_{\text{ess}} \in H^{1/2}(\Gamma_{\text{ess}})$ has one and only one solution.

*Proof.* The proof is basically an application of the theorem of Lax–Milgram. It states that there exists exactly one solution to

$$a(\varphi, \psi) = \langle \tilde{f}, \psi \rangle \quad \forall \psi \in V$$

if $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ is bounded and coercive and $\tilde{f}$ is linear and bounded.

- Boundedness of $a(\cdot,\cdot)$:

$$|a(\varphi, \psi)| = K|(\nabla\varphi, \nabla\psi)_0| \leq K\|\varphi\|_V\|\psi\|_V \quad \text{, using the Cauchy–Schwarz inequality.}$$

- Coercivity of $a(\cdot,\cdot)$:

$$a(\varphi, \varphi) = (\mathbb{K}\,\nabla\varphi, \nabla\varphi)_0 = K(\varphi, \varphi)_V = K\|\varphi\|_V^2.$$

- Boundedness of the right-hand side:

$$\begin{aligned} |\langle \tilde{f}, v\rangle| &= |(\tilde{f}, v)_0| \\ &\leq \|\tilde{f}\|_0\|v\|_0 && \text{, using the Cauchy–Schwarz inequality,} \\ &\leq C\|\tilde{f}\|_0\|v\|_V < \infty && \text{, using the Poincaré inequality (11).} \end{aligned}$$

□

# 3 Coupled flow problems

This chapter deals with problems that consist of two or more subproblems and their discretization. These subproblems have their own domain which couples at the whole boundary of the domain or at a subset of it. This part of the boundary is usually called interface.

## 3.1 Stokes–Darcy problem

This section is about the coupled Stokes–Darcy system. One splits the domain $\Omega$ into two parts $\Omega_f$ and $\Omega_p$ for the Stokes system and Darcy system respectively, such that

- $\bar{\Omega} = \bar{\Omega}_f \cup \bar{\Omega}_p$,
- $\Omega_f \cap \Omega_p = \emptyset$,
- $\bar{\Omega}_f \cap \bar{\Omega}_p = \Gamma$,

with $\Gamma$ being the so called interface between $\Omega_f$ and $\Omega_p$, see Figure 1 for a sketch. One obtains the problem

$$\left\{ \begin{aligned} -\nabla \cdot \mathbb{T}(\mathbf{u}_f, p_f) &= \mathbf{f}_f && \text{in } \Omega_f, \\ \nabla \cdot \mathbf{u}_f &= 0 && \text{in } \Omega_f, \\ -\nabla \cdot \mathbb{K}\,\nabla\varphi_p &= f_p && \text{in } \Omega_p, \end{aligned} \right. \tag{24}$$

for $\mathbf{u}_f : \Omega_f \to \mathbb{R}^d$, $p_f : \Omega_f \to \mathbb{R}$ and $\varphi_p : \Omega_p \to \mathbb{R}$. In order to be able to assign boundary conditions (including conditions on the interface), we need to split $\partial\Omega$ as well: Let $\Gamma_{f,n}, \Gamma_{f,e} \subset \partial\Omega_f \setminus \Gamma$ and $\Gamma_{p,n}, \Gamma_{p,e} \subset \partial\Omega_p \setminus \Gamma$ such that

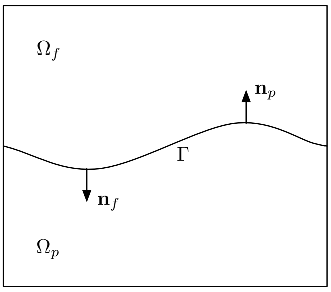


Figure 1: Sketch of the domain used in the Stokes–Darcy problem.

- $\bar{\Gamma}_{f,n} \cup \bar{\Gamma}_{f,e} = \partial\Omega_f \setminus \Gamma$,
- $\Gamma_{f,n} \cap \Gamma_{f,e} = \emptyset$,
- $\bar{\Gamma}_{p,n} \cup \bar{\Gamma}_{p,e} = \partial\Omega_f \setminus \Gamma$,
- $\Gamma_{p,n} \cap \Gamma_{p,e} = \emptyset$,
- $\text{meas}(\Gamma_{f,n}) > 0$.

Then one can impose:

$$\begin{cases} \mathbf{u}_f &= \mathbf{u}_{f,\text{ess}} \quad \text{on } \Gamma_{f,e} \text{ (essential, Dirichlet)}, \\ \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n} &= T_{f,\text{nat}} \quad \text{on } \Gamma_{f,n} \text{ (natural, Neumann)}, \\ \varphi_p &= \varphi_{p,\text{ess}} \quad \text{on } \Gamma_{p,e} \text{ (essential, Dirichlet)}, \\ (-\,\mathbb{K}\,\nabla\varphi) \cdot \mathbf{n} &= u_{p,\text{nat}} \quad \text{on } \Gamma_{p,n} \text{ (natural, Neumann)}. \end{cases} \tag{25}$$

### 3.1.1 Interface conditions

On the interface one requires several conditions in order to couple the systems:

(i) The preservation of normal velocity: Let $\mathbf{n}_f$ be the unit outer normal vector of $\Omega_f$, then

$$\mathbf{u}_f \cdot \mathbf{n}_f = \mathbf{u}_p \cdot \mathbf{n}_f = -(\mathbb{K}\,\nabla\varphi) \cdot \mathbf{n}_f. \tag{26}$$

(ii) The preservation of normal stress:

$$-\mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f = g\varphi_p, \tag{27}$$

where $g$ denotes the gravitational acceleration.

(iii) The behavior of tangential velocity, Beavers–Joseph condition: Let $\boldsymbol{\tau}_i$, $i = 1, \dots, d-1$ be pairwise orthogonal unit tangential vectors on $\Gamma$, then

$$(\mathbf{u}_f - \mathbf{u}_p) \cdot \boldsymbol{\tau}_i + \alpha\boldsymbol{\tau}_i \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f = 0$$

with

$$\alpha = \alpha_0 \sqrt{\frac{\tau\,\mathbb{K}\,\tau}{\nu g}} = \alpha_0 \sqrt{\frac{K}{\nu g}},$$

where $\alpha_0$ is a dimensionless parameter which has to be determined experimentally and describes properties of the porous medium. This condition can be simplified to the so-called Beavers–Joseph–Saffman condition

$$\mathbf{u}_f \cdot \boldsymbol{\tau}_i + \alpha \boldsymbol{\tau}_i \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f = 0, \tag{28}$$

as it turned out that the Darcy velocity $\mathbf{u}_p$ is often negligible on $\Gamma$ compared to $\mathbf{u}_f$, see Section 3 of [8], equation (3.2).

### 3.1.2 Weak formulation

To obtain the weak formulation one needs appropriate test spaces. These are similar to the test spaces used in the previous sections about the Stokes and Darcy problem, i.e.:

- Test space for the Stokes velocity:
$$V_f = \left\{ \mathbf{v} \in (H^1(\Omega_f))^d : \mathbf{v}\big|_{\Gamma_{f,e}} = 0 \right\}.$$
- Test space for the Stokes pressure:
$$Q_f = L^2(\Omega_f).$$
- Test space for the Darcy pressure:
$$V_p = \left\{ v \in H^1(\Omega_p) : v\big|_{\Gamma_{p,e}} = 0 \right\}.$$

Now multiplication with a test function $\mathbf{v} \in V_f$ of the first equation, $q \in Q_f$ for the second equation, $\psi \in V_p$ for the third equation, integration and integration by parts gives the previous results (9) and (23) plus an additional term on the interface:

$$\begin{cases} (\mathbb{T}(\mathbf{u}_f, p_f), \mathbb{D}(\mathbf{v}))_{0,\Omega_f} &= \langle \mathbf{f}_f^1, \mathbf{v} \rangle + \langle \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \rangle_\Gamma, \\ (-\nabla \cdot \mathbf{u}_f, q)_{0,\Omega_f} &= \langle f_f^2, q \rangle, \\ (\mathbb{K}\, \nabla \varphi_p, \nabla \psi)_{0,\Omega_p} &= \langle \tilde{f}_p, \psi \rangle + \langle -\, \mathbb{K}\, \nabla \varphi_p \cdot \mathbf{n}_f, \psi \rangle_\Gamma, \end{cases}$$

with $\mathbf{n}_f = -\mathbf{n}_p$ on $\Gamma$ and

$$\begin{aligned} \mathbf{f}_f^1 &\in V_f', \, \langle \mathbf{f}_f^1, \mathbf{v} \rangle = (\mathbf{f}_f, \mathbf{v})_{0,\Omega_f} + \langle T_{f,\text{nat}}, \mathbf{v} \rangle_{\Gamma_{f,N}}, \\ f_f^2 &\in Q_f', \, \langle f_f^2, q \rangle = (\nabla \cdot \mathbf{u}_{f,\text{ess}}, q)_{0,\Omega_f}, \\ \tilde{f}_p &\in V_p', \, \langle f_p, \psi \rangle = (\mathbb{K}\, \nabla \varphi_{p,\text{ess}}, \nabla \psi)_{0,\Omega_p} - \langle u_{p,\text{nat}}, \psi \rangle_\Gamma. \end{aligned}$$

Now the Bievers–Joseph–Saffman condition (28) can be included into the first equation of the weak formulation. Therefore decompose $\mathbf{v}$ into its normal and tangential components

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{n}_f) \cdot \mathbf{n}_f + \sum_{i=1}^{d-1} (\mathbf{v} \cdot \boldsymbol{\tau}_i) \cdot \boldsymbol{\tau}_i$$

yields

$$\begin{aligned} \langle \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \rangle_\Gamma &= \langle \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma + \sum_{i=1}^{d-1} \langle \boldsymbol{\tau}_i \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \cdot \boldsymbol{\tau}_i \rangle_\Gamma \\ &\overset{(28)}{=} \langle \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma - \sum_{i=1}^{d-1} \langle \frac{1}{\alpha} \mathbf{u}_f \cdot \boldsymbol{\tau}_i, \mathbf{v} \cdot \boldsymbol{\tau}_i \rangle_\Gamma. \end{aligned}$$

Rearranging terms leads to the following form:

$$\begin{cases} a_f(\mathbf{u}_f, \mathbf{v}) + b_f(\mathbf{v}, p_f) - \langle \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma &= \langle \mathbf{f}_f^1, \mathbf{v} \rangle, \\ b_f(\mathbf{u}_f, q) &= \langle f_f^2, q \rangle, \\ a_p(\varphi_p, \psi) + \langle \mathbb{K} \nabla \varphi_p \cdot \mathbf{n}_f, \psi \rangle_\Gamma &= \langle \tilde{f}_p, \psi \rangle, \end{cases} \tag{29}$$

with

- $a_f : V_f \times V_f \to \mathbb{R}$ and similar, as in (9):

$$a(\mathbf{u}, \mathbf{v}) = (2\nu\, \mathbb{D}(\mathbf{u}), \mathbb{D}(\mathbf{v}))_{0,\Omega_f} + \sum_{i=1}^{d} \frac{1}{\alpha} \langle \mathbf{u} \cdot \boldsymbol{\tau}_i, \mathbf{v} \cdot \boldsymbol{\tau}_i \rangle_\Gamma,$$

- $b_f : V_f \times Q_f \to \mathbb{R}$ with

$$b_f(\mathbf{v}, p) = -(\nabla \cdot \mathbf{v}, p)_{0,\Omega_f},$$

- $a_p : Q_p \times Q_p \to \mathbb{R}$ with

$$a_p(\varphi, \psi) = (\mathbb{K}\, \nabla \varphi, \nabla \psi)_{0,\Omega_p}.$$

Now one still has to include the remaining two interface conditions into the problems. Two different approaches are studied: Assign the conditions such that they both are Neumann conditions and obtain a "Neumann–Neumann" problem or assign a weighted linear combination of the two conditions, obtaining a "Robin–Robin" problem.

### 3.1.3 Neumann–Neumann coupling

Considering (27) as Neumann boundary condition for Stokes and (26) as a Neumann boundary condition for Darcy yields the following coupled problem: Find $(\mathbf{u}_f, p_f, \varphi_p) \in V_f \times Q_f \times Q_p$ such that for all $(\mathbf{v}, q, \psi) \in V_f \times Q_f \times Q_p$ holds, that

$$\begin{cases} a_f(\mathbf{u}_f, \mathbf{v}) + b_f(\mathbf{v}, p_f) + \langle g\varphi_p, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma &= \langle \mathbf{f}_f^1, \mathbf{v} \rangle, \\ b_f(\mathbf{u}_f, q) &= \langle f_f^2, q \rangle, \\ a_p(\varphi_p, \psi) - \langle \mathbf{u}_f \cdot \mathbf{n}_f, \psi \rangle_\Gamma &= \langle \tilde{f}_p, \psi \rangle. \end{cases} \tag{30}$$

### 3.1.4 Finite element discretization of the Neumann–Neumann problem

The finite element method is based on the Ritz-/Galerkin-Method, i.e., the spaces $V_f$, $Q_f$, $Q_p$ are discretized by $V_f^h, Q_f^h, Q_p^h$ with finite basis sets $\{\mathbf{w}_i\}_{i=1}^{N_u}$, $\{q_i\}_{i=1}^{N_p}$, $\{\psi_i\}_{i=1}^{N_\varphi}$, respectively, where $h$ stands for the refinement level and $N_u, N_p, N_\varphi$ for the number of degrees of freedom (DOF) of the finite element spaces. In these spaces one can search for a solution.

**Remark.** It is sufficient to use the basis functions of the discretized space as test functions, i.e.,

$$a(u, v) = f(v)\ \forall v \in V^h \Leftrightarrow a(u, \phi_i) = f(\phi_i)\ \forall i = 1, \dots, N,$$

because one can represent $v$ as linear combination of the basis functions $\{\phi_i\}_{i=1}^{N}$ of $V^h$

$$v = \sum_{i=1}^{N} \alpha_i \phi_i$$

for some $\alpha_i \in \mathbb{R}$ and plug this into the equation:

$$a(u, v) = \sum_{i=1}^{N} \alpha_i a(u, \phi_i) = \sum_{i=1}^{N} \alpha_i f(\phi_i) = f(v).$$

This equation holds if it holds for all basis functions. The other way around it holds for all functions from $V^h$, i.e., for the basis functions in particular, so it is equivalent if one tests just for the basis functions or for all functions of the space.

Now, taking the three equations of (30), having a look at them separately and replacing the test functions with the corresponding basis functions yields:

(i) One searches for $\mathbf{u}_f = \sum_{j=1}^{N_u} u_h^j \mathbf{w}_j$, $p_f = \sum_{k=1}^{N_p} p_h^k q_k$, $\varphi_p = \sum_{l=1}^{N_\varphi} \varphi_h^l \psi_l$. Inserting this into the first equation of (30) gives

$$\sum_{j=1}^{N_u} u_h^j a_f(\mathbf{w}_j, \mathbf{w}_i) + \sum_{k=1}^{N_p} p_h^k b_f(\mathbf{w}_i, q_k) + \sum_{l=1}^{N_\varphi} \varphi_h^l \langle g\psi_l, \mathbf{w}_i \cdot \mathbf{n}_f \rangle_\Gamma$$
$$= \langle \mathbf{f}_f^1, \mathbf{w}_i \rangle \quad \text{for } i = 1, \ldots, N_u,$$

which is equivalent to the following equation in matrix-vector form:

$$\begin{pmatrix} A & B & C_\Gamma^S \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \boldsymbol{\phi} \end{pmatrix} = \mathbf{f}_1,$$

with $A$ being a $(N_u \times N_u)$ matrix, $B$ being a $(N_u \times N_p)$ matrix, and $C_\Gamma^S$ being a $(N_u \times N_\varphi)$ matrix with entries

$$A_{i,j} = a_f(\mathbf{w}_j, \mathbf{w}_i),\ B_{i,k} = b_f(\mathbf{w}_i, q_k),\ C_{i,l}^S = \langle g\psi_l, \mathbf{w}_i \cdot \mathbf{n}_f \rangle_\Gamma$$

and vectors $\mathbf{u} = (u_h^1, \ldots, u_h^{N_u})^T$, $\mathbf{p} = (p_h^1, \ldots, p_h^{N_p})^T$, $\boldsymbol{\phi} = (\varphi_h^1, \ldots, \varphi_h^{N_\varphi})^T$. The right-hand side vector consists of $\langle \mathbf{f}_f^1, \mathbf{w}_i \rangle$ for each row $i = 1, \ldots, N_u$.

(ii) In the second equation one searches for $\mathbf{u}_f = \sum_{j=1}^{N_u} u_h^j \mathbf{w}_j$ for each basis function $q_i$, $i = 1, \ldots, N_p$. One ends up with

$$\sum_{j=1}^{N_u} u_h^j b_f(\mathbf{w}_j, q_i) = \langle f_f^2, q_i \rangle$$

which is

$$B^T \mathbf{u} = \mathbf{f}_2$$

with $B$ and $\mathbf{u}$ from (i) and $\langle f_f^2, q_i \rangle$ in each component of $\mathbf{f}_2$.

(iii) In the third equation one searches for $\varphi_p = \sum_{l=1}^{N_\varphi} \varphi_h^l \psi_l$ and $\mathbf{u}_f = \sum_{j=1}^{N_u} u_h^j \mathbf{w}_j$. Insertion gives for each $i = 1, \ldots, N_\varphi$

$$\sum_{j=1}^{N_\varphi} \varphi_h^j a_p(\psi_j, \psi_i) - \sum_{k=1}^{N_u} u_h^k \langle \mathbf{w}_k \cdot \mathbf{n}_f, \psi_i \rangle_\Gamma = \langle \tilde{f}_p, \psi_i \rangle$$

which is equivalent to

$$\begin{pmatrix} C_\Gamma^D & D \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\phi} \end{pmatrix} = \mathbf{f}_3$$

with $C_\Gamma^D$ being a $(N_\varphi \times N_u)$ matrix, $D$ being a $(N_\varphi \times N_\varphi)$ matrix with entries

$$(C_\Gamma^D)_{i,k} = \langle \mathbf{w}_k \cdot \mathbf{n}_f, \psi_i \rangle_\Gamma, \; D_{i,j} = a_p(\psi_j, \psi_i),$$

respectively. The right-hand side vector has entries $\langle \tilde{f}_p, \psi_i \rangle$.

Finally, one can put all three systems of equations together and obtain a discrete coupled system

$$\begin{array}{c} \\ N_u \\ N_p \\ N_\varphi \end{array} \begin{array}{c} \begin{array}{ccc} N_u & N_p & N_\varphi \end{array} \\ \begin{pmatrix} A & B & C_\Gamma^S \\ B^T & 0 & 0 \\ C_\Gamma^D & 0 & D \end{pmatrix} \end{array} \begin{array}{c} \\ \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \boldsymbol{\phi} \end{pmatrix} \end{array} \begin{array}{c} \\ = \end{array} \begin{array}{c} \\ \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{pmatrix} \end{array} \tag{31}$$

which represents the Neumann–Neumann problem.

### 3.1.5 Robin–Robin coupling

Robin boundary conditions are a weighted combination of Dirichlet boundary conditions and Neumann boundary conditions. One has the interface conditions (26) and (27) to distribute, which means one can chose constants $\gamma_f \geq 0$ and $\gamma_p > 0$ now, such that

$$\begin{cases} \gamma_f \mathbf{u}_f \cdot \mathbf{n}_f + \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f & = -\gamma_f (\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f - g\varphi_p \quad \text{on } \Gamma, \\ -\gamma_p \mathbf{u}_f \cdot \mathbf{n}_f + \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f & = \gamma_p (\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f - g\varphi_p \quad \text{on } \Gamma. \end{cases} \tag{32}$$

In this way one obtains Robin boundary conditions for the Stokes equations with the first equation of (32) and Robin boundary conditions for the Darcy equations with the second equation. This Robin–Robin coupling is in fact (analytically) equivalent to the Neumann–Neumann coupling with $\gamma_f = 0$ and $\gamma_p \to \infty$, but also has the possibility to weight the conditions. The problem reads then as follows:
Find $(\mathbf{u}_f, p_f, \varphi_p) \in V_f \times Q_f \times Q_p$ such that for all $(\mathbf{v}, q, \psi) \in V_f \times Q_f \times Q_p$ holds that

$$\begin{cases} a_f(\mathbf{u}_f, \mathbf{v}) + b_f(\mathbf{v}, p_f) + \langle g\varphi_p, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma & \\ \qquad + \langle \gamma_f \mathbf{u}_f \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma + \langle \gamma_f (\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma & = \langle \mathbf{f}_f^1, \mathbf{v} \rangle, \\ b_f(\mathbf{u}_f, q) & = \langle f_f^2, q \rangle, \\ a_p(\varphi_p, \psi) + \langle \frac{1}{\gamma_p} \varphi_p, \psi \rangle_\Gamma + \langle \frac{1}{\gamma_p} \mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \psi \rangle_\Gamma - \langle \mathbf{u}_f \cdot \mathbf{n}_f, \psi \rangle_\Gamma & = \langle \tilde{f}_p, \psi \rangle. \end{cases} \tag{33}$$

The restrictions of $\gamma_f$ and $\gamma_p$ ensure that the coercivity of

$$a_f(\mathbf{u}_f, \mathbf{v}) + \langle \gamma_f \mathbf{u}_f \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma$$

and

$$a_p(\varphi_p, \psi) + \langle \frac{1}{\gamma_p} \varphi_p, \psi \rangle_\Gamma$$

is preserved. As the two systems of (33) depend directly on each other, but only couple on the interface, it makes sense to introduce two interface variables $\eta_f, \eta_p$ such that the direct dependence of equations is reduced to these newly introduced variables. The interface variables are defined as follows:

$$\eta_f = -\gamma_f(\mathbb{K}\,\nabla\varphi_p)\cdot\mathbf{n}_f - g\varphi_p, \qquad\qquad \eta_p = -\mathbf{u}_f\cdot\mathbf{n}_f + \frac{1}{\gamma_p}\mathbf{n}_f\cdot\mathbb{T}(\mathbf{u}_f,p_f)\cdot\mathbf{n}_f$$

for the Stokes and Darcy interface conditions, respectively. Another form of the weak formulation is obtained:

$$\left\{\begin{array}{rl} a_f(\mathbf{u}_f,\mathbf{v}) + b(\mathbf{v},p_f) + \langle\gamma_f\mathbf{u}_f\cdot\mathbf{n}_f,\mathbf{v}\cdot\mathbf{n}_f\rangle_\Gamma + \langle\eta_f,\mathbf{v}\cdot\mathbf{n}_f\rangle_\Gamma & = \langle\mathbf{f}_f^1,\mathbf{v}\rangle, \\ b(\mathbf{u}_f,q) & = \langle f_f^2,q\rangle, \\ a_p(\varphi_p,\psi) + \langle\frac{1}{\gamma_p}\varphi_p,\psi\rangle_\Gamma + \langle\eta_p,\psi\rangle_\Gamma & = \langle\tilde{f}_p,\psi\rangle. \end{array}\right. \tag{34}$$

### 3.1.6 Finite element discretization of the Robin–Robin problem

Similarly as in the finite element discretization of the Neumann–Neumann problem, choose finite subspaces $V_f^h$, $Q_f^h$, $V_p^h$ with basis sets $\{\mathbf{w}_i\}_{i=1}^{N_u}$, $\{q_i\}_{i=1}^{N_p}$, $\{\psi_i\}_{i=1}^{N_\varphi}$ of $V_f$, $Q_f$ and $V_p$, respectively. Then replacing the test functions by the corresponding basis functions and searching for a solution as a linear combination of them yields

$$\begin{pmatrix} A_{\mathrm{rob}} & B & C_\varphi^S \\ B^T & 0 & 0 \\ C_u^D & C_p^D & D_{\mathrm{rob}} \end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{p}\\ \boldsymbol{\phi}\end{pmatrix} = \begin{pmatrix}\mathbf{f}_1\\ \mathbf{f}_2\\ \mathbf{f}_3\end{pmatrix}$$

with

$$\begin{aligned} (A_{\mathrm{rob}})_{i,j} &= a_f(\mathbf{w}_j,\mathbf{w}_i) + \langle\gamma_f\mathbf{w}_j\cdot\mathbf{n}_f,\mathbf{w}_i\cdot\mathbf{n}_f\rangle_\Gamma, & (D_{\mathrm{rob}})_{i,j} &= a_p(\psi_j,\psi_i) + \langle\frac{1}{\gamma_p}\psi_j,\psi_i\rangle_\Gamma, \\ (C_\varphi^S)_{i,j} &= \langle g\psi_j,\mathbf{w}_i\cdot\mathbf{n}_f\rangle_\Gamma, + \langle\gamma_f(\mathbb{K}\,\nabla\psi_j)\cdot\mathbf{n}_f,\mathbf{w}_i\cdot\mathbf{n}_f\rangle_\Gamma, & (B)_{i,j} &= b_f(\mathbf{w}_i,q_j), \\ (C_u^D)_{i,j} &= \langle\frac{2\nu}{\gamma_p}\mathbf{n}_f\cdot\mathbb{D}(\mathbf{w}_j)\cdot\mathbf{n}_f,\psi_i\rangle_\Gamma - \langle\mathbf{w}_j\cdot\mathbf{n}_f,\psi\rangle_\Gamma, & (C_p^D)_{i,j} &= \langle-\frac{1}{\gamma_p}q_j,\psi_i\rangle_\Gamma, \end{aligned}$$

and the same right-hand sides as in the Neumann–Neumann problem (31).
Using the formally decoupled form (34), we also search for $\eta_p = \sum_{i=1}^{N_{\eta,p}}\eta_p^i\Lambda_i$ and $\eta_f = \sum_{i=1}^{N_{\eta,f}}\eta_f^1\Lambda_i$, where $\Lambda_i$ are basis functions of the interface space. One obtains

$$\begin{pmatrix} \mathcal{D}_{\gamma_p} & 0 & 0 & 0 & \mathcal{E}_p \\ \mathcal{R}_p & -\mathbb{I} & 0 & 0 & 0 \\ 0 & \mathcal{E}_f & \mathcal{A}_{\gamma_f} & \mathcal{B} & 0 \\ 0 & 0 & \mathcal{B}^T & 0 & 0 \\ 0 & 0 & \mathcal{R}_f^1 & \mathcal{R}_f^2 & -\mathbb{I} \end{pmatrix}\begin{pmatrix}\boldsymbol{\phi}\\ \boldsymbol{\eta}_f\\ \mathbf{u}\\ \mathbf{p}\\ \boldsymbol{\eta}_p\end{pmatrix} = \begin{pmatrix}\mathbf{f}_3\\ 0\\ \mathbf{f}_1\\ \mathbf{f}_2\\ 0\end{pmatrix} \tag{35}$$

with

$$\begin{aligned} (\mathcal{E}_f)_{i,j} &= \langle\Lambda_j,\mathbf{w}_i\cdot\mathbf{n}_f\rangle_\Gamma, & (\mathcal{A}_{\gamma_f})_{i,j} &= a_f(\mathbf{w}_j,\mathbf{w}_i) + \langle\gamma_f\mathbf{w}_j\cdot\mathbf{n}_f,\mathbf{w}_i\cdot\mathbf{n}_f\rangle_\Gamma, \\ (\mathcal{E}_p)_{i,j} &= \langle\Lambda_j,\psi_i\rangle_\Gamma, & (B)_{i,j} &= b_f(\mathbf{w}_i,\mathbf{q}_j), \\ (\mathcal{R}_p)_{i,i} &= -\gamma_f(\mathbb{K}\,\nabla\psi_i)\cdot\mathbf{n}_f - g\psi_i, & (\mathcal{D}_{\gamma_p})_{i,j} &= a_p(\psi_j,\psi_i) + \langle\gamma_p^{-1}\psi_j,\psi_i\rangle_\Gamma, \\ (\mathcal{R}_f^1)_{i,i} &= \frac{2\nu}{\gamma_p}\mathbf{n}_f\cdot\mathbb{D}(\mathbf{w}_i)\cdot\mathbf{n}_f - \mathbf{w}_i\cdot\mathbf{n}_f, & (\mathcal{R}_f^2)_{i,i} &= \gamma_p^{-1}q_i. \end{aligned}$$

### 3.1.7 Gauss–Seidel method

Instead of solving the whole system at once, one might consider to solve the system iteratively as it for example is too large to solve it at once. The presented approach is the so-called Block-Gauss–Seidel method. If one considers a matrix $A \in \mathbb{R}^{d\times d}$ with the decomposition $A = M - N$ such that $M$ is non-singular, one can transform a system of equations $A\mathbf{x} = \mathbf{b}$ into the fixed point equation

$$M\mathbf{x} = \mathbf{b} + N\mathbf{x} \Leftrightarrow \mathbf{x} = M^{-1}(\mathbf{b} + N\mathbf{x}).$$

Given an initial iterate $\mathbf{x}^0 \in \mathbb{R}^d$, one can try to solve this system iteratively by

$$\mathbf{x}^{k+1} = M^{-1}(\mathbf{b} + N\mathbf{x}^k). \tag{36}$$

The following theorem gives information about the convergence of this iteration:

**3.1.1 Theorem**[Banach fixed point theorem]**.** Let $X$ be a complete metric space and let $f : X \to X$ be a contraction (i.e., $f$ is Lipschitz continuous with Lipschitz constant $L < 1$), then

$$x = f(x)$$

has a unique solution (a fixed point) and the iterative scheme

$$x^{k+1} = f(x^k),\ k = 0, 1, \ldots$$

converges to the fixed point for any initial iterate $x^0 \in X$.

*Proof.* See for example pp. 41 of [4]. □

Application of the Banach fixed point theorem tells us that this iteration converges, if

$$f(\mathbf{x}) = M^{-1}(\mathbf{b} + N\mathbf{x})$$

is Lipschitz continuous with constant $L < 1$. This result can be used to derive a condition on the iteration matrix $G = M^{-1}N$. But to do that we first need a definition and the following Lemma:

**3.1.2 Definition**[Spectral radius]**.** Let $A \in \mathbb{R}^{d\times d}$ be a matrix, then the spectral radius of $A$ is defined as

$$\rho(A) := \max\{|\lambda| : \lambda \text{ is eigenvalue of } A\}.$$

**3.1.3 Lemma**[See Lemma 4.3 of [11]]**.** Let $A \in \mathbb{R}^{d\times d}$ with $\rho(A) < 1$. Then there is for each $\varepsilon > 0$ a matrix norm $\|\cdot\|_*$ which usually depends on $A$ and $\varepsilon$, such that

$$\|A\|_* \leq \rho(A) + \varepsilon.$$

*Proof.* Let $J := S^{-1}AS$ be the Jordan normal form of $A$ with transformation matrix $S$ and set

$$D_\varepsilon := \operatorname{diag}(1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{d-1}).$$

By transformation with $SD_\varepsilon$ one gets the slightly modified Jordan form

$$(SD_\varepsilon)^{-1}A(SD_\varepsilon) = D_\varepsilon^{-1}S^{-1}ASD_\varepsilon = D_\varepsilon^{-1}JD_\varepsilon =: J_\varepsilon.$$

Let $J_k(\lambda_j) \in \mathbb{R}^{l\times l}$ be the $k$-th Jordan block of $J$ which starts at row $i$. Then it is

$$\begin{pmatrix} \varepsilon^{-i+1} & & \\ & \ddots & \\ & & \varepsilon^{-i-l+1} \end{pmatrix} \begin{pmatrix} \lambda_j & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{pmatrix} \begin{pmatrix} \varepsilon^{i-1} & & \\ & \ddots & \\ & & \varepsilon^{i+l-1} \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon^{-i+1}\lambda_j & \varepsilon^{-i+1} & & & \\ & \varepsilon^{-i}\lambda_j & \varepsilon^{-i} & & \\ & & \ddots & \ddots & \\ & & & \ddots & \varepsilon^{-i-l+1} \\ & & & & \varepsilon^{-i-l+1}\lambda_j \end{pmatrix} \begin{pmatrix} \varepsilon^{i-1} & & \\ & \ddots & \\ & & \varepsilon^{i+l-1} \end{pmatrix} = \begin{pmatrix} \lambda_j & \varepsilon & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda_j \end{pmatrix}$$

$$=: J_k^{\varepsilon}(\lambda_j)$$

and one gets

$$J_\varepsilon = \operatorname{diag}(J_{n_1}^{\varepsilon}(\lambda_1), \dots, J_{n_m}^{\varepsilon}(\lambda_m))$$

where $m$ is the number of different eigenvalues. So now one can define the vector norm

$$\|\mathbf{x}\|_* := \|(SD_\varepsilon)^{-1}\mathbf{x}\|_\infty$$

which has, for the induced matrix norm, the desired property:

$$\begin{aligned} \|A\|_* &= \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_*}{\|\mathbf{x}\|_*} = \max_{\mathbf{x}\neq 0} \frac{\|(SD_\varepsilon)^{-1}A\mathbf{x}\|_\infty}{\|(SD_\varepsilon)^{-1}\mathbf{x}\|_\infty} \\ &= \max_{\mathbf{y}\neq 0} \frac{\|(SD_\varepsilon)^{-1}A(SD_\varepsilon)\mathbf{y}\|_\infty}{\|\mathbf{y}\|_\infty} && \text{,for } \mathbf{y} := (SD_\varepsilon)^{-1}\mathbf{x} \\ &= \max_{\mathbf{y}\neq 0} \frac{\|J_\varepsilon\mathbf{y}\|_\infty}{\|\mathbf{y}\|_\infty} = \|J_\varepsilon\|_\infty \\ &\leq \max_{i=1,\dots,m} |\lambda_i| + \varepsilon = \rho(A) + \varepsilon. \end{aligned}$$

□

Now all preparations are done for the following theorem:

**3.1.4 Theorem**[Condition on the iteration matrix for convergence, see Theorem 3.3 of [10]]**.** The iteration (36) converges for any initial iterate to the unique solution if and only if

$$\rho(G) < 1.$$

*Proof.* A counterexample for $\rho(G) \geq 1$ can be found in the proof of Theorem 3.3 of [10]. The operator $f(\mathbf{x}) = G\mathbf{x} + M^{-1}\mathbf{b}$ is Lipschitz continuous, because

$$\|f(\mathbf{x}) - f(\mathbf{y})\|_* = \|G\mathbf{x} - G\mathbf{y}\|_* = \|G(\mathbf{x} - \mathbf{y})\|_* \leq \|G\|_*\|\mathbf{x} - \mathbf{y}\|_*$$

for any compatible matrix norm $\|\cdot\|_*$. If $\rho(G) < 1$, it is, using the previous lemma, possible to find an $\varepsilon > 0$ such that there is a norm with

$$\|G\|_* \leq \rho(G) + \varepsilon < 1.$$

Hence $f$ is a contraction and Banach's fixed point theorem tells us, that the fixed point iteration converges for any initial iterate. □

$$\left(\begin{array}{c|c|c|c} D_{11} & U_{12} & U_{13} & U_{14} \\ \hline L_{21} & D_{22} & U_{23} & U_{24} \\ \hline L_{31} & L_{32} & D_{33} & U_{34} \\ \hline L_{41} & L_{42} & L_{43} & D_{44} \end{array}\right)$$

Figure 2: Example of a block matrix structure.

To obtain the block Gauss–Seidel method, one considers the decomposition $A = L + D + U$ where $L$ is the strictly block-lower triangular part of $A$, $D$ is the block-diagonal and $U$ is the strictly block-upper triangular part (see for example Figure 2).
Setting $M = D + L$ and $N = -U$ in (36) leads to

$$\mathbf{x}^{k+1} = (D + L)^{-1}(\mathbf{b} - U\mathbf{x}^k). \tag{37}$$

This form however involves the inverse of $(D + L)$. Multiplication with $D + L$ gives

$$\begin{aligned} (D + L)\mathbf{x}^{k+1} &= \mathbf{b} - U\mathbf{x}^k \\ \Rightarrow \mathbf{x}^{k+1} &= D^{-1}(\mathbf{b} - L\mathbf{x}^{k+1} - U\mathbf{x}^k) \\ &= \mathbf{x}^k + D^{-1}(\mathbf{b} - L\mathbf{x}^{k+1} - (D + U)\mathbf{x}^k), \end{aligned}$$

which only contains $D^{-1}$ which is block-diagonal and hence can be inverted block-wise. One could get the idea that the current iteration step would depend on itself, but written in block-component form it turns out that one just needs a forward substitution:

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + A_{ii}^{-1}\left(\mathbf{b}_i - \sum_{j=1}^{i-1} A_{ij}\mathbf{x}_j^{k+1} - \sum_{j=i}^{n} A_{ij}\mathbf{x}_j^k\right),$$

where $\mathbf{x}_j$ and $\mathbf{b}_j$ are denoting the solution vector and right-hand side vector, respectively, to the block-component $j$. Applying this approach to the Robin–Robin problem (35), one gets

$$\begin{pmatrix} \boldsymbol{\phi}^{k+1} \\ \boldsymbol{\eta}_f^{k+1} \\ \begin{pmatrix} \mathbf{u}^{k+1} \\ \mathbf{p}^{k+1} \end{pmatrix} \\ \boldsymbol{\eta}_p^{k+1} \end{pmatrix} = \begin{pmatrix} \mathcal{D}_{\gamma_p}^{-1}(\mathbf{f}_3 - \mathcal{E}_p\boldsymbol{\eta}_p^k) \\ \mathcal{R}_p\boldsymbol{\phi}^{k+1} \\ \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B}^T & 0 \end{pmatrix}^{-1} \left(\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix} - \begin{pmatrix} \mathcal{E}_f\boldsymbol{\eta}_f^{k+1} \\ 0 \end{pmatrix}\right) \\ \mathcal{R}_f^1\mathbf{u}^{k+1} + \mathcal{R}_f^2\mathbf{p}^{k+1} \end{pmatrix}. \tag{38}$$

**3.1.5 Theorem**[Convergence]**.** If the iteration matrix $G = M^{-1}N = -(D + L)^{-1}U$ is strongly block-diagonally dominant, i.e., for each block-row $i$ it holds

$$\sum_{j=1, j\neq i}^{n} \|U_{ij}\|_\infty + \|L_{ij}\|_\infty < \|D_{ii}\|_\infty,$$

then the Gauss–Seidel method converges for each initial iterate.

*Proof.* The iteration matrix fulfills the following relation:

$$\begin{aligned}G = M^{-1}N = -(D+L)^{-1}U &\Leftrightarrow (D+L)G = -U \Leftrightarrow DG + LG = -U\\ &\Leftrightarrow DG = -LG - U\\ &\Leftrightarrow G = -D^{-1}(LG+U).\end{aligned}$$

Now let $\mathbf{z} \in \mathbb{R}^d, \mathbf{z} \neq 0$, then $i$-th block-component of the relation is given by

$$(G\mathbf{z})_i = -D_{ii}^{-1}\left(\sum_{j=1}^{i-1} L_{ij}(G\mathbf{z})_j + \sum_{j=i+1}^{n} U_{ij}\mathbf{z}_j\right).$$

Having a closer look at the first block-component yields

$$\|(G\mathbf{z})_1\|_\infty \leq \|D_{11}^{-1}\|_\infty \sum_{j=2}^{n} \|U_{1j}\|_\infty \|\mathbf{z}_j\|_\infty \leq \|\mathbf{z}\|_\infty \frac{1}{\|D_{11}\|_\infty} \sum_{j=2}^{n} \|U_{1j}\|_\infty < \|\mathbf{z}\|_\infty$$

as the considered matrix is strongly block-diagonally dominant. Now we can prove the inequality for the other rows by induction: Suppose it holds for all $j < i$, then we can estimate

$$\begin{aligned}\|(G\mathbf{z})_i\|_\infty &\leq \frac{1}{\|D_{ii}\|_\infty}\left(\sum_{j=1}^{i-1} \|L_{ij}\|_\infty \|(G\mathbf{z})_j\|_\infty + \sum_{j=i+1}^{n} \|U_{ij}\|_\infty \|\mathbf{z}_j\|_\infty\right)\\ &\leq \frac{1}{\|D_{ii}\|_\infty}\left(\sum_{j=1}^{i-1} \|L_{ij}\|_\infty \|\mathbf{z}\|_\infty + \sum_{j=i+1}^{n} \|U_{ij}\|_\infty \|\mathbf{z}\|_\infty\right) &&\text{, induction hypothesis,}\\ &< \|\mathbf{z}\|_\infty &&\text{, strong diagonal dominance,}\end{aligned}$$

and it follows with

$$\|(G\mathbf{z})_i\|_\infty < \|\mathbf{z}\|_\infty \Rightarrow \|G\mathbf{z}\|_\infty < \|\mathbf{z}\|_\infty,$$

that

$$\rho(G) \leq \|G\|_\infty = \max_{\mathbf{z}\in\mathbb{R}^d, \mathbf{z}\neq 0} \frac{\|G\mathbf{z}\|_\infty}{\|\mathbf{z}\|_\infty} < 1.$$

The first inequality holds, because for any eigenvalue - eigenvector pair $(\lambda, \mathbf{z})$ of $G$ it is

$$\begin{aligned}|\lambda|\|\mathbf{z}\|_\infty = \|\lambda\mathbf{z}\|_\infty = \|G\mathbf{z}\|_\infty \leq \|G\|_\infty \|\mathbf{z}\|_\infty\\ \Rightarrow \rho(G) \leq \|G\|_\infty.\end{aligned}$$

□

**3.1.6 Remark**[Convergence of the Stokes–Darcy system for $K, \nu \to 0$]**.** In order to be able to apply Theorem 3.1.5 to the Stokes–Darcy problem with Neumann–Neumann system matrix (31), the strong block-diagonal dominance has to hold. In the system matrix of the Darcy system, each entry depends linearly on $K$. This means that if $K \to 0$, the condition is violated as the blocks which couple the systems do not depend on $K$ at all.
For the Stokes block, there is still the bilinear form $b(\cdot,\cdot)$ which does not depend on $\nu$, hence $\nu \to 0$ does not necessarily mean that the theorem does not hold.
However, using the Robin–Robin system matrix (35) of the Stokes–Darcy problem results in an additional term in the Darcy part of the system, which does not scale with $K$. Hence, convergence might be recovered.

**3.1.7 Remark**[Block Jacobi method]**.** Another classic approach for iteratively solving a system of equations is the block Jacobi method. Set $M = D$ and $N = -(L + U)$ and the iteration matrix is $G = M^{-1}N = -D^{-1}(L + U)$. Then the fixed point equation is

$$\mathbf{x}^{k+1} = D^{-1}(\mathbf{b} - (L + U)\mathbf{x}^k) = \mathbf{x}^k + D^{-1}(\mathbf{b} - A\mathbf{x}^k).$$

In contrast to the block Gauss–Seidel method, this iteration does not need a forward substitution and hence can be parallelized. Applying this approach to the Robin–Robin problem (35) one gets

$$\begin{pmatrix} \boldsymbol{\phi}^{k+1} \\ \boldsymbol{\eta}_f^{k+1} \\ \begin{pmatrix} \mathbf{u}^{k+1} \\ \mathbf{p}^{k+1} \end{pmatrix} \\ \boldsymbol{\eta}_p^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}^{k} \\ \boldsymbol{\eta}_f^{k} \\ \begin{pmatrix} \mathbf{u}^{k} \\ \mathbf{p}^{k} \end{pmatrix} \\ \boldsymbol{\eta}_p^{k} \end{pmatrix} + \begin{pmatrix} \mathcal{D}_{\gamma_p}^{-1}(\mathbf{f}_3 - \mathcal{D}_{\gamma_p}\boldsymbol{\phi}^k - \mathcal{E}_p\boldsymbol{\eta}_p^k) \\ \mathcal{R}_p\boldsymbol{\phi}^k - \boldsymbol{\eta}_f^k \\ \begin{pmatrix} \mathcal{A}_{\gamma_f} & \mathcal{B} \\ \mathcal{B}^T & 0 \end{pmatrix}^{-1} \left( \begin{pmatrix} \mathbf{f}_1 - \mathcal{E}_f\boldsymbol{\eta}_f^k \\ \mathbf{f}_2 \end{pmatrix} - \begin{pmatrix} \mathcal{A}_{\gamma_f} & \mathcal{B} \\ \mathcal{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}^k \\ \mathbf{p}^k \end{pmatrix} \right) \\ \mathcal{R}_f^1\mathbf{u}^k + \mathcal{R}_f^2\mathbf{p}^k - \boldsymbol{\eta}_p \end{pmatrix}.$$

## 3.2 Navier–Stokes–Darcy problem

The interface conditions as well as the general setting of domain and boundary is identical to the Stokes–Darcy case. The Navier–Stokes–Darcy problem therefore reads as follows: Find $(\mathbf{u}_f, p_f, \varphi_p) : \Omega_f \times \Omega_f \times \Omega_p \to \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$ such that

$$\left\{ \begin{aligned} -\nabla \cdot \mathbb{T}(\mathbf{u}_f, p_f) + (\mathbf{u}_f \cdot \nabla) \cdot \mathbf{u}_f &= \mathbf{f}_f && \text{in } \Omega_f, \\ \nabla \cdot \mathbf{u}_f &= 0 && \text{in } \Omega_f, \\ -\nabla \cdot \mathbb{K}\,\nabla\varphi_p &= f_p && \text{in } \Omega_p, \\ \\ \mathbf{u}_f &= \mathbf{u}_{f,\text{ess}} && \text{on } \Gamma_{f,e}, \\ \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f &= T_{f,\text{nat}} && \text{on } \Gamma_{f,n}, \\ \varphi_p &= \varphi_{p,\text{ess}} && \text{on } \Gamma_{p,e}, \\ (-\,\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f &= u_{p,\text{nat}} && \text{on } \Gamma_{p,n}, \\ \\ \mathbf{u}_f \cdot \mathbf{n}_f &= -(\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f && \text{on } \Gamma, \\ -\mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f &= g\varphi_p && \text{on } \Gamma, \\ \mathbf{u}_f \cdot \boldsymbol{\tau}_i + \alpha_0\boldsymbol{\tau}_i \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f &= 0 && \text{on } \Gamma, i = 1, \ldots, d-1. \end{aligned} \right. \tag{39}$$

The only new part is the nonlinear convection-term $(\mathbf{u}_f \cdot \nabla) \cdot \mathbf{u}_f$, otherwise the problem is similar to (24). Therefore the weak formulation only differs a bit, too, as it now includes the in (16) defined trilinear form:

(i) Neumann–Neumann (see (30)): Find $(\mathbf{u}_f, p_f, \varphi_p) \in V_f \times Q_f \times Q_p$ such that

$$\left\{ \begin{aligned} a_f(u_f, \mathbf{v}) + b_f(\mathbf{v}, p_f) + c(\mathbf{u}_f, \mathbf{u}_f, \mathbf{v}) + \langle g\varphi_p, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma &= \langle \mathbf{f}_f^1, \mathbf{v} \rangle, \\ b_f(\mathbf{u}_f, q) &= \langle f_f^2, q \rangle, \\ a_p(\varphi_p, \psi) - \langle \mathbf{u}_f \cdot \mathbf{n}_f, \psi \rangle_\Gamma &= \langle f_p, \psi \rangle, \end{aligned} \right. \tag{40}$$

for all $(\mathbf{v}, q, \psi) \in V_f \times Q_f \times Q_p$.

(ii) Robin–Robin (see (33)): Find $(\mathbf{u}_f, p_f, \varphi_p) \in V_f \times Q_f \times Q_p$ such that

$$\left\{ \begin{aligned} a_f(\mathbf{u}_f, \mathbf{v}) + b_f(\mathbf{v}, p_f) + c(\mathbf{u}_f, \mathbf{u}_f, \mathbf{v}) + \langle g\varphi_p, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma & \\ + \langle \gamma_f \mathbf{u}_f \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma + \langle \gamma_f (\mathbb{K}\,\nabla\varphi_p) \cdot \mathbf{n}_f, \mathbf{v} \cdot \mathbf{n}_f \rangle_\Gamma &= \langle \mathbf{f}_f^1, \mathbf{v} \rangle, \\ b_f(\mathbf{u}_f, q) &= \langle f_f^2, q \rangle, \\ a_p(\varphi_p, \psi) + \langle \tfrac{1}{\gamma_p}\varphi_p, \psi \rangle_\Gamma + \langle \tfrac{1}{\gamma_p}\mathbf{n}_f \cdot \mathbb{T}(\mathbf{u}_f, p_f) \cdot \mathbf{n}_f, \psi \rangle_\Gamma - \langle \mathbf{u}_f \cdot \mathbf{n}_f, \psi \rangle_\Gamma &= \langle f_p, \psi \rangle, \end{aligned} \right. \tag{41}$$

for all $(\mathbf{v}, q, \psi) \in V_f \times Q_f \times Q_p$.

The finite element discretization does not change much as well, just that $\mathcal{A}$ (or $\mathcal{A}_{\text{rob}}$, $\mathcal{A}_{\gamma_f}$ respectively) now depend on $\mathbf{u}_f$ due to the nonlinearity, i.e., the $(i,j)$-th component of the matrix now has the additional term

$$c(\mathbf{u}_f, \mathbf{w}_j, \mathbf{w}_i).$$

This issue can be resolved by the fixed point iterations seen in Section 2.3.2. Applying the fixed point iterations means that one has to solve either the whole system at once iteratively, or that one approximates the nonlinearity within a block-iterative method, like the Gauss–Seidel method presented in the following section.

### 3.2.1 Iterative approach

We will only discuss the case of the interface-coupled Robin–Robin method, as the Neumann–Neumann method can be obtained with $\gamma_f = 0$ and $\gamma_p \to \infty$.
Let $\eta_f^0, \eta_p^0$ and $\gamma_f, \gamma_p$ be given, as well as an initial guess $\boldsymbol{\phi}_p^0$ , e.g., the zero vector. Then the sequential or Gauss–Seidel type iteration looks like follows: Set $k = 0$, then following (38):

(i) Solve the Darcy problem with Robin boundary data $\eta_p^k$:

$$\mathcal{D}_{\gamma_p} \boldsymbol{\phi}^{k+1} = \mathbf{f}_3 - \mathcal{E}_p \boldsymbol{\eta}_p^k$$

and obtain $\boldsymbol{\phi}^{k+1}$.

(ii) Calculate $\boldsymbol{\eta}_f^{k+1} = \mathcal{R}_p \boldsymbol{\phi}_p^{k+1}$.

(iii) Solve the Navier–Stokes problem with given interface-boundary $\boldsymbol{\eta}_f^{k+1}$:

$$\begin{pmatrix} \mathcal{A}_{\gamma_f}(\mathbf{u}^{k+1}) & \mathcal{B}_{\text{rob}} \\ -\mathcal{B}_{\text{rob}}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_f^{k+1} \\ \mathbf{p}_f^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 - \mathcal{E}_f \boldsymbol{\eta}_f^{k+1} \\ \mathbf{f}_2 \end{pmatrix}$$

and obtain $(\mathbf{u}_f^{k+1}, p_f^{k+1})$.

(iv) Calculate $\boldsymbol{\eta}_p^{k+1} = \mathcal{R}_f^1 \mathbf{u}_f^{k+1} + \mathcal{R}_f^2 \mathbf{p}_f^{k+1}$.

(v) If not converged, go back to step (i) and increase $k$ by one.

The crucial step for the Navier–Stokes–Darcy problem is step (iii). There, one actually has to solve equations with the nonlinear term

$$(\mathcal{A}_{\gamma_f}(\mathbf{u}_f^k))_{ij} = a_f(\mathbf{w}_j, \mathbf{w}_i) + \langle \gamma_f \mathbf{w}_f \cdot \mathbf{n}_f, \mathbf{w}_i \cdot \mathbf{n}_f \rangle + c(\mathbf{u}_f^k, \mathbf{w}_j, \mathbf{w}_i).$$

This issue can be resolved in two ways:

- One can simply use the fixed-point iteration which was introduced in Section 2.3.2. Therefore one rewrites step (iii):

(iii).1 Solve the Navier–Stokes problem for $\mathbf{u}_f^{k,m}$ with given interface-boundary $\boldsymbol{\eta}_f^{k+1}$ by using the fixed point iterations, with the starting guess $\mathbf{u}_f^{k+1,0} = \mathbf{u}_f^k$ until

$$\begin{pmatrix} \mathcal{A}_{\gamma_f}(\mathbf{u}^{k+1,m}) & \mathcal{B}_{\text{rob}} \\ -\mathcal{B}_{\text{rob}}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_f^{k+1,m+1} \\ \mathbf{p}_f^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 - \mathcal{E}_f \boldsymbol{\eta}_f^{k+1} \\ \mathbf{f}_2 \end{pmatrix}$$

converges at iteration $m_0$. Now set $\mathbf{u}_f^{k+1} := \mathbf{u}_f^{k+1,m_0}$ and proceed with step (iv).

- One could also linearize only with the previous step $\mathbf{u}_f^k$. This corresponds to do just one fixed point iteration.

(iii).2 Solve the linear problem

$$\begin{pmatrix} \mathcal{A}_{\gamma_f}(\mathbf{u}^k) & \mathcal{B}_{\text{rob}} \\ -\mathcal{B}_{\text{rob}}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_f^{k+1} \\ \mathbf{p}_f^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 - \mathcal{E}_f \boldsymbol{\eta}_f^{k+1} \\ \mathbf{f}_2 \end{pmatrix}$$

for given $\mathbf{w}^k$.

In the first case one just iterates until a stopping criterion is reached. In the second case there is only one iteration per step. One could argue, that the full iterations in the first case are not needed, as the boundary conditions on the interface are not correct at least at the beginning (i.e., for small $k$).

**3.2.1 Remark**[Parallelized algorithm]**.** As it was mentioned in Remark 3.1.7, it is also possible to use a parallelized block Jacobi method instead of the sequential block Gauss–Seidel method. The algorithm then is: Set $k = 0$, then

(i) Parallel: $\begin{cases} \text{Solve the Darcy problem with use of } \boldsymbol{\eta}_p^k \text{ and obtain } \boldsymbol{\varphi}^{k+1}. \\ \text{Solve the Navier–Stokes problem with use of } \boldsymbol{\eta}_f^k \text{ and obtain } (\mathbf{u}_f^{k+1}, \mathbf{p}_f^{k+1}). \end{cases}$

(ii) Parallel: $\begin{cases} \text{Update } \boldsymbol{\eta}_f^k \text{ to } \boldsymbol{\eta}_f^{k+1}. \\ \text{Update } \boldsymbol{\eta}_p^k \text{ to } \boldsymbol{\eta}_p^{k+1}. \end{cases}$

(iii) If not converged, increase $k$ by one and continue with step (i).

Note that this algorithm actually is not a pure Jacobi method, as the second step relies on the solutions of the first step. It is rather a nested Jacobi and Gauss–Seidel method, see Section 3.1 of [5].

# 4 Numerical studies

This chapter is about the comparison between the numerical solution of the Navier–Stokes–Darcy problem and a given analytical solution, especially in Example 1, as well as the comparison between Example 2 (riverbed example) and results from the literature, see [7, 6]. Also it is going to be analyzed, how the choice of the algorithm for the nonlinear part (i.e., full iterations or just one iteration) impacts the costs to solve the whole system. For each solving process of a subsystem, a direct solver was used.
For the finite element discretization the spaces $P_2$ and $P_1$ were used for Navier–Stokes velocity and pressure respectively and $P_2$ for the Darcy pressure as in [5]. The choice of finite elements for Navier–Stokes velocity and pressure is also known as the Taylor–Hood element and preserves the discrete variant of the inf-sup condition, see Chapter VI of [3]. This condition is needed to assure uniqueness for the discrete problem, see Chapter 2 of [1].

## 4.1 Example 1

This example deals with a Navier–Stokes–Darcy problem of which the analytical solution is known. Let $\Omega_f = (0, 1) \times (1, 2)$ and $\Omega_p = (0, 1)^2$ with interface $\Gamma = \partial\Omega_f \cap \partial\Omega_p = (0, 1) \times \{1\}$.

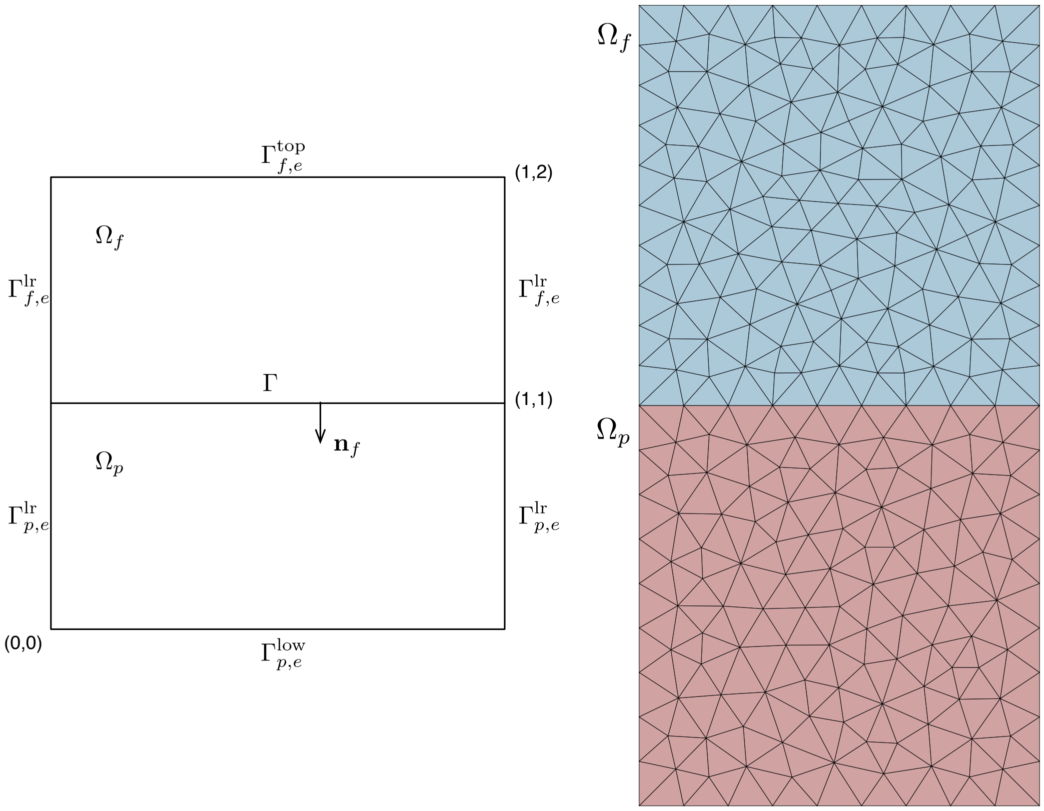

$\Gamma_{f,e}^{\mathrm{top}}$
(1,2)
$\Omega_f$
$\Gamma_{f,e}^{\mathrm{lr}}$
$\Gamma_{f,e}^{\mathrm{lr}}$
$\Gamma$
(1,1)
$\mathbf{n}_f$
$\Omega_p$
$\Gamma_{p,e}^{\mathrm{lr}}$
$\Gamma_{p,e}^{\mathrm{lr}}$
(0,0)
$\Gamma_{p,e}^{\mathrm{low}}$
$\Omega_f$
$\Omega_p$

Figure 3: Example 1: Sketch of the domain (left) and triangulation of the domain on level 1 (right).

The solution is given by

$$\begin{aligned}
\mathbf{u}_f(x,y) &= \begin{pmatrix} y^2-2y+1 \\ x^2-x \end{pmatrix}, \\
p_f(x,y) &= 2\nu(x+y-1)+\frac{g}{3K}, \\
\varphi_p(x,y) &= \frac{1}{K}\left(x(1-x)(y-1)+\frac{y^3}{3}-y^2+y\right).
\end{aligned}$$

The set of equations to solve is

$$\left\{\begin{array}{rll}
-\nabla\cdot\mathbb{T}(\mathbf{u}_f,p_f)+(\mathbf{u}_f\cdot\nabla)\cdot\mathbf{u}_f & = \begin{pmatrix} (2y-2)(x^2-x) \\ (2x-1)(y^2-2y+1) \end{pmatrix} & \text{in } \Omega_f, \\
\nabla\cdot\mathbf{u}_f & = 0 & \text{in } \Omega_f, \\
-\nabla\cdot\mathbb{K}\,\nabla\varphi_p & = 0 & \text{in } \Omega_p, \\
 & & \\
\mathbf{u}_f & = \begin{pmatrix} y^2-2y+1 \\ 0 \end{pmatrix} & \text{on } \Gamma_{f,e}^{\mathrm{lr}}=\{0,1\}\times(1,2), \\
\mathbf{u}_f & = \begin{pmatrix} 1 \\ x^2-x \end{pmatrix} & \text{on } \Gamma_{f,e}^{\mathrm{top}}=(0,1)\times\{2\}, \\
\varphi_p & = \frac{1}{K}(\frac{y^3}{3}-y^2+y) & \text{on } \Gamma_{p,e}^{\mathrm{lr}}=\{0,1\}\times(0,1), \\
\varphi_p & = \frac{1}{K}(x(x-1)) & \text{on } \Gamma_{p,e}^{\mathrm{low}}=(0,1)\times\{0\},
\end{array}\right.$$

| Type | Algorithm | Iterations | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|---|---|
| Neumann–Neumann | Jacobi | interface | 13 | 13 | 13 | 12 |
| | | nonlinear | 31 | 29 | 29 | 25 |
| | | total | 44 | 42 | 42 | 37 |
| | Gauss–Seidel | interface | 8 | 8 | 8 | 7 |
| | | nonlinear | 20 | 19 | 19 | 17 |
| | | total | 28 | 27 | 27 | 24 |
| Robin–Robin | Jacobi | interface | 42 | 42 | 42 | 40 |
| | | nonlinear | 65 | 65 | 63 | 61 |
| | | total | 107 | 107 | 105 | 101 |
| | Gauss–Seidel | interface | 22 | 22 | 21 | 21 |
| | | nonlinear | 38 | 39 | 37 | 37 |
| | | total | 60 | 61 | 58 | 58 |

Table 1: Example 1: Number of iterations of the Robin–Robin algorithm in comparison to the Neumann–Neumann algorithm, using either the Jacobi method or the Gauss–Seidel method. The number of fixed-point iterations (here denoted as "nonlinear" iterations) for solving the Navier–Stokes system had no limit.

plus the interface conditions (26), (27), the Beavers–Joseph–Saffman condition (28) on $\Gamma = (0,1) \times \{1\}$ with $\mathbf{n}_f = (0,-1)^T$ and $\{\boldsymbol{\tau}_i\}_{i=1}^{d-1} = \{(1,0)^T\}$. This example coincides mostly with Example 2 (Section 4.2) of [5] with the difference that the Stokes equations have been considered there. Therefore the right-hand side of the nonlinear Navier–Stokes equation has to change in order to obtain the same solution.

The grid consists, similar to Example 2 of [5], of triangles like shown in Figure 3.

The following tests were run for $\gamma_p = 1$, $\gamma_f = \frac{\gamma_p}{3}$, $\mathbb{K} = \mathbb{I}$ and $\nu = 1$ on four unstructured grids with 470 (level 1), 1914 (level 2), 8216 (level 3) and 33234 (level 4) cells with a total of 1690 (level 1), 6553 (level 2), 27402 (level 3) and 109613 (level 4) degrees of freedom.

For the interface iterations the following condition has been used as stopping criterion:

$$\frac{\|\mathbf{u}_f^{k+1} - \mathbf{u}_f^k\|_{\ell^2}}{\|\mathbf{u}_f^k\|_{\ell^2}} + \frac{\|p_f^{k+1} - p_f^k\|_{\ell^2}}{\|p_f^k\|_{\ell^2}} + \frac{\|\varphi_p^{k+1} - \varphi_p^k\|_{\ell^2}}{\|\varphi_p^k\|_{\ell^2}} < 10^{-10}. \tag{42}$$

The fixed-point iterations to solve the Navier–Stokes system stopped when the $\ell^2$-norm of the residual vector was smaller than $10^{-10}$.

First goal is to compare the Neumann–Neumann method (see Section 3.1.3) to the Robin–Robin method (see Section 3.1.5), as well as the performance of the Gauss–Seidel method (see Section 3.1.7) in comparison to the Jacobi method (see Remark 3.1.7).

In order to compare these things, one needs some kind of measure for the costs. The number of nonlinear iterations for solving the Navier–Stokes system were chosen for this matter. The number of solving the Darcy system is negligible as it does not change per interface iteration and hence a factorization needs to be computed only once.

In Table 1 one can see that the Neumann–Neumann method behaves better than the Robin–Robin method in terms of performed iterations. In this case, the Gauss–Seidel method is more or less twice as fast as the Jacobi method. And even though the Jacobi method can be parallelized, which is not possible with the Gauss–Seidel method due to the forward

substitution performed in each step, the parallelization can only be applied to the two coupled systems, hence there would be onyly little performance gain compared to the Gauss–Seidel method.
The Table 1 also shows the number of fixed-point iterations needed to solve the Navier–Stokes system in total, but if one has a look at for example Figure 4, there are much more fixed-point iterations for the nonlinear term in the beginning than in the end. For the other cases, the same behavior can be observed. This leads to the question whether restricting the number of iterations for the nonlinear term might lead to less iterations in total. The table also shows, that the number of iterations is independent of the mesh in all cases.

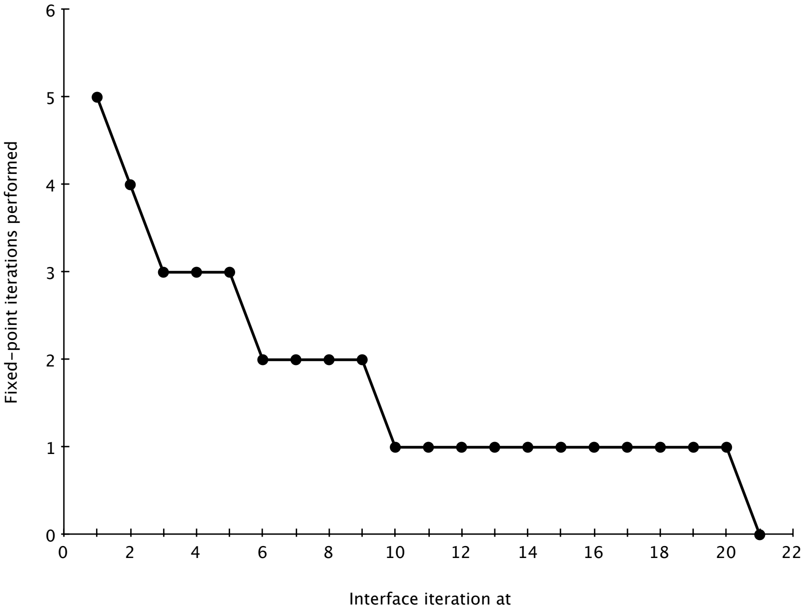


Figure 4: Example 1: The number of fixed-point iterations needed over all interface iterations. The Robin–Robin method with Gauss–Seidel algorithm on level 4 was used.

Having a look at Figure 5 shows at the example of grid level 4, that it seems to be most convenient to make just one fixed-point iteration in each step. This approach was presented in Section 3.2.1.

## 4.2 Example 2 (riverbed example)

This example deals with the flow of a river over a porous riverbed with a periodic dune-shaped (modeled as triangles) interface. It was used for example in [7, 6] to study the hydrodynamic interactions between the water flow and the underlying groundwater flow. Therefore consider the domain

$$\Omega = (0, 2L) \times (0, H_f + H_p)$$

where $H_f$ and $H_p$ denote the initial heights of the water flow domain and the groundwater flow domain respectively. The interface representing the dune consists of two triangles whose highest points are at

$$x_1 = l_D,\ x_2 = L + l_D, \text{ for } 0 < l_D < L$$

with the height (relative to the height of the interface-entrance point) $h_D$, see Figure 6.

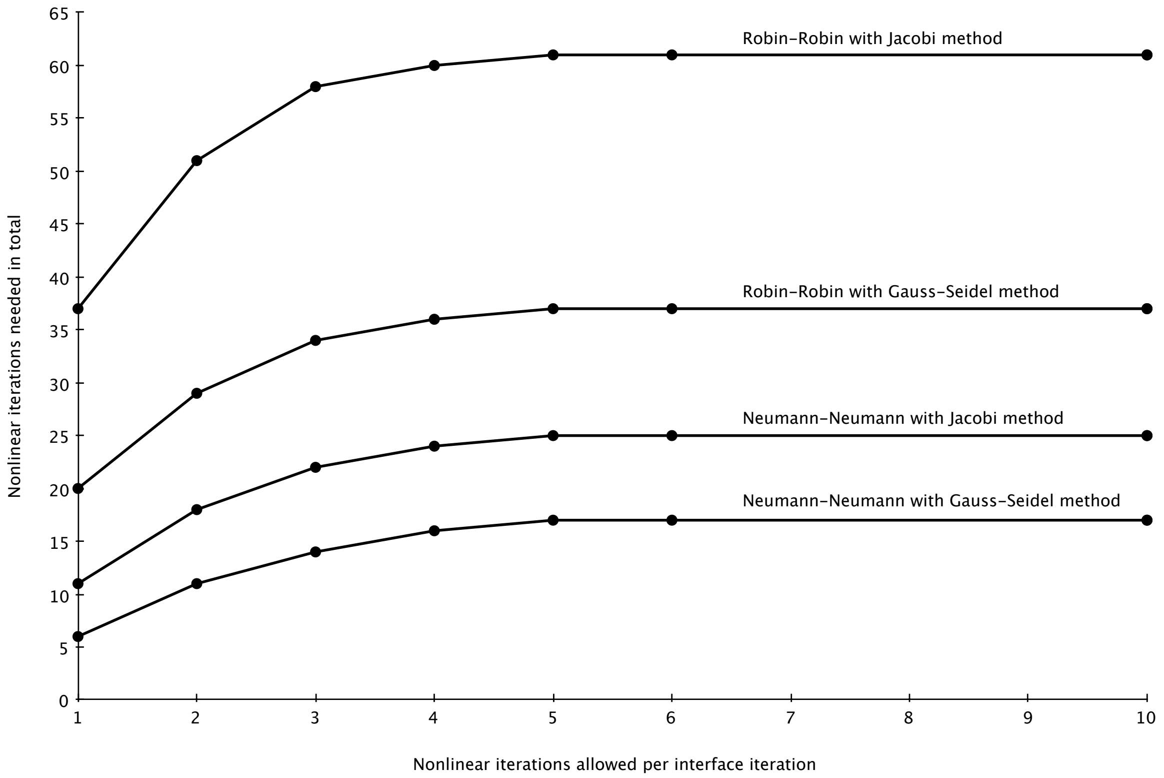


Figure 5: Example 1: This figure shows for grid level 4 the total number of nonlinear iterations needed over the number of fixed-point iterations allowed for the Naiver–Stokes system at each interface-iteration.

The boundary conditions have been chosen to match [7, 6]. Therefore the lower boundary is a no-flow boundary, i.e.,

$$(-\mathbb{K}\,\nabla\varphi_p)\cdot\mathbf{n}_f = 0 \text{ on } \Gamma_{p,n}^{\text{low}}$$

and the upper boundary is a no-slip boundary, i.e.,

$$\mathbf{u}_f = 0 \text{ on } \Gamma_{f,e}^{\text{top}}.$$

The fixed parameters are chosen as

$$\begin{aligned} &L = 1, H_f = 0.5, H_p = 1.5,\\ &l_D = 0.9, h_D = 0.1,\\ &p_0 = 10^{-3}, \mathbf{f}_f = \mathbf{0}, f_p = 0. \end{aligned}$$

The general idea of this example is to create a periodic flow, so the boundary conditions have to be periodic as well. Nevertheless, if one chose all boundary conditions to be exactly periodic (i.e., periodicity of velocity and pressure), there would be no flow at all, as there is no source term (homogeneous right-hand side). That is why one has to introduce a pressure drop between the inlet and outlet boundaries, as then one can expect the fluid to flow from the inlet boundary to the outlet boundary.

This idea yields the following boundary conditions on inlet and outlet boundaries (see Ex-

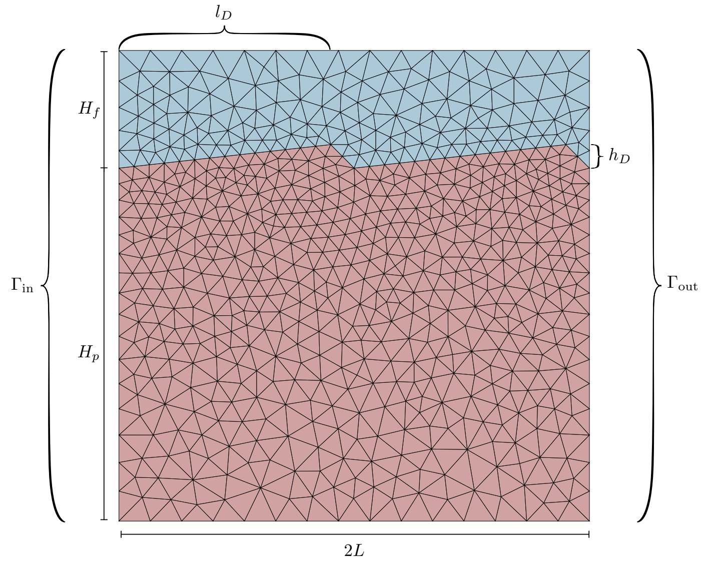


Figure 6: Example 2: Triangulation of the domain on level 1, the blue part is $\Omega_f$ and the red part is $\Omega_p$.

ample 3 of [5]):

$$\begin{cases} \mathbf{u}_f\big|_{\Gamma_{f,\text{in}}} &= \mathbf{u}_f\big|_{\Gamma_{f,\text{out}}} & \text{(essential, periodic velocity)},\\ (\mathbb{T}(\mathbf{u}_f,p_f)\cdot\mathbf{n}_f)\big|_{\Gamma_{f,\text{in}}} &= (-\,\mathbb{T}(\mathbf{u}_f,p_f)\cdot\mathbf{n}_f + p_0\mathbf{n}_f)\big|_{\Gamma_{f,\text{out}}} & \text{(natural, pressure drop)},\\ \varphi_p\big|_{\Gamma_{p,\text{in}}} &= (\varphi_p + p_0)\big|_{\Gamma_{p,\text{out}}} & \text{(essential, pressure drop)},\\ (-\,\mathbb{K}\,\nabla\varphi_p\cdot\mathbf{n}_p)\big|_{\Gamma_{p,\text{in}}} &= (\mathbb{K}\,\nabla\varphi_p\cdot\mathbf{n}_p)\big|_{\Gamma_{p,\text{out}}} & \text{(natural, periodic velocity)}.\end{cases}$$

Defining the transformation

$$\alpha : \Gamma_{\text{in}} \to \Gamma_{\text{out}}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x + 2L \\ y \end{pmatrix}$$

for $\Gamma_{\text{in}} = \Gamma_{f,\text{in}} \cup \Gamma_{p,\text{in}}$ and $\Gamma_{\text{out}} = \Gamma_{f,\text{out}} \cup \Gamma_{p,\text{out}}$, one should understand the periodic boundary conditions as follows:

$$\begin{cases} \mathbf{u}_f(\mathbf{x}) &= \mathbf{u}_f(\alpha(\mathbf{x})),\\ \mathbb{T}(\mathbf{u}_f(\mathbf{x}),p_f(\mathbf{x}))\cdot\mathbf{n}_f(\mathbf{x}) &= -\,\mathbb{T}(\mathbf{u}_f(\alpha(\mathbf{x})),p_f(\alpha(\mathbf{x})))\cdot\mathbf{n}_f(\alpha(\mathbf{x})) + p_0\mathbf{n}_f(\alpha(\mathbf{x})),\\ \varphi_p(\mathbf{x}) &= \varphi_p(\alpha(\mathbf{x})) + p_0,\\ -\,\mathbb{K}\,\nabla\varphi_p(\mathbf{x})\cdot\mathbf{n}_p(\mathbf{x}) &= \mathbb{K}\,\nabla\varphi_p(\alpha(\mathbf{x}))\cdot\mathbf{n}_p(\alpha(\mathbf{x})),\end{cases}$$

for $\mathbf{x} \in \Gamma_{f,\text{in}}$ or $\mathbf{x} \in \Gamma_{p,\text{in}}$, respectively.
For the test spaces of the weak formulation this means that

$$\begin{aligned} V_f &= \left\{\mathbf{v} \in (H^1(\Omega))^d : \mathbf{v}\big|_{\Gamma_{f,\text{in}}} = \mathbf{v}\big|_{\Gamma_{f,\text{out}}} \wedge \mathbf{v}_f\big|_{\Gamma_{f,e}^{\text{top}}} = 0\right\},\\ Q_f &= L^2(\Omega),\\ Q_p &= \left\{\psi \in H^1(\Omega) : \psi\big|_{\Gamma_{p,\text{in}}} = \psi\big|_{\Gamma_{p,\text{out}}}\right\}.\end{aligned}$$

One then searches for $\mathbf{u}_f \in V_f$, $p_f \in Q_f$ and $\varphi_p - p_0 \in Q_p$, where $p_0$ denotes an extension of $p_0$ on the inlet boundary into the interior. The right-hand side of the weak formulation of the Navier–Stokes–Darcy problem then transforms as follows:

(i) For each test function $\mathbf{v} \in V_f$, it is

$$\langle \mathbf{f}_f^1, \mathbf{v} \rangle = (\mathbf{f}, \mathbf{v})_{\Omega_f} + \langle p_0 \mathbf{n}_f(\alpha(\mathbf{x})), \mathbf{v}(\mathbf{x}) \rangle_{\Gamma_{f,\text{in}}},$$

because

$$\begin{aligned}&\langle \mathbb{T}(\mathbf{u}_f(\mathbf{x}), p_f(\mathbf{x})) \cdot \mathbf{n}_f(\mathbf{x}), \mathbf{v}(\mathbf{x}) \rangle_{\Gamma_{f,\text{n}}} \\ &\quad = \langle \mathbb{T}(\mathbf{u}_f(\mathbf{x}), p_f(\mathbf{x})) \cdot \mathbf{n}_f(\mathbf{x}), \mathbf{v}(\mathbf{x}) \rangle_{\Gamma_{f,\text{in}}} + \langle \mathbb{T}(\mathbf{u}_f(\alpha(\mathbf{x})), p_f(\alpha(\mathbf{x}))) \cdot \mathbf{n}_f(\alpha(\mathbf{x})), \mathbf{v}(\alpha(\mathbf{x})) \rangle_{\Gamma_{f,\text{in}}} \\ &\quad = \langle p_0 \mathbf{n}_f(\alpha(\mathbf{x})), \mathbf{v}(\mathbf{x}) \rangle_{\Gamma_{f,\text{in}}},\end{aligned}$$

using $\mathbf{v}\big|_{\Gamma_{f,\text{in}}} = \mathbf{v}\big|_{\Gamma_{f,\text{out}}}$ in the second step.

(ii) The second right-hand side $\langle f_f^2, q \rangle$ does not change.

(iii) For each test function $\psi \in Q_p$, it is

$$\begin{aligned}&\langle \mathbb{K}\, \nabla \varphi_p(\mathbf{x}) \cdot \mathbf{n}_p(\mathbf{x}), \psi(\mathbf{x}) \rangle_{\Gamma_{p,n}} \\ &\quad = \langle \mathbb{K}\, \nabla \varphi_p(\mathbf{x}) \cdot \mathbf{n}_p(\mathbf{x}), \psi(\mathbf{x}) \rangle_{\Gamma_{p,\text{in}}} + \langle \mathbb{K}\, \nabla \varphi_p(\alpha(\mathbf{x})) \cdot \mathbf{n}_p(\alpha(\mathbf{x})), \psi(\alpha(\mathbf{x})) \rangle_{\Gamma_{p,\text{in}}} \\ &\quad = 0.\end{aligned}$$

**4.2.1 Remark.** In the assembling process of the finite element system matrix the periodic boundary data are treated as follows:

- For each node on the inlet boundary, there is a corresponding node on the outlet boundary. The condition

  $$\mathbf{v}\big|_{\Gamma_{f,\text{in}}} = \mathbf{v}\big|_{\Gamma_{f,\text{out}}} \text{ or } \psi\big|_{\Gamma_{p,\text{in}}} = \psi\big|_{\Gamma_{p,\text{out}}}$$

  tells, that these two nodes belong to the same basis function (which now takes value 1 on both of the nodes). In order to assemble the matrix entries corresponding to these special nodes correctly, first assemble them separately and treat the periodic boundary $\Gamma_{\text{in}}$ and $\Gamma_{\text{out}}$ as Neumann boundaries without the identifying condition $\mathbf{v}\big|_{\Gamma_{f,\text{in}}} = \mathbf{v}\big|_{\Gamma_{f,\text{out}}}$ or $\psi\big|_{\Gamma_{p,\text{in}}} = \psi\big|_{\Gamma_{p,\text{out}}}$.

- With the help of basis functions of the Navier–Stokes velocity which take value 1 on the inlet or on the outlet boundary, respectively, one has to construct basis functions of $V_f$.

  Therefore consider the $(i, j)$-th and the $(k, j)$-th entry of the Stokes matrix $A$, i.e.,

  $$A_{ij} = a_f(\mathbf{w}_j, \mathbf{w}_i),\ A_{kj} = a_f(\mathbf{w}_j, \mathbf{w}_k),$$

  with test functions $\mathbf{w}_i$ and $\mathbf{w}_k$, such that the $i$-th column corresponds to a node $\mathbf{x}_i \in \Gamma_{\text{in}}$ and the $k$-th column corresponds to a node $\mathbf{x}_k \in \Gamma_{\text{out}}$ with $\alpha(\mathbf{x}_i) = \mathbf{x}_k$.

  Replacing the $i$-th row by the sum of the $k$-th row and the old $i$-th row yields due to linearity of the second argument

  $$A_{ij} = a_f(\mathbf{w}_j, \mathbf{w}_i + \mathbf{w}_k).$$

  This corresponds to take test functions $\mathbf{w}_i + \mathbf{w}_k \in V_f$.

  The $k$-th entry of the right-hand side needs to be added to the $i$-th entry of the right-hand side as well.

| $\nu$ | $K$ | $\gamma_p$ | $\gamma_f$ | total iterations on grid |  |  | nonlinear iterations on grid |  |  |
|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  | level 1 | level 2 | level 3 | level 1 | level 2 | level 3 |
| $10^{-4}$ | $10^{-3}$ | 1 | 3 | 37 | 39 | 39 | 18 | 19 | 19 |
|  |  | 10 | 30 | 35 | 39 | 37 | 17 | 19 | 18 |
|  |  | 100 | 300 | 53 | 77 | 97 | 26 | 38 | 48 |
| $10^{-4}$ | $10^{-5}$ | 1 | 3 | 37 | 39 | 39 | 18 | 19 | 19 |
|  |  | 10 | 30 | 35 | 41 | 39 | 17 | 20 | 19 |
|  |  | 100 | 300 | 35 | 41 | 39 | 17 | 20 | 19 |
| $10^{-4}$ | $10^{-7}$ | 1 | 3 | 37 | 39 | 39 | 18 | 19 | 19 |
|  |  | 10 | 30 | 35 | 41 | 39 | 17 | 20 | 19 |
|  |  | 100 | 300 | 35 | 41 | 39 | 17 | 20 | 19 |
| $\frac{1}{2} \cdot 10^{-4}$ | $10^{-7}$ | 1 | 3 | 71 | 73 | 69 | 35 | 36 | 34 |
|  |  | 10 | 30 | 71 | 73 | 73 | 35 | 36 | 36 |
|  |  | 100 | 300 | 71 | 75 | 73 | 35 | 37 | 36 |

Table 2: Example 2: Total number of interface iterations and nonlinear iterations of the Robin–Robin algorithm with the Gauss–Seidel method for different choices of $\nu$, $K$, $\gamma_p$, and $\gamma_f$. The number of nonlinear iterations is limited to 1 per interface iteration.

- Now the $k$-th row is set to zero with homogeneous right-hand side except for the $(k, k)$-th entry, which is set to 1, and the $(k, j)$-th entry, which is set to $-1$. With this, one gets the equality of velocity.
- The Darcy part of the matrix can be handled in the same way, just that one has in the last step an inhomogeneous right-hand side, as the pressure drop in this case is an essential boundary condition.

In this example only the Robin–Robin method will be considered as the Neumann–Neumann might not converge for small $K$, see Remark 3.1.6. For the domain three different unstructured triangulations were used as shown in the following table:

| Level | interface edges | cells |  | degrees of freedom |  |
|---|---|---|---|---|---|
|  |  | Navier–Stokes | Darcy | Navier–Stokes | Darcy |
| 1 | 32 | 277 | 1009 | 1392 | 2098 |
| 2 | 68 | 1285 | 4513 | 6093 | 9196 |
| 3 | 138 | 5234 | 18322 | 24181 | 36987 |

As stopping criterion the same as in Example 1 has been used, see (42).
The Robin–Robin method has parameters $\gamma_f$ and $\gamma_p$ so one could examine, which choice of these parameters on which grid level is most efficient. Following the results of [5], one could relate $\gamma_f$ with $\gamma_p$ by setting $\gamma_f = 3\gamma_p$. For measuring the costs, both the number of nonlinear iterations and the total number of iterations (i.e., nonlinear and interface iterations) are considered. Table 2 shows, that the number of iterations is mostly independent of the mesh. The algorithm performed well for all choices of $\gamma_f$ and $\gamma_p$. The number of iterations increased rapidly with a smaller choice of $\nu$. For larger $K$, smaller $\gamma_p$ and $\gamma_f$ seem to be more

convenient. Note that $\nu = \frac{1}{2} \cdot 10^{-4}$ was the smallest number possible before the nonlinear iterations did not converge anymore. In order to solve this arising instability, one has to consider the time dependent equations.

As $\nu = \frac{1}{Re}$ seems to have a great impact on the iterations, one could have a look at the restrictions of the nonlinear iterations per interface iterations. Figure 7 shows, that indeed the number of iterations grows rapidly with larger Reynolds numbers. It also shows that, when measuring the costs in terms of total iterations needed, in comparison to Figure 5, a larger Reynolds number shifts the optimal number of allowed nonlinear iterations per interface iteration to greater numbers than 1 (in this example for $Re = 2 \cdot 10^4$, a restriction of 3 nonlinear iterations is most efficient). If considering just the nonlinear iterations needed as measure of cost, this observation cannot be made.

In the papers [7, 6] it was suggested to do no interface iterations and just solve the Navier–Stokes system fully and then solve the Darcy system once. Figure 8 shows, that there is no qualitative impact. Compared to the Stokes–Darcy problem, the plots of Figure 8 show as well that the Navier–Stokes solution has eddies where the Stokes solution has not. In direct comparison with Figure 1 of [6], one can clearly see that in $\Omega_p$ there are the mentioned exchange, flow, and underflow areas.

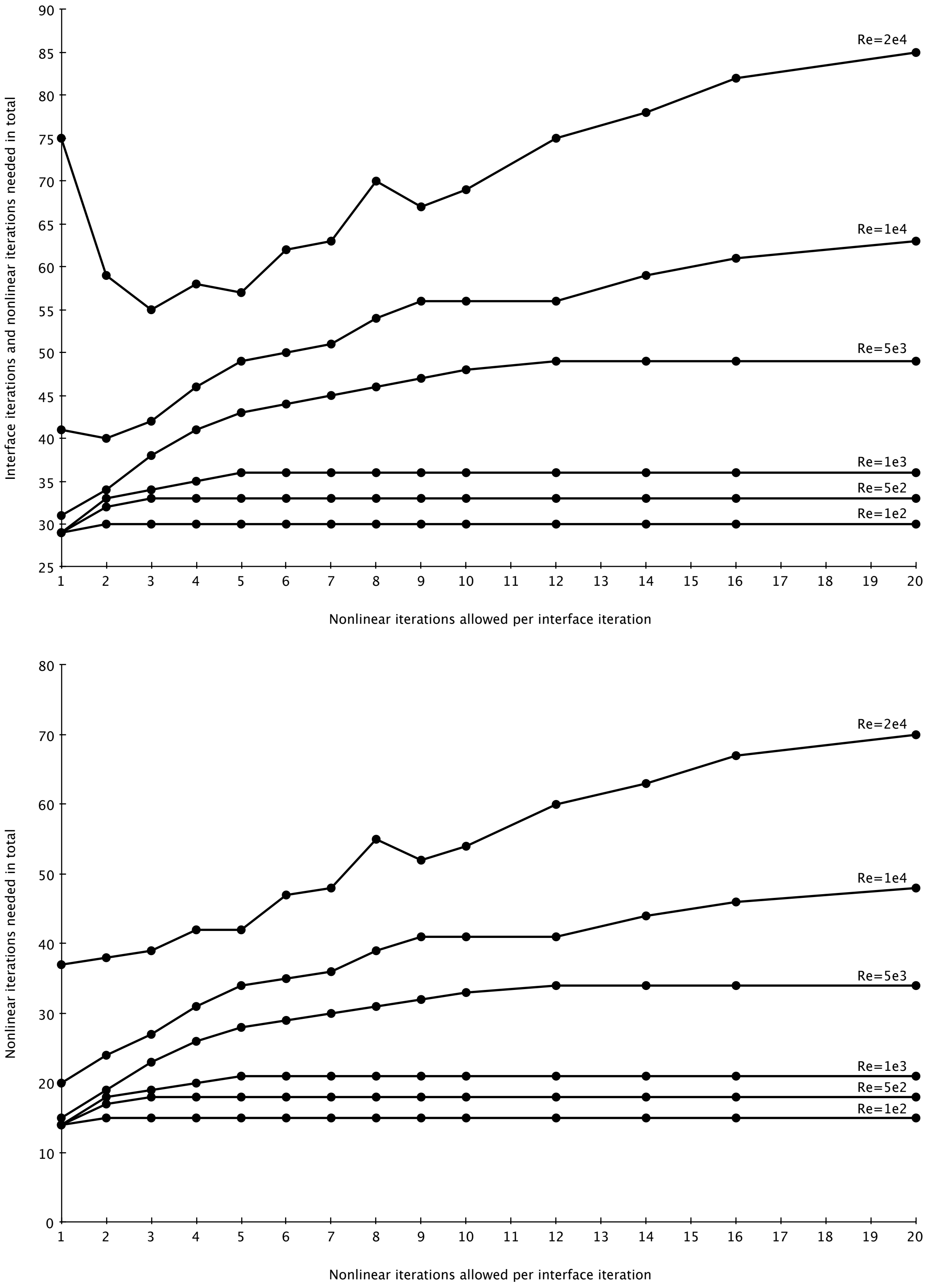
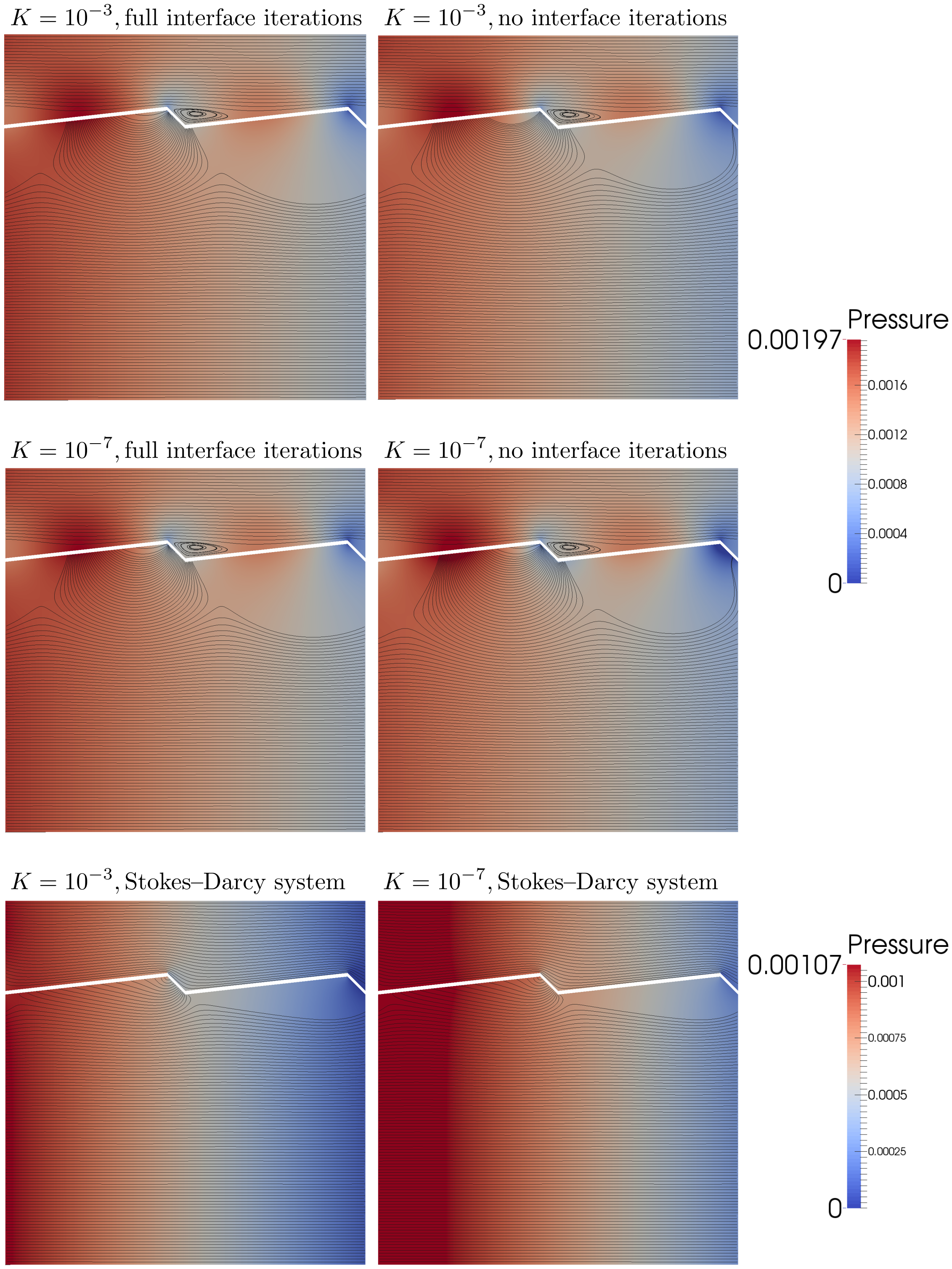


Figure 7: Example 2: This figure shows for grid level 2 with $K = 10^{-5}$, $\gamma_p = 100$ and $\gamma_f = 300$ the total number of iterations needed (upper part) and the total number of nonlinear iterations (lower part) needed, respectively, over the number of fixed-point iterations allowed for the Naiver–Stokes system at each interface-iteration.


$K = 10^{-3}$, full interface iterations

$K = 10^{-3}$, no interface iterations

$K = 10^{-7}$, full interface iterations

$K = 10^{-7}$, no interface iterations

$K = 10^{-3}$, Stokes–Darcy system

$K = 10^{-7}$, Stokes–Darcy system

Figure 8: Example 2: Numerical solution of the Navier–Stokes–Darcy system (upper four plots) and the Stokes–Darcy system (lower two plots) for $\nu = 10^{-4}$, $K \in \{10^{-3}, 10^{-7}\}$ on grid level 3. The lines represent the velocity field $\mathbf{u}_f$ and $\nabla\varphi_p$, respectively, mean flow is from left to right. All plots are colored according to the pressure field, scaled such that the minimum value is zero.

# 5 Conclusion and outlook

This thesis discussed the Navier–Stokes–Darcy problem as an extension of the Stokes–Darcy problem. In particular it compared the Gauss–Seidel method to the Jacobi method for performing the subdomain iterations which arise from the Neumann–Neumann method or the Robin–Robin method. Also it has been investigated which restrictions of nonlinear iterations per interface iterations result in the lowest overall costs needed for convergence.

When using the Gauss–Seidel method in comparison to the Jacobi method, it turns out that the Gauss–Seidel method needs only half the number of the iterations. However it cannot be parallelized. Parallelization of the Jacobi method would reduce the needed time by at most 50%. Altogether this means that the Gauss–Seidel method is the better choice in this case, as it is as fast as the parallelized Jacobi method but behaves better in terms of needed iterations and computational costs.

Using the Gauss–Seidel method, the Neumann–Neumann coupling might not be the right choice for small numbers of hydraulic permeability $K$, because Theorem 3.1.5 requires strong block-diagonal dominance in order to guarantee convergence of the method for each starting vector and the entries of the Darcy part of the system matrix linearly depend on $K$. The Robin–Robin method however has another term in the entries of the Darcy part of the system matrix, which does not depend on $K$ but linearly on $\gamma_p^{-1}$. Hence an appropriate choice of $\gamma_p$ might recover strong block-diagonal dominance.

As in the Navier–Stokes–Darcy problem also the nonlinearity of the Navier–Stokes system has to be resolved, one could ask whether a restriction of nonlinear iterations per interface iteration might give faster convergence. If one chooses to measure the costs by evaluating the number of nonlinear and interface iterations needed, the examples show that for small Reynolds numbers a restriction to one nonlinear iteration per interface iteration and for large Reynolds numbers a restriction to three nonlinear iterations per interface iteration gives optimal convergence. However it still needs to be investigated how exactly these restrictions of nonlinear iterations depend on the Reynolds numbers. If one chooses to measure the costs by just counting the number of nonlinear iterations, this observation cannot be made and a restriction to one nonlinear iteration per interface iteration gives optimal convergence in all cases. How to actually measure the costs depends on the problem. If a factorization of the Darcy system can be stored and reused, counting the nonlinear iterations only is more meaningful. If this cannot be done, one might consider to measure the costs by counting both, the nonlinear iterations and the interface iterations.

Altogether, the Stokes–Darcy problem considered in [5] has successfully been extended to the Navier–Stokes–Darcy problem. The numerical results of the riverbed example coincide qualitatively with results from [7, 6].

Further investigations could consider, e.g., different iterative solvers for the nonlinear equation in order to get (faster) convergence, especially for large Reynolds numbers. Also the optimal choice of $\gamma_f$ and $\gamma_p$ needs to be discussed. Since in the examples a direct solver is used for all subproblems, one could ask if an iterative solver is more efficient, possibly with a restricted number of iterations. When solving the subproblems, preconditioning might be of interest as well.

## References

[1] I. Babuska. On the existence, uniqueness, and approximation of saddle-point problems arising from lagrange multipliers. *Rev. Fr. Autom. Inf. Rech. Oper*, 8:129–151, 1974.

[2] S. P. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[3] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods.* Springer-Verlag New York, Inc., 1991.

[4] T. Bröcker. *Analysis II.* BI Wissenschaftsverlag, 1992.

[5] A. Caiazzo, V. John, and U. Wilbrandt. On iterative subdomain methods for the stokes–darcy problem. 2013.

[6] M. B. Cardenas and J. L. Wilson. Dunes, turbulent eddies, and interfacial exchange with permeable sediments. *Water Resources Research*, 43(8), 2007.

[7] M. B. Cardenas and J. L. Wilson. Hydrodynamics of coupled flow above and below a sediment–water interface with triangular bedforms. *Advances in water resources*, 30(3):301–313, 2007.

[8] M. Discacciati and A. Quarteroni. Navier-stokes/darcy coupling: modeling, analysis, and numerical approximation. *Revista Matemática Complutense*, 22(2), 2009.

[9] V. Girault and P.-A. Raviart. *Finite element approximation of the Navier-Stokes equations.* Springer-Verlag, Berlin, 1981.

[10] V. John. Numerical mathematics ii. 2013.

[11] C. Kanzow. *Numerik linearer Gleichungssysteme: Direkte und iterative Verfahren.* Springer-Lehrbuch. Springer London, Limited, 2005.

[12] N. G. Meyers and J. Serrin. H= w. In *Proceedings of the National Academy of Science*, volume 51, pages 1055–1056, 1964.

[13] P. Neff, D. Pauly, and K.-J. Witsch. Poincare meets korn via maxwell: Extending korn's first inequality to incompatible tensor fields. *arXiv preprint arXiv:1203.2744*, 2012.

Ich versichere, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Berlin, den