Time Space Domain Decomposition Methods for Reactive Transport — Application to CO2 Geological Storage

Florian Haeberlein

Thèse présentée

devant l'Université Paris XIII le 14 octobre 2011

pour obtenir

le grade de D ’U ´ P XIII
Mention M ´ ´

par

F H

Laboratoire d'accueil : Direction Technologie, Informatique et Mathématiques appliquées, IFP Energies nouvelles
École doctorale : Institut Galilée, Université Paris XIII

Sujet de la thèse :

Méthodes de décomposition de domaine espace temps pour le transport réactif

---

Application au stockage géologique de CO_2

Rapporteurs

Danielle H
Peter K

Directrice de thèse

Laurence H

Responsable scientifique IFP Energies nouvelles

Anthony M

Examinateurs

Martin G
Caroline J
Michel K
Roland M

A Thesis Presented to

University Paris XIII on 14 October 2011

in Partial Fulfilment for the Degree of

D ’U ´ P XIII

Mention M ´ ´

by

F H

Hosting Laboratory: Technology, Computer Science and
Applied Mathematics Division, IFP Energies nouvelles
Doctoral School: Institut Galilée, University Paris XIII

Subject:

Time Space Domain Decomposition Methods for Reactive Transport

Application to CO_2 Geological Storage

Reviewers

Danielle H
Peter K

Supervisor

Laurence H

Scientific supervisor at IFP Energies nouvelles

Anthony M

Examiners

Martin G
Caroline J
Michel K
Roland M

Acknowledgements

This thesis has been prepared at the the technology, computer science and applied mathematics division of IFP Energies nouvelles (Rueil-Malmaison, France) and at the analysis, geometry and applications laboratory (LAGA) of the University Paris XIII (Villetaneuse, France) in partial fulfilment of the requirements for receiving the Ph. D. degree. The work presented in this thesis has been carried out from November 2008 to October 2011.

Many people have provided their help and support to make this work succeed, for which I am very grateful. First of all, I like to thank my supervisors Laurence Halpern at University Paris XIII and Anthony Michel at IFP Energies nouvelles for their guidance, encouragements, fruitable discussions but especially for the liberty they gave me to follow my own interests and ideas. They provided me a very agreeable and stimulating research environment with the possibility to meet scientists all around the world.
During a long period, I had the chance to work with Filipa Caetano on many aspects of domain decomposition, theoretical mathematics and numerics. Our two-way exchange had always been highly enjoyable even if most of the time I was just a grateful pupil.
Many other people have provided their help and contributed to this work, their number is as big as my thankfulness and my apologies for not enumerating them.

A very special thank is for my colleagues at LAGA and IFP Energies nouvelles for the fun we had together, especially with my office mates Cindy and Xavier. During the last three years, we underwent the ups and downs of a life as Ph. D. student. I thank Cindy for taking the practical cooking courses with me as a teacher and I grant her officially a M. Pes. (Master of *plats en sauce*), the time we spent together was always very funny and I don't imagine the three years without her. Je tiens à remercier très chaleureusement Cindy pour son soutien dans l'élagage de l'arbre de mon jardin. Son savoir-faire pour en transformer ses fruits en confiture est inestimable.

Finally, I like to thank my friends and family for their continuous support, multiple encouragements and the mutual compassion for those of my friends who also wrote, write or will write their Ph. D. thesis. A very special thank is for my parents who always supported me during this time.

Rueil-Malmaison, October 2011
Florian Haeberlein

Vollendet das ewige Werk!
Wie im Traum ich es trug,
wie mein Wille es wies,
was bange Jahre barg
des reifenden Mannes Brust,
aus winternächtigen Wehen
der Lieb' und des Lenzes Gewalten
trieben dem Tag es zu:
Da steh' es stolz zur Schau,
als kühner Königsbau
prang' es prächtig der Welt!

Richard Wagner

Contents

Introduction

Problem Definition

In 2011, China outrun the United States of America as the most energy consuming country with nearly 50 % of the world's coal consumption. Besides coal, gas and oil play an important role in world's energy production, together the fossil based energies share more than 80 % of world's energy production (cf. [3]). They have in common that CO_2 appears as unavoidable byproduct which is mostly released into atmosphere. Even if scientists and ecologists are disagreeing on the importance, they do commonly agree to the indeed existing impact of CO_2 on the greenhouse effect. The transition from "classical" to renewable energy sources has already begun, nevertheless, they still not exceed 2 % on a global level. As a consequence, in the next decades, our energy will mainly be supplied by fossil fuels.
Carbon Capture and Storage (CCS) is seen as a promising way to ensure the transition from fossil based to renewable energy: on the one hand, it allows to keep existing infrastructures in and use the gained know-how on fossil power plants. On the other hand, it allows to provide necessary time to develop experience and build infrastructure in the relatively new field of renewable energies.
The CCS approach separates CO_2 from other gases during the energy production process and transforms it into a supercritical or liquid state. It is then eventually transported for short distances by pipelines to the injection well, where it is injected in the subsurface. There are two major targets for CO_2 geological storage: depleted oil fields and saline aquifers. In the first case, this technique can be used together with enhanced oil recovery. Saline aquifers are a more interesting target since they offer much higher storage capacities, are widespread over the world and cannot be used for drinking water abstraction. Figure 1 gives a schematic overview of the Carbon Capture and Storage (CCS) process.
During and after injection of CO_2 into saline aquifers, several physical and chemical processes appear. The injected CO_2 dissolves partially in water and changes the pH, the water becomes acid and attacks the rock matrix. This changes the geophysical system, e. g. important changes in the porosity and permeability and hence in the way of how the aquifer moves. In order to ensure the reliability of the technical processes and the consequent changes in the subsurface system, a preprocessing numerical simulation has to be undertaken in order to predict it numerically.

Reactive transport modelling describes the way of how a coupled system of physical transport processes and chemical reaction processes interact with each other. Physical transport processes can essentially be caused by two different phenomena: advection and diffusion-dispersion. Both physical processes do not change the nature, the quality or the characteristics of the transported entities. Chemical processes differ from physical transport processes in that they are local interactions between different species.
Reactive transport models can be used for large field of applications, CO_2 geological storage simulation is only one of them. Those models can be represented by a system of time-dependent transport equations modelled by partial differential equations which are coupled by nonlinear functions that represent the source terms resulting from chemical reactions.

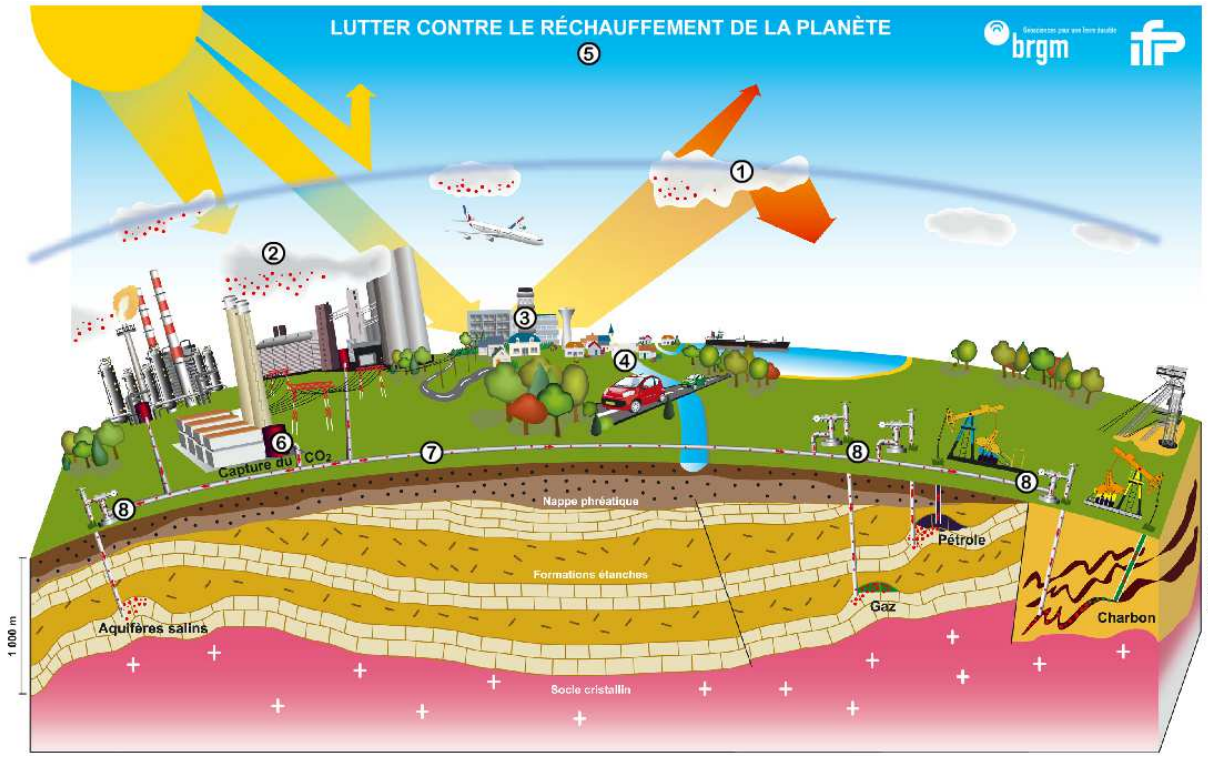


Figure 1: Carbon Capture and Storage (CCS), CO_2 is captured at electric power plants (6), transported by pipelines (7) to wells (8) where it can be injected into saline aquifers, depleted gas or oil fields or coal seams. (image source: BRGM, IFP Energies nouvelles)

The numerical simulation of such reactive transport models in the context of CO_2 geological storage is a quite challenging task. This is mainly due to two different issues. On the one hand, the aim of the simulation is to provide information on a large area within a long time period in order to ensure the reliability of the storage process. The regarded scales are tremendously large, i. e. the interesting time scale is hundreds and even thousands of years and some hundreds of kilometres in space (cf. figure 2). On the other hand, the characteristics of the problem itself which contains highly different time scales and quite different levels of numerical complexity. The time scale of the moving fluid in an aquifer is about one metre per day while chemical reactions in water can either be nearly instantaneous in the equilibrium case or quite slow in the kinetic case (10^{-12} mol s^{-1}). Besides the different time scales, there is also strong heterogeneity in space. While chemical reactions can be highly active in a certain region, most of the concerned simulation domain is in a near-equilibrium state.
The scientific community tries now for several decades to deal with this challenges on several levels: since the very beginning of reactive transport study, the mathematical formulation played an important role. Many different numerical models have been proposed, assumptions have been done and rejected and we are far from having “the reactive transport model”. Writing down a mathematical formulation is always a compromise between reality and complexity. We posses now mathematical formulations which are very powerful in terms of reality but are still manage-

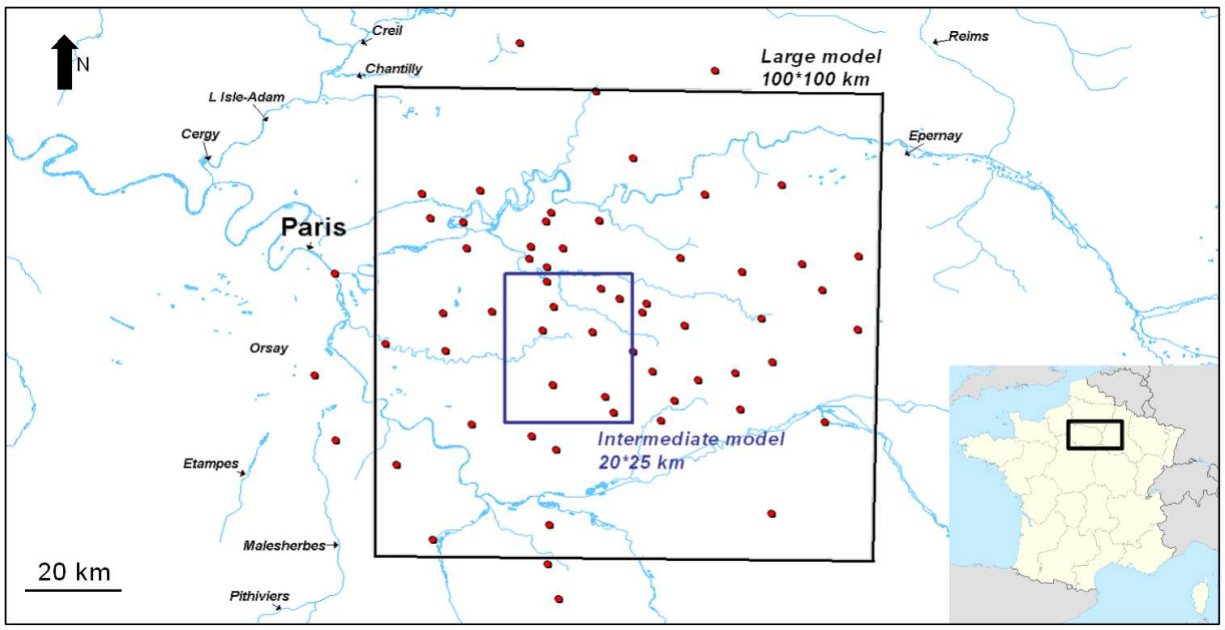


Figure 2: Location and spatial dimension of the SHPCO2 test case: the Dogger aquifer measures over 15.000 km^2, covers a large part of the Parisian basin and is a potential target for CO_2 geological storage. Red dots localise existing wells. (image source: BRGM)

able concerning complexity.
Besides the mathematical formulation, there is a more crucial and unsatisfactory issue, the numerical formulation. Many different approaches have been proposed, in the very first beginning they have been studied theoretically and some of them have been rejected. While computer power and associated numerical algorithms have tremendously increased in the last twenty years, some of the previously rejected formulations have been rediscovered and proved their performance. In the context of long-term simulations with severe chemistry interactions, the class of global implicit approaches has proven its superiority over the class of direct substitution approaches or splitting approaches. Their big advantage is to be able to simulate a quasi steady-state behaviour without being restricted to short time steps to ensure convergence or many iterations to ensure consistency. Applying a global implicit approach to a reactive transport problem results in solving the entire coupled system of partial differential equations at every time step.
Finally, this leads us to the third aspect where important advances have been reached in the last decades, the numerical algorithms for treating the problem. First, for the discretisation of partial differential equations, one has now a wide choice of methods (finite volumes, finite elements, discontinuous Galerkin or mixed/hybrid methods). They all have advantages and disadvantages and the choice of the method is often a matter of taste. Nevertheless, in subsurface modelling, finite volumes and close methods have established themselves for different reasons. On the one hand, they are easy to code, they work on very general meshes which are often based more on geological than on mathematical criteria, and they provide naturally interesting properties like flux conservation which is highly desirable in this context. Nevertheless, all different classes of meth-

ods lead to a large system of algebraic equations which are nonlinear, due to chemical coupling terms. Efficient approaches to solve those problems have been developed and tested in different applications. Robust and fast methods are now available which combine nonlinear and linear solvers and which have only frugal need of memory capacity (e. g. Jacobian-free Newton-Krylov methods).

Nonetheless, besides the development in the previously mentioned fields, there are several issues, which have not been solved. One important field is the heterogeneity in space and time. As it has been mentioned, only a small part of the simulation domain is highly reactive while most of the domain is close to an equilibrium state. The chemical reactivity reflects in strong nonlinear coupling terms which have an impact on the nonlinear solver. Usually, there are two ways out of this problem. Either, one accepts a high number of iterations in the nonlinear solver with the risk that no convergence can be achieved. Or, one adapts the time step in order to ensure faster convergence with higher probability. In practice, the second choice is done since this is the a good compromise between higher costs and convergence probability. Cutting the time step is quite easy when it is done globally. But, as the reason for cutting the time step is localised, the time step should be cut only locally. Otherwise, the part of the simulation domain, which is not dominated by strong nonlinearities is solved with a much higher precision than needed.

Objective of this Work

In this work, we are interested in applying a time space domain decomposition method to reactive transport problems in the context of CO_2 geological storage. The main objective is to benefit from the possibility to treat subdomains with different time discretisations. By this technique, it is possible to overcome the necessity of global mesh refinement in time in order to ensure convergence which is one of the speed bumps on the way towards a high performance simulation of CO_2 geological storage.

The class of waveform relaxation algorithms allows to solve coupled systems of ordinary differential equations separately, where for every equation different methods and time discretisations can be used. An interesting feature of those methods is that the approximation of the sub-equations can be done in parallel.
In the last two decades, the class of Schwarz methods have been rediscovered, studied and successfully applied to many interesting fields. They allow to split problems based on partial differential equations into several subproblems which can be solved individually, in a parallel way, as it is the case of waveform relaxation algorithms.
Both classes of methods, Schwarz methods and waveform relaxation methods, have been composed to form the class of Schwarz waveform relaxation algorithms. They allow now to decompose a time and space dependent problem into subproblems in time and space which can be solved individually, even in parallel. In the beginning, the scientific community emphasised on the parallel treatment of problems since there was and is still an increasing demand for algorithms

that can easily be used on parallel computers. An advantage of those methods is that they can be formulated continuously which means that the numerical approach as well as the discretisation (in time and space) of the subproblems is independent. A quite powerful side effect which relies to the original idea of Schwarz who laid the foundation of the method over a century ago.
In our work, we focus on the time adaptivity and apply therefore a Schwarz waveform relaxation method on a multispecies reactive transport problem in order to overcome time step restrictions on a global level.

Plan of this Work

The manuscript is organised as follows:

In the first chapter, we treat reactive transport modelling. We set up the mathematical model on which we want to apply a domain decomposition algorithm. It is a multispecies reactive transport model including mobile and fixed chemical species. They are transported by advection and diffusion-dispersion. Chemical reactions can be of kinetic type or are assumed to be in equilibrium. In this model, we do not take into account mineral species, they are introduced in chapter 6. Then, we develop a numerical formulation for this problem based on the work presented in [5]. Afterwards, based on the numerical formulation, we present two different numerical approaches using this numerical formulation, a splitting and a global implicit approach. Finally, we present and extend a reduction technique introduced in [53] that allows to reduce the primary size of the problem, decouple equilibrium conditions and group reaction rates.

In chapter 2, we present the state of the art on domain decomposition methods of Schwarz type from a geometrical point of view. We introduce the classical Schwarz algorithm in its alternating and parallel method for the steady state heat equation. We then go further by presenting the class of Schwarz waveform relaxation methods for time-dependent problems, this time on the time-dependent heat equation. Afterwards, we present the class of optimised Schwarz methods that is an extension of all previously presented Schwarz type methods. As we are interested in a performing domain decomposition method, we discuss then three different issues on convergence speed of Schwarz type domain decomposition methods. In the appendix of this chapter, we publish for the first time the entire English translation of Schwarz' original article of 1870 (cf. [75]) originally appeared in German.

The third chapter is dedicated to numerical schemes for discretising the transport operator in a domain decomposition context. We first discuss two different ways of considering a domain decomposition in the finite volumes context with its advantages and drawbacks. How to realise a Robin transmission condition is the next step since this is a necessary ingredient in high performing domain decomposition algorithms. In this context, the choice of the numerical scheme becomes an important issue, we compare therefore the behaviour of a standard and a hybrid finite volume scheme and deduce that the last one has more suitable properties without being more

complex to use. Since Ventcel transmission conditions offer a better convergence performance than Robin transmission conditions, we explain afterwards how to realise numerically such a condition with the hybrid finite volume scheme presented before. As the main objective of our work is to provide local adaptivity in time, we present then the way of how to transmit values between different time and space discretisations. In the appendix, we validate the hybrid finite volume scheme itself and its behaviour in a time space domain decomposition context. Finally, we present the features of the prototype code which is based on the two species reactive transport system studied in chapters 4 and 5.

In chapters 4 and 5, we concentrate on a two species reactive transport system which is a subsystem of the multispecies reactive transport system presented in chapter 1.
In the first part, we consider the case where a mobile and a fixed species are coupled by a linear reaction term. After defining the problem, we can state its well-posedness. We then provide results on the domain decomposition algorithm applied to this system like different transmission conditions, convergence factors, the well-posedness of the algorithm and finally we prove its convergence. Then, always in the mood of a performing algorithm, we study optimised transmission conditions on a theoretical and a numerical way.
In the second part, we consider the case where the coupling term becomes nonlinear. After defining the problem and introducing the domain decomposition algorithm, we study the numerical approaches. We recall the classical approach as an interface problem and can then develop two new approaches which are based on a technique developed in the linear case and known as Krylov accelerators or Krylov-Schwarz methods. Finally, we give numerical results in dimensions up to 3D concerning optimised transmission conditions, local adaptivity in time and the accelerating properties of the two new numerical approaches.

In chapter 6, we finally join a multispecies reactive transport system and a Schwarz waveform relaxation algorithm. We first present the model that we have used in the industrial development platform Arcane. It includes mineral species, chemical reactions can be of kinetic type or assumed to be in equilibrium. The numerical realisation uses different techniques presented in [52] and [48]. Then, we present the domain decomposition technique based on a Schwarz waveform relaxation algorithm with the windowing technique and explain some numerical concepts in the Arcane platform. We finally present two test cases and results. The first one deals with the attack of cement by CO_2 in an injection well, the second one is the benchmark case of the SHPCO2 project to which this thesis is related to.

Finally, we conclude the results of our work and give an outlook of unresolved issues and potential future works.

Essentially, all models are wrong,
but some are useful.

George E. P. Box

1

Modelling Reactive Transport

Contents

Introduction

This chapter treats reactive transport modelling and is divided into two parts:

In the first part, we are regarding a rather general reactive transport model considering one mobile phase and the underlying fixed rock matrix formed of several chemical species. In contrast to many other approaches, we do not impose restrictive conditions on the chemical systems. That is why we can introduce a reactive transport problem in a very general and didactic way in the first part of this chapter. First we state the assumptions of our model and present some notations concerning chemical reactions, then we present the governing equations of the mathematical reactive transport model.
We then develop a numerical formulation that clearly represents the underlying structure of the reactive transport problem by keeping the subproblems well distinguished.
Afterwards, we discuss two numerical approaches that differ in the way of treating the coupled phenomena numerically.

In the second part, we concentrate on reduction techniques for reactive transport modelling. Real test cases in CO_2 geological storage simulation can include several dozens of chemical species (usually up to 30). The spatial dimensions are usually very large (several hundreds or even thousands of kilometres, compare for instance the large case of the SHPCO2 benchmark in figure 2 on page 3 which covers 100 times 100 km) and therefore the number of discrete grid points can be tremendously large, i. e. several hundred thousand, even one million of discrete grid points may be necessary to obtain a fine enough spatial resolution. The price we have to pay for is that, on a numerical level, the intuitive model (i. e. one reaction transport equation per chemical species) is highly undesirable because, for every chemical species, one has one unknown and one underlying equation. One way out of the dilemma is to reformulate the reactive transport model in such a way that it has less primary unknowns, most of the coupling terms are eliminated if possible, the remaining coupling terms are as concentrated as much and, last but not least, the new model describes the same chemical system as before. A way to do so is to apply reduction techniques to the reactive transport models. We present one that claims to be optimal in the second part of this chapter and extend it for the use within a waveform relaxation context. Finally, we show how it can be used in the numerical formulation presented in the first part of the chapter.

1.1 Setting up Reactive Transport

1.1.1 Common assumptions and Notations

The mathematical model we present in this chapter describes a reactive transport system consisting of several mobile and fixed species. We consider one mobile aqueous phase where transport

is described by advection and diffusion-dispersion with a given Darcy field. For the sake of simplicity, we suppose the Darcy flow to be decoupled of the reactive transport problem. The flow field is supposed to be given and constant. Logically, the problem modelling the Darcy-flow is coupled to the reactive transport problem since they influence each other by change of porosity and permeability.

In this work, when no other assumption is done, we call a chemical reaction a "kinetic reaction" since we use kinetics to describe the reaction rate. It turns out, that in the context of reactive transport modelling, a special type of reactions is treated which is called "equilibrium reaction". Those reactions form a subclass of kinetic reactions since they have to satisfy the following assumptions:

- They are reversible reactions, i. e. they can and do occur in both directions. Reactants are at the same time products and inversely.
- The forward reaction rate and the backward reaction rate depend on at least one of the concentrations of reactants and products, respectively. For this reason, if, for instance, the forward reaction rate is higher than in the backward reaction rate, the reaction occurs in forward direction. By this procedure, the amount of reactants decrease, the forward reaction rate slows down, the amount of products increases and finally the backward reaction rate accelerates. The state where forward and backward reaction rates are equal is called the equilibrium state of the reaction.

Note that the second point does not necessarily mean that the reaction has stopped at the equilibrium state, only the net reaction rate has become zero, i. e. the forward and the backward reaction occur at the same rate.

Consider a chemical system consisting of I chemical species X_i. With c_i, $i = 1, \dots, I$ we denote the concentration of a species X_i, i. e. the amount of species in mol per volume of water in m^3. The unit mol is defined to be the number of atoms present in 12 grams of ^{12}C. One mol is given by the Avogadro constant which is approximately $6.022 \cdot 10^{23}$.
The chemical system is governed by J chemical reactions. The stoichiometry of the reactions is given by

$$\sum_{i=1,\dots,I} r_{ij} X_i \rightleftarrows \sum_{i=1,\dots,I} p_{ij} X_i, \qquad j = 1, \dots, J.$$

The species written on the left hand side are defined to be the reactants and the species written on the right hand side are defined to be the products of the reaction. The stoichiometric indices $r_{ij} \geq 0$ and $p_{ij} \geq 0$ describe in which portion of molecules the species X_i is involved in reaction j as reactants and products, respectively. We denote by R_j the reaction rate of reaction j, its unit is given by the number of mol per volume of water in m^3 and per time in s which are reacting. A reaction rate $R_j > 0$ means that the reaction is occurring "from left to right", i. e. reactants are "consumed" and products are "produced". An example of a reaction rate can be found in section 1.1.2.2. Note that the symbol $\rightleftarrows$ may imply that the reaction occurs in both directions. If

a reaction can occur only in one direction, we will model this by the reaction rate that is zero for the direction in which the reaction cannot occur but we continue to write the symbol $\rightleftarrows$.
In the following, it is convenient to write the reactions in the following way

$$\sum_{i=1,\dots,I} s_{ij} X_i \rightleftarrows 0, \qquad j = 1, \dots, J,$$

where the stoichiometric indices s_{ij} are defined to be

$$s_{ij} := \begin{cases} p_{ij} & \text{if species } X_i \text{ is a product in reaction } j, \\ -r_{ij} & \text{if species } X_i \text{ is a reactant in reaction } j, \\ 0 & \text{if species } X_i \text{ is not involved in reaction } j. \end{cases}$$

This convention gives rise to the definition of the stoichiometric matrix $S := (s_{ij})_{i=1,\dots,I,\ j=1,\dots,J}$ in which all chemical reactions are described. The rows of the matrix correspond to the chemical species, the columns correspond to the chemical reactions. Furthermore, we define $R := (R_j)_{j=1,\dots,J}$ to be the column vector of all reaction rates.

1.1.2 Governing equations

We define now the multi-species reactive transport problem as

$$\phi(x)\frac{\partial c_i}{\partial t} + \operatorname{div}\left(c_i \vec{u}(x) - \overline{\overline{D}}(x)\,\nabla c_i\right) = (S\,R(c))_i\,, \qquad i = 1, \dots, I, \tag{1.1}$$

where ϕ is the porosity of the underlying rock matrix and $c = (c_i)_{i=1,\dots,I}$ is the row vector of all concentrations. The divergence and the gradient operator vanish for fixed species since they are not transported. $\vec{u}(x)$ denotes the Darcy field and $\overline{\overline{D}}$ a diffusion-dispersion tensor. Note that the vector valued function $R(c)$ contains in its j-th entry the reaction rate of reaction j. This nonlinear reaction rate depends in general on the concentration of the species involved in this reaction. Nevertheless it is also possible that it depends on concentrations that are not affected by this reaction, this is the case when catalysts are involved. For this reason, the system is nonlinear and all equations are potentially coupled. Note that we supposed that the reactive transport process does not have any influence on the rock property like the porosity. For this reason, we suppose the diffusion-dispersion tensor and the flow field to be constant in time.
The reactive transport system consists of I equations, one for every chemical species. The primary unknowns of interest are the I concentrations.

1.1.2.1 Transport operator

It is comfortable to define a transport operator $\mathcal{L}$ for the chemical species X_i and its concentration c_i as

$$\mathcal{L}c_i := \begin{cases} \operatorname{div}\left(c_i \vec{u}(x) - \overline{\overline{D}}(x)\,\nabla c_i\right) & \text{if species } X_i \text{ is a mobile species,} \\ 0 & \text{if species } X_i \text{ is a fixed species.} \end{cases}$$

The tensor $\overline{\overline{D}}(x)$ is composed of a diffusion part and of a dispersion part

$$\overline{\overline{D}}(x) := \overline{\overline{D}}_{\text{Diffusion}} + \overline{\overline{D}}_{\text{Dispersion}}(x),$$

with

$$\begin{aligned}
\overline{\overline{D}}_{\text{Diffusion}} &:= D_i \, \mathbf{Id}_d, \\
\overline{\overline{D}}_{\text{Dispersion}}(x) &:= \underbrace{\alpha_L \|\vec{u}(x)\| \mathbf{e}_{\vec{u}(x)} \mathbf{e}^t_{\vec{u}(x)}}_{\text{longitudinal dispersion part}} + \underbrace{\alpha_T \|\vec{u}(x)\| (\mathbf{Id}_d - \mathbf{e}_{\vec{u}(x)} \mathbf{e}^t_{\vec{u}(x)})}_{\text{transversal dispersion part}} \\
&= \|\vec{u}(x)\| (\alpha_L - \alpha_T) \mathbf{e}_{\vec{u}(x)} \mathbf{e}^t_{\vec{u}(x)} + \|\vec{u}(x)\| \alpha_T \mathbf{Id}_d,
\end{aligned} \quad ,$$

where d denotes the space dimension, $\mathbf{Id}_d$ the unity matrix of dimension d and $\mathbf{e}_v := \frac{v}{\|v\|}$ the vector with direction of v and norm one. Moreover, $D_i \geq 0$ is the scalar diffusion coefficient, α_T and α_L are the transversal and longitudinal dispersion coefficients, respectively, subject to the condition $\alpha_L > \alpha_T \geq 0$. In practice, one encounters often $\alpha_L \gg \alpha_T$. Note finally, that the transport operator is linear.

1.1.2.2 Kinetic Reaction Rates

Whenever no minerals are involved in the chemical system, one can suppose idealised activities. This means that the concentration of a species is a good approximation for its activity.
One way to model the reaction rate for a kinetic reaction is to use a first order kinetic law

$$R_j(c) = k_j^f \prod_{\substack{s_{ij}<0 \\ i=1,\dots,I}} c_i^{|s_{ij}|} - k_j^b \prod_{\substack{s_{ij}>0 \\ i=1,\dots,I}} c_i^{s_{ij}}, \tag{1.2}$$

with the forward and backward reaction speeds k_j^f and k_j^b, respectively. Note that in this case, the reaction rate depends only on the concentration of the effectively involved species of the reaction.

1.1.2.3 Chemical Equilibrium Conditions

For equilibrium reactions, the reaction rate has no longer the same sense as for kinetic reactions. Indeed, the reaction rate is not infinite or indefinite as some people state, but it is exactly the rate that is necessary to maintain the reaction in equilibrium. Now, the reaction rate becomes itself an unknown of the system and can be determined by an algebraic equilibrium condition. This algebraic condition has to be verified at every discrete spatial point and at every discrete point in time. Modelling some of the reactions as equilibrium reactions can be highly desired in an effort to reduce the stiffness of the problem.
One way of modelling equilibrium reactions is given by the mass action law. The reaction j is in

equilibrium when the forward reaction rate and the backward reaction rate are equal. A nonlinear algebraic condition, the equilibrium condition

$$\prod_{i=1}^{I} c^{s_{ij}} = k_j$$

arises where k_j is the equilibrium constant.

1.2 Numerical Formulation

Reactive transport models in the form of system (1.1) are not very well suited for an identification of the subproblems because they have been mixed up in order to understand the physical and chemical interactions. In this section, we want to represent the subproblems as clearly as possible. By means of a general interface, we have an insight on the influence of chemical subproblems on the global structure. Moreover, by the relation between different subproblems we discover the nested structure and can give a procedure to treat the nested formulation numerically.

The general interface is used for the numerical formulation of the reactive transport problem and intends to give a representation of the problem in a structured way such that it is more easy to implement on computers. In the mood of object orientated programming, we define an interface of the global coupled reactive transport system. This allows to understand the global structure of the problem without struggling with the details. On a global level, this allows also to value the chemical flash as a major step of the chemical subproblems.

Note that the development and description of the numerical formulation in the following sections are based on the scope of this work: we treat only one mobile phase, two different types of chemical species are considered (mobile and fixed species) which can react within two different kind of reactions (equilibrium and kinetic reactions).

1.2.1 A General Interface for the Global System

We define a general interface for a reactive transport system including mobile and fixed species reacting in kinetic and equilibrium reactions. The interface is based on the works of Amir and Kern [5] who developed a component description of a reactive transport system with pure equilibrium reactions. The idea is to eliminate equilibrium reaction rates in the global system and replace them as local conditions by forming chemical components. The concept of chemical components is based on the ideas of Morel [67] and reduces the number of unknowns to a smaller set of primary unknowns by using the equilibrium state in an aqueous solution to access to the information of the secondary unknowns. We suppose that a mobile secondary species can be

expressed by mobile primary species only.
The extended reactive transport system is given in by

$$\begin{aligned}
\partial_t C + \partial_t F + \mathcal{L}(C) + R_{T,\text{kin}} &= 0 \qquad (\#C),\\
\partial_t W + R_{W,\text{kin}} &= 0 \qquad (\#W),\\
T - C - F &= 0 \qquad (\#T),\\
F - \Psi(T, W) &= 0 \qquad (\#F),\\
R_{T,\text{kin}} - \Theta(T, W) &= 0 \qquad (\#R_{T,\text{kin}}),\\
R_{W,\text{kin}} - \Upsilon(T, W) &= 0 \qquad (\#R_{W,\text{kin}}),
\end{aligned} \tag{1.3}$$

where the variables T and W represent the total concentrations of the heterogeneous and fixed components, respectively. C and F are the mobile and fixed part of the heterogeneous components T. $R_{T,\text{kin}}$ and $R_{W,\text{kin}}$ are the kinetic reaction rates, their assignment is done by the nonlinear operators $\Theta(T, W)$ and $\Upsilon(T, W)$. $\Psi(T, W)$ is a nonlinear operator that uses the information from a chemical flash, i. e. it gives the fixed part of the heterogeneous components T in an equilibrium state. The size of every equation is given in parentheses.

1.2.2 The Chemical Flash

The choice of components in the chemical system gives rise to the definition of primary and secondary species:

- **Mobile species:** N_c primary species denoted c and N_x secondary species denoted x.
- **Sorbed species:** N_s primary species denoted s and N_y secondary species denoted y.

Note that for the following developments in this section, c denotes no longer the entire chemical species but only the primary mobile species. We identify the chemical concentration with its name and we suppose ideal activities, i. e. the chemical activity for a species is equal to its concentration. The chemical species are related by each other with kinetic reactions and by equilibrium reactions. We assume that a mobile secondary species x is only formed of mobile primary species. The equilibrium system is described by the matrix-valued Morel tableau

	c	s
c	**Id**	
s		**Id**
x	S_c^x	
y	S_c^y	S_s^y

where S_a^b are stoichiometric matrices, **Id** is the identity matrix, missing entries represent zero matrices. The Morel tableau denotes in its rows the stoichiometric formula within primary species denoted in the columns.

The equilibrium state gives rise to the following equations which have to be satisfied. The first system describes the mass conservation of the total component amounts T, W of the mobile and sorbed components:

$$\begin{aligned} c + (S_c^x)^t x + (S_c^y)^t y &= T, \\ s \qquad\qquad + (S_s^y)^t y &= W. \end{aligned} \tag{1.4}$$

The second system describes the mass action laws of the equilibrium reactions where K_x and K_z are the equilibrium constants of the associated equilibrium reactions:

$$\begin{aligned} \ln(x) - (S_c^x)\ln(c) \qquad\qquad &= \ln K_x, \\ \ln(y) - (S_c^y)\ln(c) - (S_s^y)\ln(s) &= \ln K_y. \end{aligned} \tag{1.5}$$

The equations (1.4) and (1.5) form together the chemical flash which has to be satisfied locally. We combine the flash equations to a nonlinear function

$$H : \begin{pmatrix} c \\ s \\ x \\ y \end{pmatrix} \longmapsto \begin{pmatrix} c + (S_c^x)^t x + (S_c^y)^t y - T \\ s \qquad\qquad + (S_s^y)^t y - W \\ \ln(x) - (S_c^x)\ln(c) \qquad\qquad - \ln K_x \\ \ln(y) - (S_c^y)\ln(c) - (S_s^y)\ln(s) - \ln K_y \end{pmatrix}.$$

Performing a chemical flash consists then in finding the zero of H for a given set of total component concentration T and W. We denote this process briefly by $H^{-1}(0)$ and mark the solution with asterisks (c^*, s^*, x^*, y^*) in order to emphasise that those concentrations realise the equilibrium state.

1.2.3 Relation between the Chemical Flash and Local Physics

Basically, the kinetic reaction rates are given as function of the chemical species concentrations and not of component concentrations. Therefore, it is necessary, first, to calculate a chemical flash in order to be able to evaluate the kinetic operators. In figure 1.1 we present the corresponding scheme for the chemical flash and the subsidiary chemical operators. It is evident that the realisation of the chemical flash is a vitally important step in the realisation of a reactive transport calculation. Nevertheless, it is not the easiest component. Indeed, in practice, one encounters often difficulties in solving the chemical flash that may only be solved with heuristic adjustments of standard techniques. Concerning the existence of a solution, Weltin has proved in [78] that a unique solution exists in a certain interval when all stoichiometric coefficients are positive. The restriction to systems with only positive stoichiometric coefficients may be seen severe especially when general reduction techniques provide a Morel tableau with negative coefficients. Therefore, one is interested in numerical methods that are reliable with respect to the exact solution, robust with respect to the initial guess and fast with respect to the computation time. In [68] Morin and in [17] Carrayrou et al. present and compare several zeroth and first order approaches for solving the chemical equilibrium problem. Carrayrou et al. also propose a competitive combined zeroth-first order method using a simplex and a Newton method. A study of the chemical flash and a globalised Newton-based method in the here presented formulation has been done in [40].

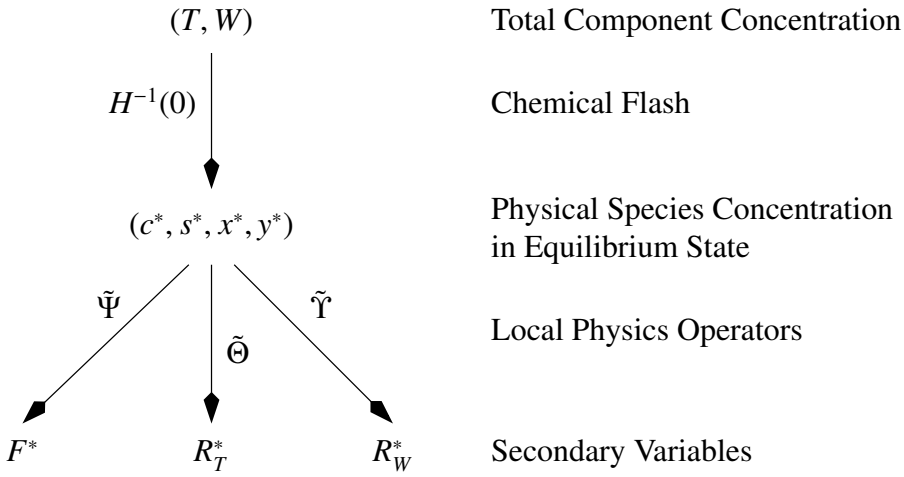


Figure 1.1: Scheme of correspondence between global and chemical variables and local physics

1.3 Numerical Approach

Approaching a reactive transport problem numerically means not only discretising the different operators and variables but also a choice of how to attack the equations and the unknowns has to be done.

Reactive transport models are coupled models that include different equations of different natures. As for different kinds of reactions exist different numerical approaches, the first intuitive approach is to solve the equations separately in what is called a splitting approach. The main advantage of splitting algorithms is that they lead to a small set of equations which can be solved with low computational costs per iteration. Moreover, from a practical viewpoint, programing a splitting approach is insofar easier as one can use already existing and sometimes highly developed modules for the two phenomena or strictly separate the two parts of the codes when programing oneself. Note that a splitting approach allows the use of basic solvers for "simpler problems" like scalar advection or diffusion solvers (scalar partial differential equation) or local chemical solvers (ordinary differential equation system). The main drawback of this method is the introduction of an additional splitting error which can lead to a large number of iterations between the two subproblems for one time step or which can restrict enormously the time step size in order to be able to achieve convergence for one time step. In fact, following the idea of Amir and Kern in [5], one can interpret iterative operator splitting methods as an application of block Gauß-Seidel type method on the coupled system. Basing on this idea, we can explain the convergence limitation of the splitting approach by extradiagonal entries like coupling terms between chemical species: the stronger the coupling influence, the smaller the time step has to be chosen in order to guarantee a diagonal domination of the matrix and hence a convergence of the splitting approach.
The more sophisticated but also more demanding approach in terms of numerical resources is a global approach. No operator splitting is applied to the model in order to solve it in a coupled

way. The advantages are clear: generally, a global implicit approach shows a faster convergence, especially for hard chemical problems. Moreover this approach is more robust and can treat stiffer problems. This allows one to use larger time steps and coarser spatial discretisations which leads to a smaller number of iterations. One drawback, that has become less important with the development of high performance methods for nonlinear and linear systems, is that a global approach leads to a huge system of equations that can include strong coupling terms. The numerical methods have to be chosen carefully and cleverly because not all standard methods are able to solve those kind of complex systems.
Finally, there is a third type of approach called direct substitution approach (DSA): in this approach, all algebraic equations resulting from the chemical equilibrium problem are directly inserted in the partial differential equations of the transport system. The resulting system consists of coupled nonlinear partial differential equations. The number of equations is tremendously reduced compared to a global implicit approach and the number of iterations to reach convergence is considerably less than for a standard iterative approach (compare Saaltink et al. [74]). Nevertheless, this approach is only possible if the primary species are carefully chosen.
Note that the numerical formulation using the operator Ψ to represent the effects of chemistry as we did in the section 1.2 allows to use both splitting and global approaches. Nevertheless, applying a global implicit method to our formulation is different from using a direct substitution approach since we do not integrate the chemical equilibrium conditions directly in the transport equations. For this reason, the choice of components is not as restricted as for a DSA and we can include for example a reduction technique as presented in section 1.4. The chemical influence of equilibrium reactions on the transport system is done by a linearised elimination instead of a direct substitution. This allows to reduce the size of the coupled system as in the DSA approach (even if the reduction is less important) while the chemical problem is kept separately. The possibility for the use of black-box solvers for chemistry is preserved even in the global implicit approach which is an important advantage since chemistry solvers have become more and more sophisticated and high-performing.

In the very beginning (about 1990) of the comparison studies of global and splitting approaches, one just compared the methods theoretically because computational power and especially memory availability restricted hardly the implementation of global approaches. In their very influential paper [79], Yeh and Tripathi came to the result that global approaches are too cost intensive and lead to excessive CPU memory and CPU times. Since then, global approaches had been priced out of the market and it took about one decade until global approaches came back to the spotlight. While both methods for solving reactive transport (global as well as splitting) and computational availability progressed enormously in the 1990s, one is now able to compare both approaches from a practical viewpoint. So have done Saaltink et al. (cf. [72] and [74]): they have proven that a standard iterative approach tends to require more iterations than a global approach. Moreover, a splitting approach seems to fail for problems with high kinetic rates and cases with a high number of flushed pore volumes. On the other hand, they have proved that a global approach is more robust in this cases. Concerning the CPU time of the splitting method, they have shown that it grows linearly up to a large number of spatial nodes, whereas the CPU time of a global approach grows with an order of 2 or 1.6 for direct or iterative linear solver, respectively.

This means that for large spatial dimensions the global approach seems to be less favourable. Nevertheless, the global implicit approach has one large practical advantage that comes to the fore when considering steady-state problems: global implicit approaches can treat very large time steps without loosing stability or introducing additional errors. In long term simulations like reservoir modelling where long time periods over several hundreds, even thousands of years, are considered, this issue becomes crucial. Beside the drawback of the huge memory size to form and store the Jacobian in global approaches, one should be conscious of the fact that new methods for solving nonlinear systems have been developed that do not need to store and form the Jacobian (cf. [5], [51] and [16]). Moreover, one should be aware of hardware development: parallel supercomputers and computer memories in general tend to grow.

For the following presentation of a splitting and a global implicit approach, we discretise the problem (1.3) by an implicit Euler scheme for a time step $t^n \to t^{n+1}$. The nonlinear function to solve is given by

$$\mathcal{F}\begin{pmatrix} C^{n+1} \\ W^{n+1} \\ T^{n+1} \\ F^{n+1} \\ R^{n+1}_{T,\mathrm{kin}} \\ R^{n+1}_{W,\mathrm{Kin}} \end{pmatrix} = \begin{pmatrix} C^{n+1} - C^n + F^{n+1} - F^n + \Delta t\, \mathcal{L}(C^{n+1}) + \Delta t\, R^{n+1}_{T,\mathrm{kin}} \\ W^{n+1} - W^n + \Delta t\, R^{n+1}_{W,\mathrm{kin}} \\ T^{n+1} - C^{n+1} - F^{n+1} \\ F^{n+1} - \Psi(T^{n+1}, W^{n+1}) \\ R^{n+1}_{T,\mathrm{kin}} - \Theta(T^{n+1}, W^{n+1}) \\ R^{n+1}_{W,\mathrm{kin}} - \Upsilon(T^{n+1}, W^{n+1}) \end{pmatrix} = 0. \tag{1.6}$$

1.3.1 Splitting Approach

Approaching the problem (1.6) by splitting type methods offers the possibility to solve the equations separately one after the other. The system sizes are reduced in the sense of classical divide and conquer techniques. The drawback of a splitting approach is the fact that they often need several iterations in order to converge for a time step. In [74] Saaltink et al. showed that they need up to five time more iterations than a global implicit approach. The order in solving the equations in the splitting approach may therefore be a less important issue whereas the association of an equation to the update of one variable is more crucial. To answer this question, it might be interesting to know the physical/chemical sense of every equation. We follow the developments of Saaltink et al. in [72] where in a first step the total mobile concentration is associated to the transport equation and in a second step the other unknowns are updated node by node. We apply here the same technique.

Let us first have a look on the transport equation for the mobile and heterogeneous components:

$$C^{n+1} - C^n + F^{n+1} - F^n + \Delta t\, \mathcal{L}(C^{n+1}) + \Delta t\, R^{n+1}_{T,\mathrm{kin}} = 0.$$

Basing on the idea that in a transport process the mobile parts of the components are principally affected, we associate this equation to the update of the mobile part C. The equation writes then

$$C^{n+1,k+1} + \Delta t\, \mathcal{L}(C^{n+1,k+1}) = C^n - F^{n+1,k} + F^n - \Delta t\, R^{n+1,k}_{T,\mathrm{kin}}.$$

Note that in the first iteration of the SIA or for a SNIA, the values of $F^{n+1,k}$ and $R^{n+1,k}_{T,\mathrm{kin}}$ on the right hand side are not known. One usually replaces therefore $-F^{n+1,k} + F^n$ by $-F^n + F^{n-1}$ and $R^{n+1,k}_{T,\mathrm{kin}}$ by $R^n_{T,\mathrm{kin}}$, i. e. the influence of the fixed part of the mobile components and the kinetic reaction rates are taken from the previous time step. Note that the right hand side includes the coupling term $-F^{n+1,k} + F^n - \Delta t\, R^{n+1,k}_{T,\mathrm{kin}}$. This coupling term becomes less important if the time step Δt is chosen to be small. For this reason, the SIA may need many iterations if the time step is chosen too large. Moreover, if the time step is chosen too large, this term can be seen as stability condition, i. e. the SIA may not converge if the time step is too large and the coupling term is too important. Then, as a matter of course, the accumulation equation for the fixed components is associated to the total fixed components:

$$W^{n+1,k+1} = W^n - \Delta t\, R^{n+1,k}_{W,\mathrm{kin}}.$$

As before, when the values $R^{n+1,k}_{W,\mathrm{kin}}$ are not known they are replaced by the values of the previous time step. Concerning the chemical closure equation for the heterogeneous components, it is associated with the update of the total concentration of the mobile and heterogeneous components

$$T^{n+1,k+1} = C^{n+1,k} + F^{n+1,k},$$

due to the fact that for the operator Ψ only the fixed part of the heterogeneous components is left:

$$F^{n+1,k+1} = \Psi(T^{n+1,k}, W^{n+1,k}).$$

Finally, the kinetic operators are updated respectively by

$$\begin{aligned} R^{n+1,k+1}_{T,\mathrm{kin}} &= \Theta(T^{n+1,k}, W^{n+1,k}), \\ R^{n+1,k+1}_{W,\mathrm{kin}} &= \Upsilon(T^{n+1,k}, W^{n+1,k}). \end{aligned}$$

The order in which we have presented the update of the variables corresponds to the order that we propose to use in a SNIA (cf. algorithm 1.1) and in a SIA (cf. algorithm 1.2).

1.3.2 Global Approach

A global approach of system (1.6) may be realised using Newton's method: in every Newton iteration $k \to k+1$, one solves the linear system

$$\mathcal{F}' \begin{pmatrix} C^{n+1,k} \\ W^{n+1,k} \\ T^{n+1,k} \\ F^{n+1,k} \\ R^{n+1,k}_{T,\mathrm{kin}} \\ R^{n+1,k}_{W,\mathrm{Kin}} \end{pmatrix} \cdot \begin{pmatrix} C^{n+1,k+1} - C^{n+1,k} \\ W^{n+1,k+1} - W^{n+1,k} \\ T^{n+1,k+1} - T^{n+1,k} \\ F^{n+1,k+1} - F^{n+1,k} \\ R^{n+1,k+1}_{T,\mathrm{kin}} - R^{n+1,k}_{T,\mathrm{kin}} \\ R^{n+1,k+1}_{W,\mathrm{Kin}} - R^{n+1,k}_{W,\mathrm{Kin}} \end{pmatrix} = \mathcal{F} \begin{pmatrix} C^{n+1,k} \\ W^{n+1,k} \\ T^{n+1,k} \\ F^{n+1,k} \\ R^{n+1,k}_{T,\mathrm{kin}} \\ R^{n+1,k}_{W,\mathrm{Kin}} \end{pmatrix}.$$

Algorithm 1.1 Standard Non-Iterative Approach for the approximation of system (1.6) using a splitting technique

INPUT: C^n, F^n, F^{n-1}, T^n, W^n, $R^n_{T,\text{kin}}$, $R^n_{W,\text{kin}}$
RETURN: C^{n+1}, F^{n+1}, T^{n+1}, W^{n+1}, $R^{n+1}_{T,\text{kin}}$, $R^{n+1}_{W,\text{kin}}$

// Solve for C^{n+1}
$C^{n+1} + \Delta t\, \mathcal{L}(C^{n+1}) = C^n - F^n + F^{n-1} - \Delta t\, R^n_{T,\text{kin}}$

// Solve for W^{n+1}
$W^{n+1} = W^n - \Delta t\, R^n_{W,\text{kin}}$

// Update T^{n+1}
$T^{n+1} = C^{n+1} + F^n$

// Solve the chemical problem in order to obtain F^{n+1}
$F^{n+1} = \Psi(T^{n+1}, W^{n+1})$

// Calculate the kinetic reaction rates
$R^{n+1}_{T,\text{kin}} = \Theta(T^{n+1}, W^{n+1})$
$R^{n+1}_{W,\text{kin}} = \Upsilon(T^{n+1}, W^{n+1})$

In the context of a finite volume approach, we exemplify for two finite volumes K and L the derivative within its bloc shape in formula (1.7). The derivatives of Ψ, Θ and Υ are described in sections 1.4.5.5 and 1.4.5.6. The coefficient a_{KL} denote transmissivity coefficients of the discretised transport operator between cells K and L.

$$\mathcal{F}' \begin{pmatrix} C^{n+1,k} \\ W^{n+1,k} \\ T^{n+1,k} \\ F^{n+1,k} \\ R^{n+1,k}_{T,\text{kin}} \\ R^{n+1,k}_{W,\text{Kin}} \end{pmatrix} = \left(\begin{array}{c||c} A_{KK} & A_{KL} \\ \hline\hline A_{LK} & A_{LL} \end{array}\right), \tag{1.7}$$

Algorithm 1.2 Standard Iterative Approach for the approximation of system (1.6) using a splitting technique

INPUT: C^n, F^n, F^{n-1}, T^n, W^n, $R^n_{T,\text{kin}}$, $R^n_{W,\text{kin}}$
RETURN: C^{n+1}, F^{n+1}, T^{n+1}, W^{n+1}, $R^{n+1}_{T,\text{kin}}$, $R^{n+1}_{W,\text{kin}}$

$k = -1$
repeat
  $k++$
  if $k = 0$ **then**
    $S = F^n - F^{n-1}$
    $R = R^n_{T,\text{kin}}$
    $F = F^n$
  else
    $S = F^{n+1,k} - F^n$
    $R = R^{n+1,k}_{T,\text{kin}}$
    $F = F^{n+1,k}$
  end if
  // Solve for $C^{n+1,k+1}$
  $C^{n+1,k+1} + \Delta t\, \mathcal{L}(C^{n+1,k+1}) = C^n - S - \Delta t\, R$

  // Update $W^{n+1,k+1}$
  $W^{n+1,k+1} = W^n - \Delta t\, R$

  // Update $T^{n+1,k+1}$
  $T^{n+1,k+1} = C^{n+1,k+1} + F$

  // Solve the chemical problem in order to obtain F^{n+1}
  $F^{n+1,k+1} = \Psi(T^{n+1,k+1}, W^{n+1,k+1})$

  // Calculate the kinetic reaction rates
  $R^{n+1,k+1}_{T,\text{kin}} = \Theta(T^{n+1,k+1}, W^{n+1,k+1})$
  $R^{n+1,k+1}_{W,\text{kin}} = \Upsilon(T^{n+1,k+1}, W^{n+1,k+1})$

until $\frac{\|C^{(n+1)(k+1)} - C^{(n+1)(k)}\|}{\|C^{(n+1)(k+1)}\|} + \frac{\|F^{(n+1)(k+1)} - F^{(n+1)(k)}\|}{\|F^{(n+1)(k+1)}\|} + \frac{\|W^{(n+1)(k+1)} - W^{(n+1)(k)}\|}{\|W^{(n+1)(k+1)}\|} + \frac{\|R^{(n+1)(k+1)}_{T,\text{kin}} - R^{(n+1)(k)}_{T,\text{kin}}\|}{\|R^{(n+1)(k+1)}_{T,\text{kin}}\|} + \frac{\|R^{(n+1)(k+1)}_{W,\text{kin}} - R^{(n+1)(k)}_{W,\text{kin}}\|}{\|R^{(n+1)(k+1)}_{W,\text{kin}}\|} < \varepsilon$

$C^{(n+1)} = C^{(n+1)(k+1)}$
$F^{(n+1)} = F^{(n+1)(k+1)}$
$T^{(n+1)} = T^{(n+1)(k+1)}$
$W^{(n+1)} = W^{(n+1)(k+1)}$
$R^{n+1}_{T,\text{kin}} = R^{n+1,k+1}_{T,\text{kin}}$
$R^{n+1}_{W,\text{kin}} = R^{n+1,k+1}_{W,\text{kin}}$

where the diagonal blocks are given by

$$A_{KK} = \left(\begin{array}{c|c|c|c|c|c} (\phi_K + \Delta t a_{KK})\,\mathbf{Id} & 0 & 0 & \phi_K \mathbf{Id} & \Delta t \mathbf{Id} & 0 \\ \hline 0 & \mathbf{Id} & 0 & 0 & 0 & \Delta t \mathbf{Id} \\ \hline -\mathbf{Id} & 0 & \mathbf{Id} & -\mathbf{Id} & 0 & 0 \\ \hline 0 & -\frac{\partial \Psi(T,W)}{\partial W} & -\frac{\partial \Psi(T,W)}{\partial T} & \mathbf{Id} & 0 & 0 \\ \hline 0 & -\frac{\partial \Theta(T,W)}{\partial W} & -\frac{\partial \Theta(T,W)}{\partial T} & 0 & \mathbf{Id} & 0 \\ \hline 0 & -\frac{\partial \Upsilon(T,W)}{\partial W} & -\frac{\partial \Upsilon(T,W)}{\partial T} & 0 & 0 & \mathbf{Id} \end{array}\right),$$

and the extradiagonal blocks are given by

$$A_{KL} = \left(\begin{array}{c|c|c|c|c|c} \Delta t a_{KL} \mathbf{Id} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

1.4 Reduction Techniques

The reactive transport model that we have introduced in chapter 1.1 is easy to understand on a mathematical level because it provides one equation for every chemical species. Exactly for this reason, it may be hard to use it directly in a numerical approach. In fact, reactive transport simulations need huge amounts of computational resources already in two dimensions, even more in three dimensions when realistic simulations are done in the context of reservoir, atmospheric or surface water applications. Different approaches to reduce the problem size or to split the reactive transport problem into smaller subproblems are possible:

The problem size and its numerical difficulty are essentially determined by the quantity and quality of physical and chemical processes that are taken into account. Moreover, both are not coupled: small problems may be difficult and large problems can be easy on a numerical level. For this reason, on the one hand, one can reduce the number of chemical species that one want to simulate to a minimum. More than that, physical and chemical effects can be limited to the most influential one. On the other hand, there is always a desire of detail and accuracy: numerical simulation becomes often interesting when the results are as close to nature as possible.

An intuitive way is to get to the root of the problem: as reactive transport models are coupled models of transport and chemical phenomena, one can decouple both problems by solving them separately. This procedure called Sequential Non Iterative Approach (SNIA) has been used in the first numerical simulators. A problematic drawback is the introduction of additional consistency errors at the solution which can only be controlled by time step restrictions.
The Sequential Iterative Approach (SIA) iterates between the two submodels until consistency is reached for the actual time step. By this long way round consistency might be guaranteed but stability cannot always be attained. Moreover, both approaches have shown to be less performing especially in the case of severe kinetic reactions.
Both SNIA and SIA are logically more numerical approaches than reduction techniques but can be seen as those since they reduce the number of equations and unknowns per subsystem. We have detailed both numerical approaches in section 1.2 and discussed there also the numerical efficiency and problems, especially compared to other approaches.

Authentic reduction techniques distinguish from the operator splitting techniques mentioned above by not imposing a numerical approach on the model but treating it as one and only one coupled model. The choice of a numerical approach is totally free after the model reduction, both global approaches and splitting techniques may be applied even if reduction techniques aim for a facilitation of the numerical treatment in a global approach.
A classical reduction technique consists in forming linear combinations of concentrations, the so called components. The aim of reduction techniques is to reformulate the reactive transport problem as an equivalent problem which has a considerably smaller number of (primary) unknowns but keeps the information of all physical present species. The essential key of reduction techniques is that the resulting problem formulation has no further restrictions compared to the original problem and simulates exactly the same phenomena but is more suitable for numerical simulations. For this reason authentic reduction techniques distinguish also from model changes.

1.4.1 Overview

A basic observation in realistic chemical systems is that they are often overdetermined when one takes all concentrations of the physically present chemical species into account. As a consequence, one can reduce the number of given equations by reducing the number of considered species without loosing the information of the eliminated species. This technique is quite close

to systems of linear equations that are overdetermined (but feasible).
Different ways of manipulating the system in order to get a minimal size are imaginable. All proposed reduction techniques so far try more or less to achieve a minimum size. The difference between those techniques settles though on a higher level: is there a way of reducing the chemical system to a minimum size while the resulting global reactive system is as much decoupled as possible? And more: is there an intelligent way of grouping the nonlinear coupling terms in as less equations as possible?

Most of the classical reduction techniques try to reduce the chemical system on an algebraic level by forming linear combination of species/concentrations. Saaltink et al. present in [71, 73] six different ways of reducing the number of equations to the number of degrees of freedom according to thermodynamic rules. Their advantages and drawbacks are studied and two of the resulting techniques are studied in detail in the context of a standard iterative and a global implicit approach. A more general technique based on a paradigm system is proposed by Molins et al. in [66]. The system size becomes totally reduced as the number of components that need to be solved together is, at most, equal to the number of independent kinetic reactions. This offers an optimal reduction an decoupling of the system and the nonlinear coupling terms.
All those reduction techniques have in common that they allow up from the beginning to form linear combinations of mobile and fixed species forming heterogeneous components. A reduction technique that, in terms of optimality, is similar to the one proposed by Molins has been proposed by Kräutle et al. in [53]. Here, special emphasis is laid on the distinction between mobile and fixed species and in particular on not mixing up mobile and fixed species during the transformation. Moreover, even equilibrium reactions between mobile and fixed species are considered which is not the case in Molins approach. An extension to a three phase flow is given as well as the capability to include precipitation/dissolution treatment is described (cf. [52]). The advantages of this technique compared to other techniques is that it enables a decoupling of some equations without enforcing a decoupling by splitting techniques, non reacting components decouple naturally and can be treated apart. Moreover, this technique is well-suited for a global implicit approach as within the transformation the sparse shape of the Jacobian matrix is preserved.

1.4.2 An Optimal Reduction Technique

We present now the reduction technique introduced by Kräutle in [52]. It is designed for the reduction of a chemical system with several mobile and fixed chemical species. Those species can be linked to each other by different kind of reactions. The chemical system is attacked on an algebraic level by transforming the stoichiometric matrix by means of linear transformations.
The method consists in several steps, each of it has another aim. The outline of the procedure is the following:

1. Arrange the chemical species in two groups, mobile and fixed ones.

2. Distinguish between kinetic and equilibrium reactions.

3. An orthogonal basis of the reaction space allows to extract non reacting components. Those components are separated in mobile and fixed ones. They can be treated separately in a system of purely linear ODEs.

4. The basis of equilibrium reactions is manipulated in such a way that is consists of three subbases. One acting exclusively on mobile species, one acting exclusively on fixed species and one representing the heterogeneous equilibrium reactions. The resulting reduced system is formed of PDEs that are coupled by the nonlinear kinetic reaction laws.

5. Finally, one can replace a part of the reduced PDE system by a local nonlinear system of ODEs describing the mass action laws for the equilibrium reactions.

The resulting system consists of three subsystems: one system of totally decoupled non reacting components described by linear ordinary differential equations that can be solved separately and for the entire simulation time in advance. One global system of nonlinear coupled partial differential equations describing the kinetically reacting components. And finally a local system of system of nonlinear algebraic equations that determinate the equilibrium state of the chemical system. The last two systems are coupled to each other.
For the sake of readability and without loss of generality, we suppose in this section the porosity to be equal to one.

1.4.2.1 Basic principle: Extracting non reacting components

Consider a chemical system with I species. Those species are related to each other by J reactions. A reaction j is described by a stoichiometric coefficient s_{ij} for the $i = 1, \dots, I$ species. The resulting structure is the stoichiometric matrix S for which a column corresponds to a reaction and a row corresponds to a chemical species. The matrix is therefore of size (I, J).

For the following procedure, we need the concept of a pseudoinverse matrix that is the extension of the concept of inverse matrices in the rectangular case. Note that different concepts of pseudoinverse matrices exist that have different properties. We treat here only the case of the widely-used Moore-Penrose-inverse and (as common in the mathematical community) use only the term "pseudoinverse" to design this type of pseudoinverse.

Definition 1.1 (*Pseudoinverse*)

Consider a matrix $T \in \mathbb{R}^{m \times n}$, with T^T we denote the transposed matrix of T. The matrix $T^\dagger$ with the following properties

$$
\begin{aligned}
T^\dagger T T^\dagger &= T^\dagger, \\
T T^\dagger T &= T, \\
(T^\dagger T)^T &= T^\dagger T, \\
(T T^\dagger)^T &= T T^\dagger,
\end{aligned}
$$

is called the pseudoinverse matrix.

One can show, that the Moore-Penrose pseudoinverse defined in 1.1 is unique while this is not the case for a "standard" pseudoinverse matrix which has to satisfy only the first two conditions. Note that if T is invertible, its pseudoinverse and its inverse are equal. In the following, we are especially interested in the case where T has a full column rank, i. e. the columns of T are linearly independent.

Theorem 1.1 (*Exact formula of pseudoinverse for matrices with full column rank*)

Consider a matrix $T \in \mathbb{R}^{m\times n}$ with full column rank. Then, $(T^T T)$ is invertible and the pseudoinverse $T^\dagger$ of T is given by

$$T^\dagger := (T^T T)^{-1} T^T. \tag{1.8}$$

Proof 1.1

As T has full column rank, we obtain by the rank theorem that $\ker T = \{0\}$. As a consequence $(T^T T)$ is invertible and it is trivial to verify that (1.8) satisfies definition 1.1. ⊠

Note that equation (1.8) is well defined since for T with full column rank the matrix $(T^T T)$ is invertible. Moreover, the proof is trivial since equation (1.8) verifies Definition 1.1.

In order to illustrate the basic principle of the reduction technique, consider the following disordered reactive transport system where all species are mobile

$$\partial_t c + \mathcal{L}(c) = S R(c),$$

where c is the concentration vector of the I species, $\mathcal{L}$ is a linear transport operator, S the stoichiometric matrix and R the vector of reaction speeds.
The first step consists in extracting the non reacting components by combining linearly the equations of the reactive transport system. Suppose $S^\star$ to be the stoichiometric matrix that is a result of reducing the stoichiometric matrix S until full column rank is obtained. $S^\star$ can be obtained by choosing a maximal set of linear independent columns of S. We now search a matrix A such that

$$S = S^\star A. \tag{1.9}$$

The columns of A contain the coefficients needed to reform the stoichiometric matrix S by means of the basis matrix $S^\star$. A is defined in a unique way by the pseudoinverse of $S^\star$. As $S^\star$ has full column rank, theorem 1.1 holds and multiplying equation (1.9) with the pseudoinverse of $S^\star$ yields

$$A := {S^\star}^\dagger S.$$

The reactive transport system can be written as

$$\partial_t c + \mathcal{L}c = S^\star A R(c).$$

The so obtained basis of the reactive system, described by the columns of the matrix $S^\star$, can be completed by a maximal set of vectors which are all orthogonal on this basis. Those vectors form the columns of the so defined matrix $S^\square$. Together, the columns of $S^\star$ and $S^\square$ form a basis of the entire reaction space of all possible stoichiometries, in other words, a basis of $\mathbb{R}^I$. The orthogonal basis vectors that complete the initial basis allow us to define non reacting components.
In practice, one can complete the initial basis defined by $S^\star$ by vectors of the canonical basis of $\mathbb{R}^I$ where I is the number of chemical species. Then, by an orthogonalisation method (Gram-Schmidt for example), one orthogonalises this entire basis respecting the order of the vectors. Finally, one extracts only the last $I - \text{rank}(S^\star)$ nonzero vectors to form the matrix $S^\square$.
Note that, by construction, the following properties hold:

$$S^\star S^\square = 0,$$
$$S^\square S^\star = 0.$$

Now, one can treat the reactive transport system my multiplying it once with ${S^\star}^\dagger$ and once with ${S^\square}^\dagger$:

$${S^\square}^\dagger \partial_t c + {S^\square}^\dagger \mathcal{L}(c) = {S^\square}^\dagger S^\star A R(c),$$
$${S^\star}^\dagger \partial_t c + {S^\star}^\dagger \mathcal{L}(c) = {S^\star}^\dagger S^\star A R(c),$$

where ${S^\square}^\dagger S^\star = ({S^\square}^T S^\square)^{-1} {S^\square}^T S^\star = 0$ (by construction of $S^\square$) and ${S^\star}^\dagger S^\star A = A$ (by property of ${S^\star}^\dagger$). It is useful in the following to define the new species η and ξ as follows:

$$\eta := {S^\square}^\dagger c,$$
$$\xi := {S^\star}^\dagger c.$$

We suppose in the following that the transport operator is linear and equal for all species. Note that this assumption is useful for the sake of readability since then ${S^\square}^\dagger \mathcal{L}(c) = \mathcal{L}(\eta)$ and ${S^\star}^\dagger \mathcal{L}(c) = \mathcal{L}(\xi)$ hold. Nevertheless, this assumption has no influence on the conceptual developments that follow.
The reduced system can now be written as

$$\partial_t \eta + \mathcal{L}(\eta) = 0$$
$$,\partial_t \xi + \mathcal{L}(\xi) = A R(c).$$

Note that one formed nothing else than a linear combination of equations of the initial system. The procedure can be seen as linear transformation for the variable c into new variables (η, ξ) where η are the non reacting components (of size $\text{rank}(S^\square)$) and ξ the reacting components (of size $\text{rank}(S^\star)$). The non reacting components can be seen as invariants within chemical reactions. They behave like physical tracers which do not react at all and are only transported. The difference lies in the fact that possibly all chemical species in the system may react but some of their stoichiometric combinations are invariant in the system.
Note finally, that we write the reaction rate functions $R(c)$ as a function depending on physical variables c and not on the new transformed variables (η, ξ) since the reaction rates are usually defined within physically present species and not mathematically created entities.

1.4.2.2 Treating mobile and fixed species by sub-systems

The previously presented reduction technique allows to decouple non reacting components. As the transport operator vanishes for fixed species, one is held not to form components that melt mobile and fixed species. Up from now, mobile species c are distinguished from the fixed species $\bar{c}$ which are marked by a bar.
Consider a chemical system with I mobile species and $\bar{I}$ fixed species. The species are numbered such that the first I species are mobile and the last $\bar{I}$ species are fixed. According to that $c_{i=1,\dots,I}$ corresponds to the first I rows of S and $\bar{c}_{i=I+1,\dots,I+\bar{I}}$ to the following $\bar{I}$ rows of S. Respecting this order, the stoichiometric matrix can be written as

$$S := \left(\frac{S_1}{S_2}\right), \tag{1.10}$$

where S_1 contains the coefficients for the mobile species and S_2 the coefficients for the fixed species. The corresponding reactive transport system can then be written as

$$\begin{aligned} \partial_t c + \mathcal{L}(c) &= S_1 R(c,\bar{c}), \\ \partial_t \bar{c} \qquad\quad &= S_2 R(c,\bar{c}). \end{aligned} \tag{1.11}$$

In order not to mix mobile and fixed species during the extraction of non reacting components, up from now, one treats the system separately by projecting S into the subspace of mobile species (described by S_1) and into the subspace of fixed species (described by S_2).
Applying the previously presented extraction technique of non reacting components, one obtains the basis matrices $S_i^{\star}$ with their orthogonal supplement $S_i^{\square}$ and the recombination matrices A_i, for $i = 1, 2$. One now multiplies equations (1.11) with $S_i^{\square\dagger}$ and $S_i^{\star\dagger}$. This means that one forms linear combinations of the equations of the reactive transport system without melting mobile and fixed species. To sum up, we form a variable transformation of the unknowns c and $\bar{c}$ to the new variables (η, ξ) and $(\bar{\eta}, \bar{\xi})$. One obtains for the mobile species

$$\begin{aligned} (S_1^{\square T} S_1^{\square})^{-1} S_1^{\square T} \partial_t c \;+\; (S_1^{\square T} S_1^{\square})^{-1} S_1^{\square T} \mathcal{L}(c) &= (S_1^{\square T} S_1^{\square})^{-1} \underbrace{S_1^{\square T} S_1}_{=0} R(c,\bar{c}) \\ (S_1^{\star T} S_1^{\star})^{-1} S_1^{\star T} \partial_t c \;+\; (S_1^{\star T} S_1^{\star})^{-1} S_1^{\star T} \mathcal{L}(c) &= (S_1^{\star T} S_1^{\star})^{-1} S_1^{\star T} S_1 R(c,\bar{c}), \end{aligned}$$

and for the fixed species

$$\begin{aligned} (S_2^{\square T} S_2^{\square})^{-1} S_2^{\square T} \partial_t \bar{c} &= (S_2^{\square T} S_2^{\square})^{-1} \underbrace{S_2^{\square T} S_2}_{=0} R(c,\bar{c}) \\ (S_2^{\star T} S_2^{\star})^{-1} S_2^{\star T} \partial_t \bar{c} &= (S_2^{\star T} S_2^{\star})^{-1} S_2^{\star T} S_2 R(c,\bar{c}). \end{aligned}$$

As before, it is useful to define the new species

$$\begin{aligned} \eta &:= S_1^{\square\dagger} c, \\ \bar{\eta} &:= S_2^{\square\dagger} \bar{c}, \\ \xi &:= S_1^{\star\dagger} c, \\ \bar{\xi} &:= S_2^{\star\dagger} \bar{c}. \end{aligned} \tag{1.12}$$

The transformed system is finally written as

$$\begin{aligned}
\partial_t \eta + \mathcal{L}(\eta) &= 0, \\
\partial_t \bar{\eta} &= 0, \\
\partial_t \xi + \mathcal{L}(\xi) &= A_1 R(c, \bar{c}), \\
\partial_t \bar{\xi} &= A_2 R(c, \bar{c}).
\end{aligned}$$

This systems if formed of two subsystems. The first one is totally decoupled from chemistry and the second one treats the chemical components that are affected by chemistry. The numerical solution of the non reacting components η and $\bar{\eta}$ can be done in advance and the concentration of the components $\bar{\eta}$ is even constant.

1.4.2.3 The general method for mixed kinetics-equilibrium systems

We have seen previously that it is possible to reduce purely kinetic systems. When equilibrium reactions are considered, the reaction rates of the equilibrium reactions should be eliminated to reduce the stiffness of the system. We now describe the entire method used to reduce mixed systems.

Formatting the initial system In order to eliminate the equilibrium reaction rates, we have to group as much as possible those reaction rates.
Therefore, we distinguish equilibrium reactions and kinetic reactions. Kinetic reactions are described by reaction rate models. Equilibrium reactions have reaction rates that are themselves also unknowns and whose solution is determined by an additional algebraic condition. Suppose that we have J_{eq} equilibrium reactions and J_{kin} kinetic reactions. We group the stoichiometric matrix in the following way

$$S := \left(\begin{array}{c|c} S_{1,\text{eq}} & S_{1,\text{kin}} \\ \hline S_{2,\text{eq}} & S_{2,\text{kin}} \end{array} \right), \tag{1.13}$$

where $\begin{pmatrix} S_{1,\text{eq}} \\ \hline S_{2,\text{eq}} \end{pmatrix}$ describes the coefficients for the equilibrium reactions, $\begin{pmatrix} S_{1,\text{kin}} \\ \hline S_{2,\text{kin}} \end{pmatrix}$ describes the coefficients for the kinetic reactions, $\left(\begin{array}{c|c} S_{1,\text{eq}} & S_{1,\text{kin}} \end{array} \right)$ describes the coefficients for the mobile species and finally $\left(\begin{array}{c|c} S_{2,\text{eq}} & S_{2,\text{kin}} \end{array} \right)$ describes the coefficients for the fixed species. Note that the matrix S has possibly not full column rank. In this case, there are two or more linear dependent reactions in the system. In the following, we will represent all reactions by a basis of reactions and condense linear dependent reactions.

Construction of an equilibrium basis In order to eliminate the equilibrium reaction rates, it is necessary that the reaction basis has a certain shape. Indeed, it consists of two subbases. The first subbasis is a basis for all equilibrium reactions. The second subbasis completes the first one

in order to form together a basis for the entire reaction system. By this way, a kinetic reaction is represented by the basis reactions of the kinetic and/or the equilibrium part meanwhile an equilibrium reaction is uniquely represented by the equilibrium reaction basis. This arrangement allows us to decouple a subsystem of equilibrium reactions and a subsystem of purely kinetic reactions.
Moreover, in order to decouple the mobile and fixed part as in section 1.4.2.2, the equilibrium reaction base is supposed to have a special shape such that it can be divided into three subbases:

1. The first subbasis acts only on mobile species: $\left(\frac{S^\star_{1,\mathrm{mob}}}{0}\right)$
2. The second subbasis is heterogeneous, i. e. it acts on both mobile and fixed species: $\left(\frac{S^\star_{1,\mathrm{het}}}{S^\star_{2,\mathrm{het}}}\right)$
3. The third subbasis acts only on fixed species : $\left(\frac{0}{S^\star_{2,\mathrm{immo}}}\right)$

The heterogeneous subbasis is supposed to verify the following two properties:

- Its projection on the mobile species $S^\star_{1,\mathrm{het}}$ is linearly independent from the first subbasis acting only on mobile species $S^\star_{1,\mathrm{mob}}$ and
- its projection on the fixed species $S^\star_{2,\mathrm{het}}$ is linearly independent from the third subbasis acting only on fixed species $S^\star_{2,\mathrm{immo}}$.

To put into a nutshell, the equilibrium basis has the following shape

$$S^\star_{\mathrm{eq}} = \left(\begin{array}{c|c|c} S^\star_{1,\mathrm{mob}} & S^\star_{1,\mathrm{het}} & 0 \\ \hline 0 & S^\star_{2,\mathrm{het}} & S^\star_{2,\mathrm{immo}} \end{array}\right), \tag{1.14}$$

with the columns of $S^\star_{1,\mathrm{mob}}$ being linearly independent from the columns of $S^\star_{1,\mathrm{het}}$ and the columns of $S^\star_{2,\mathrm{het}}$ being linearly independent from the columns of $S^\star_{2,\mathrm{immo}}$ and the additional property that all equilibrium reactions can be uniquely represented by linear combination of this basis of the equilibrium reactions. In practice, based on a stoichiometric equilibrium matrix, the construction of such an equilibrium basis can be proceed by the following substeps: first, pick out from the stoichiometric matrix all vectors that have already a zero projection into the mobile species part and chose a basis of them. In the next step, the columns which have a non-zero projection in the mobile part are treated. By a column-wise Gauß-elimination their projection on the mobile part is reduced to full column rank. Those vectors who now have a zero projection to the mobile part are added to the basis of the immobile part if they are linearly independent of the already existing base, otherwise they are dropped. The sub-basis of immobile reactions is finished. In the last step, the sub-basis of mobile reactions is extracted by reducing them with column-wise Gauß-elimination and grouping them such that mobile reactions and heterogeneous reactions are together. In algorithm 1.3 we give the entire algorithm that allows to reduce an arbitrary stoichiometric matrix to the desired shape in detail.

Algorithm 1.3 Construction of the equilibrium basis (1.14)

Input: $S_{1,\mathrm{eq}}, S_{2,\mathrm{eq}}$

Return: $S^\star_{1,\mathrm{mob}}, S^\star_{1,\mathrm{het}}, S^\star_{2,\mathrm{het}}, S^\star_{2,\mathrm{immo}}$

// Create a matrix of shape $\begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$ *that is a permutation/reduction of* S *with* C *having maximum rank*

for $i = 1, \ldots, J_{\mathrm{eq}}$ **do**
  if $(S_{1,\mathrm{eq}})_{:,i} = 0$ **then**
    if $\operatorname{rank} C < \operatorname{rank}\left(\begin{array}{c|c} C & (S_{2,\mathrm{eq}})_{:,i} \end{array} \right)$ **then**
      $C := \left(\begin{array}{c|c} C & (S_{2,\mathrm{eq}})_{:,i} \end{array} \right)$
    end if
  else
    $A := \left(\begin{array}{c|c} A & (S_{1,\mathrm{eq}})_{:,i} \end{array} \right)$
    $B := \left(\begin{array}{c|c} B & (S_{2,\mathrm{eq}})_{:,i} \end{array} \right)$
  end if
end for

// Create $\begin{pmatrix} A_1 & 0 \\ B_1 & C \end{pmatrix}$ *such that* A_1 *and* C *have both maximum rank*

for $i = 1, \ldots,$ # columns of A **do**
  if $\operatorname{rank} A_1 < \operatorname{rank}\left(\begin{array}{c|c} A_1 & A_{:,i} \end{array} \right)$ **then**
    $A_1 := \left(\begin{array}{c|c} A_1 & A_{:,i} \end{array} \right)$
    $B_1 := \left(\begin{array}{c|c} B_1 & B_{:,i} \end{array} \right)$
  else
    $v := B_{:,i} - B_1 (A_1^T A_1)^{-1} A_1^T A_{:,i}$
    if $\operatorname{rank} C < \operatorname{rank}\left(\begin{array}{c|c} C & v \end{array} \right)$ **then**
      $C := \left(\begin{array}{c|c} C & v \end{array} \right)$
    end if
  end if
end for

// Create the final matrix

$S^\star_{2,\mathrm{immo}} := C$

for $i = 1, \ldots,$ # columns of A_1 **do**
  if $\operatorname{rank}\left(\begin{array}{c|c} S^\star_{2,\mathrm{het}} & S^\star_{2,\mathrm{immo}} \end{array} \right) < \operatorname{rank}\left(\begin{array}{c|c|c} S^\star_{2,\mathrm{het}} & S^\star_{2,\mathrm{immo}} & (B_1)_{:,i} \end{array} \right)$ **then**
    $S^\star_{1,\mathrm{het}} := \left(\begin{array}{c|c} S^\star_{1,\mathrm{het}} & (A_1)_{:,i} \end{array} \right)$
    $S^\star_{2,\mathrm{het}} := \left(\begin{array}{c|c} S^\star_{2,\mathrm{het}} & (B_1)_{:,i} \end{array} \right)$
  else
    $\widetilde{v} := A_{1:,i}$
    $$- \left(\begin{array}{c|c} S^\star_{1,\mathrm{het}} & 0 \end{array} \right) \left(\left(\begin{array}{c|c} S^\star_{2,\mathrm{het}} & S^\star_{2\mathrm{immo}} \end{array} \right)^T \left(\begin{array}{c|c} S^\star_{2,\mathrm{het}} & S^\star_{2\mathrm{immo}} \end{array} \right) \right)^{-1} \left(\begin{array}{c|c} S^\star_{2,\mathrm{het}} & S^\star_{2\mathrm{immo}} \end{array} \right)^T B_{1:,i}$$
    $S^\star_{1,\mathrm{mob}} := \left(\begin{array}{c|c} S^\star_{1,\mathrm{mob}} & \widetilde{v} \end{array} \right)$
  end if
end for

Adding kinetic reactions to the problem In the following, we only consider the projection of the basis matrix into its two subspaces of mobile and fixed species. The fact that we have decoupled the heterogeneous basis of the homogeneous basis allows us, on the one hand, to inherit the independence of the basis vectors even in the projected subspaces of mobile and fixed species, and, on the other hand, to obtain the same number of mobile and fixed components that are formed by heterogeneous reactions.
Due to the developments of the previous paragraph, we suppose that the stoichiometric matrix has the following form

$$S = \left(\begin{array}{c|c|c||c} S^\star_{1,\mathrm{mob}} & S^\star_{1,\mathrm{het}} & 0 & S_{1,\mathrm{kin}} \\ \hline 0 & S^\star_{2,\mathrm{het}} & S^\star_{2,\mathrm{immo}} & S_{2,\mathrm{kin}} \end{array}\right),$$

with the associated reaction rates

$$R = \left(\begin{array}{c|c|c||c} R_{\mathrm{mob}} & R_{\mathrm{het}} & R_{\mathrm{immo}} & R_{\mathrm{kin}} \end{array}\right)^T.$$

Now, one completes the basis matrix of equilibrium reactions to a basis matrix of all the reactions, including kinetics, by still keeping the projection into the subspaces of mobile and fixed species. We complete the matrix $\left(\begin{array}{c|c} S^\star_{1,\mathrm{mob}} & S^\star_{1,\mathrm{het}} \end{array}\right)$ with the maximum number of columns of the matrix $S_{1,\mathrm{kin}}$ such that the result has full column rank. One obtains then

$$S^\star_1 := (\underbrace{S^\star_{1,\mathrm{mob}}}_{\#J_{\mathrm{mob}}} \mid \underbrace{S^\star_{1,\mathrm{het}}}_{\#J_{\mathrm{het}}} \parallel \underbrace{S^\star_{1,\mathrm{kin}}}_{\#J'_{1,\mathrm{kin}}}).$$

In the same way, one completes the projection on the fixed species and obtains

$$S^\star_2 := (\underbrace{S^\star_{2,\mathrm{het}}}_{\#J_{\mathrm{het}}} \mid \underbrace{S^\star_{2,\mathrm{immo}}}_{\#J_{\mathrm{immo}}} \parallel \underbrace{S^\star_{2,\mathrm{kin}}}_{\#J'_{2,\mathrm{kin}}}).$$

Note that the completion with the projections of the columns of the kinetic reactions on the mobile and fixed species is done independently. For this reason, the matrices $S^\star_{1,\mathrm{kin}}$ and $S^\star_{2,\mathrm{kin}}$ will in general not have the same number of columns.

Proposition 1.1 (*Representation of the stoichiometric system*)

The projections of the reactions of the original system can be represented by

$$S_1 = S^\star_1 A_1, \qquad S_2 = S^\star_2 A_2,$$

with the recombination matrices

$$A_1 = \left(\begin{array}{c|c|c||c} \mathrm{Id} & 0 & 0 & A_{1,\mathrm{mob}} \\ \hline 0 & \mathrm{Id} & 0 & A_{1,\mathrm{het}} \\ \hline 0 & 0 & 0 & A_{1,\mathrm{kin}} \end{array}\right), \qquad A_2 = \left(\begin{array}{c|c|c||c} 0 & \mathrm{Id} & 0 & A_{2,\mathrm{het}} \\ \hline 0 & 0 & \mathrm{Id} & A_{2,\mathrm{immo}} \\ \hline 0 & 0 & 0 & A_{2,\mathrm{kin}} \end{array}\right).$$

System reduction by orthogonalisation In the same way as in equation (1.12), we define the new species

$$\begin{aligned}\eta &:= S_1^{\square\dagger} c,\\ \overline{\eta} &:= S_2^{\square\dagger} \overline{c},\\ \xi &:= S_1^{\star\dagger} c,\\ \overline{\xi} &:= S_2^{\star\dagger} \overline{c}.\end{aligned}$$

By using the bloc shape of the matrices $S_1^\star$ and $S_2^\star$, we can split the vectors ξ and $\bar{\xi}$ into $(\xi_{\text{mob}}, \xi_{\text{het}}, \xi_{\text{kin}})$ and $(\bar{\xi}_{\text{het}}, \bar{\xi}_{\text{immo}}, \bar{\xi}_{\text{kin}})$ of sizes $(J_{\text{mob}}, J_{\text{het}}, J'_{1,\text{kin}})$ and $(J_{\text{het}}, J_{\text{immo}}, J'_{2,\text{kin}})$, respectively. We obtain finally the reduced system:

$$\begin{aligned}
\partial_t \eta &+ \mathcal{L}(\eta) &&= 0, && && (1.15a)\\
\partial_t \bar{\eta} &&&= 0, && && (1.15b)\\
\partial_t \xi_{\text{mob}} &+ \mathcal{L}(\xi_{\text{mob}}) &&= R_{\text{mob}} &+& A_{1,\text{mob}} R_{\text{kin}}(c, \bar{c}), && (1.15c)\\
\partial_t \xi_{\text{het}} &+ \mathcal{L}(\xi_{\text{het}}) &&= R_{\text{het}} &+& A_{1,\text{het}} R_{\text{kin}}(c, \bar{c}), && (1.15d)\\
\partial_t \xi_{\text{kin}} &+ \mathcal{L}(\xi_{\text{kin}}) &&= &+& A_{1,\text{kin}} R_{\text{kin}}(c, \bar{c}), && (1.15e)\\
\partial_t \bar{\xi}_{\text{het}} &&&= R_{\text{het}} &+& A_{2,\text{het}} R_{\text{kin}}(c, \bar{c}), && (1.15f)\\
\partial_t \bar{\xi}_{\text{immo}} &&&= R_{\text{immo}} &+& A_{2,\text{immo}} R_{\text{kin}}(c, \bar{c}), && (1.15g)\\
\partial_t \bar{\xi}_{\text{kin}} &&&= &+& A_{2,\text{kin}} R_{\text{kin}}(c, \bar{c}). && (1.15h)
\end{aligned}$$

Elimination of the equilibrium reaction rates One the one hand, one is interested in eliminating the equilibrium reaction rates. On the other hand, one has to be able to integrate the equilibrium constraints imposed by equilibrium equations into the system. In order to do so, the equilibrium reaction rates have to appear in exactly $J_{\text{mob}} + J_{\text{het}} + J_{\text{immo}}$ equations. In system (1.15) they appear in $J_{\text{mob}} + 2 \cdot J_{\text{het}} + J_{\text{immo}}$ equations. One can reduce their appearance by forming, for the first time ever, heterogeneous components of mobile and fixed species. This is done by replacing equation (1.15d) by ((1.15d) − (1.15f)):

$$\begin{aligned}
\partial_t \eta &+ \mathcal{L}(\eta) &&= 0, && && (1.16a)\\
\partial_t \bar{\eta} &&&= 0, && && (1.16b)\\
\partial_t \xi_{\text{mob}} &+ \mathcal{L}(\xi_{\text{mob}}) &&= R_{\text{mob}} &+& A_{1,\text{mob}} R_{\text{kin}}(c, \bar{c}), && (1.16c)\\
\partial_t (\xi_{\text{het}} - \bar{\xi}_{\text{het}}) &+ \mathcal{L}(\xi_{\text{het}}) &&= &+& (A_{1,\text{het}} - A_{2,\text{het}}) R_{\text{kin}}(c, \bar{c}), && (1.16d)\\
\partial_t \xi_{\text{kin}} &+ \mathcal{L}(\xi_{\text{kin}}) &&= &+& A_{1,\text{kin}} R_{\text{kin}}(c, \bar{c}), && (1.16e)\\
\partial_t \bar{\xi}_{\text{het}} &&&= R_{\text{het}} &+& A_{2,\text{het}} R_{\text{kin}}(c, \bar{c}), && (1.16f)\\
\partial_t \bar{\xi}_{\text{immo}} &&&= R_{\text{immo}} &+& A_{2,\text{immo}} R_{\text{kin}}(c, \bar{c}), && (1.16g)\\
\partial_t \bar{\xi}_{\text{kin}} &&&= &+& A_{2,\text{kin}} R_{\text{kin}}(c, \bar{c}). && (1.16h)
\end{aligned}$$

Recalling that for every equilibrium reaction that is described by by the equilibrium reaction basis $S_{\text{eq}}^\star$ a mass action law is available, one obtains therefore $J_{\text{mob}} + J_{\text{het}} + J_{\text{immo}}$ nonlinear additional

equations. At the same time, one has $J_{\text{mob}} + J_{\text{het}} + J_{\text{immo}}$ equations (1.16c), (1.16f) and (1.16g) for the concentrations and the equilibrium reaction rates. One can now replace the equations on the concentrations and the equilibrium reaction rates by the same number of mass action law equations. By this way, one eliminates the equilibrium reaction rates. The mass action laws are described by the equilibrium reactions of the basis of the equilibrium reactions in (1.14). Together with this mass action laws, one can state the entirely reduced system by

$$\partial_t \eta + \mathcal{L}(\eta) = 0, \tag{1.17a}$$
$$\partial_t \bar{\eta} = 0, \tag{1.17b}$$
$$\partial_t(\xi_{\text{het}} - \bar{\xi}_{\text{het}}) + \mathcal{L}(\xi_{\text{het}}) = (A_{1,\text{het}} - A_{2,\text{het}})R_{\text{kin}}(c, \bar{c}), \tag{1.17c}$$
$$\partial_t \xi_{\text{kin}} + \mathcal{L}(\xi_{\text{kin}}) = A_{1,\text{kin}} R_{\text{kin}}(c, \bar{c}), \tag{1.17d}$$
$$\partial_t \bar{\xi}_{\text{kin}} = A_{2,\text{kin}} R_{\text{kin}}(c, \bar{c}), \tag{1.17e}$$
$$Q(c, \bar{c}) = 0, \tag{1.17f}$$

where $Q(c, \bar{c})$ describes the mass action laws.

1.4.3 Example: Hard Test Case of the GDR MoMaS Reactive Transport Benchmark

The MoMaS reactive transport benchmark ([8]) provides three different chemical test cases with different degrees of difficulty. We apply the previously presented reduction technique to the most difficult case of this test series as it has been treated only by one of the eleven participating working groups (cf. [63]).

The equilibrium reactions of the chemical system are described in a Morel tableau. In this framework, a certain number of primary species (originally called components) is already chosen (denoted X_i for mobile and S for the only primary fixed species). The underlying representation principle supposes that every secondary species (called C_j for mobile and CS_k for fixed secondary species) can be represented in a unique way by one chemical reaction within the primary species. The stoichiometric indices of the underlying reactions are noted in the rows of the Morel tableau. In the last column of every row, one notes the equilibrium constant of the associated equilibrium reaction.
Morel described in his book about aquatic chemistry (cf. [67]) one schematic way to describe closed chemical systems. His developments are based on the ideas of Gibbs (cf. [37]), who has defined the "components" as those entities with which a chemical system and every possible change of its configuration can be entirely defined. Note that Gibbs allowed also to use ions and electrical charges to be components. However, Morel proposed a procedure to select components only within chemical species to define the system and not to mix up different kind of conservation principles (mass, electrical charges, ...). The associated manipulations and reduced representations of chemical equilibrium systems are based on principles of linear algebra such as basis reduction. The link to the works of Kräutle et al. (cf. [52] and associated articles) is

the following: while Morel allowed only chemical species as components, Kräutle extended the ideas of Gibbs to use different entities describing the chemical system, namely linear combinations of chemical species, together with the linear algebra formulations of Morel to represent the chemical system. Moreover, he used not only the basis reduction principle but also basis transformation as tools of the linear algebra.

The Morel tableau for the hard test case is given by:

	X_1	X_2	X_3	X_4	X_5	S	K
C_1	0	−1	0	0	0	0	10^{-12}
C_2	0	1	1	0	0	0	1
C_3	0	−1	0	1	0	0	1
C_4	0	−4	1	3	0	0	10^{-1}
C_5	0	4	3	1	0	0	10^{35}
C_6	0	10	3	0	0	0	10^{32}
C_7	0	−8	0	2	0	0	10^{-4}
CS_1	0	3	1	0	0	1	10^{6}
CS_2	0	−3	0	1	0	2	10^{-1}
CP_1	0	3	1	0	0	0	$8 \cdot 10^{10}$
CP_2	0	1	0	0	1	0	20
Cc	0	−3	0	1	0	0	

Besides the equilibrium reactions, a fixed species Cc is formed within the heterogeneous reaction

$$3C_3 \leftrightarrows Cc + 2X_4.$$

This reaction is described using the following kinetic formulation

$$\frac{dCc}{dt} = \left(0.2\frac{C_3^3}{X_4^2} - 1\right)k \quad \text{with } Cc \geq 0 \text{ and } k = \begin{cases} 10^{-2} & \text{if } 0.2\frac{C_3^3}{X_4^2} \geq 1 \\ 10 & \text{else} \end{cases}.$$

Species X_5 is unstable and dissociates to

$$X_5 \rightarrow 2X_2 + X_3.$$

The reaction rate for this homogeneous reaction depends on the concentration of CP_2

$$\frac{dX_5}{dt} = 0.05X_5 + 5CP_2.$$

In order to obtain a reduced and decoupled system, we apply the previously described method and obtain the following system:

$$\begin{aligned}
\partial_t \eta &+ \mathcal{L}(\eta) &&= 0, \\
\partial_t \overline{\eta} & && = 0, \\
\partial_t (\xi_{\text{het}} - \overline{\xi_{\text{het}}}) &+ \mathcal{L}(\xi_{\text{het}}) && = (A_{1,\text{het}} - A_{2,\text{het}}) R_{\text{kin}}(c, \overline{c}), \\
\partial_t \xi_{\text{kin}} &+ \mathcal{L}(\xi_{\text{kin}}) && = A_{1,\text{kin}} R_{\text{kin}}(c, \bar{c}), \\
\partial_t \overline{\xi_{\text{kin}}} & && = A_{2,\text{kin}} R_{\text{kin}}(c, \bar{c}), \\
& && Q(c, \overline{c}) = 0,
\end{aligned}$$

with the following chemical components

$$\begin{aligned}
\eta &= (\eta_1, \eta_2)^T, \\
\overline{\eta} &= \overline{\eta_1}, \\
\xi_{\text{het}} &= (\xi_{\text{het1}}, \xi_{\text{het2}})^T, \\
\overline{\xi_{\text{het}}} &= (\overline{\xi_{\text{het1}}}, \overline{\xi_{\text{het2}}})^T, \\
\xi_{\text{kin}} &= (\xi_{kin1}), \\
\overline{\xi_{\text{kin}}} &= (\overline{\xi_{kin1}}).
\end{aligned}$$

$Q(c, \bar{c})$ denotes the equilibrium conditions for the equilibrium reactions.

1.4.4 Extension: Extraction of Components Influenced by Slow Kinetic Laws

Supposing a certain knowledge of the orders of magnitude of the kinetic reaction rates (which can be not always easy, cf. the two kinetic reactions of the MoMaS hard test case in section 1.4.3), one is then able to extract a group of kinetic components that are conquered only by slow kinetics. We use the ideas of Kräutle presented in section 1.4.2.1 and apply them in an extended way on the kinetic reaction rates in order to group them.
Referring to the reduced system (1.17), one may now rewrite equations (1.17c), (1.17d) and (1.17e) in a vectorial form

$$\partial_t \begin{pmatrix} \xi_{\text{het}} - \bar{\xi}_{\text{het}} \\ \xi_{\text{kin}} \\ \bar{\xi}_{\text{kin}} \end{pmatrix} + \mathcal{L} \begin{pmatrix} \xi_{\text{het}} \\ \xi_{\text{kin}} \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} A_{1,\text{het}} - A_{2,\text{het}} \\ A_{1,\text{kin}} \\ A_{2,\text{kin}} \end{pmatrix}}_{:=A} R_{\text{kin}}(c, \bar{c}).$$

One permutes now $R_{\text{kin}}(c,\bar{c})$ such that the first $i \geq 1$ components of the vector belong to slow kinetics and the last $j \geq 1$ components to fast kinetics. Permuting the matrix A in the same way gives

$$\partial_t \begin{pmatrix} \xi_{\text{het}} - \bar{\xi}_{\text{het}} \\ \xi_{\text{kin}} \\ \bar{\xi}_{\text{kin}} \end{pmatrix} + \mathcal{L} \begin{pmatrix} \xi_{\text{het}} \\ \xi_{\text{kin}} \\ 0 \end{pmatrix} = \left(\begin{array}{c|c} A^{\text{fast}} & A^{\text{slow}} \end{array} \right) \left(\begin{array}{c} R^{\text{fast}}_{\text{kin}}(c,\bar{c}) \\ \hdashline R^{\text{slow}}_{\text{kin}}(c,\bar{c}) \end{array} \right),$$

which is equivalent to

$$\partial_t \begin{pmatrix} \xi_{\text{het}} - \bar{\xi}_{\text{het}} \\ \xi_{\text{kin}} \\ \bar{\xi}_{\text{kin}} \end{pmatrix} + \mathcal{L} \begin{pmatrix} \xi_{\text{het}} \\ \xi_{\text{kin}} \\ 0 \end{pmatrix} = A^{\text{fast}} R^{\text{fast}}_{\text{kin}}(c,\bar{c}) + A^{\text{slow}} R^{\text{slow}}_{\text{kin}}(c,\bar{c}). \tag{1.18}$$

As the matrix A^{fast} has possibly not full column rank, one might, referring to chapter 1.4.2.1, extract a full rank matrix $A^{\star\text{fast}}$ and give a recomposition matrix B^{fast} such that

$$A^{\text{fast}} = A^{\star\text{fast}} B^{\text{fast}}.$$

As the matrix $A^{\star\text{fast}}$ has a full full column rank which is lower than its number of rows, one can now complete its column space by a orthogonal space described in the columns of $A^{\square\text{fast}}$ with the property

$$A^{\square\text{fast}} A^{\star\text{fast}} = 0.$$

One forms now linear combinations of equation (1.18) by multiplying it once with $A^{\square\text{fast}\dagger}$ and once with $A^{\square\text{fast}\star}$ which holds

$$A^{\square\text{fast}\dagger} \left(\partial_t \begin{pmatrix} \xi_{\text{het}} - \bar{\xi}_{\text{het}} \\ \xi_{\text{kin}} \\ \bar{\xi}_{\text{kin}} \end{pmatrix} + \mathcal{L} \begin{pmatrix} \xi_{\text{het}} \\ \xi_{\text{kin}} \\ 0 \end{pmatrix} \right) = A^{\square\text{fast}\dagger} A^{\text{slow}} R^{\text{slow}}_{\text{kin}}(c,\bar{c}), \tag{1.19a}$$

$$A^{\star\text{fast}\dagger} \left(\partial_t \begin{pmatrix} \xi_{\text{het}} - \bar{\xi}_{\text{het}} \\ \xi_{\text{kin}} \\ \bar{\xi}_{\text{kin}} \end{pmatrix} + \mathcal{L} \begin{pmatrix} \xi_{\text{het}} \\ \xi_{\text{kin}} \\ 0 \end{pmatrix} \right) = B^{\text{fast}} R^{\text{fast}}_{\text{kin}}(c,\bar{c}) + A^{\star\text{fast}\dagger} A^{\text{slow}} R^{\text{slow}}_{\text{kin}}(c,\bar{c}). \tag{1.19b}$$

In equation (1.19a) we have thus created chemical components which are only conquered by slow kinetics. They can now be treated numerically in a different way, using larger time steps in a waveform context may be possible for instance. The chemical components defined in (1.19b) are conquered by both slow and fast kinetics.

1.4.5 Chemical Subproblems in the Context of the Numerical Formulation

The reduced system (1.17) can easily be used with the numerical formulation presented in section 1.2. We show now with the help of lumping and delumping operators one way to combine both reduction technique and numerical formulation.

1.4.5.1 Lumping/Delumping

In the following, we will make use of the following variable transformation proposed by Kräutle (cf. equations (2.24) and (2.37) in [52]):

$$\begin{aligned}
\eta &:= (S_1^{\square T} S_1^{\square})^{-1} S_1^{\square T} c, \\
\overline{\eta} &:= (S_2^{\square T} S_2^{\square})^{-1} S_2^{\square T} \overline{c}, \\
\xi &:= (S_1^{\star T} S_1^{\star})^{-1} S_1^{\star T} c, \\
\overline{\xi} &:= (S_2^{\star T} S_2^{\star})^{-1} S_2^{\star T} \overline{c},
\end{aligned}$$

$$\begin{aligned}
c &= S_1^{\star}\xi + S_1^{\square}\eta = S_{1,\mathrm{mob}}^{\star}\xi_{\mathrm{mob}} + S_{1,\mathrm{het}}^{\star}\xi_{\mathrm{het}} + S_{1,\mathrm{kin}}^{\star}\xi_{\mathrm{kin}} + S_1^{\square}\eta, \\
\overline{c} &= S_2^{\star}\overline{\xi} + S_2^{\square}\overline{\eta} = S_{2,\mathrm{het}}^{\star}\overline{\xi_{\mathrm{het}}} + S_{2,\mathrm{immo}}^{\star}\overline{\xi_{\mathrm{immo}}} + S_{2,\mathrm{kin}}^{\star}\overline{\xi_{\mathrm{kin}}} + S_2^{\square}\overline{\eta},
\end{aligned}$$

This variable transformation gives rise to the definition of two operators, namely the lumping and the delumping operator. The lumping operator realises a transformation up from the "species space" described by c, $\overline{c}$ to the "component space" described by η, $\overline{\eta}$, ξ, $\overline{\xi}$.

Definition 1.2 (*Lumping Operator*)

The linear lumping operator is defined as

$$A_{\mathrm{Lumping}} := \begin{pmatrix} S_1^{\square\dagger} & 0 \\ S_1^{\star\dagger} & 0 \\ 0 & S_2^{\square\dagger} \\ 0 & S_2^{\star\dagger} \end{pmatrix}. \tag{1.20}$$

The delumping operator realises the inverse transformation up from the "component space" η, $\overline{\eta}$, ξ, $\overline{\xi}$ to the "species space" c, $\overline{c}$.

Definition 1.3 (*Delumping Operator*)

The linear delumping operator is defined as

$$\begin{aligned}
A_{\mathrm{Delumping}} &:= \begin{pmatrix} S_1^{\square} & S_1^{\star} & 0 & 0 \\ 0 & 0 & S_2^{\square} & S_2^{\star} \end{pmatrix} \\
&= \begin{pmatrix} S_1^{\square} & S_{1,\mathrm{mob}}^{\star} & S_{1,\mathrm{het}}^{\star} & S_{1,\mathrm{kin}}^{\star} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_2^{\square} & S_{2,\mathrm{het}}^{\star} & S_{2,\mathrm{immo}}^{\star} & S_{2,\mathrm{kin}}^{\star} \end{pmatrix}.
\end{aligned} \tag{1.21}$$

Now, the variable transformation can be expressed as

$$\begin{pmatrix} \eta \\ \xi \\ \overline{\eta} \\ \overline{\xi} \end{pmatrix} = \begin{pmatrix} \eta \\ \xi_{\text{mob}} \\ \xi_{\text{het}} \\ \xi_{\text{kin}} \\ \overline{\eta} \\ \overline{\xi_{\text{het}}} \\ \overline{\xi_{\text{immo}}} \\ \overline{\xi_{\text{kin}}} \end{pmatrix} = A_{\text{Lumping}} \begin{pmatrix} c \\ \overline{c} \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} c \\ \overline{c} \end{pmatrix} = A_{\text{Delumping}} \begin{pmatrix} \eta \\ \xi \\ \overline{\eta} \\ \overline{\xi} \end{pmatrix}.$$

It is now clear that a formulation of a function in variables of the component space or in variables of the species space is equivalent since the variable transformation is linear and invertible. The transformation between the two formulations is given in the next section in the case of the equilibrium conditions.

1.4.5.2 Equilibrium Conditions

By means of the lumping and delumping operator we are now able to state the equilibrium conditions $Q(c, \overline{c})$ that interact in the reactive transport problem (1.17f) in component space variables.
We first write the equilibrium equations in a logarithmic form

$$\ln K_{\text{eq}} + {S^{\star}_{\text{eq}}}^{T} \ln\left(\frac{c}{\overline{c}}\right) = 0.$$

Applying the lumping operator (1.20) holds

$$\ln K_{\text{eq}} + {S^{\star}_{\text{eq}}}^{T} \ln \begin{pmatrix} S^{\star}_{1,\text{mob}} \xi_{\text{mob}} + S^{\star}_{1,\text{het}} \xi_{\text{het}} + S^{\star}_{1,\text{kin}} \xi_{\text{kin}} + S^{\square}_{1} \eta \\ S^{\star}_{2,\text{het}} \overline{\xi_{\text{het}}} + S^{\star}_{2,\text{immo}} \overline{\xi_{\text{immo}}} + S^{\star}_{2,\text{kin}} \overline{\xi_{\text{kin}}} + S^{\square}_{2} \overline{\eta} \end{pmatrix} = 0. \tag{1.22}$$

This defines the equilibrium conditions $\widetilde{Q}(\eta, \xi, \overline{\eta}, \overline{\xi}) = 0$ in the transformed variables.

1.4.5.3 New global variables T, W, C, F

After the transformation of the problem variables, we follow now the developments of [5] and define a reactive transport problem in the global variables T, for the total concentration of the mobile and heterogeneous components with C being their mobile part and F being their fixed part and finally W the total concentration of the fixed components. The global variables are defined by

$$T := \underbrace{\begin{pmatrix} \eta \\ \xi_{\text{kin}} \\ \xi_{\text{het}} \end{pmatrix}}_{:=C} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ -\overline{\xi_{\text{het}}} \end{pmatrix}}_{:=F}, \qquad W := \begin{pmatrix} \overline{\eta} \\ \overline{\xi_{\text{kin}}} \end{pmatrix}. \tag{1.23}$$

Problem (1.17) is then equivalent to (1.3). In order to finish the problem transformation, we need to give the exact definition of the functions $\Psi(T, W)$, $\Theta(T, W)$ and $\Upsilon(T, W)$.

1.4.5.4 Chemical Flash *H*

We define now the chemical flash H. Starting from the definition of the total concentrations in equation (1.23) and the mass action laws in equation (1.22), we write now the equations for the chemical flash:

$$\begin{pmatrix} \begin{pmatrix} \eta \\ \xi_{\text{kin}} \\ \xi_{\text{het}} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\overline{\xi_{\text{het}}} \end{pmatrix} - T \\ \begin{pmatrix} \overline{\eta} \\ \overline{\xi_{\text{kin}}} \end{pmatrix} - W \\ \ln K_{\text{eq}} + {S^{\star}_{\text{eq}}}^{T} \ln \begin{pmatrix} S^{\star}_{1,\text{mob}} \xi_{\text{mob}} + S^{\star}_{1,\text{het}} \xi_{\text{het}} + S^{\star}_{1,\text{kin}} \xi_{\text{kin}} + S^{\square}_{1} \eta \\ S^{\star}_{2,\text{het}} \overline{\xi_{\text{het}}} + S^{\star}_{2,\text{immo}} \overline{\xi_{\text{immo}}} + S^{\star}_{2,\text{kin}} \overline{\xi_{\text{kin}}} + S^{\square}_{2} \overline{\eta} \end{pmatrix} \end{pmatrix} = 0. \tag{1.24}$$

The solution of problem (1.24) gives the individual concentrations of all components for a given pair (T, W).
Up from now, we can write (1.24) in matrix form in the mass action law part

$$\begin{pmatrix} \begin{pmatrix} \eta \\ \xi_{\text{kin}} \\ \xi_{\text{het}} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\overline{\xi_{\text{het}}} \end{pmatrix} - T \\ \begin{pmatrix} \overline{\eta} \\ \overline{\xi_{\text{kin}}} \end{pmatrix} - W \\ \ln K_{\text{eq}} + {S^{\star}_{\text{eq}}}^{T} \ln \begin{pmatrix} S^{\star}_{1,\text{mob}} \xi_{\text{mob}} + \begin{pmatrix} S^{\square}_{1} & S^{\star}_{1,\text{kin}} & S^{\star}_{1,\text{het}} \end{pmatrix} \begin{pmatrix} \eta \\ \xi_{\text{kin}} \\ \xi_{\text{het}} \end{pmatrix} \\ S^{\star}_{2,\text{het}} \overline{\xi_{\text{het}}} + S^{\star}_{2,\text{immo}} \overline{\xi_{\text{immo}}} + \begin{pmatrix} S^{\square}_{2} & S^{\star}_{2,\text{kin}} \end{pmatrix} \begin{pmatrix} \overline{\eta} \\ \overline{\xi_{\text{kin}}} \end{pmatrix} \end{pmatrix} \end{pmatrix} = 0. \tag{1.25}$$

For the solution of (1.25) one may easily eliminate the variables $\begin{pmatrix} \eta & \xi_{\text{kin}} & \xi_{\text{het}} \end{pmatrix}^{T}$ and $\begin{pmatrix} \overline{\eta} & \overline{\xi_{\text{kin}}} \end{pmatrix}^{T}$ by replacing them by the mass conservations via the total concentrations T and W. We define the function H by

$$H \begin{pmatrix} \xi_{\text{mob}} \\ \overline{\xi_{\text{immo}}} \\ \overline{\xi_{\text{het}}} \end{pmatrix} := \ln K_{\text{eq}} + {S^{\star}_{\text{eq}}}^{T} \ln \begin{pmatrix} S^{\star}_{1,\text{mob}} \xi_{\text{mob}} + \begin{pmatrix} S^{\square}_{1} & S^{\star}_{1,\text{kin}} & S^{\star}_{1,\text{het}} \end{pmatrix} \left[T + \begin{pmatrix} 0 \\ 0 \\ \overline{\xi_{\text{het}}} \end{pmatrix} \right] \\ S^{\star}_{2,\text{het}} \overline{\xi_{\text{het}}} + S^{\star}_{2,\text{immo}} \overline{\xi_{\text{immo}}} + \begin{pmatrix} S^{\square}_{2} & S^{\star}_{2,\text{kin}} \end{pmatrix} W \end{pmatrix}. \tag{1.26}$$

The chemical equilibrium problem is now entirely defined by finding the zero of the function H

$$H\begin{pmatrix} \xi^*_{\text{mob}} \\ \overline{\xi^*_{\text{immo}}} \\ \overline{\xi^*_{\text{het}}} \end{pmatrix} \stackrel{!}{=} 0. \tag{1.27}$$

1.4.5.5 Equilibrium Operator $\Psi(T, W)$

By means of the solution of the chemical flash $H^{-1}(0)$, the equilibrium operator writes now

$$\Psi(T,W) = F^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\mathbf{Id} \end{pmatrix} H^{-1}(0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\mathbf{Id} \end{pmatrix} \begin{pmatrix} \xi_{\text{mob}}{}^* \\ \overline{\xi_{\text{immo}}}^* \\ \overline{\xi_{\text{het}}}^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\overline{\xi_{\text{het}}}^* \end{pmatrix},$$

where $H^{-1}(0)$ depends only on the parameters T and W. The asterisk attached to a variable means that this value is obtained by solving the chemical equilibrium problem (1.27).

As the operator Ψ interacts in the formulation of the nonlinear reactive transport problem, it is necessary to calculate its derivative in order to be able to apply a Newton-type method with analytical Jacobian. The derivative can easily be derived by means of the implicit function theorem.

$$\begin{aligned} \frac{\partial H}{\partial (T,W)} &= \begin{pmatrix} \frac{\partial H}{\partial \xi_{\text{mob}}{}^*} & \frac{\partial H}{\partial \overline{\xi_{\text{immo}}}^*} & \frac{\partial H}{\partial \overline{\xi_{\text{het}}}^*} \end{pmatrix} \begin{pmatrix} \frac{\partial \xi_{\text{mob}}{}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{immo}}}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{het}}}^*}{\partial (T,W)} \end{pmatrix} + \begin{pmatrix} \frac{\partial H}{\partial T} & \frac{\partial H}{\partial W} \end{pmatrix} \begin{pmatrix} \frac{\partial T}{\partial T} & \frac{\partial T}{\partial W} \\ \frac{\partial W}{\partial T} & \frac{\partial W}{\partial W} \end{pmatrix} \\ &= H' \begin{pmatrix} \frac{\partial \xi_{\text{mob}}{}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{immo}}}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{het}}}^*}{\partial (T,W)} \end{pmatrix} + \begin{pmatrix} \frac{\partial H}{\partial T} & \frac{\partial H}{\partial W} \end{pmatrix} \begin{pmatrix} \mathbf{Id} & 0 \\ 0 & \mathbf{Id} \end{pmatrix} = H' \begin{pmatrix} \frac{\partial \xi_{\text{mob}}{}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{immo}}}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{het}}}^*}{\partial (T,W)} \end{pmatrix} + \begin{pmatrix} \frac{\partial H}{\partial T} & \frac{\partial H}{\partial W} \end{pmatrix} \end{aligned}$$

There exists a neighbourhood of the equilibrium point where the following relation holds

$$\begin{pmatrix} \frac{\partial \xi_{\text{mob}}{}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{immo}}}^*}{\partial (T,W)} \\ \frac{\partial \overline{\xi_{\text{het}}}^*}{\partial (T,W)} \end{pmatrix} = -\left(H' \begin{pmatrix} \xi_{\text{mob}}{}^* \\ \overline{\xi_{\text{immo}}}^* \\ \overline{\xi_{\text{het}}}^* \end{pmatrix} \right)^{-1} \begin{pmatrix} \frac{\partial H}{\partial T} & \frac{\partial H}{\partial W} \end{pmatrix}.$$

The derivative of the operator Ψ writes then

$$\frac{\partial \Psi(T,W)}{\partial (T,W)} = \begin{pmatrix} 0 \\ 0 \\ -\frac{\partial \overline{\xi_{\text{het}}}^*(T,W)}{\partial (T,W)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \left(H' \begin{pmatrix} {\xi_{\text{mob}}}^* \\ {\overline{\xi_{\text{immo}}}}^* \\ \overline{\xi_{\text{het}}}^* \end{pmatrix} \right)^{-1} \begin{pmatrix} \frac{\partial H}{\partial T} & \frac{\partial H}{\partial W} \end{pmatrix} \end{pmatrix}.$$

1.4.5.6 Kinetic reaction rates $\Theta(T,W)$, $\Upsilon(T,W)$

Given a formula for the reaction rates $R_{\text{kin}}(c, \overline{c})$, one can directly deduce a formula for the operators Θ and Υ:

$$\Theta(T,W) = \begin{pmatrix} 0 \\ A_{1,\text{kin}} \\ A_{1,\text{het}} - A_{2,\text{het}} \end{pmatrix} R_{\text{kin}}(c, \overline{c}),$$

$$\Upsilon(T,W) = \begin{pmatrix} 0 \\ A_{2,\text{kin}} \end{pmatrix} R_{\text{kin}}(c, \overline{c}).$$

The transition form the variable space T, W to the variable space c, $\overline{c}$ is realised in the following way: the solution of (1.26) for a given pair of (T,W) allows to find the variables ξ_{mob}, $\overline{\xi_{\text{immo}}}$ and $\overline{\xi_{\text{het}}}$. The variables η, ξ_{kin}, ξ_{het} and $\overline{\eta}$, $\overline{\xi_{\text{kin}}}$ are now obtained by the mass conservation equations using the information of T and W. Finally, the application of the delumping operator (1.21) allows now the use of the variables c and $\overline{c}$.

Conclusion

In this first chapter, we have attempted to introduce on a general and didactic way the formulations and numerical approaches for a reactive transport problem. We supposed only one mobile phase in our model. The stated system treats several mobile and fixed species that can react between each other by different types of reactions (equilibrium and kinetic).
The reactive transport system contains several nested subproblems that are coupled to each other. In order to state a numerical formulation, we defined a general interface for coupled multispecies reactive transport problems that allows the users/software engineers to keep a general viewpoint of the coupled problem without struggling with the details of the nested subproblems. Anyway we developed the detailed relationships between the subproblems and the global formulation for implementation purpose. Due to the comprehensive formulation on the global interface level, the numerical formulation of a global and a splitting approach are easy to state.

The straight through modelling of the reactive transport system provides a clear vision the phenomena but implies a great number of coupled unknowns and equations. Meanwhile, it is possible to reformulate the system in a condensed way without losing its character. For this reason, we applied in the second part of this chapter a reduction technique that minimises the number of unknowns and eliminate/concentrate the coupling terms introduced by kinetic and equilibrium reactions as much as possible. The reduction technique that we applied has two main advantages: first, it reduces the chemical system to an optimal size, i. e. the number of resulting equations is optimal in terms of size. Second, the technique is suitable for very general systems without imposing additional conditions on the underlying chemical reaction system, a very strong point compared to many other reduction techniques. For this reason, we can finally present the relationship between the totally reduces system and our numerical formulation presented in the previous part of this chapter.

The challenge of this very first part was to introduce a problem that includes all the phenomena we want to take into account and that is also suitable for a numerical approach. The idea was to proceed a top-down strategy. First define a model that is easy to state but possibly useless for numerical use. Then, propose a numerical use of the model in order to keep the details as much hidden as possible but as visible as necessary. The resulting application of the numerical approaches is straight through and can be understood without knowing the reduction technique presented afterwards. Finally, presented the reduction technique and showed how it can be used in the context of the numerical interface.
A huge advantage of this strategy is that, at every step of this chapter, one can rise up to the didactic and general model stated at the first section, in order to understand the real phenomena visible for human's eyes. Inversely, the more one advances in the chapter, the closer one approaches to a final numerical use in real world codes.

In a global context, this first chapter allowed to prepare the following chapters as it provides not only the main model and its numerical formulation and approach that we will use in the application of domain decomposition methods to multi-species reactive transport problems but also several useful results: first, we have seen that, after reducing a general reactive transport problem, the most interesting and challenging equations in this context are equations of mobile and fixed species that are coupled by chemical reaction. This fact will invite us to study such type of equations on a more detailed level in chapter 4 for linear and in chapter 5 for nonlinear coupling terms in the context of domain decomposition. Second, we have seen that, for a performing numerical approach of reactive transport models, many work has to be put in the modelling phase in order not to collapse afterwards: the most sophisticated high-level algorithms will failure when the subsidiary low-level models and procedures do.

Divide et impera.

Niccolò Machiavelli

2

Schwarz Type Domain Decomposition

Contents

Introduction

Schwarz type domain decomposition is a class of domain decomposition methods that are based on the original idea of H. A. Schwarz's proof of the Dirichlet principle for nonstandard geometrical domains established in 1870 (cf. [75]). He used a divide and conquer technique together with the maximum principle to show the existence of a regular solution of the heat equation on a geometric domain composed of a circle overlapping with a square. While the problem was unsolvable by a direct method at the time, he split it in two easier subproblems where the solution was already known to exist. By an iterative process and the maximum principle he was able to design a converging sequence in the subdomains. The limit was called the solution of the original problem.
Originally established as a method of proof, the process has been criticised as being nonconstructive: on the one hand, the existence of the solution of the original problem was proven, but, on the other hand, it is not possible to access this limit in an analytical way. Note that at the time, the concept of numerical methods and solutions in the way we have today did not exist. It was perhaps for this reason that the paper felt down in a over one hundred year long hibernation until it has been awoken in the late 1980s by P.-L. Lions when he rediscovered it as an indeed constructive method in the numerical context by studying it under several points of view. Since then, not only the class of Schwarz-type methods, but domain decomposition methods in general have been developed intensively and provide today many high performing algorithms for different classes of problems.

In this chapter, we try to retrace the class of Schwarz type domain decomposition and its derivatives with a special emphasis on the geometrical viewpoint. We found orientation in the article of Gander [33] where Schwarz methods are presented and discussed over the course of time. Besides the geometrical domain decomposition methods, there is a huge class of algebraic domain decomposition methods. Without any doubt, the link between both classes of methods is highly visible but we do not venture to establish an entire state of the art work of all so far developed domain decomposition methods. For a general introduction to the different classes of domain decomposition methods, we refer to the books of Quarteroni and Valli [70], Toselli and Widlund [77] and finally Smith, Bjørstad and Gropp [76].
In the first section we will describe the classical Schwarz domain decomposition method, both in its alternating and parallel version, for the steady-state heat equation. In the second section, we will consider time-dependent problems and the associated Schwarz waveform relaxation methods (SWR). Then, we explain the so far most sophisticated class of Schwarz-type domain decomposition methods, the so called optimised Schwarz methods. Finally, we will discuss some special issues on convergence for Schwarz type domain decomposition methods and conclude afterwards.

2.1 Classical Schwarz Domain Decomposition

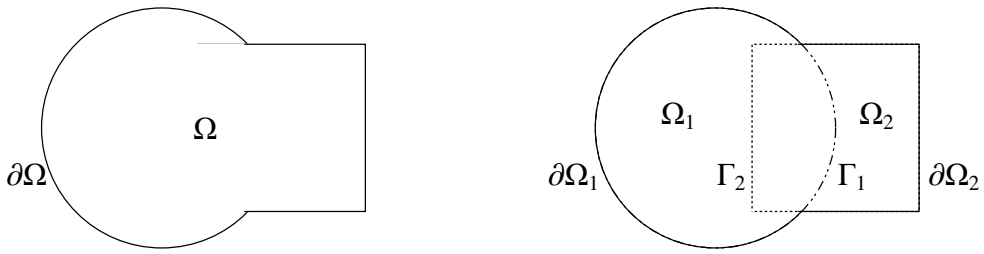
In 1869, Hermann Amandus Schwarz sent an essay to Leopold Kronecker concerning a method of proof of the existence of a harmonic function $u(x, y)$, described by the partial differential equation

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \tag{2.1}$$

on arbitrary geometrical domains $(x, y) \in \Omega$ in 2D together with given Dirichlet boundary conditions

$$u(x, y) = g(x, y),$$

on $\partial\Omega$ for a given function g.
At the time, the existence of such harmonic functions has already been proved for "easy" domains like circles or squares. Nevertheless, for arbitrarily geometrical shaped domains Ω, no general proof has been established yet at the time.
The idea of Schwarz was simple: while the original geometrical domain may have a complicated shape, one can decompose it in two (or more) subdomains that have a more convenient shape. In his original paper, he considered a geometrical domain Ω as it represented in figure 2.1a. He decomposed it into two overlapping subdomains (cf. figure 2.1b), a circle Ω_1 and a square Ω_2. The part of the global domain that belongs to both subdomains is called the overlap $\Omega_1 \cap \Omega_2$. The physical boundary $\partial\Omega$ of Ω decomposes into the part $\partial\Omega_1 \setminus \Gamma_1$ that is the part of the boundary of the circle which is common to the physical boundary and into the part $\partial\Omega_2 \setminus \Gamma_2$ that is the part of the boundary of the square that is common to the physical boundary. The boundary parts of the subdomains Γ_1 and Γ_2 are the boundary parts of the subdomains that are not common to the physical boundary of the domain Ω, they are called the interfaces.



(a) Undecomposed domain Ω (b) Decomposed domain $\Omega = \Omega_1 \cup \Omega_2$

Figure 2.1: Shape of the original Schwarz domain decomposition method

2.1.1 Alternating Method

The original method proposed by Schwarz is based on an alternating process and the use of the maximum principle. Today, it is called "Schwarz method" or "alternating Schwarz method".

The Schwarz alternating method proceeds as follows:
First, one has to possess an initial guess for the boundary values on Γ_1. Schwarz proposed to chose the value $\min(g(x,y))$ for (x,y) on $\partial\Omega$ in order to be able to apply the maximum principle. Together with the boundary values given for the physical boundary on $\partial\Omega_1 \setminus \Gamma_1$, one can solve equation (2.1) on the disk Ω_1 imposing the physical boundary conditions on its physical boundary part and the initial guess on the interface Γ_1.
Then, for the square domain, one can solve equation (2.1) imposing the just now calculated values on Γ_2 provided by the values of the solution in Ω_1 while the given boundary condition is imposed on the physical boundary part $\partial\Omega_2 \setminus \Gamma_2$.
Finally, the process is iterated while for one subdomain, the boundary values on the interface are always the values of the solution in the complementary subdomain at the previous iteration.
The entire algorithm is given in algorithm 2.1.

Algorithm 2.1 Alternating Schwarz method for the steady-state heat equation

$$u_2^0 = \min_{(x,y)\in\partial\Omega} g(x,y)$$

$$\begin{array}{ll} \Delta u_1^{k+1} = 0 & \text{on } \Omega_1 \\ u_1^{k+1}(x,y) = g(x,y) & \text{on } \partial\Omega_1 \setminus \Gamma_1 \\ u_1^{k+1}(x,y) = u_2^k(x,y) & \text{on } \Gamma_1 \end{array}$$

$$\begin{array}{ll} \Delta u_2^{k+1} = 0 & \text{on } \Omega_2 \\ u_2^{k+1}(x,y) = g(x,y) & \text{on } \partial\Omega_2 \setminus \Gamma_2 \\ u_2^{k+1}(x,y) = u_1^{k+1}(x,y) & \text{on } \Gamma_2 \end{array}$$

It is now clear why the method is also called "alternating Schwarz method", the process iterates in an alternating way between the subdomains. Schwarz motivated the name "alternating process" by the analogy to a two-piston vacuum pump which arises in the proof of the method. The maximum principle as basic ingredient of the alternating Schwarz method can be stated as follows:

Theorem 2.1 (*Maximum principle*)

Let $u : \Omega \to \mathbb{R}$ (where $\Omega \subseteq \mathbb{R}^d$) be a harmonic function, i. e. $\Delta u = 0$. Then u attains its extremal values on any compact $K \subseteq \Omega$ on the boundary ∂K of K. If u attains an extremal value anywhere in the interior of K, then it is constant.

We will see in the proof of the method that the maximum principle is used for the error function defined on the compact subdomains.

Proof 2.1 (*Convergence proof for the alternating Schwarz method*)

We denote $\underline{u}$ the minimum of $g(x,y)$ on $\partial\Omega$. In the same matter, $\overline{u}$ denotes the maximum of $g(x,y)$ on $\partial\Omega$.

For the initial guess on Γ_1, impose the value $\underline{u}$. The boundary values for the first iterate are now set and the calculation of the first iterate u_1^1 on Ω_1 is proceeded. This is equivalent to the first pumping of the first chamber of the vacuum pump. The values obtained by u_1^1 on Γ_2 are now fixed in order to complete the boundary values for Ω_2, this is equivalent to closing the valve of the second chamber. The iterate u_2^1 is now obtained by considering the physical boundary and the interface boundary values. The second chamber is pumping.
The interesting observation is now that the difference $u_2^1 - u_1^1$ or even $u_2^1 - \underline{u}$ on Γ_1 is less than $G := \overline{u} - \underline{u}$. This result is a direct consequence of the maximum principle.
Imposing now the values of u_2^1 on Γ_1, a new iterate u_1^2 can be calculated (second valve closed, first chamber is pumping). The difference $u_1^2 - u_1^1$ along Γ_2 is now by a factor $q_1 < 1$ smaller than G since the maximum principle holds. It holds then $u_1^2 - u_1^1 < Gq_1$ on Γ_2.
For the subdomain Ω_2, the next iterate u_2^2 holds a similar result: u_2^2 is obtained by the iterative process and on Γ_1 we have $u_2^2 - u_2^1 < Gq_1q_2$ with another factor $q_2 < 1$.
A linearity argument explains that the quantities q_1 and q_2 are the same for all iterations. Moreover, by induction, an infinite sequence of iterates u_1^n and u_2^n is obtained that converge uniformly to limit functions defined by

$$u_1 = u_1^1 + (u_1^2 - u_1^1) + (u_1^3 - u_1^2) + \cdots + (u_1^{n+1} - u_1^n) + \cdots \text{ in inf.},$$
$$u_2 = u_2^1 + (u_2^2 - u_2^1) + (u_2^3 - u_2^2) + \cdots + (u_2^{n+1} - u_2^n) + \cdots \text{ in inf.}$$

Indeed, the series of functions $(u_1^{k+1} - u_1^k)$ and $(u_2^{k+1} - u_2^k)$ converge uniformly in Ω_1 and Ω_2, respectively, because

$$(u_1^{n+1} - u_1^n) < G(q_1q_2)^{n-1}, \qquad (u_2^{n+1} - u_2^n) < G(q_1q_2)^{n-1}q_1,$$

with the property that $q_1q_2 < 1$.
The observation is now that u_1 and u_2 agree both on Γ_1 and Γ_2 and have therefore to be identical on the overlap $\Omega_1 \cap \Omega_2$. The final conclusion is then that u_1 and u_2 must be values of the same function u satisfying equation (2.1) on Ω. ⊠

Over one hundred years after the original article of Schwarz, Pierre-Louis Lions studied in a series of articles ([57], [58] and finally [59]) Schwarz type methods within many details, theoretical, technical as well as practical. He clarified many aspects that have not been considered by Schwarz as the lack of maximum principle within a region close to the intersection points of the boundaries of the two subdomains. He also considered the extension to more than two subdomains by emphasising the straight-through technique within paying attention that always the newest information of complementary subdomains has to be transferred on the interfaces. Finally, he proposed already in the first paper ([57]) a parallel extension that we will present in the following.

2.1.2 Parallel Method

Schwarz himself proposed the alternating method more as a tool of proof than as a numerical algorithm. In the 1980s supercomputers became more and more available and studies for parallel algorithms or at least algorithms that can easily be scheduled in parallel have been in the centre of interest. In the first paper of the series about Schwarz methods, Lions already proposed a parallel extension of Schwarz's classical method. The algorithm is given in algorithm 2.2.

Algorithm 2.2 Parallel Schwarz method for the steady-state heat equation

$$u_1^0 = \min_{(x,y)\in\partial\Omega} g(x,y)$$
$$u_2^0 = \min_{(x,y)\in\partial\Omega} g(x,y)$$

$$\begin{aligned}
&\Delta u_1^{k+1} = 0 && \text{on } \Omega_1\\
&u_1^{k+1}(x,y) = g(x,y) && \text{on } \partial\Omega_1 \setminus \Gamma_1\\
&u_1^{k+1}(x,y) = u_2^k(x,y) && \text{on } \Gamma_1
\end{aligned}$$

$$\begin{aligned}
&\Delta u_2^{k+1} = 0 && \text{on } \Omega_2\\
&u_2^{k+1}(x,y) = g(x,y) && \text{on } \partial\Omega_2 \setminus \Gamma_2\\
&u_2^{k+1}(x,y) = u_1^k(x,y) && \text{on } \Gamma_2.
\end{aligned}$$

The only difference compared to the alternating method is that in the interface transmission condition on Γ_2, we use the information on the previous iterate on the complementary subdomain instead of the actual. By this manipulation, one can easily perform the calculations in both subdomains in parallel. In contrast to the alternating method, it is called "parallel Schwarz method" or "additive Schwarz method".

In the simple case of two subdomains the subsequence of every second iterate of the parallel method is equivalent to the iterates of the alternating method. For this reason, the parallel method does not converge faster than the alternating method does on parallel structures. Nevertheless, the parallel method becomes interesting within domain decompositions with more than two subdomains, in this case, no simple subsequences that are identical to the alternating method exist. But there is still a critical issue in the case of several subdomains: suppose that for one interface, one has more than one overlapping subdomain (cf. figure 2.2). In this case, it is not clear from which subdomain the values have to be taken into account since for the dashed interface parts values from two different complementary subdomains are available.
One possibility is to exclude such situations, i. e. for every interface, there is one and only one overlapping subdomain. Such a restriction, as it has been considered by Lions in [57], is not always practicable.
One possible way to overcome those restrictions and to use the parallel Schwarz method within multiple subdomains is to use black-red colouring: colour all subdomains that do not have to communicate with the same colour (black and red for instance) and update all subdomains marked

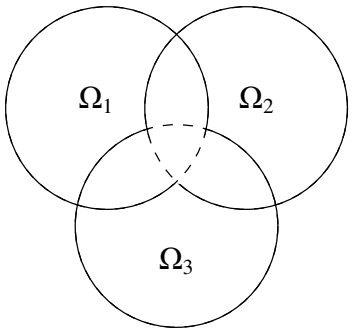


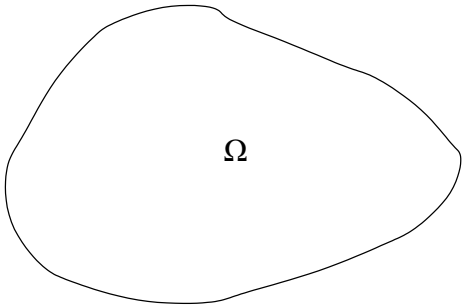
Figure 2.2: Multiple overlapping subdomains

by the same colour at the same time. Note that for the example in figure 2.2, one would need three different colours since every subdomain has to communicate with both other subdomains in order to access to the information on its interfaces.
Another possible way to deal with this situation is to use the information from both subdomains by creating a linear combination of interface values.

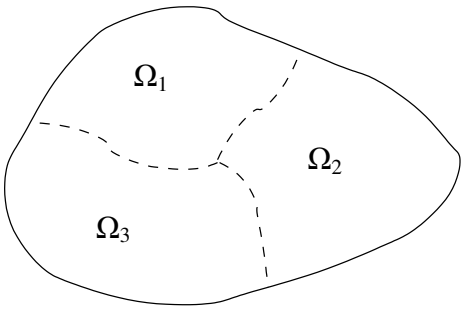
Finally, it is interesting to see how it is possible to create an overlapping domain decomposition: suppose, that the geometrical domain Ω (cf. figure 2.3a) has been decomposed into a sequence of nonoverlapping subdomains Ω_i, $i = 1, 2, \ldots$ with respect to certain criteria, cf. figure 2.3b. Starting up from a nonoverlapping decomposition, one extends every subdomain Ω_i around its interface part $\partial\Omega_i \setminus \partial\Omega$ in such a way that the new larger subdomain $\widetilde{\Omega}_i$ completely includes the former subdomain Ω_i. This means that the new interface $\partial\widetilde{\Omega}_i \setminus \partial\Omega$ and the former interface $\partial\Omega_i \setminus \partial\Omega$ have no points in common and $\widetilde{\Omega}_i \supsetneq \Omega_i$, i. e. the former subdomain is completely covered by the new subdomain. The points where the interface touches the physical boundary have also to be moved in order to prevent the overlap to become minimal or vanishing in this region. This can be seen as a "blow-up" of the nonoverlapping subdomains in order to create overlapping subdomains (cf. figure 2.3c).

2.2 Schwarz Waveform Relaxation Methods

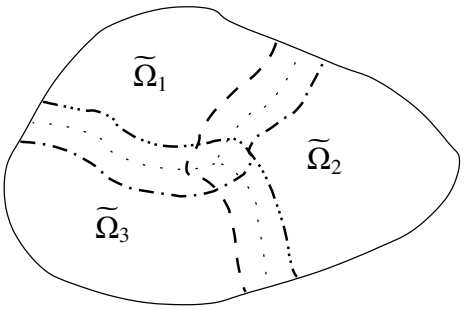
Waveform relaxation methods have been invented in order to solve large systems of coupled ordinary differential equations in the context of integrated electrical circuit simulation (see [54] for the original article). The idea is to split the system of equations in several subsystems that can be solved separately or even in parallel. After having solved the subsystems, the information between the coupled systems is exchanged and one iterates the process until convergence. The advantage is that within this technique one can solve efficiently huge coupled systems, especially on parallel architectures. The drawback is that the convergence of the method is sensible to several parameters like the coupling strength between the subequations or the length of the time interval. While the convergence is only linear for large time intervals or problems with large



(a) Undecomposed domain Ω



(b) Nonoverlapping decomposition Ω_i



(c) Overlapping decomposition $\widetilde{\Omega}_i$

Figure 2.3: Creating an overlapping domain decomposition based on a nonoverlapping one

Lipschitz constants, the algorithm has superlinear convergence behaviour for short time intervals (see [42] and references therein for an overview of the connections concerning convergence issues between waveform relaxation and Schwarz waveform relaxation methods).

In the case of time dependent partial differential equations, one can extend the Schwarz type domain decomposition methods in several ways. We consider, for instance, two strategies that differ in the order of discretisation and application of a domain decomposition strategy:
The first one uses, for an implicit scheme, the globally in time discretised problem (within a continuous or discrete formulation in space) and applies at every time step iteration a classical Schwarz type method (see [13], [14] or [64] for instance). This method is the straight through application of (algebraic or geometric) domain decomposition methods on the rising global problem at every time step. This approach has several disadvantages: first, it is no longer possible to use different time discretisations in the subdomains, which can be highly desirable especially in a nonlinear context where time step restrictions can be localised in the subdomains and have an influence on the number of iterations of the nonlinear solver. The second drawback is the fact that at every iteration only few information has to be exchanged between the subdomains. The overhead cost for the transmission of few information is not negligible and moreover this has to be done very often. Finally, it may appear that the numerical effort for doing one calculation in the subdomains may vary tremendously within the subdomains, i. e. some subdomains proceed

their calculations quite fast while other may need much more time. Now, the fast subdomains have to wait for the pending slow subdomains to have finished their calculations in order to be able to access the new information. This is a huge drawback since, *a priori*, one has to wait until all subdomains have finished their calculations of the actual time step before information can be exchanged and the iteration proceeded.
The second strategy discretises the problem first in space while the time derivative is kept on a continuous level. The resulting problem consists of a system of coupled ordinary differential equations (in time) where the unknowns are the discretised in space and continuous in time functions. One applies then a waveform relaxation method to the huge system of ordinary differential equations in order to obtain an approach for the fully discretised (in space and time) unknowns. The main problem of this approach is that the essential character of the equations, namely the spatial connectivity resulting from the partial differential equations, is lost when one separates the subsystems during the waveform relaxation method.

Schwarz waveform relaxation methods proceed a different strategy: they do not discretise at all but apply a Schwarz type domain decomposition method to the continuous in time and space problem. The resulting problems in the subsystems are global in time and therefore at every iteration of the domain decomposition algorithm the information has to be exchanged over all the time period. The problem of minimal information exchange at every iteration that we stated above is therefore eliminated and the character of special connectivity is kept. Moreover, in all subdomains, the solvers can proceed until the end of time interval without being forced to wait the other subdomains to have finished the actual time step. To exemplify the method consider the time dependent version of equation (2.1) in 2D

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0 \tag{2.2}$$

on arbitrary geometrical domains $(t, x, y) \in [0, T] \times \Omega$ where T is the end of the considered time period together with given Dirichlet boundary conditions

$$u(t, x, y) = g(t, x, y)$$

on $[0, T] \times \partial\Omega$ for a given function g. The initial condition is given by

$$u(0, x, y) = u_0(x, y)$$

for a given function u_0 on Ω. The alternating Schwarz waveform relaxation method is described in algorithm 2.3.

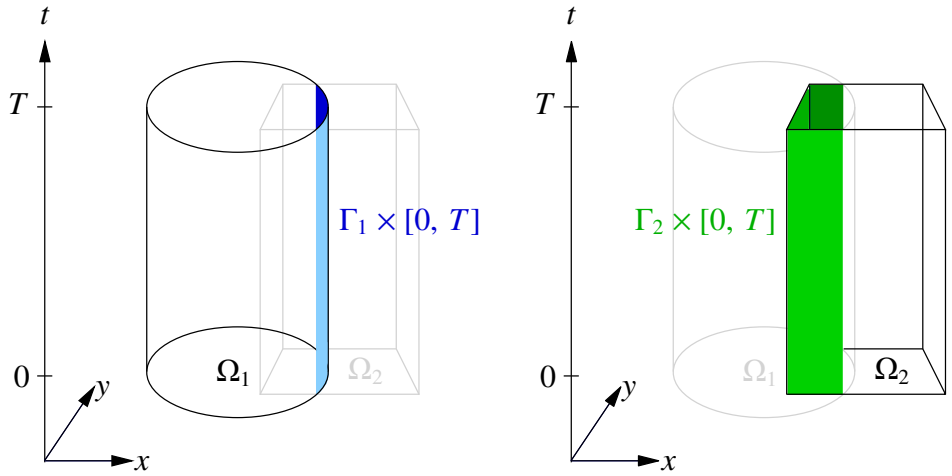
The information exchange between the subdomains is done globally in time, i. e. at every iteration, every subdomain proceeds one calculation for the entire time period $[0, T]$. After the calculation, the information is exchanged on the interface for the entire time period $[0, T]$. The information itself is global in time (cf. figure 2.4).
It is evident that the numerical treatment in the subdomains can be totally different: not only

Algorithm 2.3 Alternating Schwarz waveform relaxation method for the time-dependent heat equation

$$u_2^0(t,x,y) = u_{\text{guess}}, \quad \text{on } [0,T]\times\Gamma_1$$

$$\begin{aligned}
&\frac{\partial u_1^{k+1}}{\partial t} - \frac{\partial^2 u_1^{k+1}}{\partial x^2} - \frac{\partial^2 u_1^{k+1}}{\partial y^2} = 0 && \text{on }]0,T]\times\Omega_1\\
&u_1^{k+1}(0,x,y) = u_0(x,y) && \text{on } \partial\Omega_1\setminus\Gamma_1\\
&u_1^{k+1}(t,x,y) = g(x,y,t) && \text{on } [0,T]\times\partial\Omega_1\setminus\Gamma_1\\
&u_1^{k+1}(t,x,y) = u_2^{k}(t,x,y) && \text{on } [0,T]\times\Gamma_1
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial u_2^{k+1}}{\partial t} - \frac{\partial^2 u_2^{k+1}}{\partial x^2} - \frac{\partial^2 u_2^{k+1}}{\partial y^2} = 0 && \text{on }]0,T]\times\Omega_2\\
&u_2^{k+1}(0,x,y) = u_0(x,y) && \text{on } \partial\Omega_2\setminus\Gamma_1\\
&u_2^{k+1}(t,x,y) = g(x,y,t) && \text{on } [0,T]\times\partial\Omega_2\setminus\Gamma_2\\
&u_2^{k+1}(t,x,y) = u_1^{k+1}(t,x,y) && \text{on } [0,T]\times\Gamma_2
\end{aligned}$$



(a) Interface $\Gamma_1 \times [0,\ 1]$

(b) Interface $\Gamma_2 \times [0,\ T]$

Figure 2.4: Global in time information exchange for Schwarz waveform relaxation methods

the discretisation in space and time can be chosen individually but also different numerical algorithms can be applied to the subdomain problems. Moreover, with this approach it is even possible to couple different models (different chemical systems for instance).
The name Schwarz waveform relaxation method comes from the Schwarz type domain decomposition in space and the consideration of time-dependent subproblems that are integrated over the whole time interval as in the waveform relaxation methods. Those methods have been developed in [80], [32], [36], and independently in [38].

An interesting way of applying a Schwarz waveform relaxation algorithm to problems with a long time integration interval without loosing the superlinear convergence character has been proposed by Martin in [62]: the time integration interval $[0, T]$ is split into n time-windows $[0, T_1], [T_1, T_2], \ldots, [T_{n-1}, T]$. The Schwarz waveform algorithm 2.3 is first applied on the first time window $[0, T_1]$ and iterated until convergence. Afterwards, one applies the algorithm on the next time window $[T_{n_{i-1}}, T_{n_i}]$ imposing as initial condition for $t = T_{n_{i-1}}$ the solution of the converged iterate of the end of the previous time window and proceeds in such a way until all time windows have been treated. The number of time windows can be chosen freely. Even time windows with different size can be chosen. This technique is known as windowing in the context of waveform relaxation approaches.

Finally, a different strategy to treat the time dimension is done in the parareal Schwarz algorithm. While until now the time-dimension was treated separately — justified by its different character —, in the parareal Schwarz algorithm it is treated like an ordinary dimension that is purely advective (i. e. information propagates always in positive direction). The parareal algorithm was introduced by Lions et al. in [55], its purpose is that one can time-integrate an equation on a time-window before the previous one has converged. The starting points are predicted by a coarse grid approximation at the very beginning and are corrected at every iteration afterwards. Nevertheless, after only few iterations, the overall-accuracy is comparable to a classical sequential approach on a fine time-discretisation. Combining now the parareal algorithm in time with a parallel Schwarz method in space leads to a parareal Schwarz method. The difference is that the time windows can be chosen independently in the space-subdomains and do not have to match, moreover, the parallelism appears not only in space but also in time. The resulting algorithm can be seen as full domain decomposition in time and space.

2.3 Optimised Schwarz Methods

The class of classical Schwarz type domain decomposition methods (alternating, parallel and Schwarz waveform relaxation methods) suffers from two important drawbacks: they do need an overlap of the subdomains in order to converge and even with overlap they converge only slowly. In order to overcome those drawbacks, a new class of Schwarz type methods has been developed: the so-called optimised Schwarz methods. They are a direct extension of all so-far

presented methods in the sense that they conserve the basic principles but are much more performing without any major additional cost compared to classical approach. The crucial ingredient of improvement lies in the change of the transmission conditions on the interface. Indeed, when one studies the convergence behaviour of classical Schwarz methods with Dirichlet transmission conditions, one can give, for an idealised problem on an infinite global domain, an analytical expression of the convergence factor of the algorithm depending on the size of the overlap and the frequencies of the error of the initial guess on the interface in Fourier space. The exact formula will depend on the considered problem type but the general form is always

$$\rho(\omega, L) = \mathrm{e}^{-s(\omega)L}, \tag{2.3}$$

where ρ is the reduction factor of the error between the domain decomposition iterates and the exact solution within two iterates, ω represents the frequencies of the error of the initial guess in tangential interface direction (in space and possibly in time), s is a function that is characteristic to the considered type of problem and finally L is the size of the overlap. One sees clearly that, the larger the overlap L, the smaller the reduction factor, the faster the convergence of the algorithm. But, if the domains do not overlap, i. e. $L = 0$, the reduction factor is one which means that the algorithm does not converge at all.

Now, an interesting discovery is that one can replace the classical Dirichlet transmission conditions by other ones that behave more favourably in terms of convergence. More than that, there is one type of transmission condition that makes the algorithm converging within two iterations. This type of transmission condition is called the "optimal transmission condition". For two subdomains, it is very easy to define the optimal interface condition in the idealised case — but only in the Fourier space. The problem is that the optimal transmission condition would result in a pseudodifferential operator in the original space. Therefore, the optimal transmission condition is hard to realise in a numerical approach.
Nevertheless, there are approaches that try to establish the optimal operator for different classes of problems on different geometrically shaped domains in the original space. This approach is known as Dirichlet-to-Neumann map. It consists in artificial boundary conditions that allow to compute a solution of the equation on a restricted domain such that the solution coincides with the solution on the global nonrestricted domain. Similar principles such as transparent boundary conditions or perfectly matched layers are available for many types of equations.

In the class of optimised Schwarz methods, one tries to approach the optimal transmission condition with operators that can be represented by polynomials in Fourier space. The basic ideas for the development can be found in [26] where Engquist and Majda established absorbing boundary conditions for the simulation of waves and in [41] where Halpern established a family of artificial boundary conditions for the advection-diffusion equation. The resulting transmission operator in the original space can be represented as a local and differential operator which is easy to use in a numerical context. The easiest approach of the optimal transmission condition is of zeroth order. The resulting transmission operator is of Robin type

$$B(u) = \frac{\partial u}{\partial n} + pu,$$

where n is the normal coordinate on the interface of the subdomain and p is a positive and real constant that has to be chosen. It is clear that Robin boundary conditions are as easy to use as standard boundary conditions like Dirichlet or Neumann, since they are just a weighted combination of the two. The advantage of Robin boundary conditions is double: they allow for all choices of $p > 0$ a convergence independently of the presence of the overlap. More important, one can chose the parameter p such that it "optimises" the convergence speed of the method within this class of transmission conditions.

Finally, there are more sophisticated approximations in the Fourier space like second order (also called Ventcel) transmission conditions. In [49] Japhet developed for the first time optimised second order transmission conditions for convection-diffusion problems in the context of Schwarz methods. In this case one approaches the optimal transmission condition by a first order polynomial in Fourier space, now two different parameters p and q have to be chosen. The name *second order* transmission conditions is due to the fact that the conditions include second order (but local) derivatives in the original space. In chapter 4.2.1 we develop different types of optimised transmission conditions including Robin and Ventcel conditions for a toy-problem consisting of a linear coupled two species reactive transport problem.

2.4 Convergence Issues for Schwarz Type Domain Decomposition Methods

The convergence and the convergence speed of Schwarz type domain decomposition methods depend on several issues. We discuss in this section three of the major influences and exemplify them by numerical results.

2.4.1 Overlap

In his original method, Schwarz considered two overlapping subdomains with Dirichlet transmission conditions. The convergence factor of this algorithm is given by equation (2.3). The fact that the subdomains do overlap is crucial in the proof of convergence for his method. Indeed, one could also imagine two nonoverlapping subdomains where the two interfaces coincide to a single interface $\Gamma := \Gamma_1 = \Gamma_2$ (cf. figure 2.5). In this case, using Dirichlet conditions as he did, one would pass always the initial guess from one subdomain to another during the iteration and, suppose the initial guess on the interface Γ to be different from the exact solution, one would not converge.
Supposing some idealised conditions, one can perform a theoretical study of the convergence factor of the Schwarz method (cf. chapter 4.2.2 where this is done in detail for a linear coupled two species reactive transport system) and it turns out that, for equation (2.1), it has the form

$$\rho(\xi, L) = \mathrm{e}^{-|\xi|L}, \tag{2.4}$$

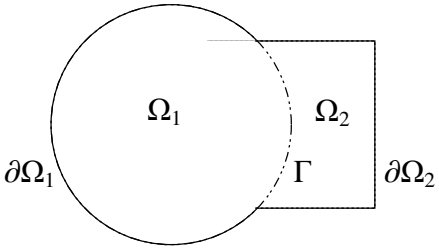


Figure 2.5: Non-overlapping Schwarz domain decomposition geometry

where ρ is the reduction factor of the error between two successive iterates and the global solution, ξ represents the dual variable of the tangential interface direction y and L is the size of the overlap. In the case of Dirichlet transmission conditions, the convergence factor depends only on two factors:

- The presence of an error in the initial guess: if there is no error in the initial guess, the method converges since the first calculated iterates are equal to the values of the exact solution. If the initial guess is not the exact solution, the lowest present error frequency determines the convergence speed.
- The size of the overlap of the two subdomains: the larger the overlap, the faster decreases the exponential term, the better the error reduction factor, the faster the convergence. If there is no overlap, i. e. $L = 0$ there is no convergence at all (with Dirichlet conditions) except for the initial guess to be the exact solution.

In practice, one has more or less influence on the convergence speed:

First, the present frequencies in the error of the initial guess cannot be controlled. The crucial issue is that one does not know the behaviour of the solution: suppose the initial guess to be smooth, one might think that the error contains only low frequencies. Now, if the problem itself contains high frequency behaviour the error contains also high frequencies. Inversely, there are high frequency problems (i. e. geostatistical or financial models) for which it is nearly impossible to give an initial guess which does not contain negligible low error frequencies.
In practice, the frequencies of the error of the initial guess are spread over a bounded range since the space is discrete, for this reason, the error frequencies cannot be arbitrarily high or small. This *a priori* information can be used to give a worst case convergence behaviour of the algorithm. This technique is used in the class of optimised transmission conditions where one tries to optimise the convergence factor of the algorithm for all worst possible error frequencies. Then, there are two different ways of treating special ranges of frequencies. Concerning low frequencies, especially when many subdomains are used, coarse grid approximations over the global domain are used in order to keep the convergence speed independent of the number of subdomains. Using this technique, one performs a projection of the subdomain problems on a global coarse grid in space and, for time-dependent problems, also in time and uses the obtained

solution as an estimation of the solution on the fine grid (see Toselli and Widlund [77] or Japhet et al. [50] for instance). Concerning high frequencies, overlapping subdomains are used since overlap attenuates especially high frequencies. The explanation lies in the exponential term of the convergence factor (2.4) which, for high frequencies ξ, becomes small when overlap is present.

Second, often, the user is free to chose the subdomain configuration and hence the overlap size. This entices the user to let the subdomains overlap as much as possible. Inversely, numerical efficiency limits the overlap since the larger the overlap, the larger the subdomains, the larger the subproblems. The "worst case" for numerical efficiency and the "best case" for convergence speed is the case where both subdomains are identical to the global domain. Here, one does not need to iterate since every subdomain can calculate the global exact solution without communicating with the other subdomain but the subproblems are as difficult as the global problem.
The effort for solving a problem with an alternating domain decomposition method using two subdomains may roughly be estimated by

$$\#\text{iterations} \cdot \sum_i (\text{cost of solving subdomain } i) + \text{synchronisation cost.} \tag{2.5}$$

Note that the number of iterations depends on the size of the overlap, the larger the overlap, the faster the convergence, less iterations are needed to reach a certain precision. The cost of solving a subdomain depends essentially on its size, the larger the overlap, the larger the problem size and the more effort has to be put in to solve it. Finally, the more the subdomains overlap the more data has to be synchronised.
It is now interesting to study if there is a good compromise between fast convergence of the Schwarz algorithm due to large overlap and numerical efficiency due to small subproblems. Therefore, we try to minimise equation (2.5) over different overlap sizes $L \geq 0$. We solve the steady-state heat equation on the square $[0,1]^2$ in 2D with the Schwarz alternating method using classical Dirichlet transmission conditions. The subdomains $\Omega_1 = [0, 0.5 + \frac{L}{2}] \times [0,1]$ and $\Omega_2 = [0.5 - \frac{L}{2}, 1] \times [0,1]$ do overlap over a length of $L \in]0,1]$ where the limit case $L = 0$ means nonoverlapping subdomains and the case $L = 1$ means that both subdomains are equal to the global domain. Imposing a random initial guess on the interface, we proceed the Schwarz algorithm until the error norm between the iterates and a global monodomain solution on the interfaces is less than 10^{-12}. We measure the CPU time in seconds for different overlap sizes since this illustrates the numerical efficiency of the overall procedure. In figure 2.6 we plot the overlap size versus CPU time for a problem with $N_x = N_y = 100$ grid cells in every space direction. The plot shows clearly that, for performance reasons, one should use the maximum overlap size since the performance behaviour is the best even if the gain stagnates for large overlap in this case.
However, this result should not be generalised since it depends also on the refinement level of the discretisation and on the way how the subdomain problems are solved.

In real applications, the choice of the subdomains are not always free. Until now, we supposed that the user is able to chose the subdomains freely. Often, when one adds Schwarz methods to already existing numerical codes, it is difficult to realise overlapping subdomains since the

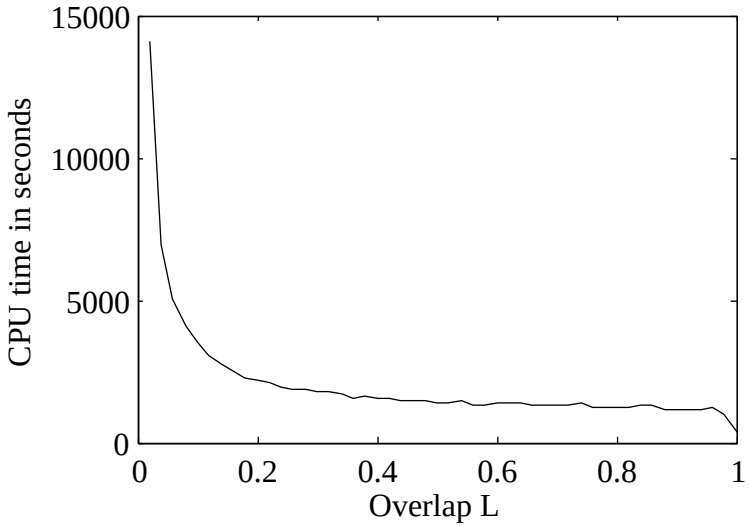


Figure 2.6: Overlap size L versus performance of the classical Schwarz method, $N_x = N_y = 100$

existing conceptional structures are mostly optimised for global calculations and are not flexible enough for overlapping submeshes. Nevertheless and even if in modern or from scratch codes it is possible to chose the subdomains to overlap as much as needed, the overlap is chosen to be minimal but present, i. e. one layer of grid cells. The reason is simple, on the one hand, in our days, Schwarz methods are essentially used for parallel computation on distributed structures (processors and/or memories). In order to be as much scaleable as possible for a high number of subdomains, the subdomains should be as distinct as possible, i. e. as less overlap as possible. On the other hand, overlap, even if it is small, can reduce drastically the number of iterations in Schwarz-type methods. In many applications coarse grid calculations are used. The initial guess after a coarse grid calculation contains therefore more high than low frequencies which can easily be eliminated by using overlap which acts as a high frequency filter on the error.

Finally, we have to mention that the influence of the overlap is not a question that can be settled conclusively without taking into account the type of transmission condition. In fact, taking into account only Dirichlet conditions is quite unfair since they do not converge for non-overlapping subdomains and hence overlap is the only way to improve convergence. As soon as one allows the use of optimised transmission conditions things change drastically: they do converge also for nonoverlapping subdomains and moreover they suffer much less from the sensitivity of the size of the overlap. In order to exemplify the situation we perform the same test "overlap size versus CPU time" for the heat equation as before but this time with optimised Robin conditions. In figure 2.7 we plot the results for the case of $N_x = N_y = 100$. One can clearly see that this time the overlap size is much less sensitive to the numerical performance compared to the case of classical Dirichlet transmission conditions. There is a minimum at $L^* = 0.14$ for which the performance is the best if the case $L = 1$ is not taken into account. But this is not the crucial point. It is more important to see three things: first, using overlap, independently of the size, is important since it makes the algorithm more performing (less than 1410 seconds CPU for

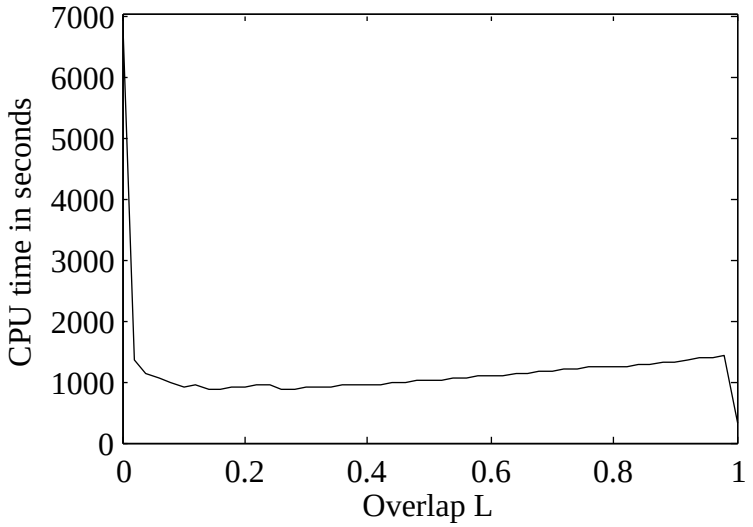


Figure 2.7: Overlap size L versus performance of the Schwarz method with optimised Robin transmission conditions, $N_x = N_y = 100$

all overlapping subdomain tests, about 6755 seconds CPU for the nonoverlapping case). Then, using too much overlap is contradictory to the domain decomposition strategy and its parallel nature. And last but not least, the question of the choice of the overlap size is overrated since with the use of optimised transmission conditions there is on the one hand much less sensibility to the overlap size if it is indeed present and on the other hand the reduction of computation time by using optimised transmission conditions is much higher (CPU times about around 1200 seconds independently of the size of the present overlap) than by keeping classical Dirichlet transmission conditions and using large overlap (CPU times about 2000 seconds). This means that using an optimised Robin condition with a small overlap is still better than a classical Dirichlet condition with large overlap.
To conclude, one can say that in the case of classical transmission conditions (Dirichlet, Neumann) the choice of the overlap size is an important point since it is the only way to improve performance. As soon as one takes into account other performing tools like optimised transmission conditions, things change and always a rather heuristic compromise has to be found for the overlap size as long as overlap is really done. If no overlap is used, for which reason ever, only optimised transmission conditions are feasible since classical do not lead to converging algorithms.

2.4.2 Transmission Conditions

Classical Schwarz methods use Dirichlet transmission conditions in order to pass the information from one subdomain to another. This strategy is quite intuitive and handy for theoretical studies which allowed already Schwarz to prove the convergence of the method. Nevertheless, standard conditions like Dirichlet or Neumann conditions suffer from two major drawbacks: first, the

method does not converge if the subdomains do not overlap and second, even if the subdomains do overlap they lead only to a slow convergence. For this reasons it is interesting to use other kind of transmission conditions that do have better properties. One strategy that has been proposed by Lions in [58] is to replace the standard Dirichlet conditions by conditions or Robin type

$$\frac{\partial u}{\partial n} + pu,$$

where n is the unit outward normal of the subdomain on the interface and $p > 0$ being a real parameter. With this small change Lions was able to prove the convergence of Schwarz's method in the case of several non-overlapping subdomains.
Numerically seen, the Robin transmission condition is not much more difficult to use than standard Dirichlet and Neumann conditions since it is only a linear combination of both. The remarkable difference is twice: first, the lack of convergence for non-overlapping subdomains is fixed and second the method converges faster than with classical conditions. The last fact can be explained in a simple way: choosing $p = 0$ in the Robin conditions, one obtains Neumann conditions. Choosing $p \to \infty$, one would obtain Dirichlet conditions. For p to be well chosen one obtains a condition that transmits the information of both the trace value and the derivative on the interface. Roughly spoken, the value of the transmitted information is much higher than for Dirichlet or Neumann conditions solely. This fact can be illustrated having a look at the convergence of the heat equation. As in equation (2.4), one obtains with Robin conditions

$$\rho(\xi, L, p) = \frac{(|\xi| - p)^2}{(|\xi| + p)^2} \, \mathrm{e}^{-|\xi|L} \, . \tag{2.6}$$

The difference compared to Dirichlet conditions is that a fractional term depending on the parameter p of the Robin condition and the present frequencies ξ in the error of the initial guess appears. One can see that for $L = 0$ only the fractional term rests and for all choices of $p > 0$ the error reduction factor is less than 1 under the condition that the error frequencies are bounded. The method does now converge without overlap. For general $L \geq 0$ the method converges always faster than with Dirichlet or Neumann conditions. For the case where $p = 0$ or $p \to \infty$, the convergence rate (2.6) degenerates, the fractional term tends to one and the error reduction factor is then equal to (2.4).
We exemplify this behaviour by once again using the heat equation on a grid with $N_x = N_y = 20$ grid cells in each direction with an overlap of $L = 0.1$. In figure 2.8, we plot the parameter p versus the number of iterations needed to achieve an error tolerance of 10^{-12}. One can see that for $p = 0$, i. e. Neumann transmission conditions, the number of iterations is 49. For $p^* \approx 8.5$ there is a minimum of 7 iterations needed to achieve convergence. For larger parameter p the number of iterations rises. The number of iterations for $p \to \infty$, i. e. the Robin condition degenerates to a Dirichlet condition, is 52. The best possible Robin condition needs therefore about 86 % less iterations than classical transmission conditions and this with the same cost per iteration in both cases, Dirichlet or Robin.

The interesting question is now: is there an *a priori* optimal choice for the parameter p? The answer is yes. Indeed, one can "optimise" the convergence rate with respect to p over a range

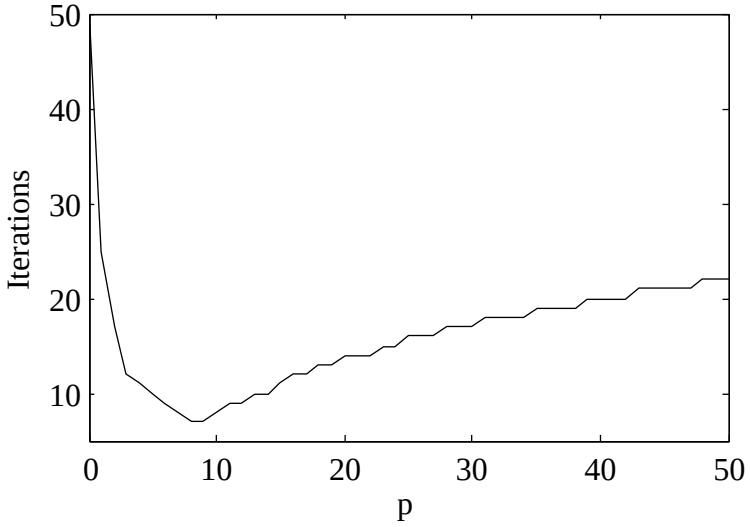


Figure 2.8: Parameter p of the Robin condition versus number of iterations until convergence is reached

of frequencies. This strategy is the key ingredient of the so-called class of optimised Schwarz methods: the optimal parameters are developed under idealised conditions (decomposition in two half-planes) in Fourier space and depend on the discretisation and physical parameters. For asymptotic cases analytical formula exist for different kind of problems and finally, those optimal parameters have shown to be quite close to the optimal choice for the parameter p. An overview of references concerning the development and study of this approach can be found in Gander [33] and references therein.
More than that, one can use even more sophisticated transmission conditions. The general approach is the following: the problem to solve has a characteristic root $\lambda(\xi)$ being a function of the tangential frequencies on the interface. For general problems, it has the form $\lambda(\xi) = \sqrt{f(\xi)}$ with f a polynomial function in ξ. By choosing a transmission condition that has the form

$$\frac{\partial u}{\partial n} + S\,u,$$

where S is a linear operator that has the Fourier symbol σ, the convergence rate has the form

$$\rho = \frac{(\lambda(\xi) - \sigma(\xi))^2}{(\lambda(\xi) + \sigma(\xi))^2}\, \mathrm{e}^{-\lambda(\xi)L}.$$

Indeed, the best choice for a transmission condition would be that one with $\sigma = \lambda$ since then the error reduction factor would be 0 and hence the algorithm converges, independently of the initial guess and a present overlap, within two iterations. The drawback is that λ is not a polynomial in the Fourier variables but a square root of a polynomial in the Fourier variables. The retransformation to the original space of $\sigma = \lambda$ as transmission operator would result in a pseudodifferential operator. Even if the development and implementation of pseudodifferential operators are possible in practice for simple problems with optimal circumstances, even the development is too

difficult for more complex problems. They may therefore be considered as unhandy. The more convenient strategy is to approach λ by an operator that is indeed a polynomial in the Fourier variables and whose retransformation results hence in a local differential operator that is easy to use. In the case of Robin conditions, one tries to approach λ by a constant p. A more sophisticated way is to use higher order approaches for example a first order approach in Fourier variables that would be $\lambda = \sqrt{f(\xi)} \approx p + qf(\xi)$ with real variables p and q. The resulting type of transmission condition is called Ventcel condition or second order condition since for many problems including second order operators like heat or diffusion type equations the resulting transmission condition includes second order derivatives along the interface.
Concerning the choice of the parameters p in the Robin and p, q in the Ventcel conditions, the most sophisticated strategy is to chose them such that the error reduction factor is optimised for all possible frequencies. A min-max problem results that may need many efforts to be solved. A different strategy is to optimise not over all possible frequencies but only over the lowest or over the highest frequencies. In this case, the resulting optimised parameter problem is only a minimisation problem which is cheaper to solve than a min-max problem.
Besides the question of the choice of the optimised parameters, one can state that Ventcel conditions are more performing and more robust than Robin conditions in a practical use. The reason is that Ventcel conditions approach also the tangential behaviour along the interface while Robin condition do not take into account this and concentrate only on the normal behaviour on the interface. For this reason, the difference in performance using Robin and Ventcel conditions can be crucial especially when the subdomains are chosen such that the equation has high tangential and less orthogonal behaviour on the interfaces. We will come back and exemplify this issues in chapter 4.2.1 where we develop Ventcel conditions for a linear coupled two species reactive transport system in the context of Schwarz waveform relaxation algorithms.

To conclude, it is important to retain that the choice of the transmission condition is a crucial issue since they determine not only the convergence rate drastically but can also make the method fail when the subdomains do not overlap. Moreover, in the case of sophisticated transmission conditions like Robin or Ventcel conditions it is important to use optimal parameters otherwise the convergence behaviour can degenerate to the one of classical conditions like Dirichlet or Neumann.

2.4.3 Krylov Accelerators

For linear problems, Krylov subspace methods like GMRES are widely used to accelerate the convergence of Schwarz type domain decomposition methods. They apply directly to classical Schwarz methods, optimised Schwarz methods and Schwarz waveform relaxation methods. We exemplify the procedure applied to equation (2.1) with Robin interface conditions on two

nonoverlapping subdomains where Γ denotes the common interface.
We define the operator

$$\mathcal{M}_j : \ (\lambda, g) \mapsto u_j \text{ solution of } \begin{cases} \dfrac{\partial^2 u_j}{\partial x^2} + \dfrac{\partial^2 u_j}{\partial y^2} & = 0 \quad \text{in } \Omega_j \\ u_j & = g \quad \text{on } \partial\Omega_j \setminus \Gamma \\ \dfrac{\partial u_j}{\partial n_j} + p u_j & = \lambda \quad \text{on } \Gamma \end{cases},$$

for $j = 1, 2$.
We define now the interface value symbol λ_j at iteration k by

$$\lambda_j^k := \left(\frac{\partial}{\partial n_j} + p\right) u_j^k,$$

which is related to the complementary $\widetilde{j}$ domain by the domain decomposition algorithm

$$\lambda_j^k = \left(\frac{\partial}{\partial n_j} + p\right) u_{\widetilde{j}}^{k-1},$$

and can be reformulated to

$$\lambda_j^k = -\left(-\frac{\partial}{\partial n_j} + p\right) u_{\widetilde{j}}^{k-1} + 2p u_{\widetilde{j}}^{k-1},$$

referring to the complementary outward normal by

$$\lambda_j^k = -\left(\frac{\partial}{\partial n_{\widetilde{j}}} + p\right) u_{\widetilde{j}}^{k-1} + 2p u_{\widetilde{j}}^{k-1},$$

finally, owing to the definition of $\lambda_{\widetilde{j}}$, written as

$$\lambda_j^k = -\lambda_{\widetilde{j}}^{k-1} + 2p\, \mathcal{M}_{\widetilde{j}}(\lambda_{\widetilde{j}}^{k-1}, g).$$

The alternating Schwarz method can therefore be rewritten as

$$\begin{aligned} \lambda_1^k &= -\lambda_2^{k-1} + 2p\, \mathcal{M}_2(\lambda_2^{k-1}, g), \\ \lambda_2^k &= -\lambda_1^k + 2p\, \mathcal{M}_1(\lambda_1^k, g), \end{aligned}$$

and the parallel Schwarz method as

$$\begin{aligned} \lambda_1^k &= -\lambda_2^{k-1} + 2p\, \mathcal{M}_2(\lambda_2^{k-1}, g), \\ \lambda_2^k &= -\lambda_1^{k-1} + 2p\, \mathcal{M}_1(\lambda_1^{k-1}, g). \end{aligned}$$

Owing to the linearity of the operator $\mathcal{M}_j$ in both arguments, the presented form gives rise to the formulation of a linear system of the form

$$\begin{pmatrix} \mathrm{Id} & \mathrm{Id} - 2p\,\mathcal{M}_2(\cdot,0) \\ \mathrm{Id} - 2p\,\mathcal{M}_1(\cdot,0) & \mathrm{Id} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 2p\,\mathcal{M}_2(0,g) \\ 2p\,\mathcal{M}_1(0,g) \end{pmatrix}, \tag{2.7}$$

which is called the interface problem. The alternating Schwarz method can be seen as an application of a Gauß-Seidel method and the parallel Schwarz method as application of a Jacobi method to the interface problem (2.7). Krylov subspace acceleration can be done by solving the linear interface problem (2.7) by a Krylov type method like GMRES instead of an splitting method like Jacobi or Gauß-Seidel. The higher performance is obvious since Krylov-type methods converge much faster than splitting methods, especially on large problems.
We want to exemplify the better performance of Krylov methods compared to the standard alternating Schwarz method. We use again the equation (2.1) with non-overlapping subdomains and Robin transmission conditions. In figure 2.9 we plot the residual error of the interface problem against the number of iterations for both a Gauß-Seidel and a GMRES method. One sees clearly

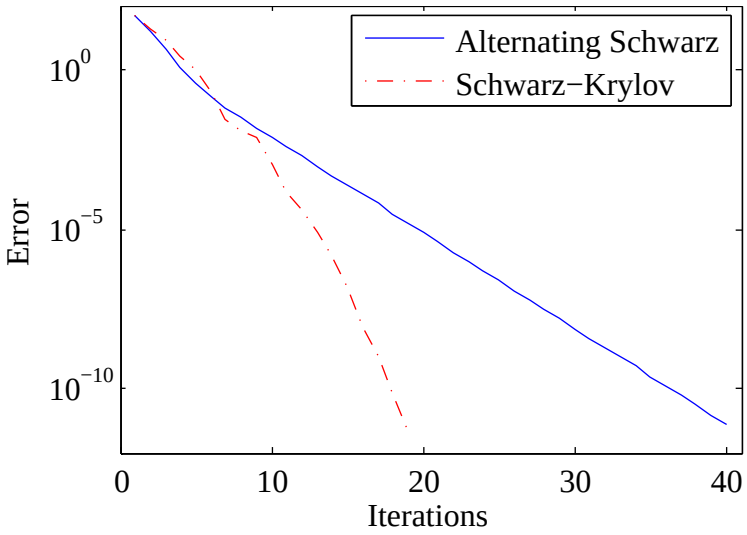


Figure 2.9: Iterations versus residual error of the interface problem with optimised Robin transmission conditions for alternating Schwarz method (e. g. Gauß-Seidel) and Krylov-Schwarz method (e. g. GMRES)

that the alternating Schwarz method and the Krylov solver have both a linear convergence behaviour but the Krylov solver converges much faster.
Note that the interface problem can also be formulated when the subdomains do overlap or when other transmission condition operators are used as long as they are linear operators. The way we developed it here is special since we used non-overlapping subdomains with Robin conditions and eliminated the reconstruction of normal derivatives on the interface.

Krylov accelerators can be seen in two ways: on the one hand, they can be seen as accelerators for the Schwarz algorithm as we presented here. On the other hand, inversely, one can

see Schwarz methods as preconditioners for the global problem solved by a Krylov-type method. The second viewpoint is the most spread in the scientific community. Krylov-accelerators have been used longtime before optimised transmission conditions have been in the spotlight. Brakkee and Wilders have studied in [9] the influence of interface conditions on the convergence of Krylov-Schwarz domain decomposition methods and showed that an application of a Krylov-type method on the interface problem has no overhead compared to a standard approach but accelerates significantly the convergence speed of the algorithm for all types of considered transmission condition. More than that, Krylov-accelerators have become very popular because they are easy to implement and have no significant overhead compared to a standard approach for the Schwarz domain decomposition algorithm. Finally, in many environments, ready-to-use Krylov-type methods like GMRES or BiCGStab are available and already optimised.

In the case where nonlinear problems are considered, it is also possible to add a Krylov accelerator. In chapter 5.3.2 we propose two different ways to add Krylov-accelerators for the Schwarz waveform relaxation method and give results that show the accelerating property.

Conclusion

In this chapter, we have given an overview of Schwarz type domain decomposition methods with a special emphasis on the geometrical viewpoint. Starting from the classical Schwarz method for steady-state problems we passed to the extension to time-variant problems and presented finally the so-far most sophisticated class of optimised Schwarz methods.
The following discussion about convergence issues for Schwarz type methods has been exemplified by the classical problem of the heat equation treated also by Schwarz. We have seen, that there are three major ingredients that make Schwarz-type methods high-performing: overlapping subdomains, optimised interface conditions and Krylov-accelerators. All together, Schwarz type domain decomposition provides a class of promising algorithms that have proved their capability in real-world problems.

2.A "Über einen Grenzübergang durch alternierendes Verfahren" — "On a limit process by an alternating method"

In this section, we give a translation of Schwarz's original article (cf. [75]) appeared in 1870. The original article is written in German, the translation is as close to the original as possible, no remark is added or omission is done. Footnotes of the original print are included directly in the text at the moment when they appear.

On a limit process by an alternating method.
H. A. Schwarz

Trimestrial scripture of the Natural Science Studying Society in Zürich, volume 15, pages 272-286. Excerpt of a talk hold on may 30th 1870 at the Natural Science Studying Society.

The method of conclusion that is known by the name Principle of Dirichlet, which, in a certain way, has to be seen as the basis of the branch of the theory of analytical functions developed by Riemann, succumbs, as it has now probably been accorded in general, to the, concerning the rigour, well-founded objections whose entire removals, as far as I know, have not been realised by the efforts of the mathematicians.

By continuation of some investigations that concern certain kinds of mapping problems and of which a part (See page 65 of this journal.) has been published in the 70th volume of Borchardt's Journal and in the essays "On the theory of maps" (See page 108 of this journal.) that attend the program of the federal polytechnic school for the winter semester 1869-70, I encountered a method of proof, by which, as I believe to have convinced myself, all theorems, whose proofs Riemann tried to produce by the Principle of Dirichlet in his published essays, can be proved with stringency.

The following note is essentially an excerpt of an essay that I have communicated in November of the last year to Mister Kronecker and some other mathematicians concerning the integration of the partial differential equation $\Delta u = 0$.

Basically, the idea is to produce the proof of existence of a function u that satisfy on a certain given domain T the partial differential equation $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ for the independent real variables x and y together with given boundary and discontinuity conditions.

For the sake of brevity, I limit myself to the case where the side conditions are only boundary conditions in which is demanded that the function u has only finite values and, along the boundary of T, a given set of finite values that belongs to one or more continuous sequences. The general case may now, by the following method, be reduced to this case.

It is in no way necessary for the applicability of the designed method of proof to do the assumption that the boundary of T has only a finite number of corners and, in general, in every point has only a certain finite radius of curvature, as have done

Messrs. Weber and Carl Neumann at their studies with the same aim. (Cf. Borchardt's Journal, ed. 81, page 29 and the records of the mathematical-physical class of the Royal Saxon society of science, meeting of April 21th 1870.) It is not even the continuity of the change of direction of the tangent of the boundary line demanded; more than that, it is sufficient to know that it is possible to divide the boundary into a finite number of domains so that the change of the direction of the tangent inside those domains may always be in the same sens, even if the change may be in an infinitely discontinuous way, so that the boundary line can possess infinitely many corners.

Even vertices of the boundary are not excluded. For such vertices, that appear by the contact of two analytical curves which have in the vicinity of the contact point the character of algebraic curves, I carried out investigations; In order not to amplify here, we do not consider vertices here.

To manage the proof whose basic idea is communicated here, we finally need the following lemma:

The boundary line of an area T, for which it is possible to integrate the partial differential equation $\Delta u = 0$ with respect to the boundary conditions, will be divided into a finite number of segments. Those segments might be arranged in two groups such as there is at least one segment in each group. Depending on the belonging to the first or the second group, the segments are associated with an odd or even ordinal number and the points which divide the segments with odd and even ordinal numbers are denoted as P. In the interior of T one might have a finite number of analytical lines L which might have either no points or only ending points P in common with segments of odd ordinal numbers without being tangential to them in this points.

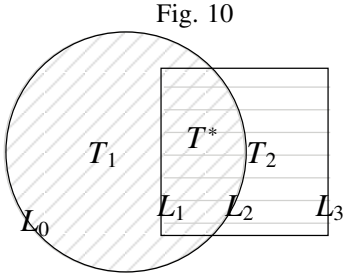
One might think of a function u on the area T which respects the partial differential equation $\Delta u = 0$ and that has on every point of the boundary of T the value 0 or 1 depending on the fact if the ordinal number of the segment in whose inner the concerned point lies is even or odd. Then, the upper bound, or rather the maximum of all values of the function u along the segments L, is a positive number q that is strictly less than 1.

If one sets for the same area T with the same subdivision of the boundary in segments into segments with even or odd ordinal number and the same lines L a function u_1 that respects the differential equation $\Delta u_1 = 0$, that has on the boundary of T on segments with even ordinal numbers a zero value and on segments with odd ordinal numbers an arbitrary given value that does not exceed the absolute value g, then the absolute values of the function u_1 on points of the lines L is nowhere larger than gq where q has the previously intended meaning, i. e. is less than 1.

There is no difficulty to integrate the partial differential equation $\Delta u = 0$ with given boundary conditions on the area of a circle and on all simply connected areas for which a conformal mapping on the area of a circle is known. Concerning this exercise it might be allowed to refer to a published article in this journal (pages 113 - 128 of the actual volume) (See page 175 of this journal.); Therein, for the sake of shortness, discontinuities for the set of values along the boundary given for the func-

tion u are formally excluded; Whereas the there developed conclusions are *mutatis mutandis* still valid when there is a discontinuity of the set of boundary values in a finite number of points.
Having shown that it is possible for a number of simple areas to integrate the differential equation $\Delta u = 0$ with respect to arbitrary boundary conditions, it is now time to show that is is possible to do so for a less simple area consisting of a union of areas for which it is possible to integrate the differential equation with respect to arbitrary boundary conditions. For the proof of this theorem, one might use a limiting process that is close to the process of the creation of an air diluted space by the use of a two-piston air pump. The period of operations consists in both cases of two alternating single operations which have the same effect but, concerning the way of doing it, are not identical but rather symmetrical in a certain way.
Such a limiting process might be called limiting process by alternating process.
One might have two areas T_1 and T_2 which have one or more common areas T^* and whose boundary lines are not tangential. (In the schematic figure 10, T_1 is the area of a circle. T_2 is the area of a square.) The total of all segments of the boundary of

Fig. 10



T_1, which lie in the outside of T_2 is denoted by L_0, the total of all other segments that lie inside of T_2 is denoted by L_2.
In the same way, the boundary of T_2 decomposes into the parts L_1 and L_3 where L_1 denotes all the segments inside the area of T_1 and L_3 denotes all the segments outside the area of T_1.
We suppose that it is possible for the area T_1 as well as for the area T_2 to integrate the partial differential equation $\Delta u = 0$ with respect to arbitrary boundary conditions; it is now to show that this is also possible for the area $T_1 + T_2 - T^* = T$ that contains the areas T_1 and T_2 as a part and for which the area T^* that is common to T_1 and T_2 is counted only once.
For the area T_1 and the segment L_1 as well as for the area T_2 and the segment L_2 the conditions of the former lemma are satisfied; in the first case, the line L_0, in the second case the line L_3 might be at the place of the group of lines with even ordinal number. Therefore, it is possible to find two numbers q_1 and q_2 that represent the role of the number q in the lemma and which are both less than 1.
Keeping the mentioned analogy, the recipient of the pump corresponds to the area T^*, the two pistons correspond to the two areas $T_1 - T^*$, $T_2 - T^*$, the valves correspond to the lines L_1 and L_2.

For the boundary of T, i. e. along L_0 and L_3, the values of the function u might be given arbitrarily: g might be the upper bound, k might be the lower bound of this values: the difference $g - k$ is denoted by G.
One supposes the values along L_2 to be arbitrary, e. g. k in all points along L_2, and one determines for the area T_1 a function u_1 that has on L_0 the given values, along L_2 the value k and that satisfies in the inner of T_1 the partial differential equation $\Delta u_1 = 0$. Concerning the assumptions that have been done on the area T_1 such a function do exist. (First traction of the first piston.)
One might fix the values of the function u_1 along L_1 and determines now for the area T_2 a function u_2 that has the given values along L_3, that matches the function u_1 along L_1 and that satisfies $\Delta u_2 = 0$. Concerning the assumptions that have been done on the area T_2 such a function do exist. (First traction of the second piston.)
The value of $u_2 - u_1$ or of $u_2 - k$ along L_2 is smaller than $g - k = G$.
One determines now for the area T_1 a function u_3 that has the given values along L_0 and that matches u_2 along L_2 and for which $\Delta u_3 = 0$ is satisfied. (Second traction of the first piston.)
The difference $u_3 - u_1$ in the inner of T_1 is in no point negative; the absolute value of the difference $u_3 - u_1$ is less than G along L_1 but due to the former lemma less than Gq_1, because $u_3 - u_1$ has the value 0 along L_0 and is less than G along L_2.
One might fix the values of the function u_3 along L_1 and determines now for the area T_2 a function u_4 that matches u_3 along L_1, that has the given values along L_3 and that satisfies $\Delta u_4 = 0$. (Second traction of the second piston.)
The difference $u_4 - u_2$ has the value 0 along L_3 and is along L_1, where it matches $u_3 - u_1$, positive and less than Gq_1; that is why $u_4 - u_2$ is nowhere negative and consistently less than Gq_1 in the inner of T_2 but strictly less than Gq_1q_2 along L_2.
By continuing this alternating method one obtains a series of a infinite number of functions with even and odd indices. The one are described for the area T_1, the other are described for the area T_2 in such a way that they have the given values along L_0 and L_3 and that in the inner they satisfy the partial differential equation $\Delta u = 0$.
For the area T^* there are functions defined with both odd and even index and more than that, they match in an alternating way along L_1 and along L_2. Along L_1 we have $u_{2n-1} = u_{2n}$ and along L_2 we have $u_{2n+1} = u_{2n}$.
It is now simple to show that the functions with odd and even index tend with rising indices to certain limit functions u' and u'' in an unlimited way, which are described by the following equations

$$u' = u_1 + (u_3 - u_1) + (u_5 - u_3) + \cdots + (u_{2n+1} - u_{2n-1}) + \cdots \text{ in inf.}$$
$$u'' = u_2 + (u_4 - u_2) + (u_6 - u_4) + \cdots + (u_{2n+2} - u_{2n}) + \cdots \text{ in inf.}$$

The series on the right hand side converge unconditionally and for all possible pairs x, y in the same way: namely it is

$$(u_{2n+1} - u_{2n-1}) < G(q_1q_2)^{n-1} \text{ and}$$
$$(u_{2n+2} - u_{2n}) < G(q_1q_2)^{n-1}q_1.$$

Along L_1 as well as along L_2 it is $u' = u''$. In the inner of T_1 it is $\Delta u' = 0$, in the inner of T_2 it is $\Delta u'' = 0$, therefore for every point in T^* it is $u' = u''$ because along the entire boundary of T^* both functions match.
Therefore, both functions u' and u'' are values of the same function u which is defined for the entire domain $T = T_1 + T_2 - T^*$ and which for the same domain satisfies the partial differential equation $\Delta u = 0$ and which has the given values along the boundary $L_0 + L_3$.
By this, the proof for the exactness of the claimed thesis is indicated: under the given assumptions it is possible to integrate the partial differential equation $\Delta u = 0$ subject to arbitrarily given boundary conditions also on domain T. —
By repeated application and adequate modification of the noted limit process by alternating method, the existence of a function u for a given domain may even be shown if, besides boundary conditions, also discontinuity conditions or, as like for Abel's integrals, only discontinuity conditions are given. In the last case, Riemann has claimed its existence in his papers and tried to prove it by use of the principle of Dirichlet.
The given method of proof is not only valid for the case where the simply or multiply connected areas of Riemann type describing the domain T in a geometrical way lie in the entire way on a single plane or on the same spherical surface but it is also essentially valid for the case where this domain is formed by one or several plane or spherical areas on the surface of a polyhedron.
By this extension one may show amongst others that it is possible to map a simply connected domain on a polyhedral surface in a conformal way on the area of a circle if the domain has a closed border line or on the spherical surface if the domain is a simply connected and closed area.
An answer on the question of the possibility to find constants for the conformal mapping of a simply connected surface of a polyhedron that is bounded by plane areas on a spherical surface is hereby given. (Confirm Borchardt's Journal, ed. 70, page 119 (See page 52 of this journal.))
A special case of the just mentioned mapping problem appears by simply connected areas of a polygon that is limited by straight lines that are mapped on circle areas in a conformal way if the area of the polygon is entirely in the finite range or if the infinite distanced point lies once or several times in the inner of the area; even twist points in the inner are not excluded. For this problem the only difficulty in the proof of the method lies in finding a several number of partially real or partially complex conjugated constants on which the mapping function depends in such a way that all assumptions are satisfied.
This difficulty may be overcome by the application of the method developed by Weierstrass. The application of the above mentioned limit process offers a new tool for overcoming this difficulty.
The proof of the possibility of the determination of the constants for the conformal map of a simply connected area limited by circle-lines on the area of a circle is similarly reduced. (l. c., page 117) (See page 79 of this journal.)

Even in this case there might be a twist point or the infinite distanced point in the inner of the area.

Appendix. Recently, I got to know three essays of Mister Christoffel (Annali di Matematica diretti da Brioschi e Cremona. tomo IV. page 1-9; News of the Royal Society of Science of Göttingen, Volume 1870, pages 283-298 and 359-369) that give rise to some remarks concerning the above communication.

On page 1 of the fourth volume of the Annali, one can read: "... the determination of the steady state temperature on a rectangular area F is not a problem, the associated problem of the steady state temperature stays completely inaccessible by the finally used method..."

And on page 284 of the mentioned news: "...(one) gets then to a strange family of problems which offer so visible difficulties that the solution of such a kind of problem has been done only once and that in a thoroughly trivial case where the boundary of B_1 (— the in all directions infinite area that rests when a simply connected, finite piece B is cut out —) is a circle. Namely, we find in the following the explication why all efforts failed to treat one of the special, preferentially interesting cases where B_1 is delimited by a straight-lined figure.

Across from this statement, it might be adequate to point to some easy examples which might provide interest even to those, for whom they are not new.

1. In a plane, whose points represent the complex value z on a geometrical way, a parable is given whose focal point is $z = 0$ and whose apex is $z = +1$. By the function $Z = \mathrm{tg}^2(\frac{1}{4}\pi\sqrt{z})$ the inner and by the function $Z_1 = \frac{2}{\sqrt{w}} - 1$ the outer of the parabola becomes connected respectively, and is mapped in the smallest pieces similar to the area of a circle whereby, as it is well-known, the concerned heat-exercise for the outer and the inner of the parabola might be seen as solved.

2. Concerning the the equation of an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ given by $a^2 - b^2 = 1$. By the function $Z = \sin\mathrm{am}(\frac{2K}{a}$ are sin $z)$, $q = (\frac{a-b}{a+b})^2$, the inner and by the function $Z_1 = \frac{z-\sqrt{z^3-1}}{a-b}$ the outer of the ellipse is mapped in a conformal way on the area of a circle whereby the heat-exercise might also for this case be seen as solved.

3. Given a square whose corners are the four points $z = +1, +\mathrm{i}, -1, -\mathrm{i}$. By the function $Z = \sin\mathrm{am}\, Kz$, $k = i$ the inner of the square is mapped in a conformal way on the area of a square and by the function

$$z = -\frac{1}{C}\int^{Z_1} \frac{\sqrt{1 - Z_1^4}}{Z_2^2} dZ_1, \quad C = \int_0^{\frac{\pi}{2}} \sqrt{\sin\varphi} d\varphi$$

reciprocally the area of the circle with radius 1 around 0 in the complex area is mapped in a conformal way on the area of the above square if the integration variable Z_1 is limited on the circle area and the integration constant is determined with respect to the condition that $\lim(z - \frac{1}{CZ_1})$ is 0 for $Z_1 = 0$. By the given maps the heat-exercise for the inner and the outer of the square might be seen as solved. (Confirm Borchardt's Journal, Ed. 70, p. 115) (See page 77 of this journal.)

The number of examples might be multiplied. —
It should not be unmentioned that one might, as it seems to me, also object against the other content of the mentioned papers.
For a complete proof, one might without any doubt claim the proof of the possibility of the determination of all constants in such a way that they satisfy the constraints of the exercise.
Moreover, one has to note that in both papers of the Göttinger Nachrichten the study is formally limited to the cases where the concerned area that has to be mapped is bounded by a boundary line that is given by a ”non reducible equation“. Hereby all cases where the boundary line persists of several different segments of analytical lines as well as the case where this line has at no point the character of an algebraic curve is a priori excluded.
Finally, one should not miss that in a larger number of cases, including the case of the mapping of the inner of an ellipse on the inner of a circle, the claimed formula of conclusion are still attached by some unsolved difficulties; A circumstance which casts the proof of the possibility of the solution of the general exercises in doubt.
Zurich, August 1870.

If you wish to make an apple pie from scratch, you must first create the universe.

Carl Sagan

3

Numerical Schemes for Discretising the Transport Operator and Prototyping

Contents

Introduction

Numerical prototyping is a procedure to implement numerical algorithms on a small scale problem under idealised or restricted conditions. The aim is to study, validate and test different numerical ingredients in order to be able to do several choices concerning the underlying structures, parameters and algorithms to be implemented afterwards on a large scale. Prototyping can be done from scratch or based on an already existing code.
In the thesis work, a large part of the coding work has been spent on the numerical prototyping. Starting from scratch, we developed a first attempt of Schwarz-type domain decomposition for a coupled two-species reactive transport system (cf. chapters 4 and 5) in the context of an underlying finite volume structure.
The choice of finite volumes as underlying structure is insofar justified since many of the codes in subsurface simulation, industrial as well as academical, are of finite volume type. The reasons are multiple: easy and high-performing implementation, well-suited for flow and mass transport problems, mass conservativity properties and some more. In this context, the experience gained during the prototype phase can directly be exploited during the large-scale implementation in an already existing reactive transport code based on a finite volume structure.

As it has been mentioned, several choices concerning the realisation of domain decomposition methods in the finite volume context have to be made during the prototyping: the first choice concerns the numerical realisation of transmission conditions and the associated way how to choose the subdomains. While transmission conditions are easy to handle in a continuous study, they may become more demanding on a discrete level. In the first part of this chapter, we discuss the way to chose the subdomains on a discrete level especially in the context of Robin transmission conditions.

Sophisticated transmission conditions like Robin or Ventcel transmission conditions will use a flux information to pass information from one subdomain to another. Using the flux information in the transmission conditions will not lead to a satisfactory result in all cases. One fundamental condition in domain decomposition methods is that the iterates resulting from a converged domain decomposition algorithm form a solution on the global domain that is identical to the global monodomain solution. On a continuous level, the presented algorithms respect this principle. On a discrete level, additional conditions may appear in order to be able to respect this principle at least when the discretisations are equal in all subdomains and a comparison with a discrete monodomain solution is possible. Now, having chosen to use an upwind discretisation of the advection, one faces the problem that within a standard finite volume approach this condition may not always be satisfied. The reason lies in the introduction of additional face unknowns on the interface faces in the domain decomposition approach that have not been present in the global monodomain approach. The way of establishing the flux in both approaches is different and even when the domain decomposition algorithm has converged the solution is different from the global monodomain approach. We exemplify by a counterexample why the standard finite volume approach is badly-suited for domain decomposition methods including a combined

upwind-advection and two-point diffusion discretisation. One comfortable way out is the use of hybrid finite volume schemes: the fundamental approach is similar to classical finite volumes but the difference lies in the introduction of well-suited boundary values on every face. Now, the way of establishing a flux information in a global monodomain approach and in a domain decomposition approach is equal in both cases and the stated basic principle is respected also on a discrete level.

Based on the hybrid finite volume scheme we explain then one way to realise the tangential flux information along the interface — a basic ingredient for the realisation of Ventcel conditions in 2D and 3D.

Finally, we discuss the time-space decomposition character of Schwarz and Schwarz waveform type algorithms on a numerical level. Schwarz type domain decomposition allows not only to use different numerical methods in the subdomains but also different space discretisations and, in the case of Schwarz waveform type algorithms for time-dependant problems, different time discretisations. On the discrete level, the information between the subdomains has to be projected between different space and time grids. We present two algorithms for the space and time projection and discuss their realisation issues.

3.1 Finite Volumes and Flux Information — Realisation of a Robin Transmission Condition

The Robin transmission condition is a key ingredient in the numerical realisation of domain decomposition methods since they are the easiest way to couple overlapping or non-overlapping subdomains in a high-performing way. Suppose the original problem is to find a function $u(x)$ for $x \in \Omega$ that is defined by

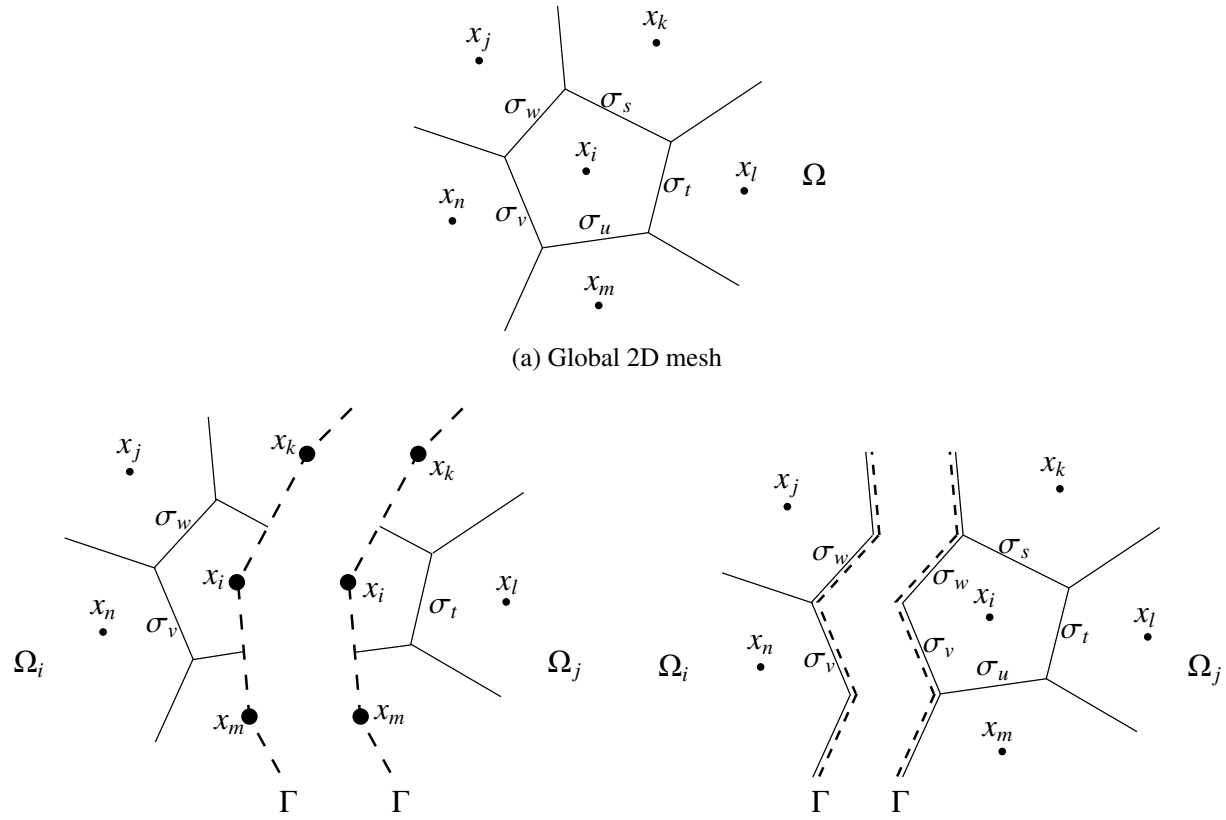
$$\operatorname{div}(\vec{f}(u)) = 0, \quad \vec{f}(u) = -a\vec{\nabla}u + \vec{b}u, \tag{3.1}$$

and some boundary conditions on $\partial\Omega$, where f is the flux function of the advection-diffusion problem. The Robin transmission operator for this model problem has the continuous formulation

$$Bu = \left(-\vec{f} \cdot \vec{n} + p\right) u, \tag{3.2}$$

where $\vec{n}$ is the outgoing unit normal on the considered interface Γ of the subdomain Ω_i and p is a real parameter.

If one wants to realise the domain decomposition on a discrete level, the first question to ask is how to subdivide the global domain in a finite volume context. Finite volumes are artificial control volumes recovering the entire global domain Ω (cf. figure 3.1a). Every finite volume has a centre x_i and a border consisting of several faces $\sigma_s, \ldots, \sigma_w$.



(a) Global 2D mesh

(b) Decomposition within the cell centres

(c) Decomposition within the faces

Figure 3.1: Domain decomposition choices in the finite volume context

One way to realise the domain decomposition in the finite volume context is to place the interface Γ on a layer that is defined by the centres of cells (cf. figure 3.1b). As a result, a subdomain consists of a number of entire cells and a layer of divided interface cells where the transmission conditions have to be realised in the centre of cells. This approach may appear unnatural since by the splitting of entire cells into interface cells some important properties of the geometry like the starshapedness may get lost. Nevertheless, this approach has been tested in some codes. While the geometric decomposition into two subdomains is more or less comfortable to realise, a problem arises when four or more subdomains meet at cornerpoints: additional conditions for the parameter p arise and the transmission conditions become more complicated (cf. the works of Gander and Kwok [35]).

Robin transmission conditions are easier to realise in the finite volume context when the interface Γ lies on a layer of faces (cf. figure 3.1c) since the major numerical and mathematical ingredients appear naturally. The major ingredient in finite volumes is the concept of discrete flux that is the amount of mass or concentration transported during a certain time interval over the border of the cells, called faces. The flux arises when one applies the divergence theorem

during the development of the finite volume discretisation. Integrate therefore equation (3.2) over a control volume K with centre x_K for example, i. e.

$$\int_K \operatorname{div}(\vec{f}(u))\mathrm{d}V = \int_{\partial K} \vec{f}(u)\cdot\vec{n}\mathrm{d}s = \sum_{\sigma\in\partial K}\int_\sigma \vec{f}(u)\cdot\vec{n}\mathrm{d}s,$$

where $\vec{n}$ is the outgoing unit normal vector on the boundary ∂K of cell K. The (outflowing) flux in the sense of finite volumes of the unknown u at a volume K over its boundary face σ with its centre x_σ can be expressed by

$$F_{K\sigma}u = \int_\sigma \vec{f}(u)\cdot\vec{n}_{K\sigma}\mathrm{d}s,$$

where $\vec{n}_{K\sigma}$ is the outgoing unit normal vector of K at σ.
Now, the realisation of a Robin transmission condition in the discrete context is straightforward by using the discrete flux over an interface face

$$\left(-\vec{f}(\cdot)\cdot\vec{n}+p\right)u \approx -\frac{1}{\mathrm{m}(\sigma)}F_{K\sigma}u + pu(x_\sigma)$$

as a boundary condition. Note, that no additional effort has to be put in the realisation of Robin transmission conditions since the discrete flux has already to be formed for the cell balance equation and hence no additional reconstruction is necessary.
In our prototype, we use this way to realise Robin transmission conditions. There is a problematic issue concerning the type of finite volume schemes which can be used in order to obtain consistency between a converged domain decomposition solution and a global monodomain solution. We will stress this issue in the following section.
Finally, we have to mention that the theoretical development of a Robin condition does not exactly let arise the flux function as we stated in equation (3.2). The exact Robin condition for an advection-diffusion-like problem is developed in chapter 4.2.1 and has the form

$$B_{\text{exact}} = \frac{\partial u}{\partial\vec{n}} - \frac{\vec{b}\cdot\vec{n}-\tilde{p}}{2a}u. \tag{3.3}$$

As the Robin condition is a linear condition, one can transform it linearly without loosing the major properties: setting $p = \frac{\vec{b}\vec{n}+\tilde{p}}{2}$ in formulation (3.2) and using $\frac{B}{a}$ leads to formulation (3.3).

3.2 Discretisation of the Transport Operator: Standard vs. Hybrid Finite Volume Schemes

While in global calculations the choice of the numerical scheme for discretising the partial differential equations is more or less a matter of context, accuracy and performance, the situation

changes when domain decomposition is used. Domain decomposition algorithms on a continuous level provide several basic properties, the most important one is: the converged domain decomposition solution is equal to the global monodomain solution independently of the shape of the decomposition. This principle has also to be verified on a discrete level: all discrete schemes of the subdomains together should be equal to the global monodomain scheme when the domain decomposition algorithm has converged, independently of the shape of the domain decomposition.
Basically, all discrete schemes, whether they are of finite element, finite volume or of other type, could be used in the domain decomposition context. Unfortunately, some are more easily applicable than others and finally there are discrete schemes, that work well in a monodomain context but are not sophisticated enough to work satisfactorily in a domain decomposition context without fundamental modifications in the scheme.
Basing on a standard cell-centred finite volume scheme with a two-point discretisation for diffusion and an upwind discretisation for advection we exemplify with a counterexample why, in a domain decomposition context, this scheme fails. Then, we show how to extend the scheme without any major differences in the underlying numerical data structure such that it works perfectly in a domain decomposition context.

3.2.1 A Counterexample

Suppose the 1D domain $\Omega = [-1,\ 1]$ with a discretisation into two cells K_1 and K_2 whose centres lie at $x_{K_1} = -0.5$ and $x_{K_2} = 0.5$. The three faces σ_l, σ_0 and σ_r lie at $x_{\sigma_l} = -1$, $x_{\sigma_0} = 0$ and $x_{\sigma_r} = 1$, respectively (cf. figure 3.2). We want to solve the steady-state equation

$$\mathrm{div}(-a\vec{\nabla}u + \vec{b}u) = 0,$$

with parameters $a > 0$ and $\vec{b}$, such that $\vec{b}\cdot\vec{x} > 0$, i. e. the advection speed is in positive x-direction. We impose Dirichlet boundary conditions with values $u(x = -1) = 1$ and $u(x = 1) = 0$.

The first attempt is to solve the problem globally. The discrete unknowns are u_{K_1}, u_{K_2} for the two cells and u_{σ_l}, u_{σ_r} for the two boundary faces. We do not proceed a direct elimination of the boundary face unknowns. Note that the face σ_0 has no unknown.
The first step is to establish an expression for the discrete fluxes along the three faces. On σ_l we impose the discrete flux as

$$F_{K_1,\sigma_l} = (T_{K_1,\sigma_l} + b^{\oplus}_{K_1,\sigma_l})u_{K_1} - (T_{K_1,\sigma_l} + b^{\ominus}_{K_1,\sigma_l})u_{\sigma_l},$$

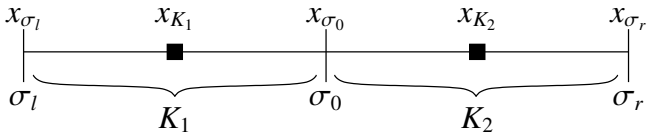


Figure 3.2: 1D mesh with two cells and three faces

where the diffusive transmissivity coefficient is defined by

$$T_{K\sigma} = \frac{a}{|x_K - x_\sigma|},$$

and the advection upwind transmissivity coefficient is defined by

$$b^{\oplus}_{K\sigma} = \begin{cases} \left| \int_\sigma \vec{b} \cdot \vec{n}_{K\sigma} \, \mathrm{d}\sigma \right| & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} \geq 0 \\ 0 & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} < 0 \end{cases}, \qquad b^{\ominus}_{K\sigma} = \begin{cases} 0 & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} \geq 0 \\ \left| \int_\sigma \vec{b} \cdot \vec{n}_{K\sigma} \, \mathrm{d}\sigma \right| & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} < 0 \end{cases},$$

with $\vec{n}_{K\sigma}$ being the unit outward normal of K on σ.
Finally, using the information on the upwind discretisation the discrete flux simplifies to

$$F_{K_1,\sigma_l} = (T_{K_1,\sigma_l})u_{K_1} - (T_{K_1,\sigma_l} + b^{\ominus}_{K_1,\sigma_l})u_{\sigma_l}.$$

In the same way, we can express the flux on the other boundary face

$$F_{K_2,\sigma_r} = (T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})u_{K_2} - (T_{K_2,\sigma_r})u_{\sigma_r}.$$

The boundary conditions write

$$u_{\sigma_l} = 1,$$
$$u_{\sigma_r} = 0.$$

Concerning the face σ_0, we establish the flux expression in a different way. As σ_0 is an inner face, no explicit unknown on the face is available. We can express the outgoing flux of cell K_1 on face σ_0 by using the information of the two cells K_1 and K_2:

$$\begin{aligned} F_{K_1,K_2} &= (T_{K_1,K_2} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} - (T_{K_1,K_2} + b^{\ominus}_{K_1,\sigma_0})u_{K_2} \\ &= (T_{K_1,K_2} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} - (T_{K_1,K_2})u_{K_2}. \end{aligned}$$

In the same way, the outgoing flux of cell K_2 on face σ_0 writes:

$$\begin{aligned} F_{K_2,K_1} &= (T_{K_2,K_1} + b^{\oplus}_{K_2,\sigma_0})u_{K_2} - (T_{K_2,K_1} + b^{\ominus}_{K_2,\sigma_0})u_{K_1} \\ &= (T_{K_2,K_1})u_{K_2} - (T_{K_2,K_1} + b^{\ominus}_{K_2,\sigma_0})u_{K_1}. \end{aligned}$$

We will see afterwards that it is exactly this flux which causes the incoherence between the global monodomain and the domain decomposition approach since it uses the upwind cell value of cell K_1 for the advective part of the flux. In the global approach, this is not problematic since the value is accessible but in the domain decomposition approach, it is not and has to be replaced by another value.

In the concept of finite volumes, one establishes one equation per finite volume that represents a flux balance of inflowing and outflowing fluxes of type $F_{\text{in}}+F_{\text{out}} = 0$. We state the two equations for cell K_1 and cell K_2 by

$$\begin{aligned} F_{K_1,\sigma_l} + F_{K_1,K_2} &= (T_{K_1,\sigma_l} + T_{K_1,K_2} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} + (-T_{K_1,K_2})u_{K_2} + (-T_{K_1,\sigma_l} - b^{\ominus}_{K_1,\sigma_l})u_{\sigma_l} = 0, \\ F_{K_2,K_1} + F_{K_2,\sigma_r} &= (-T_{K_2,K_1} - b^{\ominus}_{K_2,\sigma_0})u_{K_1} + (T_{K_2,K_1} + T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})u_{K_2} + (-T_{K_2,\sigma_r})u_{\sigma_r} = 0, \end{aligned}$$

which form, together with the two boundary conditions, a well-defined linear system that is solved by

$$u_{K_1} = \frac{(T_{K_1,\sigma_l} + b^{\ominus}_{K_1,\sigma_l})(T_{K_2,K_1} + T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})}{(T_{K_2,K_1} + T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})(T_{K_1,\sigma_l} + T_{K_1,K_2} + b^{\oplus}_{K_1,\sigma_0}) - (T_{K_1,K_2})(T_{K_2,K_1} + b^{\ominus}_{K_2,\sigma_0})}, \qquad u_{\sigma_l} = 1,$$

$$u_{K_2} = \frac{(T_{K_1,\sigma_l} + b^{\ominus}_{K_1,\sigma_l})(-T_{K_2,K_1} - b^{\ominus}_{K_2,\sigma_0})}{(T_{K_2,K_1} + T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})(T_{K_1,\sigma_l} + T_{K_1,K_2} + b^{\oplus}_{K_1,\sigma_0}) - (T_{K_1,K_2})(T_{K_2,K_1} + b^{\ominus}_{K_2,\sigma_0})}, \qquad u_{\sigma_r} = 0.$$

Setting $a = 1$ and $\vec{b} = (1)$ for instance, one obtains

$$u_{K_1} = \frac{7}{6} \approx 0.8571, \qquad u_{\sigma_l} = 1,$$
$$u_{K_2} = \frac{3}{6} \approx 0.4286, \qquad u_{\sigma_r} = 0.$$

We proceed now a domain decomposition solution with the two non-overlapping subdomains $\Omega_1 = [-1, 0]$ and $\Omega_2 = [0, 1]$. We impose the Robin transmission conditions as we presented in section 3.1 with a positive parameter p.
For Ω_1 we establish the two flux expressions as

$$F_{K_1,\sigma_l} = (T_{K_1,\sigma_l})u_{K_1} - (T_{K_1,\sigma_l} + b^{\ominus}_{K_1,\sigma_l})u_{\sigma_l},$$
$$F_{K_1,\sigma_0} = (T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} - (T_{K_1,\sigma_0})u_{\sigma_0,1},$$

where $u_{\sigma_0,1}$ is the additional unknown at the face σ_0 belonging to subdomain Ω_1. In the same way, we can establish the two flux expressions for Ω_2 as

$$F_{K_2,\sigma_r} = (T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r})u_{K_2} - (T_{K_2,\sigma_r})u_{\sigma_r},$$
$$F_{K_2,\sigma_0} = (T_{K_2,\sigma_0})u_{K_2} - (T_{K_2,\sigma_0} + b^{\ominus}_{K_2,\sigma_0})u_{\sigma_0,2},$$

where $u_{\sigma_0,2}$ is the additional unknown at the face σ_0 belonging to subdomain Ω_2. This new unknown is used for the upwind value of discrete flux F_{K_2,σ_0} on σ_0 since the value u_{K_1}, which is used in the global monodomain approach, is not accessible since it lies in the complementary subdomain.
Finally, the Robin transmission conditions are

$$-F_{K_1,\sigma_0} + pu_{\sigma_0,1} = -(T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} + (T_{K_1,\sigma_0})u_{\sigma_0,1} + pu_{\sigma_0,1} = g_l,$$

for subdomain Ω_1 and

$$-F_{K_2,\sigma_0} + pu_{\sigma_0,2} = -(T_{K_2,\sigma_r})u_{K_2} + (T_{K_2,\sigma_r} + b^{\ominus}_{K_2,\sigma_0})u_{\sigma_0,2} + pu_{\sigma_0,2} = g_r,$$

for subdomain Ω_r.
Suppose the domain decomposition algorithm to have converged, then we have the property

$$g_l = F_{K_2,\sigma_0} + pu_{\sigma_0,2},$$
$$g_r = F_{K_1,\sigma_0} + pu_{\sigma_0,1}.$$

Once again, we can state the linear systems to solve in both subdomains by imposing first the balance equation for the cell, than the Dirichlet boundary condition equation and finally the Robin transmission condition equation:

$$
\begin{aligned}
&(T_{K_1,\sigma_l} + T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} + (-T_{K_1,\sigma_l} - b^{\ominus}_{K_1,\sigma_l})u_{\sigma_l} + (-T_{K_1,\sigma_0})u_{\sigma_0,1} = 0,\\
&u_{\sigma_l} = 1,\\
&-(T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} + (T_{K_1,\sigma_0})u_{\sigma_0,1} + pu_{\sigma_0,1} = -(T_{K_2,\sigma_r})u_{K_2} - (T_{K_2,\sigma_r} + b^{\ominus}_{K_2,\sigma_0})u_{\sigma_0,2} + pu_{\sigma_0,2},
\end{aligned}
$$

for subdomain Ω_1 and

$$
\begin{aligned}
&(T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r} + T_{K_2,\sigma_0})u_{K_2} + (-T_{K_2,\sigma_r})u_{\sigma_r} - (T_{K_2,\sigma_r} + b^{\ominus}_{K_2,\sigma_0})u_{\sigma_0,2} = 0,\\
&u_{\sigma_r} = 0,\\
&-(T_{K_2,\sigma_0})u_{K_2} + (T_{K_2,\sigma_0} + b^{\ominus}_{K_2,\sigma_0})u_{\sigma_0,2} + pu_{\sigma_0,2} = (T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0})u_{K_1} + (-T_{K_1,\sigma_0})u_{\sigma_0,1} + pu_{\sigma_0,1},
\end{aligned}
$$

for subdomain Ω_2.
Both systems can be solved at the same time independently of the choice of $p > 0$. Defining the coefficients

$$
\begin{aligned}
c &:= (T_{K_1,\sigma_l}),\\
d &:= (-T_{K_1,\sigma_l} - b^{\ominus}_{K_1,\sigma_l}),\\
e &:= (T_{K_1,\sigma_0} + b^{\oplus}_{K_1,\sigma_0}),\\
f &:= (-T_{K_1,\sigma_0}),\\
g &:= (T_{K_2,\sigma_r} + b^{\oplus}_{K_2,\sigma_r}),\\
h &:= (T_{K_2,\sigma_0}),\\
i &:= (-T_{K_2,\sigma_0} - b^{\ominus}_{K_2,\sigma_0}),
\end{aligned}
$$

the solution is given by

$$
\begin{aligned}
u_{\sigma_0,1} = u_{\sigma_0,2} &= \frac{ed(g+h)}{ig(c+e) + cf(g+h)},\\
u_{K_1} &= -\frac{d}{c+e} - \frac{d}{c+e}u_{\sigma_0,1},\\
u_{K_2} &= -\frac{i}{g+h}u_{\sigma_0,1},\\
u_{\sigma_l} &= 1,\\
u_{\sigma_r} &= 0.
\end{aligned}
$$

Setting once again $a = 1$ and $\vec{b} = (1)$ for instance, one obtains

$$
\begin{aligned}
u_{K_1} &= \frac{57}{65} \approx 0.8769, \qquad u_{\sigma_l} = 1, \qquad u_{\sigma_0,1} = \frac{9}{13} \approx 0.6923,\\
u_{K_2} &= \frac{27}{65} \approx 0.4154, \qquad u_{\sigma_r} = 0, \qquad u_{\sigma_0,2} = \frac{9}{13} \approx 0.6923.
\end{aligned}
$$

The converged domain decomposition solution is different from the monodomain solution even if the transmission conditions are verified and both interface values as well as both fluxes along

the interface are equal, respectively. The problem lies in the introduction of the additional face unknown on the interface together with the upwind advection discretisation and the presence of diffusion. While the introduction of the additional unknown would not have caused a problem when only advection or only diffusion is present, the mixture of both discretisations, upwind advection and two-point diffusion causes the problem.
There are different ways to overcome this problem. First, taking into account the interface value $u_{\sigma_0,2}$ for the statement of the flux F_{K_2,σ_0} is problematic since it reflects not the right upwind value that should have been provided by the value of u_{K_1}. But this value lies in the complementary subdomain Ω_1 and is therefore not accessible. Introducing a minimum overlap with accessibility to values in the inner of other subdomains may solve this problem. Nevertheless, this approach may be seen as disaccording to the local data accessibility restriction on the interface in the domain decomposition approach.
We show in the following a way to modify slightly the numerical discretisation scheme. The resulting scheme is based on the same numerical complexity concerning data structures and number of unknowns but has solved the inconsistency of the standard finite volume discretisation exemplified in this section.

3.2.2 A Weighted Hybrid Finite Volume Scheme

For the following developments, we extended the ideas of Eymard, Gallouët and Herbin in [27] where a compromise between hybrid finite volume schemes and nonconforming finite element schemes is introduced for diffusion problems by keeping only face unknowns where they are necessary. The original ideas have been developed by Droniou and Eymard in [23]. We use their ideas for the discrete approximation of the equation

$$\operatorname{div}(-a\vec{\nabla}u + \vec{b}u) = 0, \tag{3.4}$$

on a domain Ω with its boundary $\partial\Omega$ with the unknown u. We use a control volume approach for a cell K and obtain

$$\sum_{\sigma\in\varepsilon_K}\int_{\sigma}(-a\vec{\nabla}u_K + \vec{b}u_K)\cdot\vec{n}_{K\sigma}\mathrm{d}s = 0,$$

where u_K is an approximation of $u(x_K)$, x_K is the centre and $m(K)$ is the measure of the control volume K, ε_K the set of faces of the cell K and $\vec{n}_{K\sigma}$ the outgoing normal of K on face σ with centre x_σ.
We are now interested in approximating the flux

$$F_{K\sigma} = \int_{\sigma}(-a\vec{\nabla}u_K + \vec{b}u_K)\cdot\vec{n}_{K\sigma}\mathrm{d}s$$

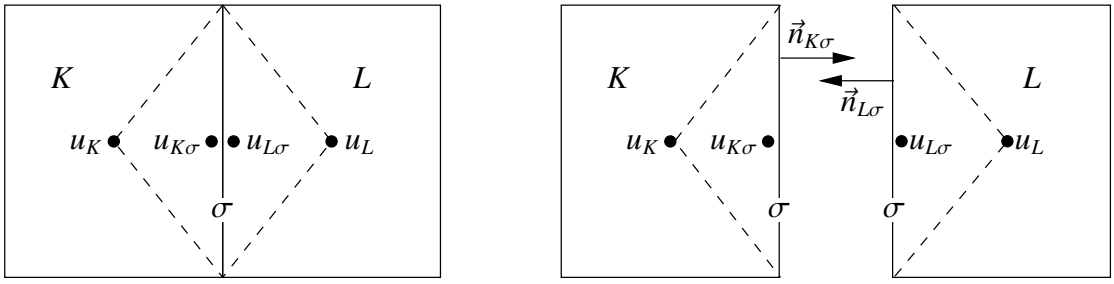
by a numerical approximation called $\widetilde{F_{K\sigma}}$ such that the flux is continuous, i. e. for two cells K and L connected by the face σ we have the property

$$\widetilde{F_{K\sigma}} = -\widetilde{F_{L\sigma}}.$$

Moreover, we want the numerical flux to be consistent, i. e.

$$|\widetilde{F_{K\sigma}} - F_{K\sigma}| \to 0$$

when $(\text{size}(\tau)) \to 0$, where $\text{size}(\tau)$ is the size of the spatial discretisation. We restrict our developments to the case of orthogonal meshes and plot in figure 3.3 a scheme of a 2D mesh with cell K and its neighbouring cell L. In a standard finite volume case, one defines the approximation



(a) Global point of view, face unknowns and coupling equations are directly eliminated.

(b) Domain decomposition point of view, an artificial interface on σ is introduced and coupling conditions (face value equity and flux equity) are introduced.

Figure 3.3: 2D mesh grid with face unknowns for global and domain decomposition point of view in the hybrid finite volume scheme.

of an unknown cellwide, e. g. u_K for cell K. In our hybrid scheme, we add for every face σ two additional values, one calls $u_{K\sigma}$ the value on face σ viewed from K and $u_{L\sigma}$ the value on the same face but viewed from L. We define now $F_{K\sigma}$ the outgoing flux of K on σ. As we want to use a two-point approximation of the fluxes, we define

$$\widetilde{F_{K\sigma}} = \alpha_K u_K - \alpha_{K\sigma} u_{K\sigma}$$

the approximation of the flux $F_{K\sigma}$ by the values of the cell centre u_K and the face centre $u_{K\sigma}$ viewed from cell K. The approximation coefficients α_K and $\alpha_{K\sigma}$ are defined as the sum of the classical finite volume discretisation coefficients for the advective and the diffusive part. The advective part is described by the advection upwind coefficients

$$b^{\oplus}_{K\sigma} = \begin{cases} \left| \int_\sigma \vec{b} \cdot \vec{n}_{K\sigma} \, \mathrm{d}\sigma \right| & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} \geq 0 \\ 0 & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} < 0 \end{cases}, \qquad b^{\ominus}_{K\sigma} = \begin{cases} 0 & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} \geq 0 \\ \left| \int_\sigma \vec{b} \cdot \vec{n}_{K\sigma} \, \mathrm{d}\sigma \right| & \text{if } \vec{b} \cdot \vec{n}_{K\sigma} < 0 \end{cases}.$$

The diffusive part is approximated by the transmissivity coefficients

$$T_{K\sigma} = \frac{a(x_K)}{|x_K - x_\sigma|}.$$

The advection-diffusion coefficients are finally defined as

$$\alpha_K = T_{K\sigma} + b^{\oplus}_{K\sigma}, \qquad \alpha_{K\sigma} = T_{K\sigma} + b^{\ominus}_{K\sigma}.$$

Note that this approach is equivalent to the classical approach for boundary faces. The difference is now, that we use this approach in our hybrid scheme also for faces that do not lie on the boundary or on artificial discontinuities:
As before, we define the outgoing flux from L on the face σ by

$$\widetilde{F_{L\sigma}} = \alpha_L u_L - \alpha_{L\sigma} u_{L\sigma}.$$

Note that we can write the flux approximation as

$$\widetilde{F_{K\sigma}} = T_{K\sigma}(u_K - u_{K\sigma}) + b_{K\sigma}(u_{K\sigma})^{\oplus},$$

where $b_{K\sigma}$ is the normal face velocity value and $(u_{K\sigma})^{\oplus}$ is the upwind value of u.

If we are on an inner face, we are not interested in face unknowns, we can eliminate them since we imposed rectangular meshes and isotropic diffusion. Note that the face unknown elimination is no longer possible for general meshes or if anisotropic diffusion tensors are taken into account. We refer for instance to the work of Enchéry et al. in [25] where the same ideas of flux and trace continuity are used but this time, the trace value results in nonlinear conditions, for this reason, it is also not possible to eliminate the face unknown directly.
First, we want to ensure the trace continuity on σ:

$$u_{L\sigma} = u_{K\sigma} =: u_\sigma.$$

Then, we recall the flux continuity

$$\widetilde{F_{K\sigma}} = -\widetilde{F_{L\sigma}}.$$

By assembly of the two equations, given values on u_K and u_L and owing to the relation $b^{\ominus}_{L\sigma} = b^{\oplus}_{K\sigma}$, u_σ is given by

$$u_\sigma = \frac{\alpha_L u_L + \alpha_K u_K}{\alpha_{K\sigma} + \alpha_{L\sigma}}$$

when $\alpha_{K\sigma} + \alpha_{L\sigma} \neq 0$ and u_σ is defined by

$$u_\sigma = \frac{u_L + u_K}{2}$$

when $\alpha_K + \alpha_L = 0$.
Defining finally the relative coefficients as

$$\theta_K := \begin{cases} \dfrac{\alpha_{K\sigma}}{\alpha_{K\sigma} + \alpha_{L\sigma}} & \text{if } \alpha_{K\sigma} + \alpha_{L\sigma} \neq 0, \\ \dfrac{1}{2} & \text{otherwise,} \end{cases} \quad \text{and} \quad \theta_L := \begin{cases} \dfrac{\alpha_{L\sigma}}{\alpha_{K\sigma} + \alpha_{L\sigma}} & \text{if } \alpha_{K\sigma} + \alpha_{L\sigma} \neq 0, \\ \dfrac{1}{2} & \text{otherwise,} \end{cases}$$

one can express the flux approximations as

$$\widetilde{F_{K\sigma}} = \alpha_K(1 - \theta_K)u_K - \alpha_L \theta_K u_L,$$
$$\widetilde{F_{L\sigma}} = \alpha_L(1 - \theta_L)u_L - \alpha_K \theta_L u_K,$$

where no more face unknowns are used.
In the case where σ is a boundary or interface face, the unknown u_σ is part of the primary unknowns and no special treatment is needed besides the additional boundary condition equation. Finally, we discretise the continuous problem by setting discrete unknowns on the cell centres and the boundary/interface faces. The discrete equations are the balance equations on every cell and the boundary conditions on every boundary face. Moreover, we implicitly use the trace continuity and the flux continuity on every inner face.

With this hybrid scheme, the convergence of the both interface values of different subdomains to the face value of the monodomain solution is ensured if the domain decomposition algorithm has converged. This property can easily be seen in the following way for non-overlapping subdomains: suppose K to be the interface cell of domain Ω_1 and L to be the interface cell of domain Ω_2. Now, in the converged domain decomposition algorithm with non-overlapping subdomains, the two transmission conditions verify

$$\begin{aligned} -\frac{1}{\mathrm{m}(\sigma)}\widetilde{F_{K\sigma}} + p_1 u_{K\sigma} &= \frac{1}{\mathrm{m}(\sigma)}\widetilde{F_{L\sigma}} + p_1 u_{L\sigma} \\ -\frac{1}{\mathrm{m}(\sigma)}\widetilde{F_{L\sigma}} + p_2 u_{L\sigma} &= \frac{1}{\mathrm{m}(\sigma)}\widetilde{F_{K\sigma}} + p_2 u_{K\sigma} \end{aligned}$$

at the same time. Defining the jump of the flux as $\left[\!\left[\widetilde{F_\sigma}\right]\!\right] := \widetilde{F_{K\sigma}} + \widetilde{F_{L\sigma}}$ and the jump of the trace value as $[\![u_\sigma]\!] := u_{K\sigma} - u_{L\sigma}$, one can write the linear system in matrix form as

$$\begin{pmatrix} -\frac{2}{\mathrm{m}(\sigma)} & p_1 \\ -\frac{2}{\mathrm{m}(\sigma)} & -p_2 \end{pmatrix} \begin{pmatrix} \left[\!\left[\widetilde{F_\sigma}\right]\!\right] \\ [\![u_\sigma]\!] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This linear system is invertible iff $p_1 \neq -p_2$. The resulting solution is $\left[\!\left[\widetilde{F_\sigma}\right]\!\right] = 0$ (i. e. flux continuity) and $[\![u_\sigma]\!] = 0$ (i. e. trace continuity). The resulting conditions in the converged domain decomposition approach are the same as in the global monodomain approach.
In the case of overlapping subdomains Ω_1 and Ω_2, the proof becomes different but not much more difficult: we denote $\Gamma_1 := \partial\Omega_1 \setminus \partial\Omega$ the interface of Ω_1 and by $\Gamma_2 := \partial\Omega_2 \setminus \partial\Omega$ the interface of Ω_2. Suppose the domain decomposition algorithm to have converged. The first step is to see that both solutions verify the same scheme on the overlap $\Omega_1 \cap \Omega_2$ and on the two interfaces Γ_1 and Γ_2 and are therefore equal since they verify the same boundary conditions on the boundary of the overlap (i. e. the interfaces and physical boundaries of the overlap). The second step consists of extending the overlapping part of the solution u_1 on Ω_1 with its non-overlapping part to form the entire solution on one subdomain using the trace and flux continuity on their common boundary (which is the interface of the complementary subdomain Γ_2). The same holds for the extension of u_2 on Ω_2. Together, we have therefore two solutions u_1 on Ω_1 and u_2 on Ω_2 which verify together the same conditions as the global scheme. Since the global scheme provides existence and unicity, the converged domain decomposition solution is the same as the monodomain solution. Note that in the overlapping case, no additional condition on the parameters p_1 and p_2 of the transmission condition are imposed, the proof works for Dirichlet, Neumann, Robin and

also for Ventcel conditions. Even asymmetric combinations are possible like Dirichlet-Neumann or Neumann-Ventcel for instance.

Concerning the convergence properties of the hybrid finite volume scheme we refer to the work of Droniou et al. in [24] where a unified approach of mimetic finite difference, hybrid finite volume and mixed finite volume approaches is established and convergence results are developed. Furthermore, we refer to the work of da Veiga et al. in [21] where a unified handling of convection terms in hybrid finite volumes is described.
We recall two major results, the first (confirm [21, Theorem 3.7, page 18]) concerns the convergence of the discrete solution:

Theorem 3.1 (*Convergence of the discrete solution and the discrete gradient*)

Let $u \in H_0^1(\Omega)$ be the weak solution of problem (3.4) on Ω with Dirichlet boundary conditions on $\partial\Omega$. Ω is a bounded, open, polygonal subset of $\mathbb{R}^d$ with $d \geq 1$. Let $a : \Omega \to M_d(\mathbb{R})$ a bounded, measurable, symmetric and uniformly elliptic tensor. A possible right hand side term of (3.4) has to be in $L^2(\Omega)$. $\vec{b} \in C^1(\overline{\Omega})^d$ is such that $\mathrm{div}(\vec{b}) \geq 0$. Let $(u_h, \widetilde{F}_h)$ be the numerical solution as presented in this chapter and according to some regularity conditions on the family of meshes presented in [21, section 3.1.2, page 18]. Then, for the the meshsize $h \to 0$ there holds that

1. $u_h \to u$ in $L^r(\Omega)$ for all $r < \frac{2d}{d-2}$,
2. $\mathbf{v}_h(\widetilde{F}_h) \to \nabla u$ in $L^2(\Omega)^d$ where $\mathbf{v}_h(\widetilde{F}_h) : \Omega \to \mathbb{R}^d$ is the piecewise-constant function equal to the constant approximation of F on cell K.

The second result (confirm [21, Corollary 3.13, page 29]) concerns the convergence order, in the case of advection-diffusion problems the general convergence is of order 1 for the discrete solution u_h:

Theorem 3.2 (*Convergence order of the method*)

Under the hypotheses of [21, Theorem 3.11, page 25] it holds

$$\|u^I - u_h\|_{L^r(\Omega)} \lesssim h\|u\|_{H^2(\Omega)},$$

where $r = \frac{2d}{d-2}$ if $d > 2$ and $r < +\infty$ if d=2 and u^I is the sequence of interpolated fields of u on the mesh elements for a sequence of meshes with $h \to 0$.

Note that for the general method including advection and diffusion, only a first order convergence is given. Nevertheless, the numerical convergence in convection-dominated cases is at least of superconvergence type (cf. [21]). In the case of pure diffusion and under certain regularity conditions a second order convergence can be proven (confirm [10]). Note that these theoretical results are in accordance with the numerical results of the prototype code presented in appendix 3.B.

Remark that the design of the numerical scheme implicitly introduces a notion of domain decomposition on every face: we introduce an artificial interface on every face. The coupling conditions are the flux equity and the face value equity. Coupling conditions are directly used, therefore, this scheme might be seen as a substructuring method on every face-interface. Note furthermore that the number of primary unknowns does not change compared to the classical approach. Moreover, the underlying numerical structures are equal. Extending a classical finite volume code such that it uses the hybrid finite volume scheme is therefore straightforward whenever orthogonal meshes with a two-point discretisation and diagonal diffusion tensors are supposed.

In this work, we restrict ourselves to orthogonal meshes. For the case of non-orthogonal meshes, we would have to use a multi-point approximation rather than a two-point approximation. An appropriate Multi Point Flux Approximation Scheme (MPFA Scheme) has to be introduced for diffusion approximation and high order approximation schemes for advection are imaginable. We refer to [27] for the discussion of MPFA schemes.

3.3 Tangential Flux Information Along the Interface — Realisation of a Ventcel Condition

Ventcel conditions can be developed when approaching the optimal interface condition by a first order polynomial in Fourier space. They are a direct extension of Robin conditions by using not only the normal flux information on the interface but also the tangential derivative information. The Ventcel condition for the steady-state advection-diffusion operator has the continuous form

$$B(u,v) = -\vec{f}(u)\cdot\vec{n} + pu + q\left(-a\Delta_\tau u + \vec{b}_\tau\cdot\vec{\nabla}_\tau u\right), \tag{3.5}$$

where $\vec{f}(u)$ is the flux function, $\vec{n}$ is the outgoing normal vector and τ the associated tangential span on the interface of the considered subdomain, $p > 0$ and $q > 0$ are two real parameters. a is the diffusion coefficient, $\vec{b} \in \mathbb{R}^d$ is the Darcy field vector for space dimension $d = 1, 2, 3$.
The challenging issue in the realisation of a Ventcel condition of type (3.5) is how to realise numerically the tangential second order derivatives

$$-a\Delta_\tau u + \vec{b}_\tau\cdot\vec{\nabla}_\tau u,$$

when $d > 1$. We found inspiration by Halpern and Hubert in [43]: the basic idea is to establish $2(d-1)$ tangential fluxes along the interface face at the centres of the $2(d-1)$ edges connecting the considered interface face with other interface faces or the physical boundary of the numerical domain. With the help of the sum of two outflowing tangential fluxes (i. e. the first order derivatives) per tangential direction one can establish a finite difference approximation of the second order derivatives.

We exemplify this procedure for $d = 2$ (cf. figure 3.4). We introduce for every face σ_{K_i} that is

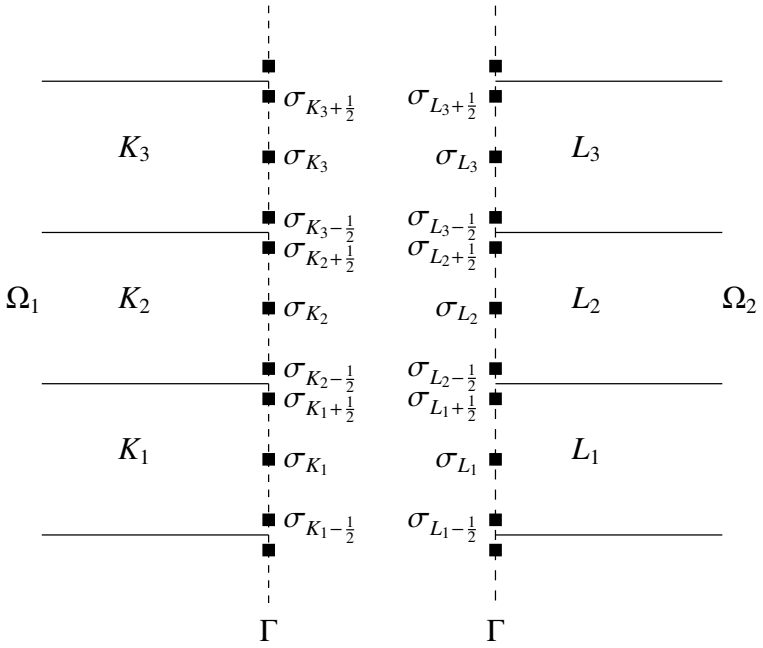


Figure 3.4: 2D mesh grid with face unknowns and additional edge unknowns for the transversal flux reconstruction.

an interface face two additional unknowns, $\sigma_{K_i+\frac{1}{2}}$ and $\sigma_{K_i-\frac{1}{2}}$, that are the unknowns on the edge of the face σ_{K_i} touching the two neighbouring interface faces $\sigma_{K_{i-1}}$ and $\sigma_{K_{i+1}}$, respectively. Those unknowns are the edge unknowns viewed from cell K_i. We can now impose the same hybrid finite volume scheme that we developed in section 3.2.2 for the $(d-1)$-dimensional manifold formed of the layer of interface faces. Concretely, the former interface faces have to be seen as new cells and the former edges between interface faces have to be seen as new faces. The face unknowns $u_{\sigma_{K_i}}$ and the edge unknowns $u_{\sigma_{K_i\pm\frac{1}{2}}}$ behave in the same way.
First, we impose the edge unknown continuity

$$\begin{aligned} u_{\sigma_{K_i+\frac{1}{2}}} &= u_{\sigma_{K_{i+1}-\frac{1}{2}}}, \\ u_{\sigma_{K_i-\frac{1}{2}}} &= u_{\sigma_{K_{i-1}+\frac{1}{2}}}, \end{aligned}$$

and the flux conservativity along the edges

$$\begin{aligned} \widetilde{F}_{\sigma_{K_i}\sigma_{K_i+\frac{1}{2}}} &= -\widetilde{F}_{\sigma_{K_{i+1}}\sigma_{K_{i+1}-\frac{1}{2}}}, \\ \widetilde{F}_{\sigma_{K_i}\sigma_{K_i-\frac{1}{2}}} &= -\widetilde{F}_{\sigma_{K_{i-1}}\sigma_{K_{i-1}+\frac{1}{2}}}. \end{aligned}$$

As before, we can eliminate all edge unknowns that lie in the inner of the interface manifold, i. e. only edge unknowns that lie on physical boundary of the domain are kept. Moreover, the tangential flux between two neighbouring faces σ_{K_i} and σ_{K_j} can be expressed only by use of the face unknowns $u_{\sigma_{K_i}}$ and $u_{\sigma_{K_j}}$ without explicit use of the edge unknowns. For edge unknowns

that touch the physical boundary of the domain, we can use either the standard reconstruction formula of the tangential flux and introduce an additional boundary condition equation that is provided by the boundary condition or we can eliminate directly the boundary edge unknown.

Having introduced the notion of tangential fluxes on the interface faces we can approach the tangential derivative by

$$\operatorname{div}_\tau\left(-a\nabla_\tau u_{\sigma_{K_i}} + \vec{b}_\tau u_{\sigma_{K_i}}\right) \approx \frac{\left[(-a\nabla_\tau u + \vec{b}_\tau u)\cdot\tau\right]_{x_{\sigma_{K_i}-\frac{1}{2}}}^{x_{\sigma_{K_i}+\frac{1}{2}}}}{|x_{\sigma_{K_i}+\frac{1}{2}} - x_{\sigma_{K_i}-\frac{1}{2}}|} = \frac{\frac{1}{\mathrm{m}\left(\sigma_{K_i+\frac{1}{2}}\right)}\widetilde{F}_{\sigma_{K_i}\sigma_{K_i+\frac{1}{2}}} + \frac{1}{\mathrm{m}\left(\sigma_{K_i-\frac{1}{2}}\right)}\widetilde{F}_{\sigma_{K_i}\sigma_{K_i-\frac{1}{2}}}}{|x_{\sigma_{K_i}+\frac{1}{2}} - x_{\sigma_{K_i}-\frac{1}{2}}|},$$

where $\mathrm{m}\left(\sigma_{K_i\pm\frac{1}{2}}\right)$ is the measure of the edge $\sigma_{K_i\pm\frac{1}{2}}$.
Note that for $d = 3$ there are two tangential directions and hence we have to create four additional edge unknowns per interface face. The previously developed approximations hold for every tangential direction and the extension is straightforward.

Finally, besides the lack of complete theory concerning domains with reentrant corners, we want to stress two problematic issues concerning Ventcel conditions with corners. Suppose the interface to have corners (cf. figure 3.5). First, the reconstruction of a tangential flux at the

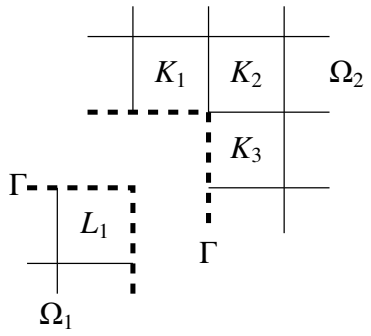


Figure 3.5: 2D mesh grid with corner at the interface.

edges of the corner of the interface in Ω_1 requires the access to the values in the inner of the complementary subdomain Ω_2. The access to an information of the complementary subdomain besides the furnished value of the transmission condition locally on the interface can be criticised to be not conform to the domain decomposition concept as it is stated on a continuous level: the information exchange is only locally done for the values of the transmission operator and it is not allowed to access directly to other information. A way to overcome this dilemma is to reconstruct a kind of "tangential flux around the corner": suppose therefore that the corner does not exist and that the two interface faces touching at the corner edge parallel. In this case, one can simply proceed as if no corner exists and reconstruct the flux information by use of the two face unknowns of the interface faces. The resulting flux may be seen as a flux that starts from one face centre to the corner edge, turns direction and finishes finally from the corner edge to the face centre. This approach has to be seen more as information about the variation between

two neighbouring interface interface faces than as tangential derivative in a strict mathematical sense.
Second, we remark that Ventcel conditions can perturb the global performance of the optimised domain decomposition algorithm by degenerating in the corner vicinity for fine meshes. We refer to Chniti et al. who have developed (cf. [18]) and tested in the context of finite elements (cf. [19]) a locally modified Ventcel condition such that the global performance of the Ventcel condition does not degenerate in the corner vicinity when the mesh is refined.

3.4 Time-Space Domain Decomposition

Schwarz-type domain decomposition methods decompose the problems on a continuous and global way. The resulting subproblems that arise in the domain decomposition algorithm are formulated continuously in time and space and globally in time. As a result, the choice of the numerical method and the space and time discretisations in the subdomains can be chosen totally independently. We are especially interested in the case, when discretisation sizes differ in the subdomains. In this case, the transfer of the interface values which represent time-space values discretised with a discretisation Δ_1 have to be used in the complementary subdomain which is discretised with discretisation Δ_2. A grid projection of the interface values between two discrete grids has to be performed. In [30], [29], [7], [45], [44], [46] different ways of how to combine Schwarz waveform relaxation methods with non-conforming time and space discretisations in various contexts have been proposed, studied and applied numerically. We present in the following, two different projection algorithms, one for the time-projection in a context of an implicit Euler discretisation and one for the space projection using a hybrid finite volume discretisation.

3.4.1 Projection Between Different Time Grids

In the case of different time grids in the subdomains Ω_1 and Ω_2, one needs to transfer interface condition values from one time grid to another. In figure 3.6 we show two different time grids t^a and t^b in a time window $[t_0, T]$. We call λ_i^a, $i = 0, \ldots, I$ the values of a function $\lambda(t)$ on the time grid t_i^a and λ_j^b, $j = 0, \ldots, J$ the values on the time grid t_j^b. Suppose the following matching:

$$\begin{aligned} t_0^a &= t_0^b = t_0, \\ t_I^a &= t_J^b = T, \\ \lambda_0^a &= \lambda_0^b. \end{aligned}$$

As to our application, for a given value of the function $\lambda(t)$ on the time grid t^a, we search the values on the time grid t^b. As we do not use an integral or variational formulation in the time discretisation scheme, we are able to use the simplest transfer strategy that is a piecewise affine

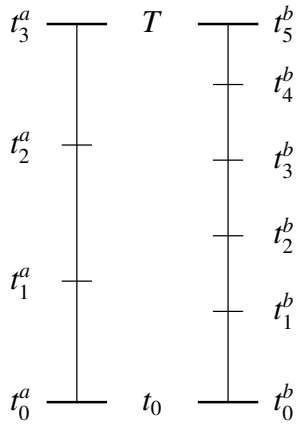


Figure 3.6: Two different time grids t^a and t^b in a time window $[t_0, T]$

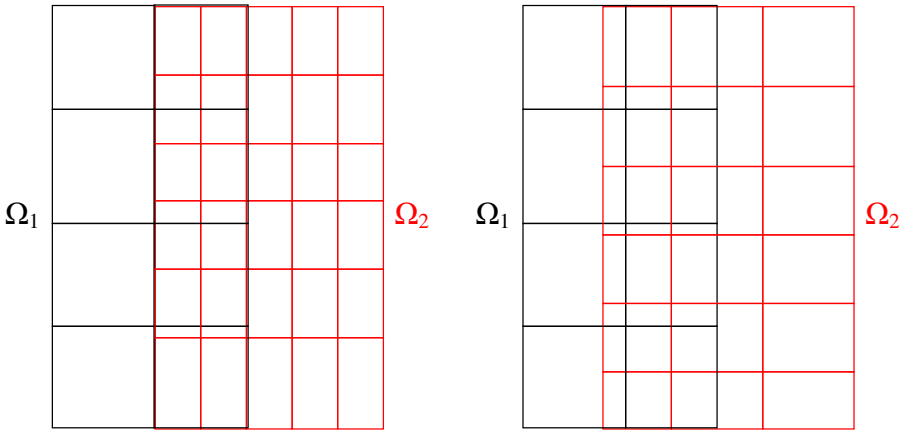
interpolation. This strategy is based on the vision of numerically solving an ordinary differential equation by an Euler method: for a function $\lambda(t)$ described by the given ordinary differential equation $\partial_t \lambda = f(\lambda, t)$ with initial value $\lambda(t^0) = \lambda^0$, we search a numerical approximation on the time grid t^a. Euler methods approach the partial derivative by $\partial_t \lambda \approx (\lambda^{n+1} - \lambda^n)/(t^{n+1} - t^n)$ which corresponds to the slope of an affine function on the interval $[t^n, t^{n+1}]$. In the explicit case one sets then $\partial_t \lambda = f(\lambda^n, t^n)$, in the implicit case $\partial_t \lambda = f(\lambda^{n+1}, t^{n+1})$, i. e. the slope of the affine discrete approximation in the interval $[t^n, t^{n+1}]$ is set to the slope of the tangent of the exact function at t^n in the explicit and at t^{n+1} in the implicit case. In this context of an affine approximation on the time grid t^a it is natural to use a projection between different time grids which is based on an affine interpolation. Given the values of the function λ_i^a and λ_{i+1}^a, we define for a time $t_j^b \in [t_i^a, t_{i+1}^a]$ the value

$$\lambda_j^b = \lambda_i^a + (t_j^b - t_i^a) \frac{\lambda_{i+1}^a - \lambda_i^a}{t_{i+1}^a - t_i^a}.$$

In a numerical context, this projection strategy can be optimally realised using two pointers indicating the actual positions in the time grids t^a and t^b, no cost-intensive iterative search has to be done.

3.4.2 Projection Between Different Space Grids

In the case of different space discretisations in the subdomains Ω_1 and Ω_2, one needs to transfer interface condition values from one space grid to another. Note that we consider only the case where the interface faces of one subdomain coincide with a layer of faces of the complementary subdomain. This is always ensured if the subdomains do not overlap and the interface faces of both subdomains lie in the same layer. In the case of overlapping subdomains, every interface face of one subdomain needs to be totally recovered by one or more faces that lie in the same plane as the interface face. This property has to hold for all interface faces of both subdomains. In figure 3.7 we exemplify this issue by two subdomain choices. While in figure 3.7a the interface



(a) Feasible discretisations (b) Unfeasible discretisations

Figure 3.7: Different space discretisations in the subdomains

faces of one subdomain lie all in a plane with inner faces of the complementary subdomain, in figure 3.7b, they do not lie in a plane with faces from the complementary subdomain.

In [4], Achdou et al. developed a method to glue non-conforming grids with Robin transmission conditions in the finite volume case with non-overlapping subdomains. The associated error analysis is based on non-overlapping subdomains but the basic idea, that we present in the following applies, under the given conditions, also to overlapping subdomains.
For one subdomain Ω_i, the transmission condition (Robin transmission condition for instance) is realised for every interface face σ_m

$$-\vec{f}(u_i^k)\cdot\vec{n}_{\sigma_m} + pu_i^k = g, \quad \text{on } \sigma_m,$$

where g is the value reconstructed in the complementary subdomain Ω_j. One can couple both by using an integral formulation

$$\int_{\sigma_m}\left(-\vec{f}(u_i^k)\cdot\vec{n}_{\sigma_m} + pu_i^k\right) = \sum_{\sigma_n\in\varepsilon_j}\int_{\sigma_m\cap\sigma_n}\left(-\vec{f}(u_j^{k-1})\cdot\vec{n}_{\sigma_m} + pu_j^{k-1}\right),$$

where ε_j is the set of faces of the mesh in Ω_j. Supposing the discrete values to be constant on every face, one can express the projection by

$$-\vec{f}(u_i^k)\cdot\vec{n}_{\sigma_m} + pu_i^k = \frac{1}{\mathrm{m}(\sigma_m)}\sum_{\sigma_n\in\varepsilon_j}\mathrm{m}(\sigma_m\cap\sigma_n)\left(-\vec{f}(u_j^{k-1})\cdot\vec{n}_{\sigma_m} + pu_j^{k-1}\right),$$

where $\mathrm{m}(\sigma_m)$ is the measure of the face σ_m. By this equation, one can easily calculate in a preliminary phase all projection weights $\alpha_{mn} := \frac{\mathrm{m}(\sigma_m\cap\sigma_n)}{\mathrm{m}(\sigma_m)}$ and perform the projection by a simple matrix-vector multiplication.

In 2D, the projection algorithm can be easily implemented, the algorithm is close to the time projection algorithm presented in the previous section. The difficult and cost-intensive task appears in the projection weights calculation in 3D: the calculation of the intersections $\sigma_m \cap \sigma_n$ of two faces of two different subdomains which are of dimension 2. Gander and Japhet developed in [31] an algorithm that realises the projection between two different triangular meshes in 2D that has a linear complexity. This algorithm is based on an intelligent way of running through the faces by using as much as information about the intersection as possible. The extension of the algorithm to rectangles or polyhedrons is straightforward, it is not necessary to cut the elements into basic triangles.

Conclusion

In this chapter, we have treated several technical issues appearing in the prototype phase: first, we have motivated the choice of finite volumes as underlying structure of the prototype code. Then, we discussed two different ways of realising a domain decomposition on the discrete context with their advantages and drawbacks. The realisation of sophisticated transmission conditions like Robin or Ventcel conditions led to use a decomposition within the faces of the finite volume geometry. This choice was justified by the natural appearance of the major numerical ingredients. We then exemplified by a counterexample why the standard finite volume approach is not directly exploitable in a domain decomposition context and proposed a hybrid finite volume scheme that offers a natural consistency between domain decomposition and global monodomain approach. Basing on this hybrid finite volume scheme we exemplified the tangential second order derivative reconstruction along the interface used for the numerical realisation of Ventcel transmission conditions. Finally, we presented two algorithms for the projection of interface values between two different grids in time and space.

The numerical prototyping provided the possibility to study several issues in the numerical realisation of a domain decomposition approach in the finite volume context. The results presented in this chapter are of technical character and can be seen as a catalogue of experiences gained during this thesis. All low-level issues concerning the discrete schemes, transmission conditions and projection algorithms have been validated numerically (see appendix). Finally, the prototype code allows to realise the validation of the high-level mathematical results concerning the domain decomposition that are presented in the following two chapters 4 and 5.

3.A Numerical Validation of the Time Integration Scheme and the Time Projection Algorithm

We treat the coupled two-species reactive transport system

$$\begin{aligned} \partial_t(\phi u) + \operatorname{div}(-a\nabla u + bu) - k(v - cu) &= f_u \quad \text{on } \Omega \times [0, T], \\ \partial_t(\phi v) \qquad\qquad\quad + k(v - cu) &= f_v \quad \text{on } \Omega \times [0, T], \end{aligned} \tag{3.6}$$

in 2D, i. e. $\Omega \subset \mathbb{R}^2$. We are interested in the convergence of the numerical approximation using an optimised Schwarz waveform relaxation method with different time discretisations in the subdomains. We used Robin transmission conditions

$$\mathcal{B}_1(u, v) = \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u, \quad \mathcal{B}_2(u, v) = \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u,$$

and Ventcel transmission conditions

$$\begin{aligned} \mathcal{B}_1(u, v) &= \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\mathbf{y}} u + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} u - kv + kcu\right), \\ \mathcal{B}_2(u, v) &= \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\mathbf{y}} u + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} u - kv + kcu\right), \end{aligned}$$

where the associated optimised parameters p^* and (p^*, q^*) are studied in chapter 4.3.

We define

$$\begin{aligned} u(x, y, t) &= \sin(t)\cos(x)\cos(y) + \cos(t)\sin(x)\cos(y) + \cos(t)\cos(x)\sin(y) + tx^2y^3, \\ v(x, y, t) &= 1 - \cos(t)\sin(x)\sin(y) - \sin(t)\cos(x)\sin(y) - \sin(t)\sin(x)\cos(y) - tx^3y^2 \end{aligned}$$

to be the exact solution of problem (3.6). The aim here is to validate the correct implementation of the presented algorithms and discretisation schemes. The right hand side source terms can be obtained by applying the left hand side operators of (3.6) to the exact solution. The partial derivatives of the exact solution are given by:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \cos(t)\cos(x)\cos(y) - \sin(t)\sin(x)\cos(y) - \sin(t)\cos(x)\sin(y) + x^2y^3, \\ \frac{\partial u}{\partial x} &= -\sin(t)\sin(x)\cos(y) + \cos(t)\cos(x)\cos(y) - \cos(t)\sin(x)\sin(y) + 2txy^3, \\ \frac{\partial^2 u}{\partial x^2} &= -\sin(t)\cos(x)\cos(y) - \cos(t)\sin(x)\cos(y) - \cos(t)\cos(x)\sin(y) + 2ty^3, \\ \frac{\partial u}{\partial y} &= -\sin(t)\cos(x)\sin(y) - \cos(t)\sin(x)\sin(y) + \cos(t)\cos(x)\cos(y) + 3tx^2y^2, \\ \frac{\partial^2 u}{\partial y^2} &= -\sin(t)\cos(x)\cos(y) - \cos(t)\sin(x)\cos(y) - \cos(t)\cos(x)\sin(y) + 6tx^2y, \\ \frac{\partial v}{\partial t} &= \sin(t)\sin(x)\sin(y) - \cos(t)\cos(x)\sin(y) - \cos(t)\sin(x)\cos(y) - x^3y^2. \end{aligned}$$

We chose the chemical equilibrium parameter as $c = 0.5$. The other parameters are chosen as $\phi = 1, a = 10^{-3}, \vec{b} = \begin{pmatrix} 10^{-2} \\ 10^{-2} \end{pmatrix}$ and $k = 10^{-2}$. We set our simulation domain to $\Omega \times [0,\ T] = [0,\ 1]^2 \times [0,\ 10]$ with 40 grid cells in both x and y direction, respectively, i. e. $\Delta x = \Delta y = 0.025$ fixed for all simulations in this section. We decompose the spatial domain into two subdomains. In the non-overlapping case, we define the subdomains by $\Omega_1 = [0,\ 0.5] \times [0,\ 1]$, $\Omega_2 = [0.5,\ 1] \times [0,\ 1]$, in the overlapping case, we define the subdomains by $\Omega_1 = [0,\ 0.5 + \Delta x] \times [0,\ 1]$, $\Omega_2 = [0.5 - \Delta x,\ 1] \times [0,\ 1]$. We apply an OSWR algorithm with Robin and Ventcel conditions until the variation of the interface conditions between two iterations is smaller than 10^{-8}. After that, we calculate the relative discrete L_2 error between the numerical solution $\tilde{u}$ and the exact solution u. We define the discrete L_2 norm for a function g that is affine in every time interval $[t^n,\ t^{n+1}]$ and constant in every space cell $K \in \mathcal{T}$ as

$$
\begin{aligned}
\mathrm{L}_2(g) :=& \left(\int_0^T \int_\Omega g^2 \,\mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} = \left(\sum_{n=1}^{N} \sum_{K \in \mathcal{T}} \int_{t^{n-1}}^{t^n} \int_K g^2 \,\mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} \\
=& \left(\sum_{n=1}^{N} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \int_{t^{n-1}}^{t^n} \left(g_K^n + \frac{t - t^{n-1}}{t^n - t^{n-1}} \left(g_K^n - g_K^{n-1} \right) \right)^2 \,\mathrm{d}t \right)^{\frac{1}{2}} \\
=& \left(\sum_{n=1}^{N} \sum_{K \in \mathcal{T}} \mathrm{m}(K) \left(\frac{1}{3} \left((t^n)^3 - \left(t^{n-1} \right)^3 \right) m^2 + \left((t^n)^2 - \left(t^{n-1} \right)^2 \right) mc + \left(t^n - t^{n-1} \right) c^2 \right) \right)^{\frac{1}{2}},
\end{aligned}
$$

with

$$
\begin{aligned}
m =& \frac{g_K^n - g_K^{n-1}}{t^n - t^{n-1}}, \\
c =& g_K^{n-1} - t^{n-1} \frac{g_K^n - g_K^{n-1}}{t^n - t^{n-1}},
\end{aligned}
$$

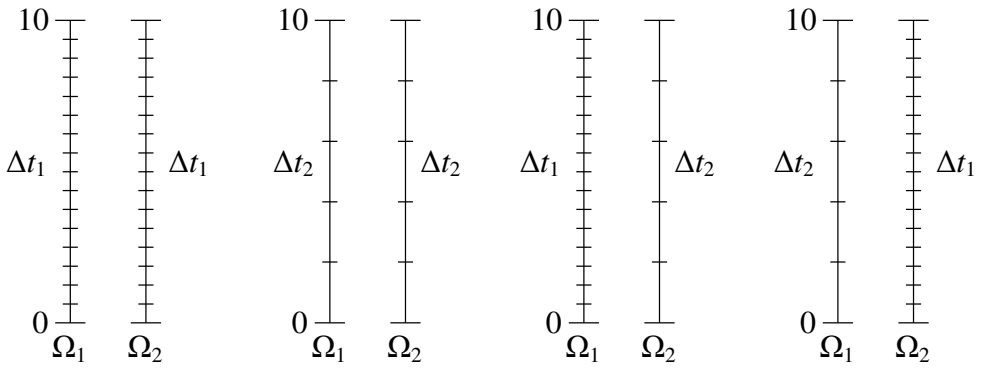
where g_K^n denotes the discrete value at cell K with measure $\mathrm{m}(K)$ of the mesh $\mathcal{T}$ at time step $t = t^n$ where $t^0 = 0$ and $t^N = T$. The relative discrete L_2 error norm is then defined by

$$
\mathrm{L}_{2\mathrm{rel}}(u, \tilde{u}) := \frac{\mathrm{L}_2(u - \tilde{u})}{\mathrm{L}_2(u)}.
$$

We compare the behaviour of this error when we refine the temporal discretisations in the two subdomains. Therefore, we use 4 different initial time grids that are refined by halving the time step in each subdomain at each refinement level. The different time grid configurations are:

- Case 1 (conforming fine grid): initial time step is $\Delta t_1 = 10/16$ in both subdomains (cf. figure 3.8a).

- Case 2 (conforming coarse grid): initial time step is $\Delta t_2 = 10/5$ in both subdomains (cf. figure 3.8b).
- Case 3 (non-conforming fine coarse): initial time step is $\Delta t_1 = 10/16$ in Ω_1 and $\Delta t_2 = 10/5$ in Ω_2 (cf. figure 3.8c).
- Case 4 (non-conforming coarse fine): initial time step is $\Delta t_2 = 10/5$ in Ω_1 and $\Delta t_1 = 10/16$ in Ω_2 (cf. figure 3.8d).



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Figure 3.8: Initial grids for time projection algorithm validation

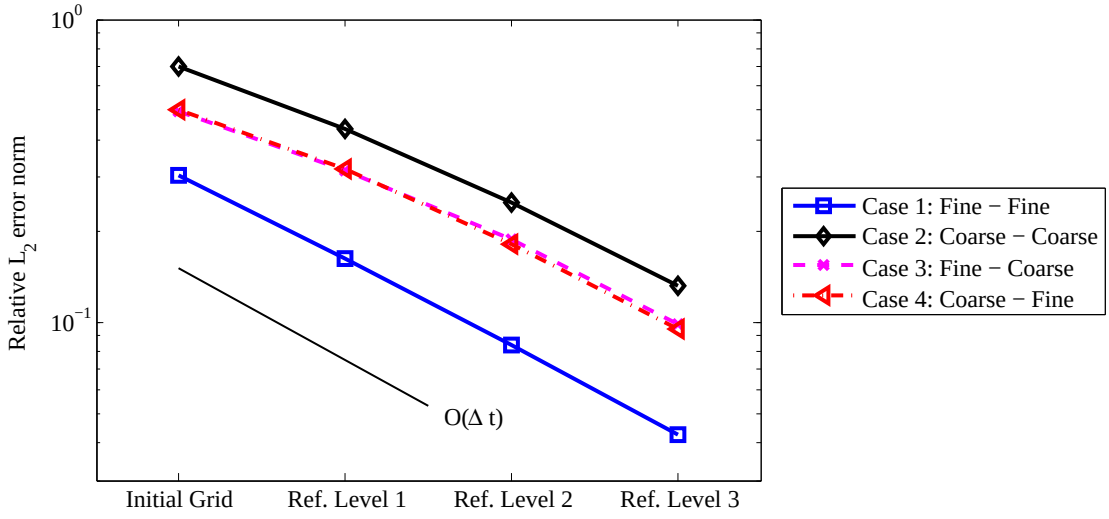


Figure 3.9: Behaviour of the relative L_2 error norm of the converged domain decomposition solution with optimised Robin interface conditions for the four different cases with three refinement levels of the initial grid without overlap.

In tables 3.1 and 3.2 we observe first that the errors of a global monodomain simulation and a converged domain decomposition are equal in the case of Robin and Ventcel transmission conditions. This justifies that the error committed by a converged domain decomposition algorithm is negligible compared to the error of the numerical scheme and the projection algorithm. Note furthermore that the projection algorithm matches identity when both time grids are the same as in case 1 and 2.
Next, we can observe that all errors in case 1 and case 2 behave like order 1, i. e. halving the step size halves the error. For the coarse grids, this reduction factor is slightly deteriorated because the asymptotic behaviour can only be expected for fine meshes.
In tables 3.1 and 3.2 we give the errors of the converged domain decomposition solution with optimised Robin and optimised Ventcel transmission conditions for case 3 and case 4 that have non-conforming time discretisations in the different subdomains. Comparing figure 3.9, one can observe that the errors in the case of non-conforming time discretisations lie between the errors of the conforming fine grid and the conforming coarse grid. Moreover, one can observe that both errors of the non-conforming grids are not only quite close to each other but they do also behave like the errors for the conforming grid cases, i. e. one observes a reduction factor that is equal to the reduction factor for the cases with conforming meshes. Finally, one observes that the errors in the subdomains with finer discretisation are in the same range as the the errors of the same discretisation in the conforming case. This shows that the advantage of locally refined subdomains is preserved. The error introduced by the projection algorithm is small compared to the discretisation error in the subdomains

In the case of Ventcel transmission conditions one observes the same behaviour.

In tables 3.5, 3.6, 3.7 and 3.8 we show the results in the overlapping case. The behaviour of the projection algorithm is the same as in the non-overlapping case.

Monodomain solution

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.92 \cdot 10^{-1}$		$3.16 \cdot 10^{-1}$		$3.04 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.88	$1.67 \cdot 10^{-1}$	1.88	$1.62 \cdot 10^{-1}$	1.88
64	64	$8.03 \cdot 10^{-2}$	1.94	$8.68 \cdot 10^{-2}$	1.94	$8.35 \cdot 10^{-2}$	1.93
128	128	$4.08 \cdot 10^{-2}$	1.97	$4.42 \cdot 10^{-2}$	1.97	$4.25 \cdot 10^{-2}$	1.97

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.92 \cdot 10^{-1}$		$3.16 \cdot 10^{-1}$		$3.04 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.88	$1.67 \cdot 10^{-1}$	1.89	$1.62 \cdot 10^{-1}$	1.88
64	64	$8.03 \cdot 10^{-2}$	1.94	$8.68 \cdot 10^{-2}$	1.93	$8.35 \cdot 10^{-2}$	1.93
128	128	$4.08 \cdot 10^{-2}$	1.97	$4.42 \cdot 10^{-2}$	1.96	$4.25 \cdot 10^{-2}$	1.97

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.92 \cdot 10^{-1}$		$3.16 \cdot 10^{-1}$		$3.04 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.88	$1.67 \cdot 10^{-1}$	1.89	$1.62 \cdot 10^{-1}$	1.88
64	64	$8.03 \cdot 10^{-2}$	1.94	$8.68 \cdot 10^{-2}$	1.93	$8.35 \cdot 10^{-2}$	1.93
128	128	$4.08 \cdot 10^{-2}$	1.97	$4.42 \cdot 10^{-2}$	1.96	$4.25 \cdot 10^{-2}$	1.97

Table 3.1: Case 1 (Fine - Fine): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a monodomain solution as well as for a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions in the non-overlapping case.

Monodomain solution

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.57 \cdot 10^{-1}$		$7.35 \cdot 10^{-1}$		$6.95 \cdot 10^{-1}$	
10	10	$4.24 \cdot 10^{-1}$	1.55	$4.47 \cdot 10^{-1}$	1.55	$4.36 \cdot 10^{-1}$	1.60
20	20	$2.40 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.77	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.91	$1.31 \cdot 10^{-1}$	1.90

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.57 \cdot 10^{-1}$		$7.35 \cdot 10^{-1}$		$6.95 \cdot 10^{-1}$	
10	10	$4.24 \cdot 10^{-1}$	1.55	$4.47 \cdot 10^{-1}$	1.64	$4.36 \cdot 10^{-1}$	1.60
20	20	$2.40 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.74	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.90	$1.31 \cdot 10^{-1}$	1.90

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.57 \cdot 10^{-1}$		$7.35 \cdot 10^{-1}$		$6.95 \cdot 10^{-1}$	
10	10	$4.24 \cdot 10^{-1}$	1.55	$4.47 \cdot 10^{-1}$	1.64	$4.36 \cdot 10^{-1}$	1.60
20	20	$2.40 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.74	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.90	$1.31 \cdot 10^{-1}$	1.90

Table 3.2: Case 2 (Coarse - Coarse): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions as well as for a monodomain solution in the non-overlapping case.

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	5	$3.14 \cdot 10^{-1}$		$6.83 \cdot 10^{-1}$		$4.95 \cdot 10^{-1}$	
32	10	$1.68 \cdot 10^{-1}$	1.87	$4.32 \cdot 10^{-1}$	1.58	$3.18 \cdot 10^{-1}$	1.56
64	20	$8.63 \cdot 10^{-2}$	1.94	$2.54 \cdot 10^{-1}$	1.70	$1.87 \cdot 10^{-1}$	1.70
128	40	$4.37 \cdot 10^{-2}$	1.98	$1.34 \cdot 10^{-1}$	1.89	$9.88 \cdot 10^{-2}$	1.89

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	5	$3.03 \cdot 10^{-1}$		$6.95 \cdot 10^{-1}$		$4.98 \cdot 10^{-1}$	
32	10	$1.61 \cdot 10^{-1}$	1.88	$4.37 \cdot 10^{-1}$	1.59	$3.20 \cdot 10^{-1}$	1.56
64	20	$8.35 \cdot 10^{-2}$	1.93	$2.56 \cdot 10^{-1}$	1.71	$1.87 \cdot 10^{-1}$	1.71
128	40	$4.26 \cdot 10^{-2}$	1.96	$1.35 \cdot 10^{-1}$	1.90	$9.91 \cdot 10^{-2}$	1.89

Table 3.3: Case 3 (Fine - Coarse): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions in the non-overlapping case.

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	16	$6.49 \cdot 10^{-1}$		$3.60 \cdot 10^{-1}$		$5.05 \cdot 10^{-1}$	
10	32	$4.19 \cdot 10^{-1}$	1.55	$1.87 \cdot 10^{-1}$	1.92	$3.20 \cdot 10^{-1}$	1.58
20	64	$2.36 \cdot 10^{-1}$	1.77	$9.49 \cdot 10^{-2}$	1.97	$1.80 \cdot 10^{-1}$	1.78
40	128	$1.24 \cdot 10^{-1}$	1.90	$4.78 \cdot 10^{-2}$	1.99	$9.43 \cdot 10^{-2}$	1.91

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	16	$6.55 \cdot 10^{-1}$		$3.42 \cdot 10^{-1}$		$5.02 \cdot 10^{-1}$	
10	32	$4.23 \cdot 10^{-1}$	1.55	$1.76 \cdot 10^{-1}$	1.94	$3.20 \cdot 10^{-1}$	1.57
20	64	$2.39 \cdot 10^{-1}$	1.77	$8.99 \cdot 10^{-2}$	1.96	$1.80 \cdot 10^{-1}$	1.78
40	128	$1.25 \cdot 10^{-1}$	1.91	$4.60 \cdot 10^{-2}$	1.95	$9.46 \cdot 10^{-2}$	1.90

Table 3.4: Case 4 (Coarse - Fine): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions in the non-overlapping case.

Monodomain solution

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.94 \cdot 10^{-1}$		$3.17 \cdot 10^{-1}$		$3.04 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.56	$1.67 \cdot 10^{-1}$	1.65	$1.62 \cdot 10^{-1}$	1.88
64	64	$8.06 \cdot 10^{-2}$	1.77	$8.67 \cdot 10^{-2}$	1.74	$8.35 \cdot 10^{-2}$	1.93
128	128	$4.09 \cdot 10^{-2}$	1.91	$4.41 \cdot 10^{-2}$	1.90	$4.25 \cdot 10^{-2}$	1.97

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.94 \cdot 10^{-1}$		$3.17 \cdot 10^{-1}$		$3.05 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.88	$1.67 \cdot 10^{-1}$	1.89	$1.62 \cdot 10^{-1}$	1.89
64	64	$8.06 \cdot 10^{-2}$	1.94	$8.67 \cdot 10^{-2}$	1.93	$8.37 \cdot 10^{-2}$	1.93
128	128	$4.09 \cdot 10^{-2}$	1.97	$4.41 \cdot 10^{-2}$	1.97	$4.25 \cdot 10^{-2}$	1.97

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	16	$2.94 \cdot 10^{-1}$		$3.17 \cdot 10^{-1}$		$3.05 \cdot 10^{-1}$	
32	32	$1.56 \cdot 10^{-1}$	1.88	$1.67 \cdot 10^{-1}$	1.89	$1.62 \cdot 10^{-1}$	1.89
64	64	$8.06 \cdot 10^{-2}$	1.94	$8.67 \cdot 10^{-2}$	1.93	$8.37 \cdot 10^{-2}$	1.93
128	128	$4.09 \cdot 10^{-2}$	1.97	$4.41 \cdot 10^{-2}$	1.97	$4.25 \cdot 10^{-2}$	1.97

Table 3.5: Case 1 (Fine - Fine): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions as well as for a monodomain solution in the overlapping case.

Monodomain solution

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.65 \cdot 10^{-1}$		$7.39 \cdot 10^{-1}$		$6.95 \cdot 10^{-1}$	
10	10	$4.27 \cdot 10^{-1}$	1.56	$4.48 \cdot 10^{-1}$	1.65	$4.36 \cdot 10^{-1}$	1.60
20	20	$2.41 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.74	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.90	$1.31 \cdot 10^{-1}$	1.90

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.65 \cdot 10^{-1}$		$7.39 \cdot 10^{-1}$		$7.01 \cdot 10^{-1}$	
10	10	$4.27 \cdot 10^{-1}$	1.56	$4.48 \cdot 10^{-1}$	1.65	$4.37 \cdot 10^{-1}$	1.60
20	20	$2.41 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.74	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.90	$1.31 \cdot 10^{-1}$	1.90

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	5	$6.65 \cdot 10^{-1}$		$7.39 \cdot 10^{-1}$		$7.01 \cdot 10^{-1}$	
10	10	$4.27 \cdot 10^{-1}$	1.56	$4.48 \cdot 10^{-1}$	1.65	$4.37 \cdot 10^{-1}$	1.60
20	20	$2.41 \cdot 10^{-1}$	1.77	$2.58 \cdot 10^{-1}$	1.74	$2.49 \cdot 10^{-1}$	1.75
40	40	$1.26 \cdot 10^{-1}$	1.91	$1.36 \cdot 10^{-1}$	1.90	$1.31 \cdot 10^{-1}$	1.90

Table 3.6: Case 2 (Coarse - Coarse): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions as well as for a monodomain solution in the overlapping case.

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	5	$3.04 \cdot 10^{-1}$		$6.82 \cdot 10^{-1}$		$4.91 \cdot 10^{-1}$	
32	10	$1.61 \cdot 10^{-1}$	1.89	$4.34 \cdot 10^{-1}$	1.57	$3.18 \cdot 10^{-1}$	1.54
64	20	$8.22 \cdot 10^{-2}$	1.95	$2.57 \cdot 10^{-1}$	1.69	$1.88 \cdot 10^{-1}$	1.69
128	40	$4.16 \cdot 10^{-2}$	1.98	$1.36 \cdot 10^{-1}$	1.90	$9.93 \cdot 10^{-2}$	1.90

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
16	5	$3.01 \cdot 10^{-1}$		$7.22 \cdot 10^{-1}$		$5.13 \cdot 10^{-1}$	
32	10	$1.58 \cdot 10^{-1}$	1.90	$4.49 \cdot 10^{-1}$	1.61	$3.27 \cdot 10^{-1}$	1.57
64	20	$8.11 \cdot 10^{-2}$	1.95	$2.60 \cdot 10^{-1}$	1.73	$1.89 \cdot 10^{-1}$	1.73
128	40	$4.11 \cdot 10^{-2}$	1.97	$1.38 \cdot 10^{-1}$	1.88	$1.01 \cdot 10^{-1}$	1.88

Table 3.7: Case 3 (Fine - Coarse): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions in the overlapping case.

Domain decomposition solution with Robin transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	16	$6.57 \cdot 10^{-1}$		$3.45 \cdot 10^{-1}$		$5.04 \cdot 10^{-1}$	
10	32	$4.23 \cdot 10^{-1}$	1.55	$1.78 \cdot 10^{-1}$	1.93	$3.21 \cdot 10^{-1}$	1.57
20	64	$2.40 \cdot 10^{-1}$	1.76	$9.04 \cdot 10^{-2}$	1.97	$1.81 \cdot 10^{-1}$	1.77
40	128	$1.26 \cdot 10^{-1}$	1.90	$4.56 \cdot 10^{-2}$	1.98	$9.50 \cdot 10^{-2}$	1.91

Domain decomposition solution with Ventcel transmission conditions

Discretisation		relative	redu-	relative	redu-	relative	redu-
$10\Delta t_1^{-1}$	$10\Delta t_2^{-1}$	L_2 in Ω_1	ction	L_2 in Ω_2	ction	L_2 in Ω	ction
5	16	$6.70 \cdot 10^{-1}$		$3.37 \cdot 10^{-1}$		$5.09 \cdot 10^{-1}$	
10	32	$4.29 \cdot 10^{-1}$	1.56	$1.72 \cdot 10^{-1}$	1.95	$3.22 \cdot 10^{-1}$	1.58
20	64	$2.42 \cdot 10^{-1}$	1.77	$8.75 \cdot 10^{-2}$	1.97	$1.81 \cdot 10^{-1}$	1.78
40	128	$1.27 \cdot 10^{-1}$	1.90	$4.42 \cdot 10^{-2}$	1.98	$9.56 \cdot 10^{-2}$	1.90

Table 3.8: Case 4 (Coarse - Fine): relative L_2 error norm for initial grid and three refinements of initial grids in the case of a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions in the overlapping case.

3.B Numerical Validation of the Finite Volume Scheme, Transmission Conditions and the Space Projection Algorithm

We are interested in the convergence of the numerical approximation when using an optimised Schwarz waveform relaxation method with different space discretisations in the subdomains. Therefore, we define

$$\begin{aligned} u(x, y, t) &= t \sin(x) \cos(y) + t \cos(x) \sin(y) + t x^2 y^3, \\ v(x, y, t) &= 1 - t \cos(x) \sin(y) - t \sin(x) \cos(y) - t x^3 y^2 \end{aligned}$$

to be the exact solution of problem (3.6). The right hand side source terms can be obtained by applying the left hand side of (3.6) to the exact solution. The partial derivatives of the exact solution are given by:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \sin(x) \cos(y) + \cos(x) \sin(y) + x^2 y^3, \\ \frac{\partial u}{\partial x} &= t \cos(x) \cos(y) - t \sin(x) \sin(y) + 2 t x y^3, \\ \frac{\partial^2 u}{\partial x^2} &= -t \sin(x) \cos(y) - t \cos(x) \sin(y) + 2 t y^3, \\ \frac{\partial u}{\partial y} &= -t \sin(x) \sin(y) + t \cos(x) \cos(y) + 3 t x^2 y^2, \\ \frac{\partial^2 u}{\partial y^2} &= -t \sin(x) \cos(y) - t \cos(x) \sin(y) + 6 t x^2 y, \\ \frac{\partial v}{\partial t} &= -\cos(x) \sin(y) - \sin(x) \cos(y) - x^3 y^2. \end{aligned}$$

We chose the chemical equilibrium parameter as $c = 0.5$. We study the behaviour of the spatial projection technique on two different cases:

- advective case: $a = 1 \cdot 10^{-2}$, $\vec{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- diffusive case: $a = 1$, $\vec{b} = \begin{pmatrix} 1 \cdot 10^{-2} \\ 1 \cdot 10^{-2} \end{pmatrix}$

The other parameters are chosen to be equal to one, i. e. $\phi = 1$, and $k = 1$ in both cases.

Note that the exact solution is linear in time and hence the error in time of the numerical approach is zero. Moreover, as we use implicit methods, we can choose the time and space discretisations independently. We chose the simulation domain to be $\Omega \times [0, T] = [0, 1]^2 \times$

[0, 1] with one time step, i. e. $\Delta t = 1.0$ fixed for all simulations in this section. For the non-overlapping case (cf. figures 3.10a and 3.10b), we decompose the spatial domain into two subdomains $\Omega_1 = [0, 0.5] \times [0, 1]$, $\Omega_2 = [0.5, 1] \times [0, 1]$. For the overlapping case (cf. figures 3.10c and 3.10d), we decompose the spatial domain into two subdomains $\Omega_1 = [0, 0.5 + \widetilde{dx}] \times [0, 1]$, $\Omega_2 = [0.5 - \widetilde{dx}, 1] \times [0, 1]$, where $\widetilde{dx} := \max\{\Delta x_1, \Delta x_2\}$ is the greater of the two discretisation parameters in x-direction of the two subdomains. This means that, in practice, the two subdomains overlap within a layer of two coarse grid cells. Moreover, for the overlapping case, we impose semi conforming discretisations in space, i. e. in y-direction, the discretisations can be totally non conforming while in x-direction, the interfaces of one domain have to coincide with a layer of faces of the complementary domain. This can, for instance, be achieved by choosing the coarser discretisation in x to be a whole-number multiple of the finer one. We apply an OSWR algorithm with optimised Robin and Ventcel conditions until the variation of the interface conditions between two iterations is smaller than 10^{-8}. After that, we calculate the discrete L_2 error in space and the discrete L_{inf} error in space between the numerical solution and the exact solution in both subdomains separately. We compare those errors to the same errors that one obtains doing a global mono-domain solution with the associated space discretisations. Since we use only one time step and the error at the initial state $t = t^0$ vanishes, we define the discrete L_2 error as

$$L_2(u - \tilde{u}) := \left(\sum_{K \in \mathcal{T}} m(K) \left(u(x_K, t^N) - \tilde{u}_K^N \right)^2 \right)^{\frac{1}{2}},$$

and the discrete L_{inf} error as

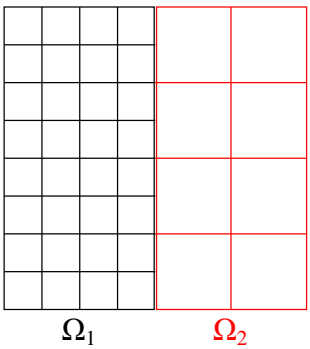
$$L_{inf}(u - \tilde{u}) := \max_{K \in \mathcal{T}} \left| u(x_K, t^N) - \tilde{u}_K^N \right|.$$

Concerning the spatial discretisations, we start with an initial grid, that is refined four times by halving the discretisation size in x- and y-direction. Finally, we use two different kinds of non-conforming grids:

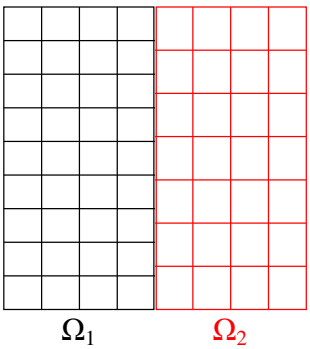
- Nested grids: the discretisation size in one domain is half the discretisation size in the other domain (cf. figures 3.10a and 3.10c).
- Non-nested grids: the discretisation sizes in both domains are of the same range but non-conforming (cf. figures 3.10b and 3.10d).

In figure 3.10, we show the initial grids in the nested and non-nested case for overlapping and non-overlapping subdomains.

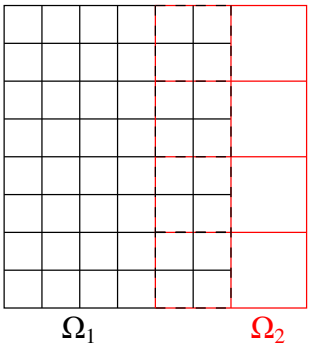
In tables 3.9 to 3.12 we give the results of the performed tests in the non-overlapping case. First of all, one observes that in all cases, advective and diffusive, for global mono-domain and domain decomposition solutions with Robin and Ventcel conditions, the numerical scheme converges in both the discrete L_2 and L_{inf} norms. In the diffusive case we obtain a second order



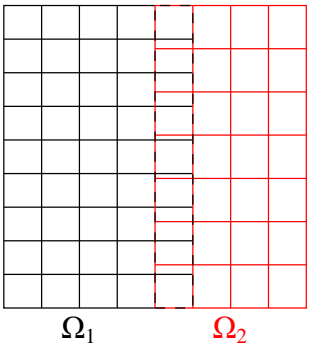
(a) Nested grids, non-overlapping subdomains



(b) Non-nested grids, non-overlapping subdomains



(c) Nested grids, overlapping subdomains



(d) Non-nested grids, overlapping subdomains

Figure 3.10: Initial grids for space projection algorithm validation

convergence (reduction factor close to 4) while in the advective case we obtain only a superconvergence. Those results are in agreement with the theoretical results presented in section 3.2.2.

In the advective case, for nested as well as non-nested grids, the discrete L_2 errors for the converged domain decomposition solutions are in the same range as the errors of the global monodomain solutions. The error for Ventcel conditions for the finest nested grid in the advective case are plotted in figure 3.11. Concerning the L_{inf} error in the subdomains, one can observe that for the most of the cases the errors are identical to those obtained with the global monodomain solution for the subdomain Ω_1, in the other subdomain Ω_2, the errors are only slightly higher but in the same range. As a consequence, the advantage of locally refined subdomains is preserved. The error introduced by the projection algorithm is small compared to the discretisation error in the subdomains and moreover the error has only a local influence (see figure 3.12).

In the diffusive case, for nested grids (see figure 3.13), the discrete L_2 errors for the converged domain decomposition solutions are in the same range as the errors of the global monodomain solutions in the coarse subdomain Ω_2 while the errors in the finer subdomain Ω_1 are higher than in the monodomain case. In fact, the error in the finer subdomain is limited by the error in the

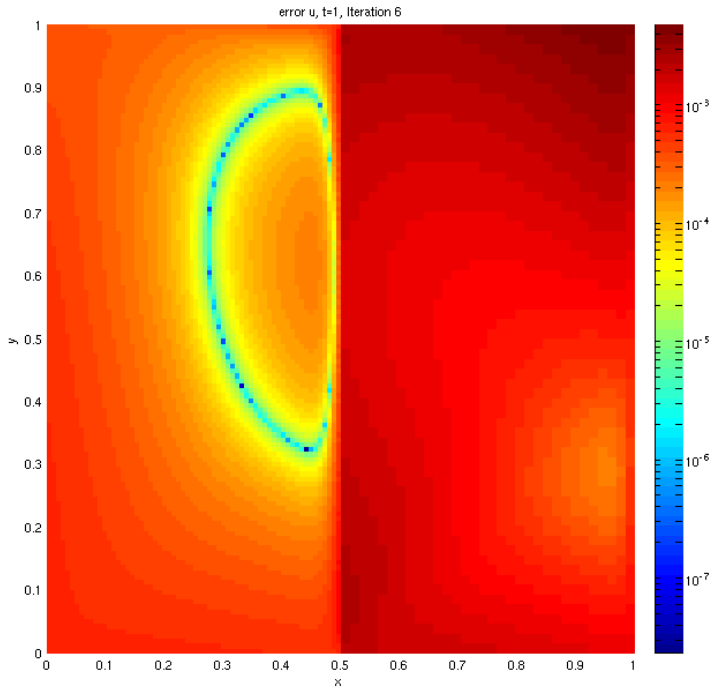


Figure 3.11: Error at the last iteration for the advective case with finest nested grids and Ventcel transmission conditions in the non-overlapping case.

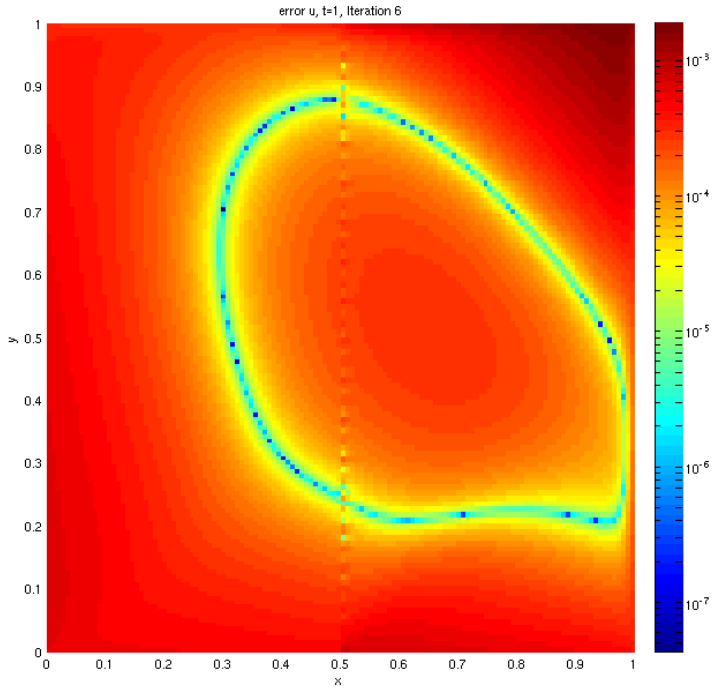


Figure 3.12: Error at the last iteration for the advective case with finest non-nested grids and Ventcel transmission conditions in the non-overlapping case.

coarser domain. Concerning the L_{inf} error, one observes for Robin conditions that the errors in the coarse subdomain Ω_2 are the same as those for a global monodomain coarse grid solution in Ω_2. Anyway, once again in the fine subdomain Ω_1, the errors are larger than the errors of a global monodomain fine grid simulation in Ω_1. This behaviour in the diffusive case with nested grids is due to the projection algorithm and the choice or the parameter p of the Robin transmission condition: according to [4, Theorem 5.3], the benefit of locally refined meshes as in the nested case is only preserved when the parameter p of the Robin transmission condition is chosen to be constant within mesh refinement. As we use here optimised parameters which are not constant within mesh refinement, the error in the fine subdomain is limited by the error in the coarse subdomain. The use of constant parameters has shown in numerical tests to be error-preserving also in the fine subdomain.
In the non-nested case or in the nested case with Ventcel conditions, the L_2 and the L_{inf} errors are not considerably influenced by the projection algorithm. Finally, we plot the error in the case of Ventcel conditions for non-nested grids 3.14.

In tables 3.13 to 3.16 we give the results of the performed tests in the overlapping case. The results are quite similar to those in the non-overlapping case. However, concerning the diffusive case, the influence of the error in the coarse subdomain on the error in the fine domain is though less visible and the convergence rate is not deteriorated as in the non-overlapping case. Indeed, using an overlap and reconstructing thus the interface values in the inner of a subdomain instead of reconstructing them at the boundary of a subdomain is less defective. To resume, one can say that using an overlap in the domain decomposition algorithm is not only faster in terms of convergence for the algorithm itself but is also less sensitive to the projection algorithm in space.

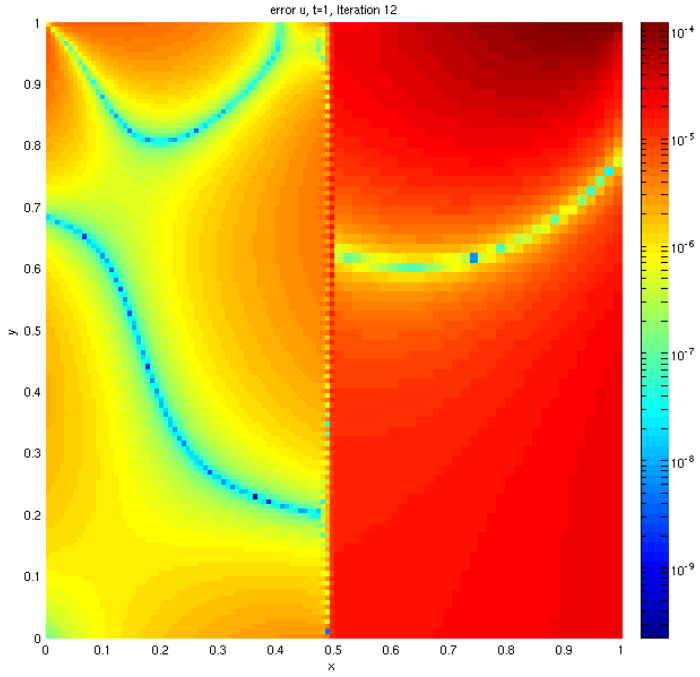


Figure 3.13: Error at at the last iteration for the diffusive case with finest nested grids and Ventcel transmission conditions in the non-overlapping case.

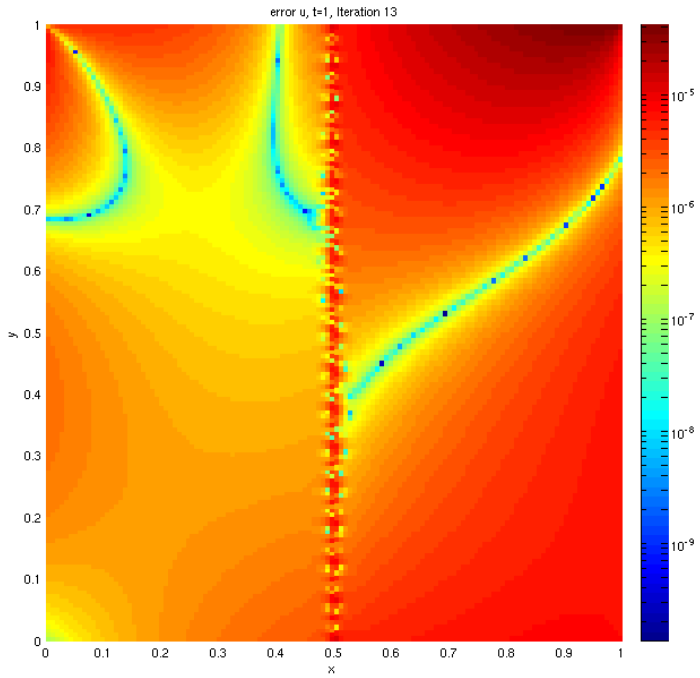


Figure 3.14: Error at at the last iteration for the diffusive case with finest non-nested grids and Ventcel transmission conditions in the non-overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	redu-	L_2 in Ω_2	redu-	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}		ction		ction		
8	8	4	4	$2.05 \cdot 10^{-2}$		$3.52 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.15 \cdot 10^{-2}$
16	16	8	8	$8.22 \cdot 10^{-3}$	2.49	$1.56 \cdot 10^{-2}$	2.26	$1.84 \cdot 10^{-2}$	$4.95 \cdot 10^{-2}$
32	32	16	16	$2.87 \cdot 10^{-3}$	2.87	$6.13 \cdot 10^{-3}$	2.54	$6.76 \cdot 10^{-3}$	$2.60 \cdot 10^{-2}$
64	64	32	32	$8.52 \cdot 10^{-4}$	3.36	$2.13 \cdot 10^{-3}$	2.88	$2.18 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$
128	128	64	64	$2.24 \cdot 10^{-4}$	3.80	$6.68 \cdot 10^{-4}$	3.18	$6.37 \cdot 10^{-4}$	$4.50 \cdot 10^{-3}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	redu-	L_2 in Ω_2	redu-	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}		ction		ction		
8	8	4	4	$2.05 \cdot 10^{-2}$		$4.01 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.98 \cdot 10^{-2}$
16	16	8	8	$8.23 \cdot 10^{-3}$	2.49	$1.84 \cdot 10^{-2}$	2.18	$1.84 \cdot 10^{-2}$	$5.34 \cdot 10^{-2}$
32	32	16	16	$2.87 \cdot 10^{-3}$	2.87	$7.73 \cdot 10^{-3}$	2.38	$6.76 \cdot 10^{-3}$	$2.77 \cdot 10^{-2}$
64	64	32	32	$8.63 \cdot 10^{-4}$	3.33	$3.02 \cdot 10^{-3}$	2.56	$2.47 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$
128	128	64	64	$2.32 \cdot 10^{-4}$	3.72	$1.14 \cdot 10^{-3}$	2.65	$1.17 \cdot 10^{-3}$	$4.82 \cdot 10^{-3}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	redu-	L_2 in Ω_2	redu-	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}		ction		ction		
8	8	4	4	$2.05 \cdot 10^{-2}$		$4.03 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.98 \cdot 10^{-2}$
16	16	8	8	$8.22 \cdot 10^{-3}$	2.49	$1.84 \cdot 10^{-2}$	2.18	$1.84 \cdot 10^{-2}$	$5.34 \cdot 10^{-2}$
32	32	16	16	$2.87 \cdot 10^{-3}$	2.87	$7.74 \cdot 10^{-3}$	2.38	$6.76 \cdot 10^{-3}$	$2.77 \cdot 10^{-2}$
64	64	32	32	$8.61 \cdot 10^{-4}$	3.33	$3.02 \cdot 10^{-3}$	2.56	$2.18 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$
128	128	64	64	$2.31 \cdot 10^{-4}$	3.72	$1.14 \cdot 10^{-3}$	2.65	$9.84 \cdot 10^{-4}$	$4.82 \cdot 10^{-3}$

Table 3.9: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Advective case with nested grids, non-overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$1.96 \cdot 10^{-2}$		$1.80 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.41 \cdot 10^{-2}$
16	18	16	14	$7.84 \cdot 10^{-3}$	2.50	$7.34 \cdot 10^{-3}$	2.46	$1.74 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$
32	36	32	28	$2.72 \cdot 10^{-3}$	2.89	$2.63 \cdot 10^{-3}$	2.79	$6.56 \cdot 10^{-3}$	$1.30 \cdot 10^{-2}$
64	72	64	56	$8.06 \cdot 10^{-4}$	3.37	$8.42 \cdot 10^{-4}$	3.13	$2.15 \cdot 10^{-3}$	$5.36 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.82	$2.75 \cdot 10^{-4}$	3.07	$6.32 \cdot 10^{-4}$	$1.89 \cdot 10^{-3}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$1.96 \cdot 10^{-2}$		$1.68 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.29 \cdot 10^{-2}$
16	18	16	14	$7.85 \cdot 10^{-3}$	2.50	$6.74 \cdot 10^{-3}$	2.50	$1.74 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$
32	36	32	28	$2.72 \cdot 10^{-3}$	2.89	$2.40 \cdot 10^{-3}$	2.80	$6.56 \cdot 10^{-3}$	$1.29 \cdot 10^{-2}$
64	72	64	56	$8.06 \cdot 10^{-4}$	3.37	$7.87 \cdot 10^{-4}$	3.05	$2.15 \cdot 10^{-3}$	$5.33 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.82	$2.74 \cdot 10^{-4}$	2.88	$6.32 \cdot 10^{-4}$	$1.88 \cdot 10^{-3}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$1.96 \cdot 10^{-2}$		$1.68 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.29 \cdot 10^{-2}$
16	18	16	14	$7.84 \cdot 10^{-3}$	2.50	$6.74 \cdot 10^{-3}$	2.50	$1.74 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$
32	36	32	28	$2.72 \cdot 10^{-3}$	2.89	$2.40 \cdot 10^{-3}$	2.81	$6.56 \cdot 10^{-3}$	$1.29 \cdot 10^{-2}$
64	72	64	56	$8.06 \cdot 10^{-4}$	3.37	$7.87 \cdot 10^{-4}$	3.05	$2.15 \cdot 10^{-3}$	$5.33 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.82	$2.74 \cdot 10^{-4}$	2.88	$6.32 \cdot 10^{-4}$	$1.88 \cdot 10^{-3}$

Table 3.10: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Advective case with non-nested grids, non-overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.91 \cdot 10^{-4}$		$3.89 \cdot 10^{-3}$		$7.98 \cdot 10^{-4}$	$1.11 \cdot 10^{-2}$
16	16	8	8	$7.68 \cdot 10^{-5}$	3.78	$1.07 \cdot 10^{-3}$	3.64	$2.73 \cdot 10^{-4}$	$4.12 \cdot 10^{-3}$
32	32	16	16	$1.94 \cdot 10^{-5}$	3.97	$2.76 \cdot 10^{-4}$	3.87	$8.29 \cdot 10^{-5}$	$1.46 \cdot 10^{-3}$
64	64	32	32	$4.71 \cdot 10^{-6}$	4.11	$6.97 \cdot 10^{-5}$	3.97	$2.30 \cdot 10^{-5}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$1.10 \cdot 10^{-6}$	4.29	$1.73 \cdot 10^{-5}$	4.02	$6.33 \cdot 10^{-6}$	$1.23 \cdot 10^{-4}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.34 \cdot 10^{-3}$		$3.96 \cdot 10^{-3}$		$8.87 \cdot 10^{-3}$	$1.12 \cdot 10^{-2}$
16	16	8	8	$6.65 \cdot 10^{-4}$	3.53	$1.07 \cdot 10^{-3}$	3.69	$3.33 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$
32	32	16	16	$1.84 \cdot 10^{-4}$	3.61	$2.75 \cdot 10^{-4}$	3.91	$1.23 \cdot 10^{-3}$	$1.45 \cdot 10^{-3}$
64	64	32	32	$4.99 \cdot 10^{-5}$	3.68	$6.84 \cdot 10^{-5}$	4.01	$4.55 \cdot 10^{-4}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$1.34 \cdot 10^{-5}$	3.74	$1.68 \cdot 10^{-5}$	4.07	$1.67 \cdot 10^{-4}$	$1.23 \cdot 10^{-4}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$8.34 \cdot 10^{-4}$		$4.05 \cdot 10^{-3}$		$4.01 \cdot 10^{-3}$	$1.13 \cdot 10^{-2}$
16	16	8	8	$1.75 \cdot 10^{-4}$	4.77	$1.10 \cdot 10^{-3}$	3.68	$1.10 \cdot 10^{-3}$	$4.13 \cdot 10^{-3}$
32	32	16	16	$3.64 \cdot 10^{-5}$	4.80	$2.83 \cdot 10^{-4}$	3.88	$2.82 \cdot 10^{-4}$	$1.46 \cdot 10^{-3}$
64	64	32	32	$7.70 \cdot 10^{-6}$	4.72	$7.14 \cdot 10^{-5}$	3.97	$7.12 \cdot 10^{-5}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$1.64 \cdot 10^{-6}$	4.71	$1.78 \cdot 10^{-5}$	4.01	$1.83 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$

Table 3.11: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Diffusive case with nested grids, non-overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$2.55 \cdot 10^{-4}$		$1.30 \cdot 10^{-3}$		$6.51 \cdot 10^{-4}$	$4.92 \cdot 10^{-3}$
16	18	16	14	$6.70 \cdot 10^{-5}$	3.81	$3.42 \cdot 10^{-4}$	3.80	$2.21 \cdot 10^{-4}$	$1.79 \cdot 10^{-3}$
32	36	32	28	$1.68 \cdot 10^{-5}$	3.98	$8.70 \cdot 10^{-5}$	3.94	$6.78 \cdot 10^{-5}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$4.07 \cdot 10^{-6}$	4.14	$2.17 \cdot 10^{-5}$	4.00	$2.20 \cdot 10^{-5}$	$1.56 \cdot 10^{-4}$
128	144	128	112	$9.37 \cdot 10^{-7}$	4.34	$5.38 \cdot 10^{-6}$	4.05	$6.44 \cdot 10^{-6}$	$4.26 \cdot 10^{-5}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$6.92 \cdot 10^{-4}$		$1.43 \cdot 10^{-3}$		$3.00 \cdot 10^{-3}$	$5.03 \cdot 10^{-3}$
16	18	16	14	$1.73 \cdot 10^{-4}$	4.00	$3.51 \cdot 10^{-4}$	4.07	$1.15 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$
32	36	32	28	$4.53 \cdot 10^{-5}$	3.82	$8.63 \cdot 10^{-5}$	4.07	$4.32 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$1.24 \cdot 10^{-5}$	3.67	$2.11 \cdot 10^{-5}$	4.09	$1.64 \cdot 10^{-4}$	$1.56 \cdot 10^{-4}$
128	144	128	112	$3.50 \cdot 10^{-6}$	3.53	$5.15 \cdot 10^{-6}$	4.10	$6.34 \cdot 10^{-5}$	$4.26 \cdot 10^{-5}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$4.17 \cdot 10^{-4}$		$1.36 \cdot 10^{-3}$		$1.53 \cdot 10^{-3}$	$4.99 \cdot 10^{-3}$
16	18	16	14	$9.23 \cdot 10^{-5}$	4.52	$3.49 \cdot 10^{-4}$	3.90	$4.99 \cdot 10^{-4}$	$1.80 \cdot 10^{-3}$
32	36	32	28	$2.05 \cdot 10^{-5}$	4.51	$8.78 \cdot 10^{-5}$	3.97	$1.45 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$4.53 \cdot 10^{-6}$	4.52	$2.18 \cdot 10^{-5}$	4.02	$4.03 \cdot 10^{-5}$	$1.57 \cdot 10^{-4}$
128	144	128	112	$9.83 \cdot 10^{-7}$	4.61	$5.38 \cdot 10^{-6}$	4.06	$1.09 \cdot 10^{-5}$	$4.26 \cdot 10^{-5}$

Table 3.12: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Diffusive case with non-nested grids, non-overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.18 \cdot 10^{-2}$		$2.62 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.15 \cdot 10^{-2}$
16	16	8	8	$8.47 \cdot 10^{-3}$	2.57	$1.37 \cdot 10^{-2}$	1.91	$1.84 \cdot 10^{-2}$	$4.95 \cdot 10^{-2}$
32	32	16	16	$2.90 \cdot 10^{-3}$	2.92	$5.79 \cdot 10^{-3}$	2.37	$6.76 \cdot 10^{-3}$	$2.60 \cdot 10^{-2}$
64	64	32	32	$8.57 \cdot 10^{-4}$	3.38	$2.07 \cdot 10^{-3}$	2.79	$2.18 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$
128	128	64	64	$2.25 \cdot 10^{-4}$	3.81	$6.62 \cdot 10^{-4}$	3.13	$6.37 \cdot 10^{-4}$	$4.51 \cdot 10^{-3}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.29 \cdot 10^{-2}$		$4.77 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.44 \cdot 10^{-2}$
16	16	8	8	$8.70 \cdot 10^{-3}$	2.63	$2.00 \cdot 10^{-2}$	2.39	$1.84 \cdot 10^{-2}$	$5.20 \cdot 10^{-2}$
32	32	16	16	$2.95 \cdot 10^{-3}$	2.95	$8.02 \cdot 10^{-3}$	2.49	$6.76 \cdot 10^{-3}$	$2.74 \cdot 10^{-2}$
64	64	32	32	$8.71 \cdot 10^{-4}$	3.38	$3.05 \cdot 10^{-3}$	2.63	$2.18 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$
128	128	64	64	$2.33 \cdot 10^{-4}$	3.74	$1.12 \cdot 10^{-3}$	2.72	$9.93 \cdot 10^{-4}$	$4.80 \cdot 10^{-3}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.29 \cdot 10^{-2}$		$4.79 \cdot 10^{-2}$		$4.39 \cdot 10^{-2}$	$8.44 \cdot 10^{-2}$
16	16	8	8	$8.69 \cdot 10^{-3}$	2.63	$2.00 \cdot 10^{-2}$	2.39	$1.84 \cdot 10^{-2}$	$5.21 \cdot 10^{-2}$
32	32	16	16	$2.94 \cdot 10^{-3}$	2.95	$8.06 \cdot 10^{-3}$	2.49	$6.76 \cdot 10^{-3}$	$2.74 \cdot 10^{-2}$
64	64	32	32	$8.70 \cdot 10^{-4}$	3.38	$3.07 \cdot 10^{-3}$	2.63	$2.18 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$
128	128	64	64	$2.32 \cdot 10^{-4}$	3.74	$1.14 \cdot 10^{-3}$	2.69	$9.33 \cdot 10^{-4}$	$4.81 \cdot 10^{-3}$

Table 3.13: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Advective case with nested grids, overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$2.07 \cdot 10^{-2}$		$1.60 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.41 \cdot 10^{-2}$
16	18	16	14	$8.04 \cdot 10^{-3}$	2.57	$6.94 \cdot 10^{-3}$	2.30	$1.74 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$
32	36	32	28	$2.75 \cdot 10^{-3}$	2.93	$2.57 \cdot 10^{-3}$	2.70	$6.56 \cdot 10^{-3}$	$1.30 \cdot 10^{-2}$
64	72	64	56	$7.89 \cdot 10^{-4}$	3.48	$8.39 \cdot 10^{-4}$	3.06	$2.10 \cdot 10^{-3}$	$5.35 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.73	$2.74 \cdot 10^{-4}$	3.07	$6.32 \cdot 10^{-4}$	$1.89 \cdot 10^{-3}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$2.07 \cdot 10^{-2}$		$1.91 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.32 \cdot 10^{-2}$
16	18	16	14	$8.04 \cdot 10^{-3}$	2.57	$7.16 \cdot 10^{-3}$	2.67	$1.74 \cdot 10^{-2}$	$2.84 \cdot 10^{-2}$
32	36	32	28	$2.75 \cdot 10^{-3}$	2.93	$2.47 \cdot 10^{-3}$	2.90	$6.56 \cdot 10^{-3}$	$1.29 \cdot 10^{-2}$
64	72	64	56	$8.09 \cdot 10^{-4}$	3.39	$7.95 \cdot 10^{-4}$	3.11	$2.15 \cdot 10^{-3}$	$5.33 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.83	$2.75 \cdot 10^{-4}$	2.90	$6.32 \cdot 10^{-4}$	$1.88 \cdot 10^{-3}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$2.07 \cdot 10^{-2}$		$1.92 \cdot 10^{-2}$		$4.09 \cdot 10^{-2}$	$5.32 \cdot 10^{-2}$
16	18	16	14	$8.04 \cdot 10^{-3}$	2.57	$7.18 \cdot 10^{-3}$	2.67	$1.74 \cdot 10^{-2}$	$2.84 \cdot 10^{-2}$
32	36	32	28	$2.75 \cdot 10^{-3}$	2.93	$2.47 \cdot 10^{-3}$	2.90	$6.56 \cdot 10^{-3}$	$1.29 \cdot 10^{-2}$
64	72	64	56	$8.09 \cdot 10^{-4}$	3.39	$7.96 \cdot 10^{-4}$	3.11	$2.15 \cdot 10^{-3}$	$5.33 \cdot 10^{-3}$
128	144	128	112	$2.11 \cdot 10^{-4}$	3.83	$2.75 \cdot 10^{-4}$	2.90	$6.32 \cdot 10^{-4}$	$1.88 \cdot 10^{-3}$

Table 3.14: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Advective case with non-nested grids, overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$4.14 \cdot 10^{-4}$		$3.34 \cdot 10^{-3}$		$1.49 \cdot 10^{-3}$	$1.11 \cdot 10^{-2}$
16	16	8	8	$8.92 \cdot 10^{-5}$	4.64	$1.03 \cdot 10^{-3}$	3.25	$3.17 \cdot 10^{-4}$	$4.12 \cdot 10^{-3}$
32	32	16	16	$2.07 \cdot 10^{-5}$	4.30	$2.73 \cdot 10^{-4}$	3.77	$8.29 \cdot 10^{-5}$	$1.46 \cdot 10^{-3}$
64	64	32	32	$4.87 \cdot 10^{-6}$	4.26	$6.93 \cdot 10^{-5}$	3.94	$2.30 \cdot 10^{-5}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$1.12 \cdot 10^{-6}$	4.36	$1.73 \cdot 10^{-5}$	4.00	$6.33 \cdot 10^{-6}$	$1.23 \cdot 10^{-4}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$2.14 \cdot 10^{-3}$		$4.02 \cdot 10^{-3}$		$1.26 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$
16	16	8	8	$3.61 \cdot 10^{-4}$	5.92	$1.10 \cdot 10^{-3}$	3.65	$2.99 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$
32	32	16	16	$6.92 \cdot 10^{-5}$	5.21	$2.84 \cdot 10^{-4}$	3.88	$7.10 \cdot 10^{-4}$	$1.46 \cdot 10^{-3}$
64	64	32	32	$1.42 \cdot 10^{-5}$	4.89	$7.13 \cdot 10^{-5}$	3.98	$1.81 \cdot 10^{-4}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$2.99 \cdot 10^{-6}$	4.74	$1.77 \cdot 10^{-5}$	4.02	$4.87 \cdot 10^{-5}$	$1.23 \cdot 10^{-4}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	8	4	4	$9.06 \cdot 10^{-4}$		$4.03 \cdot 10^{-3}$		$4.48 \cdot 10^{-3}$	$1.12 \cdot 10^{-2}$
16	16	8	8	$1.01 \cdot 10^{-4}$	8.95	$1.10 \cdot 10^{-3}$	3.66	$7.93 \cdot 10^{-4}$	$4.12 \cdot 10^{-3}$
32	32	16	16	$1.83 \cdot 10^{-5}$	5.52	$2.84 \cdot 10^{-4}$	3.87	$1.23 \cdot 10^{-4}$	$1.46 \cdot 10^{-3}$
64	64	32	32	$4.14 \cdot 10^{-6}$	4.43	$7.16 \cdot 10^{-5}$	3.97	$2.30 \cdot 10^{-5}$	$4.34 \cdot 10^{-4}$
128	128	64	64	$9.59 \cdot 10^{-7}$	4.31	$1.78 \cdot 10^{-5}$	4.01	$6.91 \cdot 10^{-6}$	$1.23 \cdot 10^{-4}$

Table 3.15: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Diffusive case with nested grids, overlapping case.

Monodomain solution

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$3.51 \cdot 10^{-4}$		$1.25 \cdot 10^{-3}$		$1.19 \cdot 10^{-3}$	$4.92 \cdot 10^{-3}$
16	18	16	14	$7.67 \cdot 10^{-5}$	4.58	$3.38 \cdot 10^{-4}$	3.69	$2.51 \cdot 10^{-4}$	$1.79 \cdot 10^{-3}$
32	36	32	28	$1.79 \cdot 10^{-5}$	4.29	$8.65 \cdot 10^{-5}$	3.90	$6.78 \cdot 10^{-5}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$4.19 \cdot 10^{-6}$	4.27	$2.17 \cdot 10^{-5}$	3.99	$2.20 \cdot 10^{-5}$	$1.56 \cdot 10^{-4}$
128	144	128	112	$9.50 \cdot 10^{-7}$	4.40	$5.37 \cdot 10^{-6}$	4.04	$6.44 \cdot 10^{-6}$	$4.26 \cdot 10^{-5}$

Domain decomposition solution with Robin transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$7.18 \cdot 10^{-4}$		$1.38 \cdot 10^{-3}$		$3.62 \cdot 10^{-3}$	$4.97 \cdot 10^{-3}$
16	18	16	14	$1.33 \cdot 10^{-4}$	5.41	$3.54 \cdot 10^{-4}$	3.90	$9.56 \cdot 10^{-4}$	$1.80 \cdot 10^{-3}$
32	36	32	28	$2.69 \cdot 10^{-5}$	4.92	$8.89 \cdot 10^{-5}$	3.99	$2.57 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$5.73 \cdot 10^{-6}$	4.70	$2.20 \cdot 10^{-5}$	4.04	$7.07 \cdot 10^{-5}$	$1.57 \cdot 10^{-4}$
128	144	128	112	$1.23 \cdot 10^{-6}$	4.68	$5.40 \cdot 10^{-6}$	4.07	$2.00 \cdot 10^{-5}$	$4.26 \cdot 10^{-5}$

Domain decomposition solution with Ventcel transmission conditions

Discretisation				L_2 in Ω_1	reduction	L_2 in Ω_2	reduction	L_{inf} in Ω_1	L_{inf} in Ω_2
Δx_1^{-1}	Δy_1^{-1}	Δx_2^{-1}	Δy_2^{-1}						
8	9	8	7	$5.20 \cdot 10^{-4}$		$1.34 \cdot 10^{-3}$		$1.95 \cdot 10^{-3}$	$4.95 \cdot 10^{-3}$
16	18	16	14	$1.02 \cdot 10^{-4}$	5.11	$3.49 \cdot 10^{-4}$	3.83	$4.24 \cdot 10^{-4}$	$1.80 \cdot 10^{-3}$
32	36	32	28	$2.19 \cdot 10^{-5}$	4.65	$8.84 \cdot 10^{-5}$	3.95	$9.22 \cdot 10^{-5}$	$5.50 \cdot 10^{-4}$
64	72	64	56	$4.85 \cdot 10^{-6}$	4.51	$2.20 \cdot 10^{-5}$	4.01	$2.20 \cdot 10^{-5}$	$1.57 \cdot 10^{-4}$
128	144	128	112	$1.05 \cdot 10^{-6}$	4.61	$5.44 \cdot 10^{-6}$	4.05	$6.44 \cdot 10^{-6}$	$4.26 \cdot 10^{-5}$

Table 3.16: Discrete error norms in different subdomains for initial grid and four refinements of initial grids in the case of a global monodomain solution, a converged domain decomposition solution with optimised Robin and optimised Ventcel conditions. Diffusive case with non-nested grids, overlapping case.

3.C Features of the Prototype Code

All numerical tests in chapters 2, 3, 4 and 5 have been performed with the prototype "Domain Decomposition for Reactive Transport" (DDTR) that has been created from scratch during the first two years of this Ph. D. thesis. The code is written in the MATLAB language and is entirely compatible with version 7.6.0.324 (R2008a) of MATLAB. The last developments have been realised in February 2011.

The prototype exists in two different forms: a 1D code based on a fixed data structure concerning the mesh and the domain decomposition. This part has been entirely re-engineered in the 2D and 3D codes which offer a proper and flexible data structure concerning the mesh and the physical properties. All codes are dimension-dependent, e. g. the 2D code does not degenerate to a 1D code and cannot handle 3D geometries.
The basic problem is a two-species reactive transport system including advection and diffusion, a chemical coupling term that is described either by given linear form (cf. chapter 4) or by a user-defined nonlinear function (cf. chapter 5).
The nonlinear problem is treated by a Newton method at every time step. The linear systems appearing at every iteration of the Newton solver in the nonlinear case and at every time step in the linear case are solved with Matlab's exact LU-solver.
The space meshes are restricted to rectangular meshes where the discretisation sizes per dimension can be set independently. Subdomains can be chosen by filter functions and can have arbitrary shape. The underlying data structure for space meshes can be extended to non-orthogonal meshes. The time mesh is set to have constant time steps, it can be extended to adaptive time-meshes.

The prototype offers different methods: global simulation of linear and nonlinear reactive transport systems, different approaches for domain decomposition algorithms in the Schwarz waveform relaxation context including four different types of transmission conditions (Dirichlet, Neumann, Robin and Ventcel). Transmission and boundary conditions can be chosen independently for every face.
Simulation data is saved in the MATLAB format `.mat`, the prototype offers different methods for the visualisation of results.

The prototype contains three different directories on the basic level for the different dimensions. In every directory, different files are available:

- **Basic testers:** Files named `test<functionality>.m` are executable functions that are able to test basic algorithms without any special treatment. Parameters concerning the tested functionality are defined in the file itself.
- **Extended studies:** files named `study<property>.m` are executable functions that study a certain property (like mesh refinement or parameter variation) based on an existing basic tester code.

- **Comparisons:** files named `compare<property/method>.m` are executable functions that compare different properties or methods based on an existing basic tester code.
- **Graphical output:** files named `plot<resulttype>.m` are functions called at the end of basic testers in order to do a graphical output of the obtained results. Different properties like separation of subdomains, type of the output, output as video and so on are defined at the beginning of the files.
- **Select subdomains:** files named `select<subdomaintypes>.m` are filter functions to select user defined subdomains called during basic testers and extended studies.
- **Get structures:** files named `get<structure>.m` are functions offering the data structures and information used in this prototype like submeshes including submesh physical properties based on select functions, face and cell projections called during basic testers and extended studies.
- **Physical data:** the file `getphysicaldata.m` plays an important role since in this file, the geometrical and physical data is defined by the user.
- **Parameter optimisation:** files named `optimise<parameters>.m` are functions that return optimised parameters for the transmission conditions in the domain decomposition context, they are called during the basic testers and extended studies.
- **Core files:** they do not follow a special name rule and cannot be called without any context. Important files are:
  - `tr.m`, `trlin.m`, `trnl.m`: this functions offer the main calculations in the linear, linearised and nonlinear case. Calculations are done globally on a given time interval, on a (sub-)mesh. Parameters concerning the calculation itself (number of Newton iterations etc.) are defined in files.
  - `reconstruct.m`: this function reconstructs globally in time and space values for different transmission conditions on the boundary or in the inner of the (sub-)mesh.
  - `darcy.m`: this function realises the calculation of a Darcy field (available only in the 2D and 3D code) for given boundary conditions and physical parameters defined in the `getphysicaldata.m` file.

All code files are commented in detail and special numerical treatments that differ from the theoretical developments are emphasised as comment.
The command `help <filename>.m` prints the purpose of the function.

The Enlightened take things Lightly.

Principia Discordia

4

A Linear Coupled Two Species Reactive Transport System

Contents

Introduction

In this chapter we study a Schwarz waveform relaxation method on a linear coupled two species reactive transport system. As the system is linear, we can study many aspects of the problem itself and the applied algorithm on a theoretical level.

In the first part, we state the problem and apply a Schwarz waveform relaxation method with different types of transmission conditions. We develop the convergence factor of the algorithm using different transmission conditions. Then, we state and prove the well-posedness of the subproblems in the case of Robin and Ventcel conditions, the well-posedness of the overlapping and non-overlapping algorithm using Robin and Ventcel conditions and of the overlapping algorithm using Dirichlet conditions. The statement and the prove of the convergence of the algorithm concludes the theoretical results. We discuss then the optimisation of transmission conditions on a theoretical, numerical and practical level and conclude finally by giving numerical results.

4.1 Problem Definition and Well-Posedness

In this chapter we consider the model problem of a coupled two species reactive transport system

$$\begin{aligned} \phi\partial_t u + \operatorname{div}(-a\vec{\nabla}u + \vec{b}u) - k(v - cu) &= 0 \qquad \text{on } \mathbb{R}^d \times (0,T), \\ \phi\partial_t v \qquad\qquad\qquad\qquad + k(v - cu) &= 0 \qquad \text{on } \mathbb{R}^d \times (0,T), \end{aligned} \tag{4.1}$$

on the spatial domain $x \in \mathbb{R}^d$, $d = 1,\ 2,\ 3$ and the time period $t \in (0, T)$. Note that for theoretical studies, it is more convenient to pose the problem on a open and unbounded spatial domain. For this reason, no boundary value problems have to be considered. Nevertheless, in the numerical results we have to restrict ourselves to closed and bounded domains.
With $\phi(x) > 0$ we denote the porosity. The mobile species u is subject to a linear transport operator $Lu := \operatorname{div}(-a\vec{\nabla}u + \vec{b}u)$ including diffusion described by a positive scalar diffusion coefficient $a > 0$ and advection described by a Darcy field vector $\vec{b} \in \mathbb{R}^d$. The fixed species v is coupled to the mobile species u by a linear coupling term $k(v - cu)$ where $k(x) \geq 0$ represents the reactive surface or, roughly spoken, the reaction speed. c denotes the stoichiometric coefficient at which the reaction $v \leftrightarrows cu$ attains the equilibrium state. We impose an initial condition for the mobile and fixed species

$$u(x,0) = u_0(x), \qquad v(x,0) = v_0(x) \qquad \text{on } \Omega. \tag{4.2}$$

This problem is the linear version of problem (5.1), (5.2), (5.3) which arises as a subproblem of multispecies reactive transport systems and which is studied in chapter 5. The linear character of the problem allows the study of the influence of the Schwarz waveform relaxation algorithm on a theoretical level without loosing the main challenge of a coupled system of equations.

We suppose that the initial condition (u_0, v_0) is in $L^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$. A weak solution of problem (4.1)-(4.2) is defined to be a function $(u, v) \in L^2(0,T;H^1(\mathbb{R}^d)) \times L^2(0,T;L^2(\mathbb{R}^d)) \cap C([0,T];L^2(\mathbb{R}^d)) \times C([0,T];L^2(\mathbb{R}^d))$ satisfying for all $(w, z) \in H^1(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$

$$\frac{\partial}{\partial t}((\phi u, w) + (\phi v, z)) + a(\nabla u, \nabla w) + \frac{1}{2}(((b \cdot \nabla)u, w) - ((b \cdot \nabla)w, u)) \\ - k((v, w) - (v, z)) + ck((u, w) - (u, z)) = 0, \text{ in } \mathcal{D}'(0,T),$$

and $u_{|t=0} = u_0, \;\; v_{|t=0} = v_0$, where $(\cdot, \cdot)$ denotes the inner product in $L^2(\mathbb{R}^d)$.

Since the second equation of (4.1) is an ordinary differential equation in the v variable, we can apply the results of Lions and Magenes [56] concerning evolution equations of parabolic type, in order to obtain a well-posedness result for the Cauchy problem (4.1)-(4.2).

Theorem 4.1 (*Well-posedness of the global linear coupled problem*)

Let $(u_0, v_0) \in L^2(\mathbb{R}^2) \times L^2(\mathbb{R}^d)$. Then problem (4.1)-(4.2) possesses a unique weak solution (u, v) such that $u \in L^2(0,T;H^1(\mathbb{R}^d)) \cap C([0,T];L^2(\mathbb{R}^d))$ and $v \in C([0,T];L^2(\mathbb{R}^d))$. We have in addition that $u \in L^\infty(0,T;H^2(\mathbb{R}^d))$.

4.2 Schwarz Waveform Relaxation Algorithm

In the following, we will use the following notation: if $d > 1$, we put $\mathbb{R}^d = \mathbb{R} \times \mathbb{R}^{d-1}$ and we denote by $(x, \boldsymbol{y})$ the coordinate in $\mathbb{R}^d$, where $\boldsymbol{y} = y$ if $d = 2$, $\boldsymbol{y} = (y, z)$ if $d = 3$. The notation $(c_x, \boldsymbol{c_y})$ holds for a given vector $c = (c_x, c_y) \in \mathbb{R}^2$ or $c = (c_x, c_y, c_z) \in \mathbb{R}^3$.

We can approach problem (4.1)–(4.2) by a Schwarz waveform relaxation algorithm. Therefore, we decompose the spatial domain $\Omega = \mathbb{R}^d$ in two possibly overlapping subdomains $\Omega_1 = (-\infty, L) \times \mathbb{R}^{d-1}$, $\Omega_2 = (0, +\infty) \times \mathbb{R}^{d-1}$ such that $\Omega = \Omega_1 \cup \Omega_2$. $L \geq 0$ is the "length" of the overlap of the two subdomains. We call $\Gamma_1 = \{L\} \times \mathbb{R}^{d-1}$ and $\Gamma_2 = \{0\} \times \mathbb{R}^{d-1}$ the interfaces. In the case of non overlapping subdomains we have $\Gamma_1 = \Gamma_2 = \{0\} \times \mathbb{R}^{d-1}$. With $\vec{n}_1 = (1, \mathbf{0})$ and $\vec{n}_2 = (-1, \mathbf{0})$ we denote the unit outward normals of Ω_1 and Ω_2 at Γ_1 and Γ_2, respectively. We use linear transmission conditions $\mathcal{B}_1$ and $\mathcal{B}_2$ to transmit the information from one subdomain to another on the interfaces Γ_1, Γ_2. Providing an initial guess (u_2^0, v_2^0) on Γ_1 and (u_1^0, v_1^0) on Γ_2, we can state the entire approach in algorithm 4.1.

4.2.1 Transmission Conditions

We define the errors of the iterates with respect to the global solution in Ω_i at iteration k by $e_{u_i}^k = u_i^k - u_{|\Omega_i}, \; e_{v_i}^k = v_i^k - v_{|\Omega_i}$. By linearity, they satisfy equations defined in algorithm (4.1) with $(u_0, v_0) = (0, 0)$. We extend $e_{u_i}^k$ and $e_{v_i}^k$ by 0 for $t < 0$ and use for the sake of readability the same

Algorithm 4.1 Parallel Schwarz waveform relaxation algorithm for the linear coupled two species reactive transport system

$$
\begin{array}{lll}
\phi\partial_t u_1^{k+1} + \operatorname{div}(-a\vec{\nabla}u_1^{k+1} + \vec{b}u_1^{k+1}) - k(v_1^{k+1} - cu_1^{k+1}) = 0 & \text{on } \Omega_1 \times (0,T) \\
\phi\partial_t v_1^{k+1} + k(v_1^{k+1} - cu_1^{k+1}) = 0 & \text{on } \Omega_1 \times (0,T) \\
(u_1^{k+1}(x,0), v_1^{k+1}(x,0)) = (u_0, v_0) & \text{on } \Omega_1 \\
\mathcal{B}_1\left(u_1^{k+1}, v_1^{k+1}\right) = \mathcal{B}_1\left(u_2^k, v_2^k\right) & \text{on } \Gamma_1 \times (0,T)
\end{array}
$$

$$
\begin{array}{lll}
\phi\partial_t u_2^{k+1} + \operatorname{div}(-a\vec{\nabla}u_2^{k+1} + \vec{b}u_2^{k+1}) - k(v_2^{k+1} - cu_2^{k+1}) = 0 & \text{on } \Omega_2 \times (0,T) \\
\phi\partial_t v_2^{k+1} + k(v_2^{k+1} - cu_2^{k+1}) = 0 & \text{on } \Omega_2 \times (0,T) \\
(u_2^{k+1}(x,0), v_2^{k+1}(x,0)) = (u_0, v_0) & \text{on } \Omega_2 \\
\mathcal{B}_2\left(u_2^{k+1}, v_2^{k+1}\right) = \mathcal{B}_2\left(u_1^k, v_1^k\right) & \text{on } \Gamma_2 \times (0,T)
\end{array}
$$

notation to denote the extended functions. We perform a Fourier transform with respect to the time and $\boldsymbol{y}$ variables. The Fourier transformation is defined by

$$\hat{w}(x,\boldsymbol{\xi},\tau) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}} \int_{\mathbb{R}^{d-1}} w(x,\mathbf{y},t) e^{-i(\mathbf{y}\cdot\boldsymbol{\xi}+\tau t)} \, \mathrm{d}\mathbf{y}\mathrm{d}t,$$

for $w \in L^2(\Omega_i \times \mathbb{R})$, where $\boldsymbol{\xi}$ and τ are respectively the dual variables of $\mathbf{y}$ and t (for $d = 2$ we have $\boldsymbol{\xi} = \xi$, for $d = 3$ we have $\boldsymbol{\xi} = (\xi, \zeta)$). The Fourier transforms $\widehat{e_{u_i}^k}$ and $\widehat{e_{v_i}^k}$ are solutions of the ordinary differential equation system in the x variable

$$
\begin{aligned}
&\phi \,\mathrm{i}\, \tau \widehat{e_{u_i}^k} - a \frac{\partial^2 \widehat{e_{u_i}^k}}{\partial x^2} + a\boldsymbol{\xi}\cdot\boldsymbol{\xi}\, \widehat{e_{u_i}^k} + b_x \frac{\partial \widehat{e_{u_i}^k}}{\partial x} + \mathrm{i}\, b_{\mathbf{y}}\cdot\boldsymbol{\xi}\, \widehat{e_{u_i}^k} - k\, \widehat{e_{v_i}^k} + kc\, \widehat{e_{u_i}^k} = 0, \\
&\phi \,\mathrm{i}\, \tau \widehat{e_{v_i}^k} + k\, \widehat{e_{v_i}^k} - kc\, \widehat{e_{u_i}^k} = 0.
\end{aligned}
$$

Supposing $\phi i\tau + k \neq 0$, one can eliminate $\widehat{e_{v_i}^k}$ from the first equation and it follows that $\widehat{e_{u_i}^k}$ is solution of a linear second order ordinary differential equation. The roots of its characteristic polynomial are

$$\lambda^+ = \frac{b_x}{2a} + \frac{\sqrt{b_x^2 + 4a\hat{z}}}{2a}, \quad \lambda^- = \frac{b_x}{2a} - \frac{\sqrt{b_x^2 + 4a\hat{z}}}{2a}, \tag{4.3}$$

with

$$\hat{z} = \hat{z}(\boldsymbol{\xi},\tau) := \phi i\tau + a\boldsymbol{\xi}\cdot\boldsymbol{\xi} + ib_{\mathbf{y}}\cdot\boldsymbol{\xi} - \frac{k^2 c}{\phi i\tau + k} + kc. \tag{4.4}$$

The complex square root is chosen to have positive real part. We can easily see that $\mathrm{Re}(\lambda^+) > 0$ and $\mathrm{Re}(\lambda^-) < 0$. Hence we must have

$$
\begin{aligned}
&\widehat{e_{u_1}^k}(x,\boldsymbol{\xi},\tau) = \alpha_1^k(\boldsymbol{\xi},\tau) e^{\lambda^+(x-L)}, \quad (x,\boldsymbol{\xi},\tau) \in]-\infty, L[\times\mathbb{R}^{d-1}\times\mathbb{R}, \\
&\widehat{e_{u_2}^k}(x,\boldsymbol{\xi},\tau) = \alpha_2^k(\boldsymbol{\xi},\tau) e^{\lambda^- x}, \quad\quad (x,\boldsymbol{\xi},\tau) \in]0, +\infty[\times\mathbb{R}^{d-1}\times\mathbb{R},
\end{aligned} \tag{4.5}
$$

where the functions α_1^k and α_2^k are defined by the boundary conditions on Γ_1 and Γ_2.

4.2.1.1 Classical Transmission Condition — Dirichlet

The classical Schwarz waveform relaxation algorithm consists in choosing for $\mathcal{B}_1(u,v)$ and $\mathcal{B}_2(u,v)$ the identity operators with respect to u which leads to Dirichlet transmission conditions. By using the recurrence relation of the interface conditions, we obtain

$$e_{u1}^{k}(L,\boldsymbol{\xi},\tau) = e_{u2}^{k-1}(L,\boldsymbol{\xi},\tau), \quad e_{u2}^{k-1}(0,\boldsymbol{\xi},\tau) = e_{u1}^{k-2}(0,\boldsymbol{\xi},\tau),$$

which yields

$$\alpha_i^k = e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2}, \tag{4.6}$$

for $i = 1,\ 2$ and for all $k \geq 2$. We deduce from relation (4.6) that if $L = 0$, i. e. the subdomains do not overlap, then the Schwarz waveform relaxation algorithm does not converge, unless the initial guess corresponds to the exact boundary condition. If not, the larger L is, the faster the algorithm converges.

4.2.1.2 Optimised Transmission Conditions — Robin and Ventcel

A more general choice for the transmission conditions is

$$\mathcal{B}_1 = \frac{\partial u}{\partial n_1} + S_1(u,v,\partial_t u,\nabla_{\mathbf{y}} u,\Delta_{\mathbf{y}} u), \quad \mathcal{B}_2 = \frac{\partial u}{\partial n_2} + S_2(u,v,\partial_t u,\nabla_{\mathbf{y}} u,\Delta_{\mathbf{y}} u), \tag{4.7}$$

where S_i is a general operator; we call σ_i the associated symbol in the Fourier variables of S_i, for $i = 1,\ 2$. By using again the recurrence relation of the interface conditions, we obtain

$$\alpha_i^k = \frac{(\sigma_1 + \lambda^-)(\sigma_2 - \lambda^+)}{(\sigma_1 + \lambda^+)(\sigma_2 - \lambda^-)} e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2}, \tag{4.8}$$

which is the analogous of equation (4.6).

If the symbols σ_i satisfied $\sigma_1 = -\lambda^-,\ \sigma_2 = \lambda^+$, algorithm 4.1 would converge in two iterations independently of the initial guess. Choosing the transmission conditions in this way would correspond to non-local operators S_i, since the functions $\lambda^\pm$ are not polynomials in the dual variables. To avoid the use of non-local operators, we follow [49] and approach the square root appearing in $\lambda^\pm$ either by a zeroth order polynomial, which leads to Robin transmission conditions, or by a first order polynomial, which leads to Ventcel transmission conditions.

Schwarz waveform relaxation approximation of order 0 The Schwarz waveform relaxation algorithm of order 0 is obtained by performing a zeroth order polynomial approximation of the square root $\sqrt{b_x^2 + 4a\hat{z}}$ appearing in λ^- and in λ^+, which leads to Robin transmission conditions. They are defined as follows: for $p \in \mathbb{R}$, $p > 0$, we consider

$$\mathcal{B}_1(u,v) = \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u, \quad \mathcal{B}_2(u,v) = \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u. \tag{4.9}$$

Schwarz waveform relaxation approximation of order 2 The Schwarz waveform relaxation algorithm of order 2 is obtained by performing a first order polynomial approximation of the square root $\sqrt{b_x^2 + 4a\hat{z}}$ appearing in λ^- and in λ^+, which leads to Ventcel transmission conditions. They are defined as follows: for $p, q \in \mathbb{R}$, $p > 0$ and $q \geq 0$, we consider $\mathcal{B}_i$ defined by

$$\begin{aligned} \mathcal{B}_1(u,v) &= \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\mathbf{y}} u + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} u - kv + kcu\right), \\ \mathcal{B}_2(u,v) &= \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\mathbf{y}} u + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} u - kv + kcu\right). \end{aligned} \tag{4.10}$$

This kind of transmission conditions is also called of second order due to the second order tangential derivative appearing in (4.10).

4.2.2 Convergence Factor of the Algorithm

We calculate in this section the convergence factor of the algorithm 4.1.

4.2.2.1 Classical Transmission Condition — Dirichlet

In the case of Dirichlet transmission conditions, we obtain equation (4.6)

$$\alpha_i^k = e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2},$$

for $i = 1,\ 2$ and $k \geq 2$, where the characteristic roots λ^- and λ^+ are defined by (4.3). The convergence factor of the classical Schwarz algorithm, in Fourier variables, is thus given by

$$\rho_D = \rho_D(\boldsymbol{\xi}, \tau) = e^{(\lambda^- - \lambda^+)L}.$$

4.2.2.2 Optimised Transmission Conditions — Robin and Ventcel

In the case of Robin transmission conditions, the transmission operators $\mathcal{B}_i(u,v)$ are defined by (4.9) and the recurrence relation of the interface conditions yields

$$\begin{aligned} \left(\lambda^+ - \frac{b_x - p}{2a}\right)\alpha_1^k &= \left(\lambda^- - \frac{b_x - p}{2a}\right)\alpha_2^{k-1}e^{\lambda^- L}, \\ \left(-\lambda^- + \frac{b_x + p}{2a}\right)\alpha_2^{k-1} &= \left(-\lambda^+ + \frac{b_x + p}{2a}\right)\alpha_1^{k-2}e^{-\lambda^+ L}, \end{aligned}$$

and thus we obtain

$$\alpha_i^k = \left(\frac{p - \sqrt{b_x^2 + 4a\hat{z}}}{p + \sqrt{b_x^2 + 4a\hat{z}}}\right)^2 e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2},$$

for $i = 1,\ 2$ and for $k \geq 2$, where $\hat{z}$ is defined by (4.4) and λ^- and λ^+ by (4.3).

In the case of Ventcel transmission conditions we can proceed in the same way with the operators $\mathcal{B}_i(u, v)$ defined by (4.10). We obtain then

$$\alpha_i^k = \left(\frac{p + q\hat{z} - \sqrt{b_x^2 + 4a\hat{z}}}{p + q\hat{z} + \sqrt{b_x^2 + 4a\hat{z}}} \right)^2 e^{(\lambda^- - \lambda^+)L} \alpha_i^{k-2},$$

for $i = 1,\ 2$ and for $k \geq 2$.

We obtain that the convergence factor of algorithm 4.1 with Robin transmission conditions is given by

$$\rho = \rho_R(\boldsymbol{\xi}, \tau, p) = \left(\frac{p - \sqrt{b_x^2 + 4ad}}{p + \sqrt{b_x^2 + 4ad}} \right)^2 e^{(\lambda^- - \lambda^+)L}, \tag{4.11}$$

and that the convergence factor of the algorithm 4.1 with Ventcel transmission conditions is given by

$$\rho = \rho_V(\boldsymbol{\xi}, \tau, p, q) = \left(\frac{p + q\hat{z} - \sqrt{b_x^2 + 4ad}}{p + q\hat{z} + \sqrt{b_x^2 + 4ad}} \right)^2 e^{(\lambda^- - \lambda^+)L}. \tag{4.12}$$

4.2.3 Well-Posedness of the Algorithm

In this section we prove the well-posedness of the subproblems appearing at every iteration of algorithm 4.1 using Robin or Ventcel transmission conditions and of the algorithm itself using different types of transmission conditions in the overlapping and non-overlapping case. By using *a priori* estimates in appropriate spaces and the Gronwall lemma we extend the results of [6] and [34] concerning the linear advection diffusion reaction equation to our case of a linear coupled reactive transport system.

4.2.3.1 Well-Posedness of the Subproblems using Robin and Ventcel Boundary Conditions

We begin by defining the Schwarz waveform relaxation algorithms in the framework of Sobolev spaces. To do so, we introduce the function spaces

$$H_s^s(\Omega_i) = \{u \in H^s(\Omega_i),\ u_{|\Gamma_i} \in H^s(\Gamma_i)\},$$

for $s \geq 1$, and set $V = H^1(\Omega_i)$, if $q = 0$, and $V = H_1^1(\Omega_i)$, if $q > 0$.
Let $\tilde{V} = L^2(\Omega_i)$, if $q = 0$, and $\tilde{V}$ be the space of functions $u \in L^2(\Omega_i)$ that possess a trace on Γ_i which is in $L^2(\Gamma_i)$, if $q > 0$.
We denote by $(\cdot, \cdot)$ the inner product in $L^2(\Omega_i)$ and by $(\cdot, \cdot)_{\Gamma_i}$ the inner product in $L^2(\Gamma_i)$ and throughout this section, without loss of generality, we consider $\phi = 1$.

Let $i \in \{1,2\}$. If $g \in L^2(0,T;L^2(\Gamma_i))$ and $p > 0,\ q \geq 0$ are given, a weak solution of the boundary value problem

$$\begin{aligned}
&\partial_t u + \operatorname{div}(-a\nabla u + bu) - k(v - cu) = 0 && \text{in } \Omega_i \times (0,T),\\
&\partial_t v \qquad\qquad\quad + k(v - cu) = 0 && \text{in } \Omega_i \times (0,T),\\
&(u,v)(\cdot,0) = (u_0(\cdot), v_0(\cdot))_{|\Omega_i} && \text{in } \Omega_i,\\
&\frac{\partial u}{\partial n_i} + \frac{p + (-1)^i b_x}{2a} u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\boldsymbol{y}} u + b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} u - kv + kcu\right) = g && \text{over } \Gamma_i \times (0,T),
\end{aligned} \tag{4.13}$$

is a function $(\tilde{u},\tilde{v}) \in L^2(0,T;V) \times L^2(0,T;\tilde{V}) \cap C(0,T;L^2(\Omega_i)) \times C(0,T;L^2(\Omega_i))$, satisfying for all $(w,z) \in V \times \tilde{V}$

$$\begin{aligned}
&\frac{\partial}{\partial t}\left((\tilde{u},w) + (\tilde{v},z)\right) + a(\nabla\tilde{u},\nabla w) + \frac{1}{2}\left((b\cdot\nabla)\tilde{u},w\right) - \left((b\cdot\nabla)w,\tilde{u}\right)\\
&\qquad\qquad - k\left((\tilde{v},w) - (\tilde{v},z)\right) + ck\left((\tilde{u},w) - (\tilde{u},z)\right)\\
&+ \frac{p}{2}(\tilde{u},w)_{\Gamma_i} + \frac{q}{2}\left(\frac{\partial}{\partial t}(\tilde{u},w)_{\Gamma_i} + a(\nabla_{\boldsymbol{y}}\tilde{u},\nabla_{\boldsymbol{y}} w)_{\Gamma_i} + (b_{\boldsymbol{y}}\cdot\nabla_{\boldsymbol{y}}\tilde{u},w)_{\Gamma_i} - (k\tilde{v},w)_{\Gamma_i} + (kc\tilde{u},w)_{\Gamma_i}\right) = (g,w)_{\Gamma_i},
\end{aligned}$$

in $\mathcal{D}'(0,T)$, and $u_{|t=0} = u_0,\ \ v_{|t=0} = v_0$.

We first prove a well-posedness result for problem (4.13). Let x_i be the x-abscissa of the interface Γ_i.

Lemma 4.1

Let $p > 0,\ q \geq 0$ be given.

1. If $q = 0$, consider $(u_0,v_0) \in H^2(\mathbb{R}^d)\times L^2(\mathbb{R}^d)$ and $g \in H^1(0,T;L^2(\Gamma_i)) \cap L^\infty(0,T;H^{\frac{1}{2}}(\Gamma_i))$. Suppose in addition that $\frac{\partial u_0}{\partial n_i}(x_i,\cdot) + \frac{p+(-1)^i b_x}{2a} u_0(x_i,\cdot) = g(\cdot,0)$. Then problem (4.13) has a unique solution (u,v) such that

$$u \in W^{1,\infty}(0,T;L^2(\Omega_i)) \cap L^\infty(0,T;H^2(\Omega_i)) \cap H^1(0,T;H^1(\Omega_i)),$$

and

$$v \in W^{1,\infty}(0,T;L^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)).$$

2. If $q > 0$, consider $(u_0,v_0) \in H^2(\mathbb{R}^d) \times H^1(\mathbb{R}^d)$ and $g \in L^2(0,T;L^2(\Gamma_i))$. Then problem (4.13) has a unique solution (u,v) such that

$$u \in L^2(0,T;H^2_2(\Omega_i)) \cap L^\infty(0,T;H^1_1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

and

$$v \in L^\infty(0,T;H^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$$

with $u_{|\Gamma_i} \in H^1(0,T;L^2(\Gamma_i))$ and $v_{|\Gamma_i} \in L^2(0,T;L^2(\Gamma_i))$ if $q > 0$.

The proof is based on energy estimates and uses the Galerkin method.

Proof 4.1

1. The case $q = 0$:
We take the inner product of equation $\partial_t u + \operatorname{div}(-a\nabla u + bu) - k(v - cu) = 0$ with u, we integrate by parts in Ω_i and we take the inner product of equation $\partial_t v + k(v - cu) = 0$ with v. We sum both equations and we apply the Cauchy-Schwarz inequality and the inequality $mn \leq \frac{m^2}{2\varepsilon} + \frac{\varepsilon}{2}n^2$, for $m,\ n \in \mathbb{R}$ and $\varepsilon \geq 0$, in the right-hand side of the resulting equation. We obtain

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\|u(t)\|^2 + \|v(t)\|^2\right) + a\|\nabla u(t)\|^2 + \frac{p}{2}\|u(t)\|^2_{\Gamma_i} + kc\|u(t)\|^2 + k\|v(t)\|^2 \\
&\quad = (k + kc)(v(t), u(t)) + (g(t), u(t))_{\Gamma_i} \\
&\quad \leq (k + kc)\,\|v(t)\|\,\|u(t)\| + \|g(t)\|_{\Gamma_i}\|u(t)\|_{\Gamma_i} \\
&\quad \leq \frac{k + kc}{2}\left(\|u(t)\|^2 + \|v(t)\|^2\right) + \frac{\|g(t)\|^2_{\Gamma_i}}{p} + \frac{p}{4}\|u(t)\|^2_{\Gamma_i}.
\end{aligned}$$

We integrate now over $(0, t)$, for $t \leq T$, both sides of the above inequality, which gives

$$\begin{aligned}
&\|u(t)\|^2 + \|v(t)\|^2 + \int_0^t \left(a\|\nabla u(s)\|^2 + \frac{p}{2}\|u(s)\|^2_{\Gamma_i} + 2kc\|u(s)\|^2 + 2k\|v(s)\|^2\right)\mathrm{d}s \\
&\qquad \leq \|u_0\|^2 + \|v_0\|^2 + \int_0^t \frac{2\|g(s)\|^2_{\Gamma_i}}{p}\,\mathrm{d}s + (k + kc)\int_0^t \left(\|u(s)\|^2 + \|v(s)\|^2\right)\mathrm{d}s.
\end{aligned}$$

Applying the Gronwall lemma yields

$$\begin{aligned}
&\|u\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \|v\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \min(a, 2kc)\|u\|^2_{L^2(0,T;H^1(\Omega_i))} + \frac{p}{2}\|u\|^2_{L^2(0,T;L^2(\Gamma))} \\
&\qquad + 2k\|v\|^2_{L^2(0,T;L^2(\Omega_i))} \leq e^{CT}\left(\|u_0\|^2 + \|v_0\|^2 + \|g\|^2_{L^2(0,T;L^2(\Gamma))}\right), \quad (4.14)
\end{aligned}$$

where C is a positive constant depending on $a,\ k,\ c$ and p. The same calculations hold for the equation satisfied by $\partial_t u$ and $\partial_t v$ and if we apply (4.14) to $\partial_t u$ and $\partial_t v$ we obtain

$$\begin{aligned}
&\|\partial_t u\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \|\partial_t v\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \min(a, 2kc)\|\partial_t u\|^2_{L^2(0,T;H^1(\Omega_i))} + \frac{p}{2}\|\partial_t u\|^2_{L^2(0,T;L^2(\Gamma))} \\
&\qquad + 2k\|\partial_t v\|^2_{L^2(0,T;L^2(\Omega_i))} \leq e^{CT}\left(\|u_{t0}\|^2 + \|v_{t0}\|^2 + \|g\|^2_{H^1(0,T;L^2(\Gamma))}\right). \quad (4.15)
\end{aligned}$$

We need to estimate $\|u_{t0}\|$ and $\|v_{t0}\|$. To do so, we take the inner product of equation $\partial_t u + \operatorname{div}(-a\nabla u + bu) - k(v - cu) = 0$ with $\partial_t u$, we integrate by parts in Ω_i and we evaluate the resulting equation at time $t = 0$. We obtain

$$\begin{aligned}
&\|u_{t0}\|^2 + a(\nabla u_0, \nabla u_{t0}) + \frac{p + (-1)^i b_x}{2}(u_0, u_{t0})_{\Gamma_i} + (b \cdot \nabla u_0, u_{t0}) - k(v_0, u_{t0}) + kc(u_0, u_{t0}) \\
&\qquad = (g(\cdot, 0), u_{t0})_{\Gamma_i}.
\end{aligned}$$

Integrating by parts the second term gives

$$\|u_{t0}\|^2 = a(\Delta u_0, u_{t0}) - a(\partial_{n_i} u_0, u_{t0})_{\Gamma_i} - \frac{p + (-1)^i b_x}{2}(u_0, u_{t0})_{\Gamma_i} - (b \cdot \nabla u_0, u_{t0}) + k(v_0, u_{t0}) - kc(u_0, u_{t0}) + (g(\cdot, 0), u_{t0})_{\Gamma_i}.$$

Since the term $-a\partial_{n_i} u_0 - \frac{p+(-1)^i b_x}{2} u_0 + g(\cdot, 0)$ vanishes, we obtain

$$\|u_{t0}\|^2 = (a\Delta u_0 - b \cdot \nabla u_0 + kv_0 - kcu_0, u_{t0}) \leq \|a\Delta u_0 - b \cdot \nabla u_0 + kv_0 - kcu_0\| \|u_{t0}\|,$$

by applying the Cauchy-Schwarz inequality, and thus

$$\|u_{t0}\| \leq \max(1, a, \|b\|_\infty, k, kc) \left(\|u_0\|_{H^2} + \|v_0\|_{L^2}\right).$$

In the same way, by taking the inner product of equation $\partial_t v + k(v - cu) = 0$ with $\partial_t v$, by evaluating the resulting equation at time $t = 0$ and by applying the Cauchy-Schwarz inequality, we obtain

$$\|v_{t0}\| \leq \| - kv_0 + kcu_0\| \leq \|u_0\|_{L^2} + \|v_0\|_{L^2}.$$

We combine the above inequality and (4.15), which gives the energy estimate

$$\|\partial_t u\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \|\partial_t v\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \min(a, 2kc)\|\partial_t u\|^2_{L^2(0,T;H^1(\Omega_i))} + \frac{p}{2}\|\partial_t u\|^2_{L^2(0,T;L^2(\Gamma))} + 2k\|\partial_t v\|^2_{L^2(0,T;L^2(\Omega_i))} \leq \tilde{C} e^{CT} \left(\|u_0\|^2_{H^2(\Omega_i)} + \|v_0\|^2_{L^2(\Omega_i)} + \|g\|^2_{H^1(0,T;L^2(\Gamma))}\right), \quad (4.16)$$

where $\tilde{C}$ is a constant that only depends on a, b, k and c. By putting together (4.14) and (4.16) we obtain by a Galerkin method a unique solution (u, v) of (4.13) such that

$$u \in W^{1,\infty}(0, T; L^2(\Omega_i)) \cap H^1(0, T; H^1(\Omega_i)),$$
$$v \in W^{1,\infty}(0, T; L^2(\Omega_i)) \cap H^1(0, T; L^2(\Omega_i)).$$

We have that $H^1(0, T; H^1(\Omega_i)) \subseteq L^\infty(0, T; H^1(\Omega_i))$ and it remains to show that $u \in L^\infty(0, T; H^2(\Omega_i))$. To do so we recall that the inclusion $H^1(0, T; L^2(\Omega_i)) \subseteq L^\infty(0, T; L^2(\Omega_i))$ holds and that $u \in H^1(0, T; H^1(\Omega_i))$ implies that $u_{|\Gamma_i} \in H^1(0, T; H^{\frac{1}{2}}(\Gamma_i)) \subseteq L^\infty(0, T; H^{\frac{1}{2}}(\Gamma_i))$. Since

$$\Delta u = \frac{1}{a}(u_t + b \cdot \nabla u - k(v - cu)) \in L^\infty(0, T; L^2(\Omega_i)),$$
$$2a\frac{\partial u}{\partial n_i} = g - (p + (-1)^i b_x)u \in L^\infty(0, T; H^{\frac{1}{2}}(\Gamma_i)),$$

we get $u \in L^\infty(0, T; H^2(\Omega_i))$, with

$$\|u\|_{L^\infty(0,T;H^2(\Omega_i))} \leq \|\Delta u\|_{L^\infty(0,T;L^2(\Omega_i))} + a\left\|\frac{\partial u}{\partial n_i}\right\|_{L^\infty(0,T;H^{\frac{1}{2}}(\Gamma_i))} \leq \tilde{\tilde{C}}\left(\|u_t\|_{L^\infty(0,T;L^2(\Omega_i))} + \|u\|_{L^\infty(0,T;H^1(\Omega_i))} + \|v\|_{L^\infty(0,T;L^2(\Omega_i))} + \|g\|_{L^\infty(0,T;H^{\frac{1}{2}}(\Gamma_i))}\right),$$

where the constant $\tilde{\tilde{C}}$ only depends on the constants $a,\ b,\ k$ and c.

2. The case $q > 0$.
The proof follows the same lines as for the case of Robin boundary conditions. We take the inner product of equation $\partial_t u + \operatorname{div}(-a\nabla u + b \cdot u) - k(v - cu) = 0$ with u, we integrate by parts in Ω_i and on Γ_i, and we take the inner product of equation $\partial_t v + k(v - cu) = 0$ with v. Since we need now further regularity for v, we apply the gradient operator to the equation $\partial_t v + k(v - cu) = 0$ and we take the inner product of the resulting equation with ∇v. We sum the three equations and we apply the Cauchy-Schwarz inequality and the inequality $mn \leq \frac{m^2}{2\varepsilon} + \frac{\varepsilon}{2}n^2$, for $m,\ n \in \mathbb{R}$ and $\varepsilon \geq 0$, in the right-hand side of the resulting equation. We obtain

$$
\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|u(t)\|^2 + \|v(t)\|^2 + \|\nabla v(t)\|^2 + \frac{q}{2}\|u(t)\|^2_{\Gamma_i}\right) + a\|\nabla u(t)\|^2 + k\|\nabla v(t)\|^2 + \frac{p}{2}\|u(t)\|^2_{\Gamma_i} \\
&\quad + \frac{q}{2}a\|\nabla_y u(t)\|^2_{\Gamma_i} + kc\|u(t)\|^2 + k\|v(t)\|^2 \\
&\quad = (k + kc)(v(t), u(t)) + \frac{qk}{2}(v(t), u(t))_{\Gamma_i} + (g(t), u(t))_{\Gamma_i} + kc(\nabla u(t), \nabla v(t)) \\
&\quad \leq \frac{k + kc}{2}(\|u(t)\|^2 + \|v(t)\|^2) + \frac{qk}{4}(\|u(t)\|^2_{\Gamma_i} + C_{\Omega_i}\|v(t)\|^2_{H^1(\Omega_i)}) \\
&\qquad\qquad + \frac{\|g(t)\|^2_{\Gamma_i} + \|u(t)\|^2_{\Gamma_i}}{2} + \frac{a}{2}\|\nabla u(t)\|^2 + \frac{k^2c^2}{2a}\|\nabla v(t)\|^2,
\end{aligned}
$$

where C_{Ω_i} is a constant depending on Ω_i. We integrate over $(0, t)$, for $t \leq T$, both sides of the above inequality, and we apply the Gronwall lemma. We obtain a first estimate which is the analogous of (4.14):

$$
\begin{aligned}
\|u\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \|v\|^2_{L^\infty(0,T;H^1(\Omega_i))} + q\|u\|^2_{L^\infty(0,T;L^2(\Gamma_i))} + \min(a, 2kc)\|u\|^2_{L^2(0,T;H^1(\Omega_i))} \\
+ \min(p, qa)\|u\|^2_{L^2(0,T;H^1(\Gamma_i))} + 2k\|v\|^2_{L^2(0,T;L^2(\Omega_i))} \\
\leq e^{CT}\left(\|u_0\|^2_{H^2} + \|v_0\|^2_{H^1} + \|g\|^2_{L^2(0,T;L^2(\Gamma))}\right), \quad (4.17)
\end{aligned}
$$

where C depends again on $a,\ k,\ c$ and b.
We now take the inner product of equation $\partial_t u + \operatorname{div}(-a\nabla u + b \cdot u) - k(v - cu) = 0$ with $\partial_t u$ and integrate by parts in Ω_i. We also take the inner product of equation $\partial_t v + k(v - cu) = 0$

with $\partial_t v$, we sum the two resulting equations and apply the Cauchy-Schwarz inequality and the inequality $mn \leq \frac{m^2}{2\epsilon} + \frac{\epsilon}{2}n^2$ in the right-hand side. We get

$$\begin{aligned}
&\|\partial_t u(t)\|^2 + \|\partial_t v(t)\|^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left(a\|\nabla u(t)\|^2 + kc\|u(t)\|^2 + \frac{p+(-1)^i b_x}{2}\|u(t)\|_{\Gamma_i} + \frac{aq}{2}\|\nabla_{\boldsymbol{y}} u(t)\|^2_{\Gamma_i}\right.\\
&\quad\left. + k\|v(t)\|^2\right) + \frac{q}{2}\|\partial_t u(t)\|^2_{\Gamma_i}\\
&= -\frac{q}{2}\left(b_{\boldsymbol{y}}\cdot\nabla_{\boldsymbol{y}} u(t), \partial_t u(t)\right)_{\Gamma_i} + \frac{qk}{2}\left(v(t), \partial_t u(t)\right)_{\Gamma_i} - \left(b\cdot\nabla u(t), \partial_t u(t)\right)\\
&\quad + kc\left(u(t), \partial_t v(t)\right) + k\left(v(t), \partial_t u(t)\right)\\
&\leq \frac{q}{2}\left(\|b\|_\infty\|\nabla_{\boldsymbol{y}} u(t)\|^2_{\Gamma_i} + \frac{\|\partial_t u(t)\|^2_{\Gamma_i}}{4}\right) + \frac{q}{2}\left(k^2 C_{\Omega_i}\|v(t)\|^2_{H^1(\Omega_i)} + \frac{\|\partial_t u(t)\|^2_{\Gamma_i}}{4}\right)\\
&\quad + \|b\|_\infty\|\nabla u(t)\|^2 + \frac{\|\partial_t u(t)\|^2}{4} + kc\left(\frac{\|u(t)\|^2}{2} + \frac{\|\partial_t v(t)\|^2}{2}\right) + k\|v(t)\|^2 + \frac{\|\partial_t u(t)\|^2}{4}.
\end{aligned}$$

By integrating both sides of the above inequality over $(0,t)$, for $t \leq T$, by applying the Gronwall lemma together with (4.17), we obtain

$$\begin{aligned}
&2(\|\partial_t u\|^2_{L^2(0,T;L^2(\Omega_i))} + \|\partial_t v\|^2_{L^2(0,T;L^2(\Omega_i))}) + \min(a, kc)\|u\|^2_{L^\infty(0,T;H^1(\Omega_i))}\\
&\qquad + k\|v\|^2_{L^\infty(0,T;L^2(\Omega_i))} + q\|\partial_t u\|^2_{L^2(0,T;L^2(\Gamma_i))} + \frac{qa}{2}\|\nabla_{\boldsymbol{y}} u\|^2_{L^\infty(0,T;L^2(\Gamma_i))}\\
&\qquad\qquad \leq e^{CT}\left(\|u_0\|^2_{H^2} + \|v_0\|^2_{H^1} + \|g\|^2_{L^2(0,T;L^2(\Gamma))}\right). \quad (4.18)
\end{aligned}$$

By putting together (4.17) and (4.18), and by using the inclusions $H^1(0,T;L^2(\Omega_i)) \subseteq L^\infty(0,T;L^2(\Omega_i))$ and $L^\infty(0,T) \subseteq L^2(0,T)$, we obtain by a Galerkin method a unique solution (u,v) of (4.13) such that

$$\begin{aligned}
u &\in L^\infty(0,T;H^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),\\
v &\in L^\infty(0,T;H^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),
\end{aligned}$$

with

$$u_{|\Gamma_i} \in L^\infty(0,T;H^1(\Gamma_i)) \cap H^1(0,T;L^2(\Gamma_i)).$$

Once again, by using equation $\Delta u = \frac{1}{a}(\partial_t u + b\cdot\nabla u + k(v - cu)) \in L^2(0,T;L^2(\Omega_i))$ and $g \in L^2(0,T;L^2(\Gamma_i))$, we get $u \in L^2(0,T;H^2(\Omega_i))$ with

$$\|u\|_{L^2(0,T;H^2(\Omega_i))} \leq \frac{1}{a}\left(\|\partial_t u\|_{L^2(0,T;L^2(\Omega_i))} + \|u\|_{L^2(0,T;H^1(\Omega_i))} + \|v\|_{L^2(0,T;L^2(\Omega_i))} + \|g\|_{L^2(0,T;L^2(\Gamma_i))}\right).$$

We use equation $\Delta_{\boldsymbol{y}} u = \frac{1}{a}\left(\frac{\partial u}{\partial n} + pu + q\partial_t u + b_{\boldsymbol{y}}\nabla u + kv - g\right)$ to obtain $\Delta_{\boldsymbol{y}} w \in L^2(0,T;L^2(\Gamma_i))$, which finishes the proof. ⊠

4.2.3.2 Well-Posedness of the Non-Overlapping Algorithm with Robin and Ventcel Transmission Conditions

In the non-overlapping case we have $L = 0$ and $\Gamma := \Gamma_1 = \Gamma_2$. As a consequence of Lemma 4.1, we obtain the following theorem concerning the well-posedness of the non-overlapping Schwarz waveform relaxation algorithms of order 0 and 2.

Theorem 4.2

Let $p > 0,\ q \geq 0$ be given and $L = 0$.

1. If $q = 0$ and if $(u_0, v_0) \in H^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ and $(g_{b_1}^0, g_{b_2}^0) \in (H^1(0,T;L^2(\Gamma)))^2 \cap (L^\infty(0,T;H^{\frac{1}{2}}(\Gamma)))^2$ are given, then algorithm 4.1 with the transmission operators defined by (4.9), defines a unique sequence of iterates $\left(\left(u_1^k, v_1^k\right), \left(u_2^k, v_2^k\right)\right)$ such that

$$u_i^k \in W^{1,\infty}(0,T;L^2(\Omega_i)) \cap L^\infty(0,T;H_1^2(\Omega_i)) \cap H^1(0,T;H^1(\Omega_i)),$$

and

$$v_i^k \in W^{1,\infty}(0,T;L^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

for $i \in \{1,2\}$.

2. If $q > 0$ and if $(u_0, v_0) \in H^2(\mathbb{R}^d) \times H^1(\mathbb{R}^d)$ and $(g_{b_1}^0, g_{b_2}^0) \in (L^2(0,T;L^2(\Gamma)))^2$ are given, then algorithm 4.1 with the transmission operators defined by (4.10) defines a unique sequence of iterates $\left(\left(u_1^k, v_1^k\right), \left(u_2^k, v_2^k\right)\right)$ such that

$$u_i^k \in L^2(0,T;H_2^2(\Omega_i)) \cap L^\infty(0,T;H_1^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

and

$$v_i^k \in L^\infty(0,T;H^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

with $u_{i\,|\Gamma}^k \in H^1(0,T;L^2(\Gamma))$ and $v_{i\,|\Gamma}^k \in L^2(0,T;L^2(\Gamma))$ $i \in \{1,2\}$.

Proof 4.2

It suffices to show that if

$$g \in \begin{cases} H^1(0,T;L^2(\Gamma)) \cap L^\infty(0,T;H^{\frac{1}{2}}(\Gamma)), & \text{if } q = 0, \\ L^2(0,T;L^2(\Gamma)), & \text{if } q > 0, \end{cases}$$

then the solution of problem (4.13) given by Lemma 4.1 is such that

$$\frac{\partial u}{\partial n_j} + \frac{p + (-1)^i b_x}{2a} u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\boldsymbol{y}} u + b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} u - kv\right) \in \begin{cases} H^1(0,T;L^2(\Gamma)) \cap L^\infty(0,T;H^{\frac{1}{2}}(\Gamma)), & \text{if } q = 0, \\ L^2(0,T;L^2(\Gamma)), & \text{if } q > 0, \end{cases} \quad (4.19)$$

where $(i, j) \in \{(1,2),(2,1)\}$.
Either if $q = 0$ or if $q > 0$, we have

$$\begin{aligned}
\frac{\partial u}{\partial n_j} &+ \frac{p + (-1)^i b_x}{2a} u + \frac{q}{2a}\left(\partial_t u - a\Delta_y u + b_y \cdot \nabla_y u - kv\right) \\
&= -\frac{\partial u}{\partial n_i} - \frac{p + (-1)^i b_x}{2a} u - \frac{q}{2a}\left(\partial_t u - a\Delta_y u + b_y \cdot \nabla_y u - kv\right) \\
&\quad + 2\frac{p + (-1)^i b_x}{2a} u + 2\frac{q}{2a}\left(\partial_t u - a\Delta_y u + b_y \cdot \nabla_y u - kv\right) \\
&= -g + 2\frac{p + (-1)^i b_x}{2a} u + 2\frac{q}{2a}\left(\partial_t u - a\Delta_y u + b_y \cdot \nabla_y u - kv\right).
\end{aligned}$$

The result of Lemma 4.1 implies then that (4.19) holds. We can thus iterate the proof over k to obtain the result of the theorem. ⊠

4.2.3.3 Well-Posedness of the Overlapping Algorithm with Dirichlet Transmission Conditions

In this section we prove that the overlapping classical Schwarz waveform relaxation algorithm with transmission conditions defined by the operators

$$\mathcal{B}_i(u, v) = u, \tag{4.20}$$

is well-defined.

Let $i \in \{1,2\}$ and consider $g \in L^2(0,T;H^{\frac{3}{2}}(\Gamma_i)) \cap H^{\frac{3}{4}}(0,T;L^2(\Gamma_i))$ and $u_0 \in H^2(\Omega_i)$ such that $g(\cdot,0) = u_0(x_i,\cdot)$, where x_i is the $x-$abscissa of the interface Γ_i. Then there exists $w \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$ such that $w = g$, over $\Gamma_i \times (0,T)$, and $w(\cdot,0) = u_0(\cdot)$ (cf. [56] for the proof of this result). We have thus that problem

$$\begin{aligned}
&\partial_t u + \operatorname{div}(-a\nabla u + bu) - k(v - cu) = 0, && \text{in } \Omega_i \times (0,T), \\
&\partial_t v + k(v - cu) = 0, && \text{in } \Omega_i \times (0,T), \\
&(u,v)(\cdot,0) = (u_0(\cdot), v_0(\cdot))_{|\Omega_i}, && \text{in } \Omega_i, \\
&u = g, && \text{over } \Gamma_i \times (0,T),
\end{aligned} \tag{4.21}$$

is equivalent to problem

$$\begin{aligned}
&\partial_t \tilde{u} + \operatorname{div}(-a\nabla \tilde{u} + b\tilde{u}) - k(v - c\tilde{u}) = \tilde{f}_1, && \text{in } \Omega_i \times (0,T), \\
&\partial_t v + k(v - c\tilde{u}) = \tilde{f}_2, && \text{in } \Omega_i \times (0,T), \\
&(\tilde{u},v)(\cdot,0) = (0, v_0(\cdot)_{|\Omega_i}), && \text{in } \Omega_i, \\
&\tilde{u} = 0, && \text{over } \Gamma_i \times (0,T),
\end{aligned} \tag{4.22}$$

where $\tilde{f}_1 = -\partial_t w - \mathrm{div}(-a\nabla w + bw) - kcw \in L^2(0,T;L^2(\Omega_i))$ and $\tilde{f}_2 = kcw$.

In fact, we define a weak solution of problem (4.22) as a function $(u,v) \in L^2(0,T;H_0^1(\Omega_i)) \times L^2(0,T;L^2(\Omega_i)) \cap C(0,T;L^2(\Omega_i)) \times C(0,T;L^2(\Omega_i))$, satisfying for all $(w,z) \in H_0^1(\Omega_i) \times L^2(\Omega_i)$

$$\frac{\partial}{\partial t}\left((u,w)+(v,z)\right) + a\left(\nabla u, \nabla w\right) + \frac{1}{2}\left(\left((b\cdot\nabla)\,u, w\right) - \left((b\cdot\nabla)\,w, u\right)\right) \\ - k\left((v,w)-(v,z)\right) + ck\left((u,w)-(u,z)\right) = (\tilde{f}_1, w) + (\tilde{f}_2, z),$$

in $\mathcal{D}'(0,T)$, and $u_{|t=0} = 0$, $v_{|t=0} = v_0$. Then (u,v) is a weak solution of problem (4.21) if $u = w + \tilde{u}$ and $(\tilde{u}, v)$ is a weak solution of (4.22).

In [56] it is proved that, if $\tilde{f} \in L^2(0,T;L^2(\Omega_i))$, then the initial and boundary value problem for the scalar reactive transport equation in $\Omega_i \times (0,T)$

$$\partial_t \tilde{u} + \mathrm{div}(-a\nabla\tilde{u} + b\tilde{u}) + kc\tilde{u} = \tilde{f},$$

with homogeneous Dirichlet boundary conditions over $\Gamma_i \times (0,T)$ and with initial condition $\tilde{u}(\cdot,0) = 0$, possesses a unique weak solution $\tilde{u} \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$. Since the second equation in (4.22) is an ordinary differential equation in the variable v, we can extend the result of [56] in order to obtain the following well-posedness result for problem (4.21):

Lemma 4.2 (*Well-posedness of the subproblems*)

Let $g \in L^2(0,T;H^{\frac{3}{2}}(\Gamma_i)) \cap H^{\frac{3}{4}}(0,T;L^2(\Gamma_i))$ and $(u_0, v_0) \in H^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ be given such that $g(\cdot,0) = u_0(x_i,\cdot)$. Then problem (4.21) has a unique solution (u,v) such that

$$u \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

and

$$v \in H^1(0,T;L^2(\Omega_i)).$$

We also obtain an analogous of Theorem 4.2 concerning the well-posedness of the classical Schwarz waveform relaxation algorithm:

Theorem 4.3 (*Well-posedness of the algorithm*)

Let $(u_0, v_0) \in H^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ and $g_{b_i}^0 \in (L^2(0,T;H^{\frac{3}{2}}(\Gamma_i))) \cap (H^{\frac{3}{4}}(0,T;L^2(\Gamma_i)))$, $i = 1,2$, be given such that $g_{b_1}^0(\cdot,0) = u_0(L,\cdot)$, $g_{b_2}^0(\cdot,0) = u_0(0,\cdot)$. Then algorithm 4.1 with the transmission operators defined by (4.20) defines a unique sequence of iterates $\left((u_1^k, v_1^k), (u_2^k, v_2^k)\right)$ such that

$$u_i^k \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$

and $v_i^k \in H^1(0,T;L^2(\Omega_i))$, $i \in \{1,2\}$.

Proof 4.3

By a trace theorem proved in [56], if $u \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$, then $u_{|\Gamma_i} \in (L^2(0,T;H^{\frac{3}{2}}(\Gamma_i)))^2 \cap (H^{\frac{3}{4}}(0,T;L^2(\Gamma_i)))^2$ and $u(x_i,\cdot,0) = u_0(x_i,\cdot)$. We can thus recursively use the result of Lemma 4.2 in order to conclude the result of the theorem. ⊠

4.2.3.4 Well-Posedness of the Overlapping Algorithm with Robin and Ventcel Transmission Conditions

In the case of an indeed present $L > 0$, we need to prove a more fine regularity result inside both domains Ω_1 and Ω_2.

We begin by recalling the following trace theorem, whose proof can be found in [56]:

Theorem 4.4

Let $s \geq 0,\ r > \frac{1}{2}$ and $u \in L^2(0,T;H^r(\Omega_i)) \cap H^s(0,T;L^2(\Omega_i))$. Then we have

1. If $j < r - \frac{1}{2}$, then $\frac{\partial^j u}{\partial {n_i}^j}$ has a trace over $\Gamma_i \times [0,T]$ and $\frac{\partial^j u}{\partial {n_i}^j} \in L^2(0,T;H^{\mu_j}(\Gamma_i)) \cap H^{a_j}(0,T;L^2(\Gamma_i))$, where $\frac{\mu_j}{r} = \frac{a_j}{s} = \frac{r-j-\frac{1}{2}}{r},\ a_j = 0$, if $s = 0$.
2. If $s > \frac{1}{2}$ and $k < s - \frac{1}{2}$, then $\frac{\partial^k u}{\partial t^k}\Big|_{t=0}$ has a trace over Ω_i and $\frac{\partial^k u}{\partial t^k}\Big|_{t=0} \in H^{p_k}(\Omega_i)$, where $p_k = \frac{r}{s}\left(s - k - \frac{1}{2}\right)$.

Furthermore, the map

$$u \longrightarrow \left\{\frac{\partial^j u}{\partial {n_i}^j}\right\}_{j<s-\frac{1}{2}} \times \left\{\frac{\partial^k u}{\partial t^k}\right\}_{k<r-\frac{1}{2}}$$

is continuous from $L^2(0,T;H^r(\Omega_i)) \cap H^s(0,T;L^2(\Omega_i))$ into

$$F := \prod_{j<s-\frac{1}{2}} L^2(0,T;H^{\mu_j}(\Gamma_i)) \cap H^{a_j}(0,T;L^2(\Gamma_i)) \times \prod_{k<r-\frac{1}{2}} H^{p_k}(\Omega_i)$$

and is onto

$$F_0 := \left\{\left\{g_j\right\} \times \{f_k\} \in F \text{ s. t. } \frac{\partial^k g_j}{\partial t^k}_{|t=0} = \frac{\partial^j f_k}{\partial {n_i}^j}_{|x=0},\ \frac{j}{r} + \frac{k}{s} < 1 - \frac{1}{2}\left(\frac{1}{r} + \frac{1}{s}\right)\right\},$$

provided that $1 - \frac{1}{2}\left(\frac{1}{r} + \frac{1}{s}\right) > 0$.

We follow now the ideas of [60] and [61], where the case of the scalar advection reaction diffusion equation is studied. We apply Theorem 4.4 with $r = 2,\ s = 1$, when $q = 0$ and with $r = 3,\ s = \frac{3}{2}$, when $q > 0$.

Let $q = 0$ and suppose that $g \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_i)) \cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))$ and $u_0 \in H^1(\Omega_i)$ are such that

$$\begin{cases} g = g_1 + \frac{p+(-1)^i b_x}{2a} g_0, \\ g_0 \in L^2(0,T;H^{\frac{3}{2}}(\Gamma_i)) \cap H^{\frac{3}{4}}(0,T;L^2(\Gamma_i)) \text{ such that } g_0(\cdot,0) = u_0(x_i,\cdot), \\ g_1 \in L^2(0,T;H^{\frac{1}{2}}(\Gamma_i)) \cap H^{\frac{1}{4}}(0,T;L^2(\Gamma_i)). \end{cases} \tag{4.23}$$

Then, Theorem 4.4 implies that there exists $w \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$ such that $w(\cdot,0) = u_0(\cdot)$ and $\frac{\partial w}{\partial n_i} + \frac{p+(-1)^i b_x}{2a} w = g$ over $\Gamma_i \times [0,T]$.

Let now $q > 0$ and suppose that $g \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_i)) \cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))$ and $(u_0, v_0) \in H^2(\Omega_i) \times H^2(\Omega_i)$ are such that

$$\begin{cases} g = g_1 + \left(\frac{p+(-1)^i b_x}{2a} + \frac{q}{2a}(\partial_t + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} - a\Delta_{\mathbf{y}} + kc) \right) g_0 - \frac{q}{2a} k v_0(x_i,\cdot), \\ g_0 \in L^2(0,T;H^{\frac{5}{2}}(\Gamma_i)) \cap H^{\frac{5}{4}}(0,T;L^2(\Gamma_i)) \text{ such that } g_0(\cdot,0) = u_0(x_i,\cdot), \\ g_1 \in L^2(0,T;H^{\frac{3}{2}}(\Gamma_i)) \cap H^{\frac{3}{4}}(0,T;L^2(\Gamma_i)) \text{ such that } g_1(\cdot,0) = \frac{\partial u_0}{\partial n_i}(x_i,\cdot). \end{cases} \tag{4.24}$$

Then, Theorem 4.4 implies that there exists $w \in L^2(0,T;H^3(\Omega_i)) \cap H^{\frac{3}{2}}(0,T;L^2(\Omega_i))$ such that $w(\cdot,0) = u_0(\cdot)$, $\frac{\partial w}{\partial n_i}(xi,\cdot,\cdot) = g_1$ and

$$\frac{\partial w}{\partial n_i} + \frac{p+(-1)^i b_x}{2a} w + \frac{q}{2a}(\partial_t w + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} w - a\Delta_{\mathbf{y}} w + kcw - kv_0) = g,$$

over $\Gamma_i \times [0,T]$ (see [60] and [61] for the details in a similar case).

We have thus that problem (4.13) is equivalent to problem

$$\begin{array}{ll} \partial_t \tilde{u} + \operatorname{div}(-a\nabla\tilde{u} + b\tilde{u}) - k(v - c\tilde{u}) = \tilde{f}_1, & \text{in } \Omega_i \times (0,T), \\ \partial_t v \qquad\qquad\qquad + k(v - c\tilde{u}) = \tilde{f}_2, & \text{in } \Omega_i \times (0,T), \\ (\tilde{u}, v)(\cdot,0) = (0, v_0(\cdot)_{|\Omega_i}), & \text{in } \Omega_i, \\ \frac{\partial \tilde{u}}{\partial n_i} + \frac{p+(-1)^i b_x}{2a} \tilde{u} + \frac{q}{2a}\left(\partial_t \tilde{u} - a\Delta_{\mathbf{y}} \tilde{u} + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} \tilde{u} - kv + kc\tilde{u}\right) = 0, & \text{over } \Gamma_i \times (0,T), \end{array} \tag{4.25}$$

where

$$\begin{cases} \tilde{f}_1 = -\partial_t w - \operatorname{div}(-a\nabla w + bw) - kcw, \\ \tilde{f}_2 = kcw, \end{cases} \tag{4.26}$$

and

$$\tilde{f}_1,\ \tilde{f}_2 \in \begin{cases} L^2(0,T;L^2(\Omega_i)), & \text{if } q = 0, \\ L^2(0,T;H^1(\Omega_i)) \cap H^{\frac{1}{2}}(0,T;L^2(\Omega_i)), & \text{if } q > 0. \end{cases}$$

In fact, a function (u,v) is a weak solution of problem (4.13) if and only if $(\tilde{u},v)$ with $u = w + \tilde{u}$ is a weak solution of problem (4.25).

We state and prove now a refinement of Lemma 4.1:

Lemma 4.3

Let $p > 0,\ q \geq 0$ be given.

1. If $q = 0$, consider $(u_0, v_0) \in H^2(\mathbb{R}^d)\times L^2(\mathbb{R}^d)$ and $g \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_i))\cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))$. Suppose in addition that (4.23) holds. Then problem (4.13) has a unique solution (u, v) such that

$$u \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),\quad v \in H^1(0,T;L^2(\Omega_i)).$$

2. If $q > 0$, consider $(u_0, v_0) \in H^2(\mathbb{R}^d)\times H^2(\mathbb{R}^d)$ and $g \in L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))\cap H^{\frac{1}{4}}(0,T;L^2(\Omega_i))$, such that (4.24) holds. Then problem (4.13) has a unique solution (u, v) such that

$$u \in L^2(0,T;H^3(\Omega_i)) \cap H^{\frac{3}{2}}(0,T;L^2(\Omega_i)),$$

and

$$v \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)).$$

Proof 4.4

1. The case $q = 0$.
We can adapt the proof of Lemma 4.1 to show that problem (4.25), with $\tilde{f_1}$ and $\tilde{f_2}$ defined by (4.26), has a solution $(\tilde{u}, v)$ such that $\tilde{u} \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$ and $v \in H^1(0,T;L^2(\Omega_i))$. First, we carry out the same analysis in order to obtain the analogous of estimate (4.14):

$$\begin{aligned}\|\tilde{u}\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \|v\|^2_{L^\infty(0,T;L^2(\Omega_i))} + \min(a,2kc)\|\tilde{u}\|^2_{L^2(0,T;H^1(\Omega_i))} + \frac{p}{2}\|\tilde{u}\|^2_{L^2(0,T;L^2(\Gamma))}\\ + 2k\|v\|^2_{L^2(0,T;L^2(\Omega_i))} \leq e^{CT}\left(\|v_0\|^2 + \|\tilde{f_1}\|^2_{L^2(0,T;L^2(\Omega_i))} + \|\tilde{f_2}\|^2_{L^2(0,T;L^2(\Omega_i))}\right). \quad (4.27)\end{aligned}$$

In order to estimate $\partial_t\tilde{u}$ and $\partial_t v$ in $L^2(0,T;L^2(\Omega_i))$, we take the inner product of equation $\partial_t\tilde{u} + \operatorname{div}(-a\nabla\tilde{u} + b\cdot\tilde{u}) - k(v - c\tilde{u}) = \tilde{f_1}$ with $\partial_t\tilde{u}$ and integrate by parts in Ω_i, we take the inner product of equation $\partial_t v + k(v - c\tilde{u}) = \tilde{f_2}$ with $\partial_t v$, we sum the two resulting equations and apply the Cauchy-Schwarz inequality and the inequality $mn \leq \frac{m^2}{2\epsilon} + \frac{\epsilon}{2}n^2$ in the right-hand side. We get

$$\begin{aligned}&\|\partial_t\tilde{u}(t)\|^2 + \|\partial_t v(t)\|^2 + \frac{1}{2}\frac{\partial}{\partial t}\left(a\|\nabla\tilde{u}(t)\|^2 + kc\|\tilde{u}(t)\|^2 + \frac{p + (-1)^i b_x}{2}\|\tilde{u}(t)\|_{\Gamma_i} + k\|v(t)\|^2\right)\\ &= -(b\cdot\nabla\tilde{u}(t),\partial_t\tilde{u}(t)) + kc\,(\tilde{u}(t),\partial_t v(t)) + k\,(v(t),\partial_t\tilde{u}(t)) + (\tilde{f_1}(t),\partial_t\tilde{u}(t)) + (\tilde{f_2}(t),\partial_t v(t))\\ &\leq \|b\|_\infty\|\nabla\tilde{u}(t)\|^2 + \frac{\|\partial_t\tilde{u}(t)\|^2}{4} + kc\frac{\|\tilde{u}(t)\|^2}{2} + \frac{\|\partial_t v(t)\|^2}{4} + k\|v(t)\|^2 + \frac{\|\partial_t\tilde{u}(t)\|^2}{4}\\ &\quad + \|\tilde{f_1}(t)\|^2 + \frac{\|\partial_t\tilde{u}(t)\|^2}{4} + \|\tilde{f_2}(t)\|^2 + \frac{\|\partial_t v(t)\|^2}{4}.\end{aligned}$$

By integrating both sides of the above inequality over $(0,t)$, for $t \leq T$ and by applying the Gronwall lemma together with (4.27), we obtain

$$\|\partial_t \tilde{u}\|^2_{L^2(0,T;L^2(\Omega_i))} + \|\partial_t v\|^2_{L^2(0,T;L^2(\Omega_i))}) + \min(a, kc)\|\tilde{u}\|^2_{L^\infty(0,T;H^1(\Omega_i))} + k\|v\|^2_{L^\infty(0,T;L^2(\Omega_i))} \leq Ce^{CT}\left(\|v_0\|^2_{H^1} + \|\tilde{f}_1\|^2_{L^2(0,T;L^2(\Omega_i))} + \|\tilde{f}_2\|^2_{L^2(0,T;L^2(\Omega_i))}\right). \quad (4.28)$$

Once again, we obtain by a Galerkin method a unique solution $(\tilde{u}, v)$ of (4.25) such that

$$\tilde{u} \in L^2(0,T;H^1(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)),$$
$$v \in H^1(0,T;L^2(\Omega_i)).$$

We now use once more the properties $\Delta\tilde{u} = \frac{1}{a}(\tilde{u}_t + b \cdot \nabla\tilde{u} - k(v - c\tilde{u})) \in L^2(0,T;L^2(\Omega_i))$ and $2a\frac{\partial\tilde{u}}{\partial n_i} = -(p + (-1)^i b_x)\tilde{u} \in L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))$, to get $\tilde{u} \in L^2(0,T;H^2(\Omega_i))$. Since $u = \tilde{u} + w$ and $w \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$, we get a unique solution (u,v) of (4.13) such that $u \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i))$.

2. The case $q > 0$.
The proof is much more technical but follows the same ideas as for the case $q = 0$. We refer to [61] and [6] for the details. ⊠

We have now that, in consequence of Lemma 4.3, for $q \geq 0$, if

$$g \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_i)) \cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_i)),$$

then the solution of problem (4.13) given by Lemma 4.3 is such that

$$\frac{\partial u}{\partial n_j} + \frac{p + (-1)^i b_x}{2a}u + \frac{q}{2a}\left(\partial_t u - a\Delta_{\mathbf{y}} u + b_{\mathbf{y}} \cdot \nabla_{\mathbf{y}} u - kv\right) \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_j)) \cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_j))$$

where $(i,j) \in \{(1,2),(2,1)\}$.

We obtain thus the analogous of Theorem 4.2:

Theorem 4.5

Let $p > 0,\ q \geq 0$ be given and $L > 0$.

1. If $q = 0$ and if $(u_0, v_0) \in H^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ and $g_{b_i}^0 \in H^{\frac{1}{4}}(0,T;L^2(\Gamma_i)) \cap L^2(0,T;H^{\frac{1}{2}}(\Gamma_i))$, $i = 1,2$, are given, then algorithm 4.1 with the transmission operators defined by (4.9) defines a unique sequence of iterates $\left((u_1^k, v_1^k), (u_2^k, v_2^k)\right)$ such that

$$u_i^k \in L^2(0,T;H^2(\Omega_i)) \cap H^1(0,T;L^2(\Omega_i)), \quad v_i^k \in H^1(0,T;L^2(\Omega_i)),$$

for $i \in \{1,2\}$.

2. If $q > 0$ and if $(u_0, v_0) \in H^2(\mathbb{R}^d) \times H^2(\mathbb{R}^d)$ and $g_{b_i}^0 \in L^2(0,T; H^{\frac{1}{2}}(\Gamma_i)) \cap H^{\frac{1}{4}}(0,T; L^2(\Gamma_i))$, $i = 1, 2$, are given, then algorithm 4.1 with the transmission operators defined by (4.10) defines a unique sequence of iterates $\left((u_1^k, v_1^k), (u_2^k, v_2^k)\right)$ such that

$$u_i^k \in L^2(0,T; H^3(\Omega_i)) \cap H^{\frac{3}{2}}(0,T; L^2(\Omega_i)),$$

and

$$v_i^k \in L^2(0,T; H^2(\Omega_i)) \cap H^1(0,T; L^2(\Omega_i)),$$

for $i \in \{1, 2\}$.

Remark 4.1

The proofs of the well-posedness of the three algorithms (with Dirichlet, Robin and Ventcel transmission conditions) can be easily extended to the case of variable parameters ϕ, a, b and k defined by continuous functions, provided that we suppose that there exists strictly positive constants m_p and M_p such that ϕ, a and k are respectively lower and upper bounded by m_p and by M_p and provided that the components of b have a constant sign.

4.2.4 Convergence of the Algorithm

In this section we prove the convergence of the SWR algorithm using different transmission conditions in appropriate spaces.

4.2.4.1 Convergence of the Non-Overlapping Algorithm with Robin and Ventcel Transmission Conditions

We begin by proving the convergence of the non overlapping algorithms. The proof is based on energy estimates (cf. [59], [22] or [47] for instance) and can be extended to the case of non constant parameters.

Theorem 4.6

The sequence $\left((u_1^k, v_1^k), (u_2^k, v_2^k)\right)$ defined by algorithm 4.1 with the transmission operators defined either by (4.9) or by (4.10) converges to $((u,v)_{|\Omega_1}, (u,v)_{|\Omega_2})$ in $\prod_{i=1}^2 (L^\infty(0,T\,; H^1(\Omega_i)) \times L^\infty(0,T\,; L^2(\Omega_i)))$ for $k \to \infty$.

Proof 4.5

For each $k > 0$, we define the errors $e_{u_i}^k = u_i^k - u_{|\Omega_i}$, $e_{v_i}^k = v_i^k - v_{|\Omega_i}$, which satisfy the equations

$$\begin{aligned}
&\partial_t e_{u_i}^k + \operatorname{div}(-a\nabla e_{u_i}^k + b e_{u_i}^k) - k(e_{v_i}^k - c e_{u_i}^k) = 0, && \text{in } \Omega_i \times (0,T),\\
&\partial_t e_{v_i}^k \qquad\qquad\qquad + k(e_{v_i}^k - c e_{u_i}^k) = 0, && \text{in } \Omega_i \times (0,T),\\
&(e_{u_i}^k, e_{v_i}^k)(\cdot, 0) = (0,0), && \text{in } \Omega_i,\\
&\mathcal{B}_i(e_{u_i}^k, e_{v_i}^k) = \mathcal{B}_i(e_{u_j}^{k-1}, e_{v_j}^{k-1}), && \text{over } \Gamma \times (0,T),
\end{aligned}$$

where $(i, j) \in \{(1, 2), (2, 1)\}$.

1. The case $q = 0$
By taking the inner product of equation $\partial_t e_{u_i}^k + \operatorname{div}(-a\nabla e_{u_i}^k + b e_{u_i}^k) - k(e_{v_i}^k - c e_{u_i}^k) = 0$ with $e_{u_i}^k$ and by integrating by parts in Ω_i, we obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{u_i}^k\right\|^2 + a\left\|\nabla e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{u_i}^k\right)_\Gamma + \left(b\cdot\nabla e_{u_i}^k, e_{u_i}^k\right) - k\left(e_{v_i}^k - c e_{u_i}^k, e_{u_i}^k\right) = 0.$$

We take the inner product of equation $\partial_t e_{v_i}^k + k(e_{v_i}^k - c e_{u_i}^k) = 0$ with $e_{v_i}^k$ in order to get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{v_i}^k\right\|^2 + k(e_{v_i}^k - c e_{u_i}^k, e_{v_i}^k) = 0,$$

and we sum the two inequalities above. We obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2\right) + a\left\|\nabla e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{u_i}^k\right)_\Gamma + \left(b\cdot\nabla e_{u_i}^k, e_{u_i}^k\right) + kc\left\|e_{u_i}^k\right\|^2 + k\left\|e_{v_i}^k\right\|^2 - (k+kc)\left(e_{u_i}^k, e_{v_i}^k\right) = 0.$$

Applying the Cauchy-Schwarz inequality and inequality $mn \le \frac{m^2}{2} + \frac{n^2}{2}$, and integrating by parts the term $(b\cdot\nabla e_{u_i}^k, e_{u_i}^k)$ yields

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2\right) + a\left\|\nabla e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{u_i}^k\right)_\Gamma + \frac{(-1)^j b_x}{2}\left\|e_{u_i}^k\right\|_\Gamma^2 + kc\left\|e_{u_i}^k\right\|^2 + k\left\|e_{v_i}^k\right\|^2 \le \frac{k+kc}{2}\left\|e_{u_i}^k\right\|^2 + \frac{k+kc}{2}\left\|e_{v_i}^k\right\|^2. \quad (4.29)$$

We now replace the boundary term using the identity

$$\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{u_i}^k\right)_\Gamma = \frac{a}{2p}\left\{\left\|\frac{\partial e_{u_i}^k}{\partial n_i} + \frac{p}{2a}e_{u_i}^k + \frac{(-1)^i b_x}{2a}e_{u_i}^k\right\|_\Gamma^2 - \left\|\frac{\partial e_{u_i}^k}{\partial n_i} - \frac{p}{2a}e_{u_i}^k + \frac{(-1)^i b_x}{2a}e_{u_i}^k\right\|_\Gamma^2\right\} - \frac{(-1)^i b_x}{2a}\left\|e_{u_i}^k\right\|^2.$$

We have $(-1)^i + (-1)^j = 0$, and thus

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2\right) + a\left\|\nabla e_{u_i}^k\right\|^2 + kc\left\|e_{u_i}^k\right\|^2 + k\left\|e_{v_i}^k\right\|^2 + \frac{a^2}{2p}\left\|\frac{\partial e_{u_i}^k}{\partial n_i} - \frac{p}{2a}e_{u_i}^k + \frac{(-1)^i b_x}{2a}e_{u_i}^k\right\|_\Gamma^2 \le \frac{k+kc}{2}\left\|e_{u_i}^k\right\|^2 + \frac{k+kc}{2}\left\|e_{v_i}^k\right\|^2 + \frac{a^2}{2p}\left\|\frac{\partial e_{u_i}^k}{\partial n_i} + \frac{p}{2a}e_{u_i}^k + \frac{(-1)^i b_x}{2a}e_{u_i}^k\right\|_\Gamma^2,$$

which yields

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\left\|e_{u_i}^{\ k}\right\|^2+\left\|e_{v_i}^{\ k}\right\|^2\right)+2a\left\|\nabla e_{u_i}^{\ k}\right\|^2+\min(2kc,2k)\left(\left\|e_{u_i}^{\ k}\right\|^2+\left\|e_{v_i}^{\ k}\right\|^2\right)$$
$$+\frac{a^2}{p}\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i}-\frac{p}{2a}e_{u_i}^{\ k}+\frac{(-1)^i b_x}{2a}e_{u_i}^{\ k}\right\|_\Gamma^2$$
$$\leq (k+kc)\left(\|u(t)\|^2+\|v(t)\|^2\right)+\frac{a^2}{p}\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i}+\frac{p}{2a}e_{u_i}^{\ k}+\frac{(-1)^i b_x}{2a}e_{u_i}^{\ k}\right\|_\Gamma^2. \quad (4.30)$$

We add now (4.30) for $i=1,2$ we integrate over $(0,t)$, for $t\leq T$, and we use the transmission condition

$$\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i}+\frac{p+(-1)^i b_x}{2a}e_{u_i}^{\ k}\right\|_\Gamma^2=\left\|\frac{\partial e_{u_j}^{\ k-1}}{\partial n_i}+\frac{p+(-1)^i b_x}{2a}e_{u_j}^{\ k-1}\right\|_\Gamma^2=\left\|\frac{\partial e_{u_j}^{\ k-1}}{\partial n_j}+\frac{-p+(-1)^j b_x}{2a}e_{u_j}^{\ k-1}\right\|_\Gamma^2$$

on the right-hand side of the resulting equation. Defining the energies

$$E(w)(t)=\|w(t)\|^2+\int_0^t\left(2a\|\nabla w(s)\|^2+\min(2kc,2k)\|w(s)\|^2\right)\mathrm{d}s,$$
$$\tilde{E}(w)(t)=\|w(t)\|^2+\int_0^t\min(2kc,2k)\|w(s)\|^2\,\mathrm{d}s,$$

and the boundary errors

$$g_i^k=\frac{\partial e_{u_i}^{\ k}}{\partial n_i}+\frac{-p+(-1)^i b_x}{2a}e_{u_i}^{\ k},$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left(E\left(e_{u1}^{\ k}\right)(t)+E\left(e_{u2}^{\ k}\right)(t)+\tilde{E}\left(e_{v1}^{\ k}\right)(t)+\tilde{E}\left(e_{v2}^{\ k}\right)(t)\right)+\frac{a^2}{p}\left(\left\|g_1^k\right\|_\Gamma^2+\left\|g_2^k\right\|_\Gamma^2\right)$$
$$\leq (k+kc)\left(E\left(e_{u1}^{\ k}\right)(t)+E\left(e_{u2}^{\ k}\right)(t)+\tilde{E}\left(e_{v1}^{\ k}\right)(t)+\tilde{E}\left(e_{v2}^{\ k}\right)(t)\right)+\frac{a^2}{p}\left(\left\|g_1^{k-1}\right\|_\Gamma^2+\left\|g_2^{k-1}\right\|_\Gamma^2\right). \quad (4.31)$$

We define the total energy at step k as a function of time t to be

$$\mathcal{E}^k(t)=E\left(e_{u1}^{\ k}\right)(t)+E\left(e_{u2}^{\ k}\right)(t)+\tilde{E}\left(e_{v1}^{\ k}\right)(t)+\tilde{E}\left(e_{v2}^{\ k}\right)(t),$$

and the total boundary error at step k to be

$$\mathcal{G}^k(t)=\frac{a^2}{p}\left(\left\|g_1^k(t)\right\|_\Gamma^2+\left\|g_2^k(t)\right\|_\Gamma^2\right),$$

in such a way that (4.31) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{E}^k(t) + \mathcal{G}^k(t) \le (k+kc)\mathcal{E}^k(t) + \mathcal{G}^{k-1}(t). \tag{4.32}$$

We sum both sides of (4.32) over $k = 0, \dots, K$, and integrate over $(0, t)$. We get the partial sums of a telescopic series and since $\mathcal{E}^k(0) = 0$, for all k, we obtain

$$\sum_{k=0}^{K} \mathcal{E}^k(t) + \int_0^t \mathcal{G}^K(s)\,\mathrm{d}s \le \int_0^t (k+kc) \sum_{k=0}^{K} \mathcal{E}^k(s)\,\mathrm{d}s + \int_0^t \mathcal{G}^0(s)\,\mathrm{d}s.$$

Applying the Gronwall lemma yields

$$\sum_{k=0}^{K} \mathcal{E}^k(t) \le e^{CT} \int_0^t \mathcal{G}^0(s)\,\mathrm{d}s, \quad \forall t \le T. \tag{4.33}$$

We deduce from (4.33) that the infinite series with general term $\mathcal{E}^k$ converges in $L^\infty(0,T)$. Therefore the general term tends to zero and $\left(u_i^k, v_i^k\right)$ converges to $(u,v)_{|\Omega_i}$ in $L^\infty(0,T;H^1(\Omega_i))\times L^\infty(0,T;L^2(\Omega_i))$.

2. The case $q > 0$.
By performing the same calculations as for the case $q = 0$, we obtain inequality (4.29).
Let $\mathcal{B}(u,v) = \left(\partial_t u - a\Delta_{\boldsymbol{y}} u + b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} u - kv\right)$. The identity

$$\begin{aligned}
-a\left(\frac{\partial e_{u_i}^{\ k}}{\partial n_i}, e_{u_i}^{\ k}\right)_\Gamma = \frac{a^2}{2p}\Bigg\{ &-\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i} + \frac{p}{2a} e_{u_i}^{\ k} + \frac{q}{2a}\mathcal{B}\left(e_{u_i}^{\ k}, e_{v_i}^{\ k}\right) + \frac{(-1)^i b_x}{2a} e_{u_i}^{\ k}\right\|_\Gamma^2 \\
&+\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i} - \frac{p}{2a} e_{u_i}^{\ k} - \frac{p}{2a}\mathcal{B}\left(e_{u_i}^{\ k}, e_{v_i}^{\ k}\right) + \frac{(-1)^i b_x}{2a} e_{u_i}^{\ k}\right\|_\Gamma^2\Bigg\} + \frac{(-1)^i b_x}{2}\left\|e_{u_i}^{\ k}\right\|^2 \\
&+ \frac{qa}{p}\left(\frac{\partial e_{u_i}^{\ k}}{\partial n_i}, \mathcal{B}\left(e_{u_i}^{\ k}, e_{v_i}^{\ k}\right)\right)_\Gamma + \frac{(-1)^i b_x}{2}\frac{q}{p}\left(\mathcal{B}\left(e_{u_i}^{\ k}, e_{v_i}^{\ k}\right), e_{u_i}^{\ k}\right)_\Gamma
\end{aligned} \tag{4.34}$$

holds. Integrating by parts over Γ yields

$$\left(\mathcal{B}\left(e_{u_i}^{\ k}, e_{v_i}^{\ k}\right), e_{u_i}^{\ k}\right)_\Gamma = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{u_i}^{\ k}\right\|_\Gamma^2 + a\left\|\nabla_{\boldsymbol{y}} e_{u_i}^{\ k}\right\|_\Gamma^2 - (k e_{v_i}^{\ k}, e_{u_i}^{\ k})_\Gamma,$$

since the term $\left(b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^{\ k}, e_{u_i}^{\ k}\right)_\Gamma$ vanishes. As we did before, we use the transmission condition

$$\begin{aligned}
\left\|\frac{\partial e_{u_i}^{\ k}}{\partial n_i} + \frac{p+(-1)^i b_x}{2a} e_{u_i}^{\ k} + \frac{q}{2a}\mathcal{B}\left(e_{u_i}^{\ k}, e_{u_i}^{\ k}\right)\right\|_\Gamma^2 &= \left\|\frac{\partial e_{u_j}^{\ k-1}}{\partial n_i} + \frac{p+(-1)^i b_x}{2a} e_{u_j}^{\ k-1}\right. \\
&\left. + \frac{q}{2a}\mathcal{B}\left(e_{u_j}^{\ k-1}, e_{v_j}^{\ k-1}\right)\right\|_\Gamma^2 = \left\|\frac{\partial e_{u_j}^{\ k-1}}{\partial n_j} + \frac{-p+(-1)^j b_x}{2a} e_{u_j}^{\ k-1} - \frac{q}{2a}\mathcal{B}\left(e_{u_j}^{\ k-1}, e_{v_j}^{\ k-1}\right)\right\|_\Gamma^2
\end{aligned}$$

in (4.34) and the notation

$$g_i^k = \frac{\partial e_{u_i}^k}{\partial n_i} + \frac{-p + (-1)^i b_x}{2a} e_{u_i}^k - \frac{q}{2a} \mathcal{B}\left(e_{u_i}^k, e_{v_i}^k\right).$$

We insert (4.34) in (4.29) and obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2\right) + a\left\|\nabla e_{u_i}^k\right\|^2 + k\left\|e_{v_i}^k\right\|^2 + kc\left\|e_{u_i}^k\right\|^2 + \frac{a^2}{2p}\left(\left\|g_i^k\right\|_\Gamma^2 - \left\|g_j^{k-1}\right\|_\Gamma^2\right)$$
$$+ \frac{qa}{p}\left(\frac{\partial e_{u_i}^k}{\partial n_i}, \mathcal{B}\left(e_{u_i}^k, e_{v_i}^k\right)\right)_\Gamma \leq \frac{(-1)^j b_x}{2}\frac{q}{p}\left(\frac{1}{2}\frac{\partial}{\partial t}\left\|e_{u_i}^k\right\|_\Gamma^2 + a\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|_\Gamma^2 - \left(k e_{v_i}^k, e_{u_i}^k\right)_\Gamma\right)$$
$$+ \frac{k+kc}{2}\left\|e_{u_i}^k\right\|^2 + \frac{k+kc}{2}\left\|e_{v_i}^k\right\|^2. \quad (4.35)$$

In order to cancel the term $\left(\frac{\partial e_{u_i}^k}{\partial n_i}, \mathcal{B}\left(e_{u_i}^k, e_{v_i}^k\right)\right)_\Gamma$ in the above inequality, we take the inner product of equation $\partial_t e_{u_i}^k + \operatorname{div}(-a\nabla e_{u_i}^k + b e_{u_i}^k) - k(e_{v_i}^k - c e_{u_i}^k) = 0$ with all the quantities appearing in $\mathcal{B}\left(e_{u_i}^k, e_{v_i}^k\right)$. After integrating by parts in Ω_i and applying the Cauchy-Schwarz inequality, the inner product with $\partial_t e_{u_i}^k$ yields

$$\left\|\partial_t e_{u_i}^k\right\|^2 + \frac{a}{2}\frac{\partial}{\partial t}\left\|\nabla e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, \partial_t e_{u_i}^k\right)_\Gamma + kc\frac{\partial}{\partial t}\left\|e_{u_i}^k\right\|^2 = -\left(b\cdot\nabla e_{u_i}^k, \partial_t e_{u_i}^k\right) + k\left(e_{v_i}^k, \partial_t e_{u_i}^k\right)$$
$$\leq \|b\|_\infty \left\|\nabla e_{u_i}^k\right\| \left\|\partial_t e_{u_i}^k\right\| + k\left\|e_{v_i}^k\right\| \left\|\partial_t e_{u_i}^k\right\| \leq 2\|b\|_\infty^2 \left\|\nabla e_{u_i}^k\right\|^2 + 2k^2\left\|e_{v_i}^k\right\|^2 + \frac{1}{4}\left\|\partial_t e_{u_i}^k\right\|^2. \quad (4.36)$$

We take now the inner product with $-a\Delta_{\boldsymbol{y}} e_{u_i}^k$, we integrate by parts in Ω_i and apply the Cauchy-Schwarz inequality in the right-hand side of the resulting equation. We obtain

$$a\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + a^2\left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, -a\Delta_{\boldsymbol{y}} e_{u_i}^k\right)_\Gamma$$
$$+ a\left(b\cdot\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k, \nabla_{\boldsymbol{y}} e_{u_i}^k\right) + akc\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 = ak\left(\nabla_{\boldsymbol{y}} e_{v_i}^k, \nabla_{\boldsymbol{y}} e_{u_i}^k\right) \leq ak\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\| \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|. \quad (4.37)$$

To estimate $\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|$ in the right-hand side of (4.37), we apply the tangential gradient operator to the equation $\partial_t e_{v_i}^k + k(e_{v_i}^k - c e_{u_i}^k) = 0$, we take the inner product of the resulting equation with $\nabla_{\boldsymbol{y}} e_{v_i}^k$ and apply the Cauchy-Schwarz inequality in the right-hand side. We obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2 + k\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2 \leq kc\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\| \left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|. \quad (4.38)$$

We add now (4.37) and (4.38) and apply again the Cauchy-Schwarz inequality. We get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \frac{1}{2}\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2\right) + a^2\left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, -a\Delta_{\boldsymbol{y}} e_{u_i}^k\right)_\Gamma + akc\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + k\left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2$$
$$\leq \frac{ak+kc}{2}\left(\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2\right) + \frac{a^2}{2}\left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \frac{\|b\|_\infty^2}{2}\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2. \quad (4.39)$$

We finally take the inner product of equation $\partial_t e_{u_i}^k + \operatorname{div}(-a\nabla e_{u_i}^k + b e_{u_i}^k) - k(e_{v_i}^k - c e_{u_i}^k) = 0$ with $b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k$ and with $e_{v_i}^k$. By arguing as before, we obtain respectively

$$\left(\partial_t e_{u_i}^k, b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right) + a\left(\nabla e_{u_i}^k, \nabla\left(b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right)\right) - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, b_{\boldsymbol{y}} \nabla_{\boldsymbol{y}} e_{u_i}^k\right)_\Gamma + \left(b \cdot \nabla e_{u_i}^k, b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right) + kc\left(e_{u_i}^k, b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right) - k\left(e_{v_i}^k, b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right) = 0 \quad (4.40)$$

and

$$\left(\partial_t e_{u_i}^k, e_{v_i}^k\right) + a\left(\nabla e_{u_i}^k, \nabla e_{v_i}^k\right) - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{v_i}^k\right)_\Gamma + \left(b \cdot \nabla e_{u_i}^k, e_{v_i}^k\right) - k\left\|e_{v_i}^k\right\|^2 + kc\left(e_{u_i}^k, e_{v_i}^k\right) = 0. \quad (4.41)$$

Integrating by parts equation (4.40) with respect to $\boldsymbol{y}$ in Ω_i shows that the terms

$$\left(\nabla e_{u_i}^k, \nabla\left(b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right)\right) \text{ and } kc\left(e_{u_i}^k, b_{\boldsymbol{y}} \cdot \nabla_{\boldsymbol{y}} e_{u_i}^k\right)$$

vanish. We obtain then

$$-a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, b_{\boldsymbol{y}} \nabla_{\boldsymbol{y}} e_{u_i}^k\right)_\Gamma \le \frac{1}{8}\left\|\partial_t e_{u_i}^k\right\|^2 + 2\max\left(2\|b\|_\infty^2, \frac{k}{2}\|b\|_\infty^2\right)\left(\left\|\nabla e_{u_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|\right). \quad (4.42)$$

In order to estimate $\left\|\nabla e_{v_i}^k\right\|$ in (4.41), we apply the gradient operator to the equation $\partial_t e_{v_i}^k + k\left(e_{v_i}^k - c e_{u_i}^k\right) = 0$, we take the inner product of the resulting equation with $\nabla e_{v_i}^k$, add equation (4.41), and apply the Cauchy-Schwarz inequality. We obtain

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\nabla e_{v_i}^k\right\|^2 + k\left\|\nabla e_{v_i}^k\right\|^2 - a\left(\frac{\partial e_{u_i}^k}{\partial n_i}, e_{v_i}^k\right)_\Gamma \le \frac{1}{8}\left\|\partial_t e_{u_i}^k\right\|^2 + 5\max\left(2, \frac{a}{2}, \frac{\|b\|_\infty}{2}, k, \frac{kc}{2}\right)\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2 + \left\|\nabla e_{u_i}^k\right\|^2 + \left\|\nabla e_{v_i}^k\right\|^2\right). \quad (4.43)$$

We add (4.36), (4.39), (4.42) and (4.43) and multiply the resulting equation by $\frac{q}{p}$. We obtain

$$C_l \frac{\mathrm{d}}{\mathrm{d}t}\left(\left\|e_{u_i}^k\right\|^2 + \left\|\nabla e_{u_i}^k\right\|^2 + \left\|\nabla e_{v_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2\right) - \frac{aq}{p}\left(\frac{\partial e_{u_i}^k}{\partial n_i}, \mathcal{B}(e_{u_i}^k, e_{v_i}^k)\right)_\Gamma + C_l\left(\left\|\partial_t e_{u_i}^k\right\|^2 + \left\|\nabla e_{v_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2 + \left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2\right) \le C_r\left(\left\|e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2 + \left\|\nabla e_{u_i}^k\right\|^2 + \left\|\nabla e_{v_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \left\|\nabla_{\boldsymbol{y}} e_{v_i}^k\right\|^2\right), \quad (4.44)$$

where C_l and C_r are positive constants depending on $a,\ p,\ q,\ b,\ k$ and c.
At this point of the proof, we introduce the energies

$$\tilde{E}(w)(t) = \|w(t)\|^2 + \|\nabla w(t)\|^2 + \|\nabla_{\boldsymbol{y}} w(t)\|^2 + \int_0^t \|w(s)\|^2 + \|\nabla w(s)\|^2 + \|\nabla_{\boldsymbol{y}} w(s)\|^2 \,\mathrm{d}s,$$

and

$$E(w)(t) = \tilde{E}(w)(t) + \int_0^t \|\partial_t w(s)\|^2 + \|\nabla\nabla_{\boldsymbol{y}} w(s)\|^2 \,\mathrm{d}s.$$

We add (4.35) with (4.44) and sum the resulting equation for $i = 1,\ 2$. We obtain

$$\begin{aligned}
&\tilde{C}_l\left(\frac{\mathrm{d}}{\mathrm{d}\,t}\left(E(e_{u1}^k)(t) + E(e_{u2}^k)(t) + \tilde{E}(e_{v1}^k)(t) + \tilde{E}(e_{v2}^k)(t)\right) + \left\|g_1^k\right\|_\Gamma^2 + \left\|g_2^k\right\|_\Gamma^2\right) \\
&\qquad \le \tilde{C}_r\left(\left(E(e_{u1}^k)(t) + E(e_{u2}^k)(t) + \tilde{E}(e_{v1}^k)(t) + \tilde{E}(e_{v2}^k)(t)\right) + \left\|g_1^{k-1}\right\|_\Gamma^2 + \left\|g_2^{k-1}\right\|_\Gamma^2\right. \\
&\qquad \left. + \left(\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{u1}^k\right\|_\Gamma^2 + \left\|\nabla_{\boldsymbol{y}} e_{u1}^k\right\|_\Gamma^2 - (e_{v1}^k, e_{u1}^k)_\Gamma\right) + \left(\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{u2}^k\right\|_\Gamma^2 + \left\|\nabla_{\boldsymbol{y}} e_{u2}^k\right\|_\Gamma^2 - (e_{v2}^k, e_{u2}^k)_\Gamma\right)\right), \quad (4.45)
\end{aligned}$$

where $\tilde{C}_l$ and $\tilde{C}_r$ are positive constants as C_l and C_r. We integrate (4.45) over $(0, t)$, for $t \le T$. We have to estimate

$$\begin{aligned}
&\int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}\,t}\left\|e_{u_i}^k(s)\right\|_\Gamma^2 + \left\|\nabla_{\boldsymbol{y}} e_{u_i}^k(s)\right\|_\Gamma^2 - (e_{v_i}^k(s), e_{u_i}^k(s))_\Gamma\right) \mathrm{d}s \\
&\qquad\qquad = \left\|e_{u_i}^k(t)\right\|_\Gamma^2 + \int_0^t \left(\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k(s)\right\|_\Gamma^2 - (e_{v_i}^k(s), e_{u_i}^k(s))_\Gamma\right) \mathrm{d}s,
\end{aligned}$$

using $\left\|e_{u_i}^k(0)\right\|_\Gamma = 0$, for $i = 1,\ 2$. To do so, we use trace inequalities and the Cauchy-Schwarz inequality in order to obtain

$$\begin{aligned}
\left\|e_{u_i}^k\right\|_\Gamma^2 &\le C\left\|e_{u_i}^k\right\|_{H^1(\Omega_i)}^2 = C\left(\left\|e_{u_i}^k\right\|^2 + \left\|\nabla e_{u_i}^k\right\|^2\right), \\
\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|_\Gamma &\le 2\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\| \left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\| \le \alpha\left\|\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2 + \frac{\left\|\nabla\nabla_{\boldsymbol{y}} e_{u_i}^k\right\|^2}{\alpha}, \\
(e_{v_i}^k, e_{u_i}^k)_\Gamma &\le \left\|e_{v_i}^k\right\|_\Gamma \left\|e_{u_i}^k\right\|_\Gamma \le C^2\left(\left\|e_{u_i}^k\right\|^2 + \left\|\nabla e_{u_i}^k\right\|^2 + \left\|e_{v_i}^k\right\|^2 + \left\|\nabla e_{v_i}^k\right\|^2\right),
\end{aligned}$$

where C is a constant depending on the domain Ω_i and α a positive constant chosen such that $\tilde{C}_l - \frac{\tilde{C}_r}{\alpha} > 0$. The second inequality is a consequence of the following equality, which we apply for each component of $\nabla_{\boldsymbol{y}} e_{u_i}^k$, available for $u \in H^1(\Omega_i)$,

$$2\,(u, \mathrm{div}\ u) = \int_{\Omega_i} \mathrm{div}\ u^2 = (-1)^j \|u\|_\Gamma^2,$$

and of the Cauchy-Schwarz inequality. We have thus that (4.45) holds with new constants $\tilde{\tilde{C}}_l$ and $\tilde{\tilde{C}}_r$ and without the last two terms in the right-hand side. We can now argue as in the case $q = 0$ to conclude the result of the theorem. ⊠

4.2.4.2 Convergence of the Overlapping Algorithm with Dirichlet Transmission Conditions

In this paragraph we prove that, under the assumption $b_x \neq 0$, the classical overlapping Schwarz waveform relaxation approximation with Dirichlet transmission conditions converges, by showing that the Fourier transforms of the errors $e_{u_i}^k$ and $e_{v_i}^k$ converge to 0.

Theorem 4.7

Let $L > 0$ and suppose that the advection speed b is such that $b_x \neq 0$. Then the sequence $\left(\left(u_1^k, v_1^k\right), \left(u_2^k, v_2^k\right)\right)$ defined by algorithm 4.1, with the transmission operators defined by (4.20) converges to $((u,v)_{|\Omega_1}, (u,v)_{|\Omega_2})$, in $\prod_{i=1}^2 (L^2(0,T\,;L^2(\Omega_i)) \times L^2(0,T\,;L^2(\Omega_i)))$ for $k \to \infty$.

Proof 4.6

As we show in Section 4.2.1, the Fourier transforms with respect to the variables t and $\boldsymbol{y}$ of the errors $e_{u_i}^k$ satisfy (4.5),

$$\begin{aligned} \widehat{e_{u1}^k}(x,\boldsymbol{\xi},\tau) &= \alpha_1^k(\boldsymbol{\xi},\tau)e^{\lambda^+(x-L)}, \quad (x,\boldsymbol{\xi},\tau) \in]-\infty, L[\times \mathbb{R}^{d-1} \times \mathbb{R}, \\ \widehat{e_{u2}^k}(x,\boldsymbol{\xi},\tau) &= \alpha_2^k(\boldsymbol{\xi},\tau)e^{\lambda^- x}, \qquad (x,\boldsymbol{\xi},\tau) \in]0, +\infty[\times \mathbb{R}^{d-1} \times \mathbb{R}, \end{aligned} \tag{4.5}$$

with

$$\alpha_i^k = e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2},$$

for $i = 1,\ 2$ and for $k \geq 2$. We have thus

$$\alpha_i^{2k} = \rho^{k-1}\alpha_i^2,\ k \geq 1, \quad \text{and} \quad \alpha_i^{2k+1} = \rho^k \alpha_i^1,\ k \geq 0,$$

where $\rho = \rho(\boldsymbol{\xi},\tau) := e^{(\lambda^- - \lambda^+)L}$. We have $|\rho| = e^{\Re\mathfrak{e}((\lambda^- - \lambda^+)L}$ and

$$\lambda^- - \lambda^+ = -\frac{\sqrt{b_x^2 + 4a\hat{z}}}{a},$$

where $\hat{z}$ is defined by (4.4). It is easy to see that $\Re\mathfrak{e}\left(\sqrt{b_x^2 + 4a\hat{z}}\right) \geq b_x^2$, which allows to conclude that $|\rho(\boldsymbol{\xi},\tau)| < 1$, $\forall(\boldsymbol{\xi},\tau)$. Hence, $\alpha_i^k(\boldsymbol{\xi},\tau) \to 0$, $\forall(\boldsymbol{\xi},\tau)$ and

$$\begin{aligned} |\alpha_1^k(\boldsymbol{\xi},\tau)e^{\lambda^+(x-L)}| &\leq |e^{\lambda^+(x-L)}|,\ \forall(x,\boldsymbol{\xi},\tau) \in]-\infty, L[\times \mathbb{R}^{d-1} \times \mathbb{R}, \\ |\alpha_2^k(\boldsymbol{\xi},\tau)e^{\lambda^- x}| &\leq |e^{\lambda^- x}|,\ \forall(x,\boldsymbol{\xi},\tau) \in]0, +\infty[\times \mathbb{R}^{d-1} \times \mathbb{R}. \end{aligned}$$

By applying the Lebesgue theorem, we conclude that $e_{u_i}^k \to 0$ in $L^2(0,T;L^2(\Omega_i))$. Since $\widehat{e_{v_i}^k} = \frac{1}{(k+i\tau)}\widehat{e_{u_i}^k}$ and $\left|\frac{1}{k+i\tau}\right| \leq \frac{1}{k}$, we also obtain that $e_{v_i}^k \to 0$ in $L^2(0,T;L^2(\Omega_i))$. ⊠

4.2.4.3 Convergence of the Overlapping Algorithm with Robin and Ventcel Transmission Conditions

We use here the same tools as for the case of the classical overlapping Schwarz waveform relaxation algorithm to prove the following theorem:

Theorem 4.8

Let $L > 0$ and suppose that the advection speed b is such that $b_x \neq 0$. Let $p > 0$ and $q = 0$ or $p > 0,\ q > 0$. Then the sequence $\left(\left(u_1^k, v_1^k\right), \left(u_2^k, v_2^k\right)\right)$ defined by algorithm 4.1, with the transmission operators defined by (4.9) or by (4.10) converges to $((u,v)_{|\Omega_1}, (u,v)_{|\Omega_2})$, in $\prod_{i=1}^{2}(L^2(0,T\,;L^2(\Omega_i)) \times L^2(0,T\,;L^2(\Omega_i)))$ for $k \to \infty$.

Proof 4.7

In the case of Robin and Ventcel transmission conditions, we obtain

$$\alpha_i^k = \left(\frac{p + q\hat{z} - \sqrt{b_x^2 + 4a\hat{z}}}{p + q\hat{z} + \sqrt{b_x^2 + 4a\hat{z}}}\right)^2 e^{(\lambda^- - \lambda^+)L}\alpha_i^{k-2},$$

for $i = 1,\ 2$ and for $k \geq 2$, where $\hat{z}$ is defined by (4.4).
We have now to prove that

$$\left|\frac{s(\hat{z}) - \sqrt{b_x^2 + 4a\hat{z}}}{s(\hat{z}) + \sqrt{b_x^2 + 4a\hat{z}}}\right|^2 |e^{(\lambda^- - \lambda^+)L}| < 1,$$

for all (ξ, τ).
We obtained in the proof of Theorem 4.7 that $|e^{(\lambda^- - \lambda^+)L}| < 1,\ \forall\ (\xi, \tau)$.

If $p > 0$ and $q = 0$, we have that

$$\left|\frac{s(\hat{z}) - \sqrt{b_x^2 + 4a\hat{z}}}{s(\hat{z}) + \sqrt{b_x^2 + 4a\hat{z}}}\right| = \frac{(p - X)^2 + Y^2}{(p + X)^2 + Y^2},$$

where $X = \Re\mathrm{e}\left(\sqrt{b_x^2 + 4a\hat{z}}\right)$ and $Y = \Im\mathrm{m}\left(\sqrt{b_x^2 + 4a\hat{z}}\right)$. Since $X > 0$, we obtain that

$$\left|\frac{s(\hat{z}) - \sqrt{b_x^2 + 4a\hat{z}}}{s(\hat{z}) + \sqrt{b_x^2 + 4a\hat{z}}}\right| < 1.$$

If $p > 0$ and $q > 0$, we have that

$$\begin{aligned}\left|\frac{s(\hat{z}) - \sqrt{b_x^2 + 4a\hat{z}}}{s(\hat{z}) + \sqrt{b_x^2 + 4a\hat{z}}}\right| &= \frac{(p + q\Re\mathrm{e}(\hat{z}) - X)^2 + (q\Im\mathrm{m}(\hat{z}) - Y)^2}{(p + q\Re\mathrm{e}(\hat{z}) + X)^2 + (q\Im\mathrm{m}(\hat{z}) + Y)^2} \\ &= \frac{(p + q\Re\mathrm{e}(\hat{z}))^2 + X^2 - 2X(p + q\Re\mathrm{e}(\hat{z})) + (q\Im\mathrm{m}(\hat{z}))^2 + Y^2 - 2Y(q\Im\mathrm{m}(\hat{z}))}{(p + q\Re\mathrm{e}(\hat{z}))^2 + X^2 + 2X(p + q\Re\mathrm{e}(\hat{z})) + (q\Im\mathrm{m}(\hat{z}))^2 + Y^2 + 2Y(q\Im\mathrm{m}(\hat{z}))}.\end{aligned}$$

We have, on the one hand, $X > 0$ and $\Re\mathrm{e}(\hat{z}) = a\boldsymbol{\xi}\cdot\boldsymbol{\xi} + \frac{kc\phi^2\tau^2}{k^2+\phi^2\tau^2} > 0$. On the other hand, we have $2XY = \Im\mathrm{m}({b_x}^2 + 4a\hat{z})$ and thus $Y = \frac{2a\Im\mathrm{m}(\hat{z})}{X}$. We conclude that $X(p + q\Re\mathrm{e}(\hat{z})) > 0$ and that $Y(q\Im\mathrm{m}(\hat{z})) > 0$ and once again we have

$$\left|\frac{s(\hat{z}) - \sqrt{b_x^2 + 4a\hat{z}}}{s(\hat{z}) + \sqrt{b_x^2 + 4a\hat{z}}}\right| < 1.$$

We conclude as in the proof of theorem (4.7). ⊠

4.3 Optimisation of the Transmission Conditions

In Section 4.2.1 we have introduced Robin and Ventcel transmission conditions for the Schwarz waveform relaxation algorithm 4.1 by approximating the Fourier symbol of the transparent boundary condition over Γ_i either by a zeroth order polynomial or by a first order polynomial in the Fourier space. The aim of this section is to establish, under some assumptions, the best polynomial that approaches this Fourier symbol. We proceed by giving a formula of the convergence rate depending on the parameters p and (p, q) for Robin and Ventcel conditions and to optimise it either numerically or analytically.

4.3.1 Numerical Optimisation

From a numerical point of view, we will consider bounded domains $\Omega_i = [x_{mi}, x_{Mi}] \times \Omega_{i\boldsymbol{y}}$, where $\Omega_{i\boldsymbol{y}}$ is a bounded interval if $d = 2$ and a bounded rectangle if $d = 3$. Hence we can consider that the frequencies τ and $\boldsymbol{\xi}$ are bounded: we have $|\tau| \in [\tau_m, \tau_M]$ and $|\boldsymbol{\xi}_i| \in [\boldsymbol{\xi}_{i_m}, \boldsymbol{\xi}_{i_M}]$, $1 \leq i \leq d-1$, where τ_m, τ_M, $\boldsymbol{\xi}_{i_m}$ and $\boldsymbol{\xi}_{i_M}$ are numerical frequencies of a discrete function on a given mesh. They can be taken as

$$\tau_m = \frac{\pi}{2T},\ \tau_M = \frac{\pi}{\Delta t},\ \boldsymbol{\xi}_{i_m} = \frac{\pi}{L_{y_i}},\ \boldsymbol{\xi}_{i_M} = \frac{\pi}{\Delta y_i},$$

where $[0, T]$ is the time interval, L_{y_i} the lengths of the space intervals and Δ_t and Δy_i the time and space steps.
We define $K = \{(\boldsymbol{\xi}, \tau)$ such that $|\tau| \in [\tau_m, \tau_M],\ |\boldsymbol{\xi}_i| \in [\boldsymbol{\xi}_{i_m}, \boldsymbol{\xi}_{i_M}]\}$.
Now, optimising the convergence factor can be interpreted as solving the following best approximation problem: we search a polynomial $s^* = p^* + q^* z$ in the space $\mathcal{P}$ of polynomials of degree less than or equal to 1 with complex coefficients, such that

$$\sup_{(\boldsymbol{\xi},\tau)\in K} \rho(\boldsymbol{\xi}, \tau, p^*, q^*) = \inf_{s=p+qz\in\mathcal{P}} \sup_{(\boldsymbol{\xi},\tau)\in K} |\rho(\boldsymbol{\xi}, \tau, p, q)|. \tag{4.46}$$

This optimisation problem can be solved numerically, which is done in section 4.4.2.
In the next paragraph, we solve this problem analytically in the particular case where the space dimension is 1, for Robin transmission conditions without overlap.

4.3.2 Analytical Solution of the Best Approximation Problem for Robin Transmission Conditions in 1D

In the case of non overlapping subdomains Ω_i, i. e. $L = 0$, in dimension 1 for Robin transmission conditions the convergence factor of algorithm 4.1 writes

$$\rho(\tau, p) = \left(\frac{p - \sqrt{b_x^2 + 4a\hat{z}}}{p + \sqrt{b_x^2 + 4a\hat{z}}} \right), \tag{4.47}$$

with $\hat{z} = \hat{z}(\tau) := \phi i\tau - \frac{k^2 c}{\phi i\tau + k} + kc$ (cf. equation (4.4)). We are now interested in solving the following best approximation problem: find $p^* \in \mathbb{C}$ such that

$$\sup_{\tau \in K} |\rho(\tau, p^*)| = \inf_{p \in \mathbb{C}} \sup_{\tau \in K} |\rho(\tau, p)|, \tag{4.48}$$

where $K = \{\tau,\ |\tau| \in [\tau_m, \tau_M]\}$.

We first show the existence of a solution to problem (4.48). To do so, we use the theory developed in [6], where the authors consider the best approximation problem that we describe below.

Let $n \in \mathbb{Z},\ n \geq 0$ and K be a compact set in $\mathbb{C}$ containing at least $n+2$ points. Let $f : K \longrightarrow \mathbb{C}$ be a continuous function such that $\Re\mathrm{e}(f(z)) > 0,\ \forall\, z \in K$. Denote by $\mathcal{P}_n$ the space of polynomials with complex coefficients of degree less than or equal to n and put

$$\delta_n = \inf_{s \in \mathcal{P}_n} \sup_{z \in K} \left| \frac{s(z) - f(z)}{s(z) + f(z)} \right|.$$

Consider the problem:

$$\text{Find } {s_n}^* \in \mathcal{P}_n \text{ such that } \sup_{z \in K} \left| \frac{{s_n}^*(z) - f(z)}{{s_n}^*(z) + f(z)} \right| = \delta_n. \tag{4.49}$$

Consider also the function $h : s \in \mathcal{P}_n \longmapsto h(s) = \sup_{z \in K} \left| \frac{s(z) - f(z)}{s(z) + f(z)} \right| \in \mathbb{R}$.
We recall the following result which is proved in [6]:

Theorem 4.9

Let $n \geq 0$. Then we have $\delta_n < 1$ and there exists a unique solution to problem (4.49). Furthermore, the following properties hold:

1. If ${s_n}^*$ is a solution of (4.49), then there exist at least $n + 2$ points $z_1, \ldots, z_{n+2} \in K$ such that
$$\left| \frac{{s_n}^*(z_i) - f(z_i)}{{s_n}^*(z_i) + f(z_i)} \right| = \sup_{z \in K} \left| \frac{{s_n}^*(z) - f(z)}{{s_n}^*(z) + f(z)} \right|,\ i = 1, \ldots, n + 2.$$
2. Let $s^* \in \mathcal{P}_n$ be a strict local minimum for h. Then s^* is the global minimum of h on $\mathcal{P}_n$.

Denote by $\mathcal{P}_n[\mathbb{R}]$ the space of polynomials with real coefficients of degree less than or equal to n. In [6] further results are proved in the symmetric case described in the next theorem:

Theorem 4.10

Let K be a compact set of $\mathbb{C}$, symmetric with respect to the real axis, containing at least $n+2$ points. Suppose that $f : K \longrightarrow \mathbb{C}$ is continuous, such that $\Re\mathrm{e}(f(z)) > 0,\ \forall\ z \in K$ and satisfying $f(\bar{z}) = \overline{f(z)},\ \forall\ z \in K$. Then the polynomial $s_n{}^*$ of best approximation of f in K has real coefficients.

Furthermore, suppose that K_1 is a compact set of $\{z \in \mathbb{C},\ \Im\mathrm{m}(z) \geq 0\}$ and let $\overline{K_1} := \{\bar{z},\ z \in K_1\}$. Then any strict local minimum $s^*_{n,\mathbb{R}}$ of

$$\left\| \frac{s(z) - f(z)}{s(z) + f(z)} \right\|_{L^\infty(K_1)}$$

in $\mathcal{P}_n[\mathbb{R}]$ is the global minimum of the complex best approximation problem in K.

We consider thus the problem (4.48). The set $K = \{\tau,\ |\tau| \in [\tau_m, \tau_M]\}$ is a compact subset of $\mathbb{C}$, symmetric with respect to the real axis, and the function $\tau \in K \longmapsto f(\tau) := \sqrt{b_x^2 + 4a\hat{z}(\tau)}$ has a strictly positive real part and satisfies $f(\bar{z}) = \overline{f(z)}$. We can thus apply Theorems 4.9 and 4.10 to obtain the following result:

Theorem 4.11

The best approximation problem (4.48) has a unique solution $p^* > 0$ and we have

$$\delta^* = \sup_{\tau \in K} |\rho(\tau, p^*)| = \inf_{p \in \mathbb{C}} \sup_{\tau \in K} |\rho(\tau, p)| < 1.$$

Furthermore the following properties hold:

1. There exist at least two points τ_1 and τ_2 such that

$$\left| \frac{p^* - \sqrt{b_x^2 + 4a\hat{z}(\tau_i)}}{p^* + \sqrt{b_x^2 + 4a\hat{z}(\tau_i)}} \right| = \sup_{\tau \in K} \left| \frac{p^* - \sqrt{b_x^2 + 4a\hat{z}(\tau)}}{p^* + \sqrt{b_x^2 + 4a\hat{z}(\tau)}} \right|,$$

   for $i = 1,\ 2$.

2. $\sup\limits_{\tau \in K} |\rho(\tau, p^*)| = \inf\limits_{p>0} \sup\limits_{\tau \in [\tau_m, \tau_M]} |\rho(\tau, p)|$.

We will give an explicit formula of the solution p^* of the best approximation problem (4.48) for sufficiently large τ_M, by studying the equioscillation properties of the problem. We begin by proving the following result:

Proposition 4.1

Suppose $\tau_M \gg 1$. Then the solution p^* of problem (4.48) is such that

$$|\rho(\tau_m, p^*)| = |\rho(\tau_M, p^*)|,$$

and there exist positive constants C_p and C_δ such that

$$p^* \sim C_p \tau_M^{\frac{1}{4}}, \quad \delta^* = 1 - C_\delta \tau_M^{-\frac{1}{4}}.$$

Proof 4.8

We introduce the set $K^+ = [\tau_m, \tau_M]$ and the function $\tau \in K^+ \longmapsto \sigma(\tau) := b_x^2 + 4a\hat{z}(\tau)$, such that $f(\tau) = \sqrt{\sigma(\tau)}$.

1. We begin by giving some geometric considerations about the curve $\tau \in K^+ \longmapsto f(\tau)$.
We denote $x := x(\tau) = \Re\mathrm{e}(f(\tau))$ and $y := y(\tau) = \Im\mathrm{m}(f(\tau))$. We have

$$\begin{cases} x^2 - y^2 = b_x^2 + 4a\frac{kc\tau^2\phi^2}{k^2+\tau^2\phi^2}, \\ 2xy = 4a\tau\phi\left(1 + \frac{k^2 c}{k^2+\tau^2\phi^2}\right), \end{cases}$$

and thus $x^2 \geq y^2$. Since we have $x > 0$, we conclude that $x \geq |y|$. Furthermore, since $\tau \in [\tau_m, \tau_M]$, we also have $y > 0$ and thus the curve $f(\tau)$ lies in the first quadrant of the complex plane below the line $y = x$. Furthermore, if $\tau \to \infty$, we have $\sqrt{b_x^2 + 4a\hat{z}(\tau)} \sim \tilde{f}(\tau) := \sqrt{b_x^2 + 4a(\phi\tau i + kc)}$ which satisfies $\Re\mathrm{e}(\tilde{f}) - \Im\mathrm{m}(\tilde{f}) = b_x^2 + 4akc$. Hence, for $\tau \to \infty$, $f(\tau)$ behaves like an hyperbola which lies below the line $y = x$.

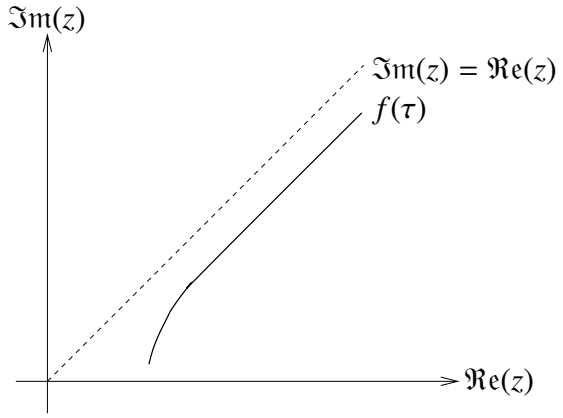


Figure 4.1: Asymptotical behaviour of $f(\tau)$.

2. We prove that $p^* \sim C_p \tau_M^{\frac{1}{4}}$, $\delta^* = 1 - C_\delta \tau_M^{-\frac{1}{4}}$.
For $p > 0$ and $\delta > 0$, we consider the set

$$C(\delta, p) = \left\{ z \in \mathbb{C}, \ \left| \frac{z - p}{z + p} \right| = \delta \right\}.$$

We have that $C(\delta, p)$ is a circle centred at $\frac{1+\delta^2}{1-\delta^2}p$ and of radius $\frac{2\delta}{1-\delta^2}p$ (cf. [6]). In order to solve the best approximation problem we must find the smallest circle $C(\delta, p)$ containing $f(K)$.

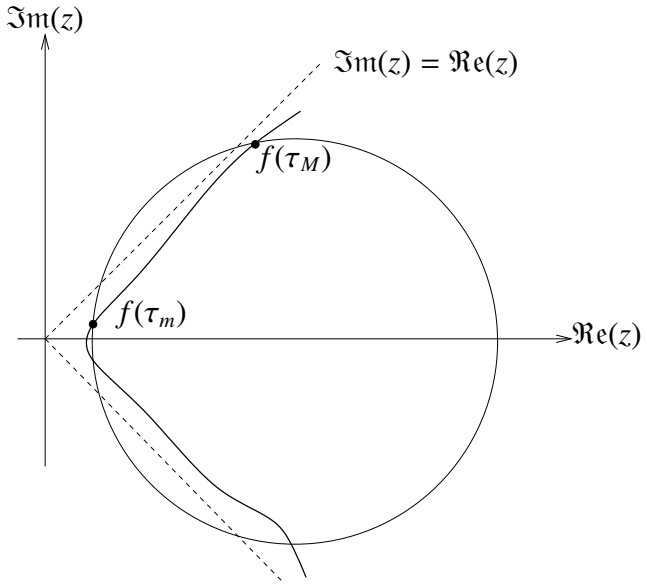


Figure 4.2: The curve $\tau \longmapsto f(\tau)$ and the optimal circle $C(\delta^*, p^*)$

Consider the circle C_0 centred in a point C_0 of the real axis and crossing the real axis in 0 and going through $f(\tau_M)$. We denote by r_0 the radius of C_0. It is easy to show that $C_0 = \{z \in \mathbb{C},\ |z - C_0| = C_0\}$, and that C_0 and r_0 satisfy $C_0 = |z_M - C_0|$, $r_0 = C_0 = \frac{|z_M|^2}{2x_M}$, where we denote $z_M = f(\tau_M)$ and $x_M = \Re\mathfrak{e}(f(\tau_M))$.
As before, we obtain for $f(\tau_M)$ as $\tau_M \to \infty$

$$\begin{cases} x_M^2 - y_M^2 \sim b^2 + 4akc, \\ 2x_M y_M \sim 4a\phi\tau_M. \end{cases} \tag{4.50}$$

By calculating $x_M^2 + y_M^2 = \sqrt{(x_M^2 + y_M^2)^2} = \sqrt{(x_M^2 - y_M^2)^2 + 4x_M y_M}$, and by combining with (4.50), we obtain

$$\begin{cases} x_M \sim \sqrt{2a\tau_M\phi}\left(1 + \frac{b^2+4akc}{8a\tau_M\phi}\right), \\ y_M \sim \sqrt{2a\tau_M\phi}\left(1 - \frac{b^2+4akc}{8a\tau_M\phi}\right). \end{cases} \tag{4.51}$$

Hence, for $\tau_M \to \infty$, z_M approaches the line $x = y$ in the first quadrant of the complex plane and the curve $f(K^+)$ lies inside the circle C_0.

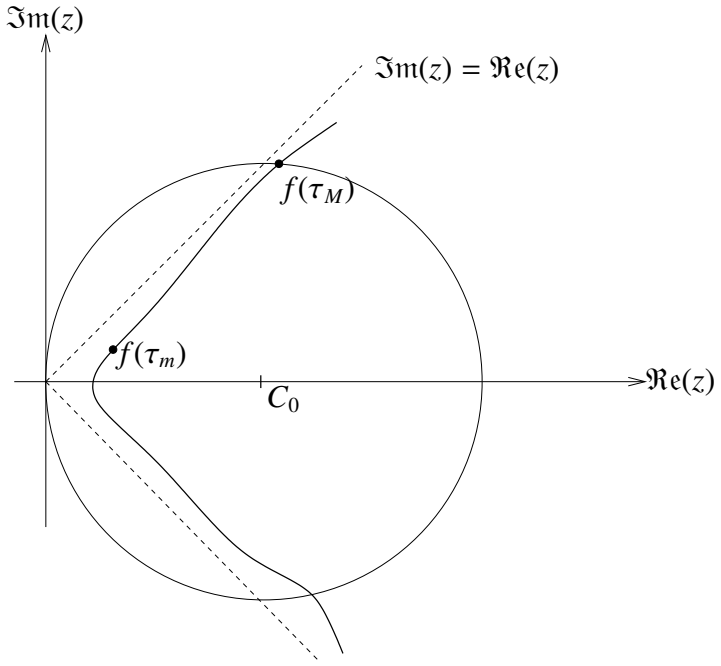


Figure 4.3: The circle C_0

We now construct a new circle C_1 still passing by $z_M = f(\tau_M)$, obtained from the circle C_0 by moving its left extreme point (which is 0 in the case of C_0) to the right on the real axis until C_1 intersects the curve $f(K^+)$. We denote by h the extreme left point on the real axis of C_1. It is easy to show that $C_1 = C(\delta_1, p_1)$, where

$$p_1 = \sqrt{\frac{h|z_M|^2 - x_M h^2}{x_M - h}}, \quad \delta_1 = \left|\frac{h - p_1}{h + p_1}\right|.$$

We have that $h < x_m,$ if $\tau_M > \tau_m$, and

$$\begin{cases} p_1 \sim \sqrt{\frac{h|z_M|^2}{x_M}} \sim \sqrt{2hx_M} \sim \left(2h\sqrt{2a\tau_M\phi}\right)^{\frac{1}{2}}, \\ \delta_1 = \left(1 - \frac{h}{p_1}\right)\frac{1}{1+\frac{h}{p_1}} \sim 1 - \frac{2h}{p_1} \sim 1 - \sqrt{\frac{2h}{x_M}} \sim 1 - \sqrt{\frac{2h}{\sqrt{2a\tau_M\phi}}}, \end{cases}$$

for $\tau_M \to \infty$. Hence, $p_1 \sim \tilde{p}_1 \tau_M^{\frac{1}{4}}$ and $\delta_1 \sim 1 - \tilde{\delta}_1 \tau_M^{-\frac{1}{4}}$, as $\tau_M \to \infty$, for some positive constants $\tilde{p}_1$ and $\tilde{\delta}_1$.
The curve $f(K^+)$ lies in the interior of the circle C_1. We have thus $\delta_1 \geq \delta^*$. The optimal circle $C(\delta^*, p^*)$ intersects the real axis on the right at the point $\frac{1+\delta^*}{1-\delta^*}p^* \geq x_M$ and on the left at the point $\frac{1-\delta^*}{1+\delta^*}p^* \leq x_m$. On the one hand, since the function $\delta \longrightarrow \frac{1-\delta}{1+\delta}$ is decreasing and since $\delta_1 < 1$, we obtain that

$$p^* \geq x_M \frac{1-\delta^*}{1+\delta^*} \geq x_M \frac{1-\delta_1}{1+\delta_1} \geq x_M \frac{1-\delta_1}{2} \gtrsim \frac{1}{2}\sqrt{\frac{2h}{x_M}} x_M = \sqrt{\frac{hx_M}{2}}.$$

On the other hand, since the function $\delta \longrightarrow \frac{1+\delta}{1-\delta}$ is increasing, we obtain that

$$p^* \leq x_m \frac{1+\delta^*}{1-\delta^*} \leq x_m \frac{1+\delta_1}{1-\delta_1} \leq x_m \frac{2}{1-\delta_1} \lesssim 2\sqrt{\frac{x_M}{2h}} x_m.$$

We conclude that

$$\begin{cases} p_1^* \tau_M^{\frac{1}{4}} \lesssim p^* \lesssim p_2^* \tau_M^{\frac{1}{4}}, \\ 1 - \rho_1^* \tau_M^{-\frac{1}{4}} \lesssim \rho^* \lesssim 1 - \rho_2^* \tau_M^{-\frac{1}{4}}, \end{cases}$$

as $\tau_M \to \infty$, where $p_1^*,\ p_2^*,\ \rho_1^*$ and ρ_2^* are positive constants.

3. We search now the extreme points of the function $\tau \longrightarrow |\rho(\tau, p)|$, for a fixed p.
In order to calculate the extreme points of the function $\tau \longrightarrow |\rho(\tau, p)|$, we calculate the zeros of the derivative of the function $\tau \longrightarrow R(\tau, p) := |\rho(\tau, p)|^2$.
We define the function $g_1(\tau^2) := \frac{k^2}{\tau^2 + k^2}$. We have

$$\begin{cases} x^2 - y^2 = b_x^2 + 4akc\phi(1 - g_1(\tau^2)), \\ 2xy = 4a\tau\phi(1 + cg_1(\tau^2)), \end{cases}$$

which implies, by taking the derivative with respect to τ,

$$\begin{cases} 2xx' - 2yy' = -4akc\phi g_1'(\tau^2) \times 2\tau, \\ 2xy' + 2x'y = 4a\phi(1 + cg_1(\tau^2) + 2c\tau^2 g_1'(\tau^2)), \end{cases}$$

with $g_1'(s) = -\frac{k^2}{(s+k^2)^2} = -\frac{g_1^2(s)}{k^2}$. We have $g_1 + 2\tau^2 g_1' = g_1(2g_1 - 1)$ and thus

$$\begin{cases} 2xx' - 2yy' = -4a\phi \frac{2c}{k} \tau g_1^2(\tau^2), \\ 2xy' + 2x'y = 4a\phi \left(1 + cg_1(\tau^2)(2g_1(\tau^2) - 1)\right), \end{cases}$$

which implies that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{4a\phi}{2(x^2+y^2)} \begin{pmatrix} x & y \\ -y & x \end{pmatrix} \begin{pmatrix} \frac{2c}{k}\tau g_1^2(\tau^2) \\ \varphi_1(\tau^2) \end{pmatrix} = \frac{4a\phi}{2(x^2+y^2)} \begin{pmatrix} \frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y \\ -\frac{2c}{k}\tau g_1^2(\tau^2)y + \varphi_1(\tau^2)x \end{pmatrix},$$

where we put $\varphi_1(\tau^2) := 1 + cg_1(\tau^2)(2g_1(\tau^2) - 1)$.
We have $R(\tau, p) = \frac{(x-p)^2+y^2}{(x+p)^2+y^2}$ and thus

$$\frac{\partial R}{\partial x} = \frac{4p(x^2 - y^2 - p^2)}{((x+p)^2 + y^2)^2}, \quad \frac{\partial R}{\partial y} = \frac{8pxy}{((x+p)^2 + y^2)^2},$$

which implies that

$$\frac{\partial R}{\partial \tau} = \frac{4a\phi 4p}{2(x^2+y^2)^2((x+p)^2+y^2)^2} S,$$

where

$$S = (x^2 - y^2 - p^2)\left(\frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y\right) + 2xy\left(-\frac{2c}{k}\tau g_1^2(\tau^2)y + \varphi_1(\tau^2)x\right).$$

We calculate now, asymptotically for $\tau_M \to \infty$, the zeros of S in $[\tau_m, \tau_M]$. We distinguish between the following two situations:

i. Suppose there is a τ which behaves like $\tau = \mathrm{O}(1)$ and which cancels S. We prove that this situation cannot happen.
We suppose $\tau = \tilde{\tau} + C\tau_M^{-\alpha}$, where $\tilde{\tau} > \tau_m$ and $\alpha > 0$. Then all of the four functions $g_1,\ \varphi_1,\ x$ and y behave at τ asymptotically as $\mathrm{O}(1)$, and thus

$$S \sim -p^2\left(\frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y\right).$$

We have $\frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y = 0$ only if $\varphi_1(\tau^2) < 0$.
Now, $\varphi_1(\tau^2) < 0$ only if

$$\tilde{\varphi}_1 := \frac{c - \sqrt{c^2 - 8c}}{4c} < g_1(\tau^2) < \frac{c + \sqrt{c^2 - 8c}}{4c} := \tilde{\varphi}_2,$$

and if $c^2 - 8c > 0$. Hence we can not have $\varphi_1(\tau^2) < 0$ if $c \leq 8$. Suppose then $c > 8$ and let us search for the zeros of $\frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y$ which lie between $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$. By multiplying the equation $\frac{2c}{k}\tau g_1^2(\tau^2)x + \varphi_1(\tau^2)y$ by x and by y and by replacing $xy = 2a\tau\phi(1 + cg_1)$ in the resulting equations, we obtain

$$\begin{cases} x^2 = -\dfrac{ak\phi}{c}\dfrac{(1 + cg_1)\varphi_1}{{g_1}^2}, \\ y^2 = -\dfrac{4ac\phi}{k}\dfrac{(1 + cg_1){g_1}^2\tau^2}{\varphi_1}. \end{cases} \tag{4.52}$$

We introduce (4.52) in the equation $x^2 - y^2 = b_x^2 + 4akc\phi(1 - g_1)$. Since we have $\tau^2 g_1 = k^2(1 - g_1)$, this equation reads

$$4ack\phi\frac{g_1(1 - g_1)(1 + cg_1)}{\varphi_1} - \frac{ak\phi}{c}\frac{(1 + cg_1)\varphi_1}{{g_1}^2} = b_x^2 + 4akc\phi(1 - g_1). \tag{4.53}$$

We put $\beta := \frac{b_x^2}{4ak\phi}$. We have that (4.53) writes

$$c{g_1}^3(1 - g_1)(1 + cg_1) - \frac{1}{4c}\varphi_1^2(1 + cg_1) - {g_1}^2\varphi_1(\beta + c(1 - g_1)) = 0.$$

Since $\varphi_1 = 2cX^2 - cX + 1$, the polynomial $S(X) = 4c^2X^3(1 - X)(1 + cX) - \varphi_1^2(1 + cX) - 4cX^2\varphi_1(\beta + c(1 - X))$ can also be written as $S(X) = -4c^2(c + 2\beta + 2)X^4 + c^2(3c + 4\beta + 8)X^3 - c(3c + 4\beta + 4)X^2 + cX - 1$. We conclude that S does not have negative zeros. We calculate $S'(X)$ and $S''(X)$, which is the polynomial of degree 2

$S''(X) = -48c^2(c + 2\beta + 2)X^2 + 6c^2(3c + 4\beta + 8)X - 2c(3c + 4\beta + 4)$. If $c > 8$, $S''(X)$ does not have any real root. Hence, $S'(X)$ has only one real root.
We study now $S(X)$, for $X \in [0, 1]$. We remark that $g_1(\tau^2) \in [0, 1]$. We have $S(0) = -1$ and $S(1) < 0$. Furthermore, if $X = \tilde{\varphi}_1$ or $X = \tilde{\varphi}_2$, we have $\varphi_1(\tau^2) = 0$ and $S(g_1) = 4c^2{g_1}^3(1 - g_1)(1 + cg_1) > 0$. We conclude that S has two roots lying respectively in $]0, \tilde{\varphi}_1[$ and in $]\tilde{\varphi}_2, 1[$, and has no root between $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$. Hence, we cannot have a zero of S that behaves as O(1), as $\tau_M \to \infty$.

ii. S has a root that behaves as $\tilde{\tau} = \mathrm{O}(\tau_M^{\alpha})$, with $\alpha > 0$.
We obtain, as in (4.51), $x \sim \sqrt{2a\tau}$, $y \sim \sqrt{2a\tau}$, as $\tau \to \infty$. Furthermore we have $g_1 \sim \frac{k^2}{\tau^2}$ and $\varphi_1 \sim 1$, as $\tau \to \infty$. Hence, $S \sim (-p^2 y + 4a\tau x)$ and, as $\tau \to \infty$, we have $(-p^2 y + 4a\tau x) = 0$ if $\tau \sim \frac{p^2}{4a}$.

We conclude that the extreme points of $\tau \longmapsto R(\tau, p^*)$ in the compact set $K^+ = [\tau_m, \tau_M]$ might be:

i. Either τ_m, in this case we have $R(\tau_m, p^*) \sim 1 - 4\frac{x_m}{p^*}$.

ii. Either τ_M, in this case we have $R(\tau_M, p^*) \sim 1 - 2\frac{p^*}{x_M}$.

iii. Or $\tilde{\tau} \sim {p^*}^2$, in this case we have $R(\tilde{\tau}, p^*) \sim \frac{2-\sqrt{2}}{2+\sqrt{2}}$.

We have that $\tilde{\tau}$ must be a minimum of R if $\tau_M \gg 1$, since $R(\tau_m, p^*) \to 1$, $R(\tau_M, p^*) \to 1$, as $\tau_M \to \infty$, and $R(\tilde{\tau}, p^*) \to \frac{2-\sqrt{2}}{2+\sqrt{2}}$.
Hence

$$\sup_{\tau \in K^+} |\rho(\tau, p^*)| = \max\{|\rho(\tau_m, p^*)|,\ |\rho(\tau_M, p^*)|\},$$

where $p^* = \mathrm{O}(\tau_M^{\frac{1}{4}})$.

4. We conclude by satisfying the equioscillation property.
We can now easily see that $p \longrightarrow |\rho(\tau_m, p)|$ is an increasing function of p and that $p \longrightarrow |\rho(\tau_M, p)|$ is a decreasing function of p.
Hence, since the equioscillation property of Theorem 4.11 holds, we have that the solution of the best approximation problem is given by

$$\inf_{p>0} \sup_{\tau \in K^+} |\rho(\tau, p)| = |\rho(\tau_m, p^*)| = |\rho(\tau_M, p^*)|,$$

with $p^* = O(\tau_M^{\frac{1}{4}})$. ⊠

As a consequence of the proof of Proposition 4.1, we have $R(\tau_m, p^*) \sim 1 - 4\frac{x_m}{p^*}$ and $R(\tau_M, p^*) \sim 1 - 2\frac{p^*}{x_M}$, as $\tau_M \to \infty$. We can thus conclude that $|\rho(\tau_m, p^*)| = |\rho(\tau_M, p^*)|$ if $p^* = \sqrt{2x_m x_M}$, for $\tau_M \gg 1$. We then obtain the following final result:

Theorem 4.12

If τ_M is sufficiently large, the solution of the best approximation problem

$$\inf_{p>0} \sup_{\tau \in K^+} |\rho(\tau, p)|$$

is given by $p^* = \sqrt{2x_m x_M} := \sqrt{2\Re\mathrm{e}\left(\sqrt{b_x^2 + 4ad(\tau_m)}\right)\Re\mathrm{e}\left(\sqrt{b_x^2 + 4ad(\tau_M)}\right)}$, and asymptotically we have

$$p^* = \mathrm{O}\left(\tau_M^{\frac{1}{4}}\right), \quad \delta^* = 1 - \mathrm{O}\left(\tau_M^{-\frac{1}{4}}\right).$$

4.4 Numerical Results

In this section, we present different numerical results in order to illustrate and validate the results of the previous sections.

4.4.1 Performance of Different Transmission Conditions

For the numerical results we fix the time period $t \in [0, 1]$ and the global domain $\Omega = [0, 1] \times [0, 1] \in \mathbb{R}^2$. Discrete steps are $\Delta t = \Delta x = \Delta y = 2 \cdot 10^{-2}$. The diffusion parameter is $a = 1$, advection is $(b_x, b_y) = (1 \cdot 10^{-2}, 5 \cdot 10^{-2})$, the reactivity coefficient is set to $k = 5$ with an equilibrium parameter of $c = 10$. Defining the function

$$f(x, y, t) = (\sin(\pi x)\cos(\pi y)\cos(\pi t) + \cos(\pi x)\sin(\pi y)\cos(\pi t) + \cos(\pi x)\cos(\pi y)\sin(\pi t) + 1)/2,$$

we can set the initial values to $u_0 = f(x, y, 0)$, $v_0 = f(x, y, 0)/c$ for $(x, y) \in \Omega$ and we impose Dirichlet boundary conditions with values set to $u(x, y, t) = f(x, y, t)$ for $(x, y) \in \partial\Omega$. The function f provides a heterogeneity in space and time so that we can ensure that the exact solution that we reconstruct numerically does not degenerate to a stationary problem.

In a first time, we want to illustrate the performance of different transmission conditions used in the Schwarz algorithm. Therefore, we decompose Ω into non-overlapping and overlapping subdomains. The non-overlapping case is $\Omega_1 = [0, 0.5] \times [0, 1]$ and $\Omega_2 = [0.5, 1] \times [0, 1]$. In the overlapping case we decompose Ω into $\Omega_1 = [0, 0.5 + \Delta x] \times [0, 1]$ and $\Omega_2 = [0.5, 1] \times [0, 1]$ using a minimal overlap of size Δx. We impose a random initial guess on the interface Γ_1 and perform a Schwarz algorithm using different transmission conditions: Dirichlet, optimised Robin and optimised Ventcel conditions in the overlapping case and optimised Robin and optimised Ventcel conditions in the non-overlapping case. Optimised in this context means that we use the parameter(s) p and (p, q) resulting from a numerical optimisation of the theoretical convergence

factor of the advection-diffusion-reaction system, i. e. we solver numerically (4.46). Note that Dirichlet conditions in the non-overlapping case do not converge according to section 4.2.1. In figure 4.4 we plot the number of iterations versus the error of the interface values compared to the global monodomain solution in the maximum norm. One can see that the classical Schwarz

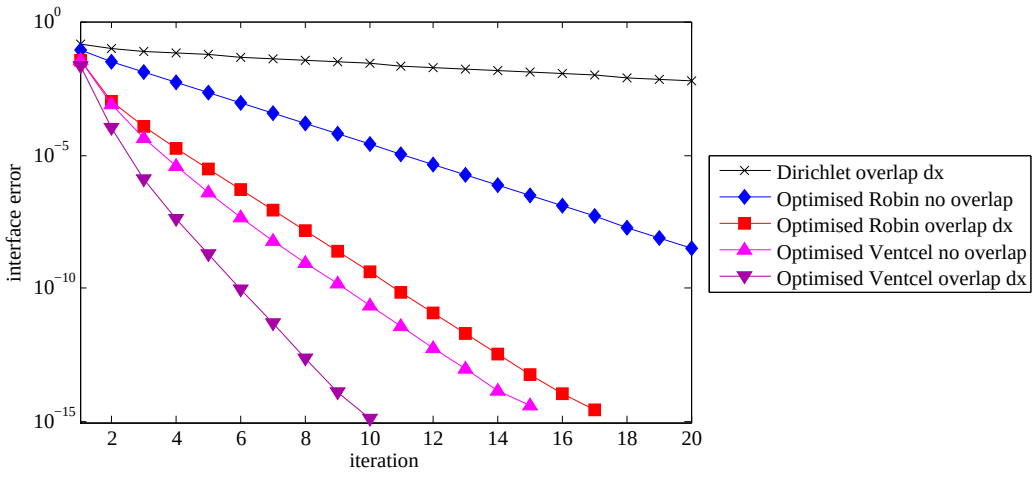


Figure 4.4: Iterations versus error of the domain decomposition iterates

algorithm using Dirichlet conditions converge very slowly while the use of optimised Robin conditions without overlap let the algorithm converge much faster. Optimised Ventcel conditions without overlap converge slightly faster than optimised Robin conditions with overlap. The best convergence behaviour is obtained with optimised Ventcel conditions with overlap which reach the error precision of 10^{-14} in only 10 iterations.

4.4.2 Optimal vs. Optimised Transmission Conditions

As described in section 4.3, the optimised transmission conditions are found by solving the min-max-problem defined in equation (4.46). The error frequencies ξ and τ can now be discretised with a sufficient small discretisation and the theoretical error reduction rate can be evaluated at all discrete points. A numerical optimisation algorithm is applied to the function that returns the maximum of the error reduction factor for those discrete error frequencies in order to find the optimised parameter or parameters that minimise the error reduction factor. We call the parameters p or (p, q) obtained by this procedure the theoretically optimised parameters since they are optimised for the theoretical and idealised problem.
It is now interesting to compare the real behaviour of the Schwarz algorithm using Robin and Ventcel conditions with different parameters p and (p, q) with the theoretically optimised parameters. It is clear that for the development of the error reduction factor many assumptions have to be done that cannot be fulfilled in numerical codes. The next two examples illustrate that the theoretical developments and the real numerical behaviour of the transmission conditions do still

accord. We use the same non-overlapping subdomains as before and vary the parameters p and (p, q) for Robin and Ventcel transmission conditions, respectively, and perform a fixed number of iterations (10 for Robin and 4 for Ventcel conditions) of the Schwarz algorithm imposing always the same initial random guess on the interface for the first iteration. We plot in figure 4.5 for Robin and in figure 4.6 for Ventcel transmission conditions the variation of the parameter(s) versus the error at the fixed iteration. One observes that in the case of the Robin and Ventcel

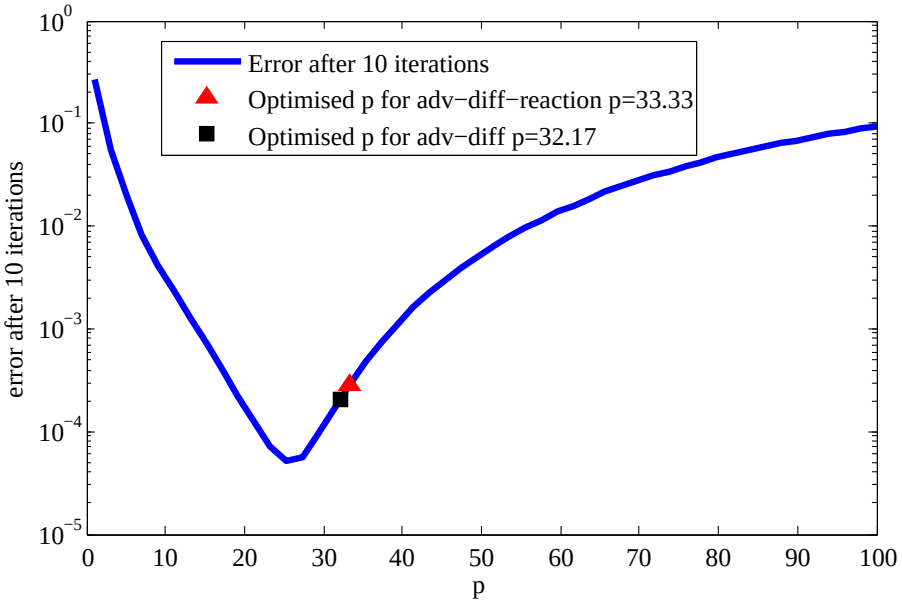


Figure 4.5: Variation of the parameter p of the Robin transmission condition versus the error of the 10th iterates. The triangle locates the numerically optimised parameter of the advection-diffusion-reaction system ($p = 33.33$), the square locates the optimised parameter of the advection-diffusion equation ($p = 32.17$).

transmission conditions the theoretically optimised parameters are very close the parameters that offer the best performance of the Schwarz algorithm using Robin and Ventcel transmission conditions, respectively. Moreover, the parameters of the advection-diffusion-reaction system and the advection-diffusion equation are quite close in the 2D case.

4.4.3 Sensitivity of Optimised Transmission Conditions to the Coupling Term Strength

In order to do a first step towards the nonlinear problem, we are, in a first time, interested in the sensitivity of the optimised parameters with respect to the parameter k resulting from the reaction coupling term. Note that for $k = 0$ and for $k \to \infty$ the error reduction factor degenerates to the error reduction factor of an advection-diffusion type (with a different parameter for the porosity in the case $k \to \infty$) that has been widely studied and for which an analytical formula exists

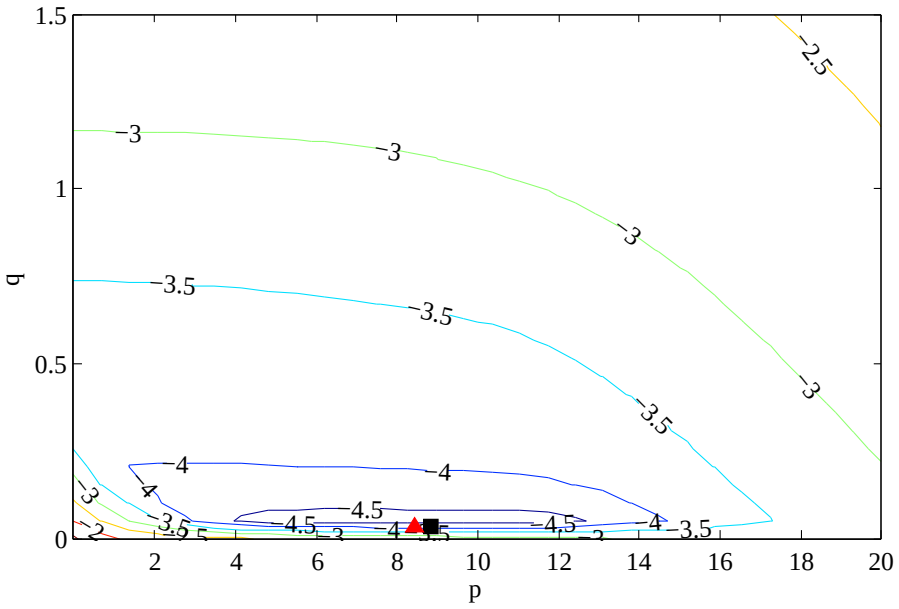


Figure 4.6: Variation of the parameters (p, q) of the Ventcel transmission condition versus the logarithm of the error of the 4th iterates. The triangle locates the numerically optimised parameter of the advection-diffusion-reaction system ($p = 8.8507$, $q = 0.0322$), the square locates the optimised parameter of the advection-diffusion equation ($p = 8.4184$, $q = 0.0329$).

(cf. Bennequin et al. [6] for the 1D case). The important question is: is it necessary to take into account the coupling term in order to find optimised parameters or can we neglect the reaction and use optimised parameters for the simpler single equation of advection-diffusion type?

We study this question first in 1D. We use the same test parameters as before (reduced to 1D in x-direction in the non-overlapping case) and study the theoretically optimised parameters by varying the coupling parameter k. In figure 4.7 we plot the theoretically optimised parameters for different k of the coupled advection-diffusion-reaction system together with the theoretically parameter for a single equation of advection-diffusion type. One can see that the parameters for $k = 0$ are equal since the coupled system degenerates to a single equation of advection-diffusion type. For $k \to \infty$ there is no more sensitivity visible since the equation converges also to a single equation of advection-diffusion type (with a different parameter for the porosity). The variation of the parameters for intermediate k is highly visible.
But how important is the impact of the parameter on the error reduction factor of the Schwarz algorithm for different k? We can compare, for different k, the maximum of the error reduction factor for the coupled advection-diffusion-reaction system using different parameters: the theoretically optimised parameter for the coupled advection-diffusion-reaction system and the theoretically optimised parameter for the single equation of advection-diffusion type. In figure 4.8 we plot the error reduction factors using different parameters versus the coupling factor k. One observes that the error reduction factor for $k = 0$ and for $k \to \infty$ using the theoretically optimised parameter for the advection-diffusion-reaction system (solid blue line) is around 0.4 while

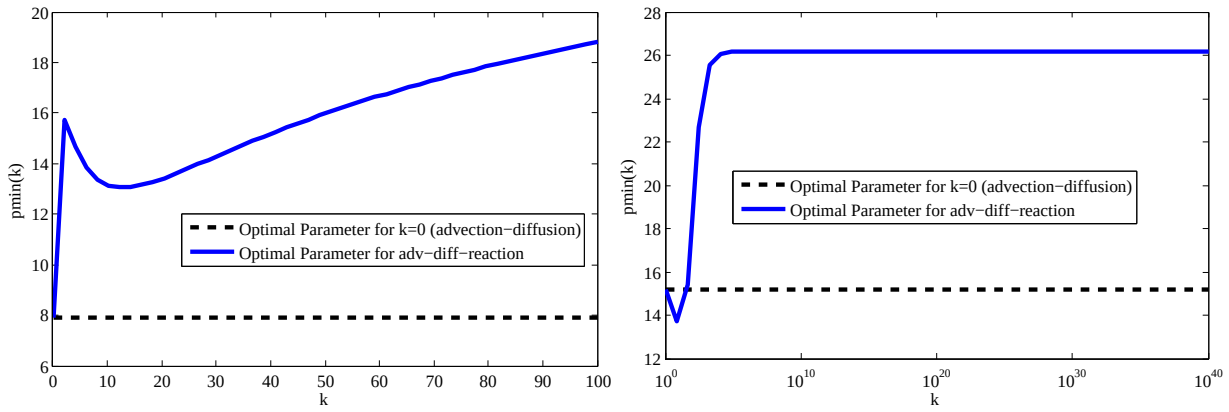


Figure 4.7: Variation of the theoretically optimised parameter for the Robin transmission condition versus different k in 1D.

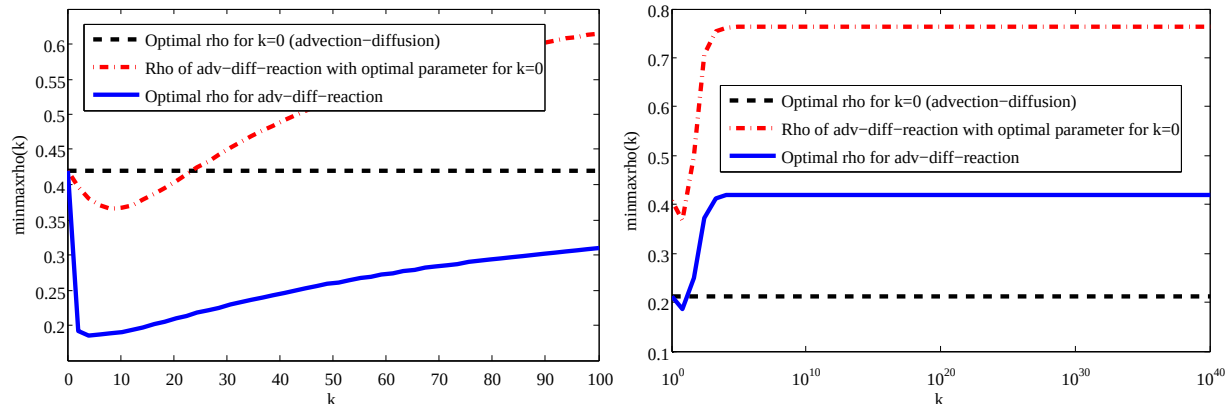


Figure 4.8: Error reduction factor of the coupled advection-diffusion-reaction system using different parameters for the Robin condition versus different k in 1D.

for intermediate k it is lower (down to 0.2). If one uses the theoretically optimised parameter of the single equation of advection-diffusion type (dash-dot line in red), the error reduction factor for $k = 0$ is the same but double for intermediate k and large k. This means, even for small k (and especially for intermediate and large k), the error reduction factor is tremendously different. To exemplify this crucial issue, suppose the initial guess to have an error of 1. If one wants to iterate the Schwarz algorithm until the error is 10^{-12} one would need about 30 iterations with the parameter taking into account the reaction (error reduction factor of 0.4) but about 124 iterations if one uses the parameter of the advection-diffusion equation (error reduction factor of 0.8).
We consider now Ventcel conditions in 1D and perform the same tests. In figures 4.9 and 4.10 we plot the same results as for Robin conditions. One observes that both parameters p and q are nearly insensitive to the parameter k. Moreover, the error reduction factor is about 0.11 for the optimised parameter taking into account the reaction and 0.13 when not taking into account the reaction. Assuming the same error properties as for Robin conditions, one would need about

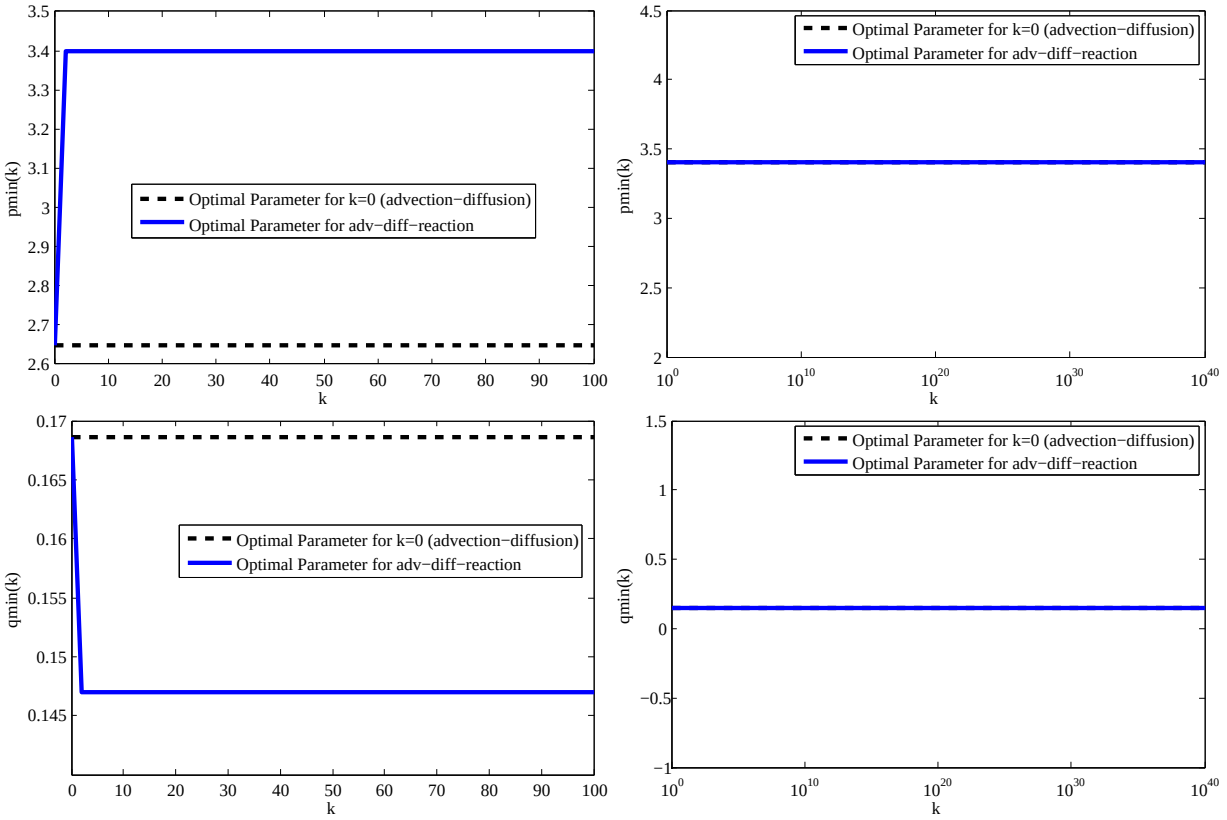


Figure 4.9: Variation of the theoretically optimised parameters for the Ventcel transmission condition versus different k in 1D.

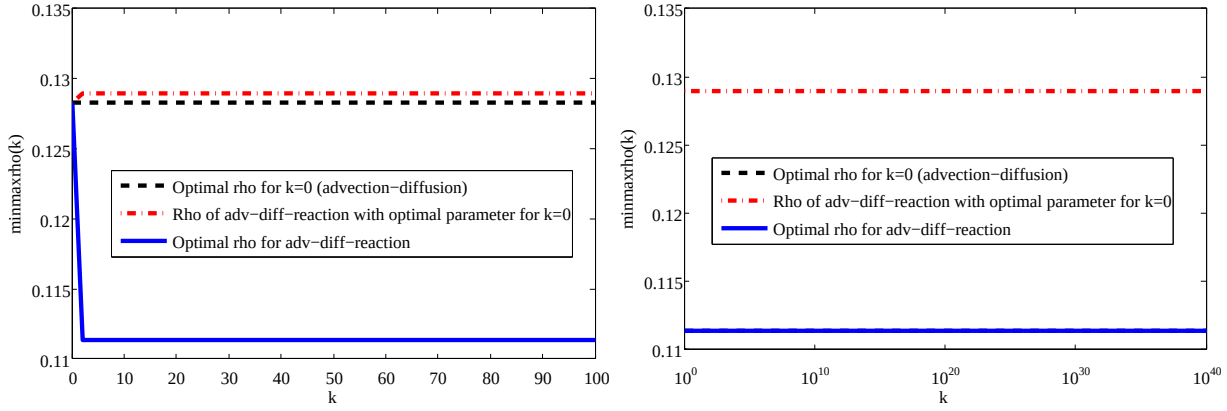


Figure 4.10: Error reduction factor of the coupled advection-diffusion-reaction system using different parameters for the Ventcel conditions versus different k in 1D.

13 iterations using the optimised parameter taking into account reaction and about 14 iterations when not taking into account reaction.

In 2D, things change. We perform the same tests as before, this time in 2D and plot in figures 4.11 and 4.12 the results of the same tests as for Robin conditions in 1D. One observes that now

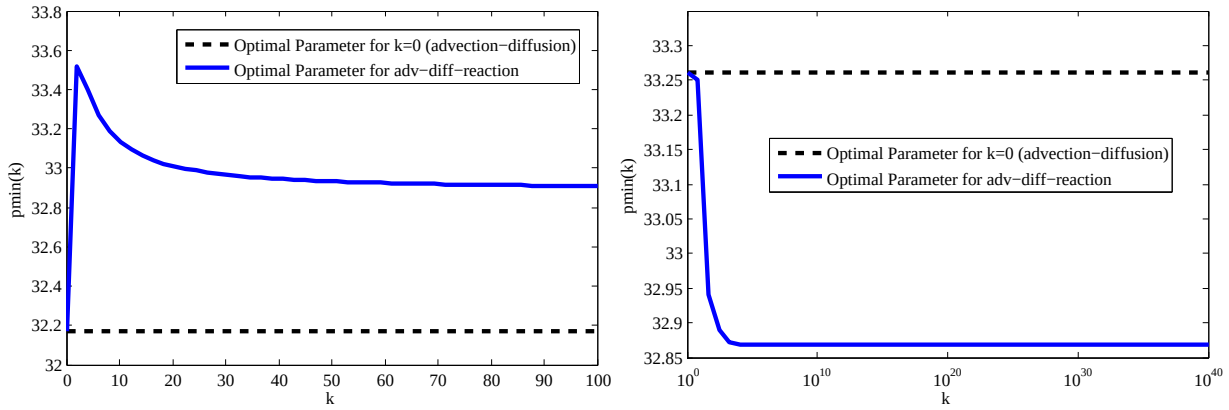


Figure 4.11: Variation of the theoretically optimised parameter for the Robin transmission condition versus different k in 2D.

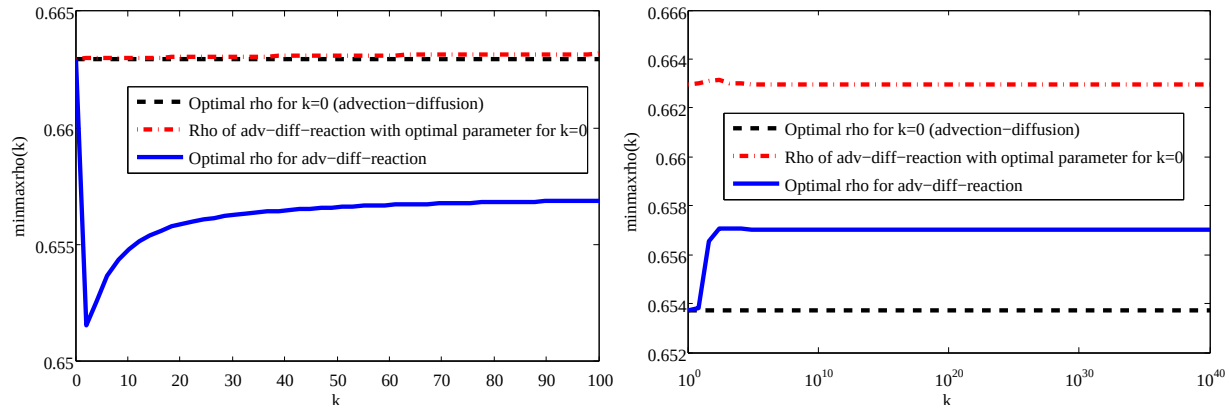


Figure 4.12: Error reduction factor of the coupled advection-diffusion-reaction system using different parameters for the Robin condition versus different k in 2D.

neither the optimised parameters nor error reduction factors have a significant sensitivity to the coupling factor k. For Ventcel conditions in 2D similar results are obtained.
Now, we are interested in the question if the insensitivity of the optimal error reduction factor with respect to the choice of either the optimised parameter of the advection-diffusion equation or the optimised parameter of the advection-diffusion-reaction system is kept when the mesh is refined. Therefore we can calculate the difference of the reduction factor of the advection-diffusion-reaction system using two different parameters: the optimised parameter of the advection-diffusion

equation and the optimised parameter of the advection-diffusion-reaction system. We are interested in the maximum difference over a large range of different coupling parameters k. The maximum difference is defined by

$$\max_{k\in[k_0,k_{\max}]}\left(\max_{\tau,\xi}\left(\rho_{\text{adv-diff-react}}(p^*_{\text{adv-diff-react}},\xi,\tau)\right)-\max_{\tau,\xi}\left(\rho_{\text{adv-diff}}(p^*_{\text{adv-diff}},\xi,\tau)\right)\right),$$

where $\rho_{\text{adv-diff-react}}$ is the error reduction factor of the coupled advection-diffusion-reaction system, $p*_{\text{adv-diff-react}}$ is the optimised parameter of the coupled advection-diffusion-reaction system and $p^*_{\text{adv-diff}}$ is the optimised parameter of the advection-diffusion single-equation. We study the maximum difference with respect to different discretisation sizes in y and t. In figure 4.13 we plot the variation of this maximum difference with respect to different discretisation sizes. One observes

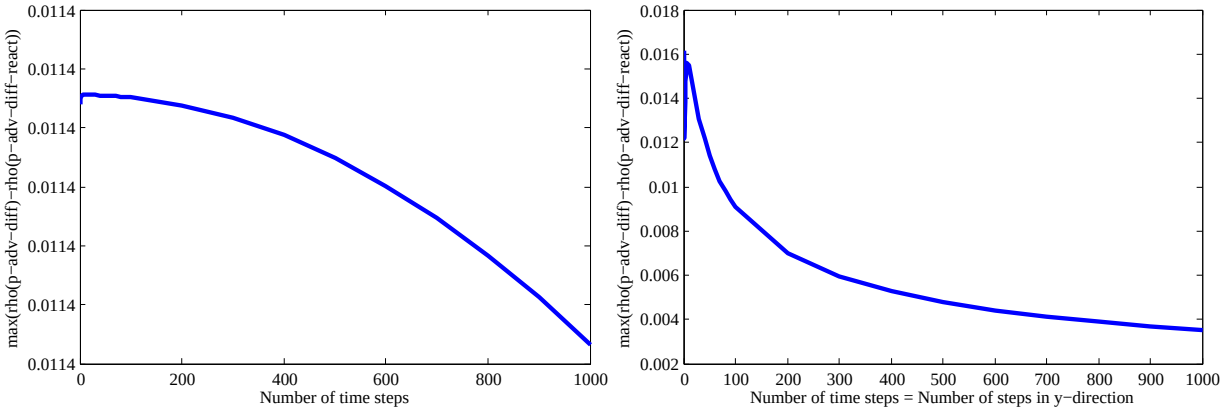


Figure 4.13: Maximum difference of the convergence factor using different optimised parameters. Left: $N_y = 50$ points in y-direction and varying the number of points in t-direction. Right: variation of discretisation points with $N_y = N_t$.

that for a variation of the discretisation only in t and for t and y at the same time the difference does not exceed 0.2. The insensitivity of the error reduction factor with respect to the use of the two different optimised parameters in the case of mesh refinement is therefore conserved.
Since we know that the optimised parameter of the advection-diffusion equation behaves asymptotically like $p^*_{\text{adv-diff}} \sim O(\tau_{\max}^{\frac{1}{4}})$ and therefore a constant $C_{\text{adv-diff}}$ exists such that $p^*_{\text{adv-diff}} = C_{\text{adv-diff}}\tau_{\max}^{\frac{1}{4}}$ when $\tau_{\max} \to \infty$. We study therefore $\frac{p^*_{\text{adv-diff}}}{\tau_{\max}^{\frac{1}{4}}}$ and $\frac{p^*_{\text{adv-diff-react}}}{\tau_{\max}^{\frac{1}{4}}}$ in both cases when refining asymptotically the mesh in time keeping the number of points in y-direction constant at 50 and plot the results in figure 4.14. One can see that for both cases the ratio converges to a constant value when the time mesh is refined. The optimised parameters for both cases behave as supposed and the constants are only slightly different.
Finally, we are interested in the comparison of the real behaviour of the Schwarz waveform relaxation algorithm using three different parameters in 1D: the numerically and asymptotically optimised parameters for the reactive transport system and the numerically optimised parameters for the advection-diffusion equation. Therefore, we set the same test parameters as before in

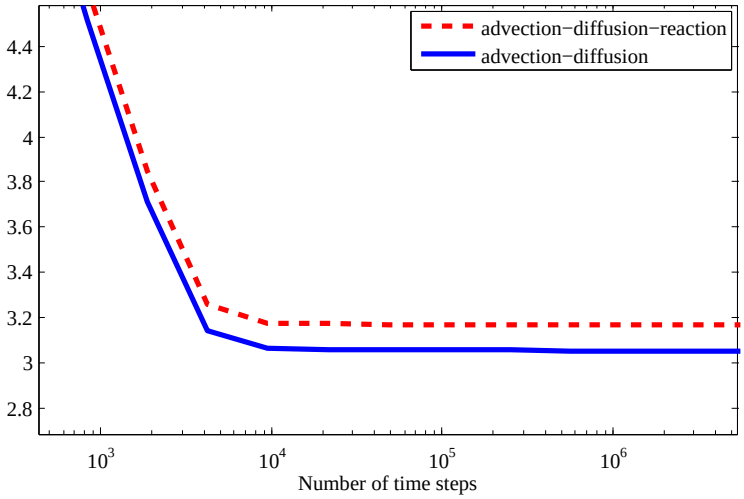


Figure 4.14: Asymptotic behaviour of $\frac{p^*}{\tau_{\max}^{\frac{1}{4}}}$ for advection-diffusion-reaction system and advection-diffusion equation.

1D using $k = 100$ with Dirichlet conditions $u(x = 0) = 1$, $u(x = 1) = 0$ and initial conditions $u_0 = v_0 = 0$. We calculate first the three different parameters for this problem using different discretisations in time and plot in figure 4.15 their variation when refining asymptotically the time discretisation. One observes that the asymptotically and numerically optimised parameters

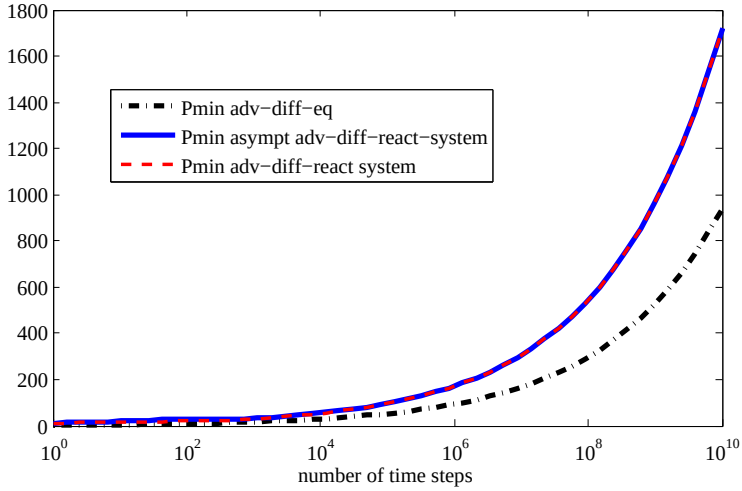


Figure 4.15: Asymptotic behaviour of the numerically (dashed red line) and asymptotically (solid blue line) optimised parameters for the reactive transport system and the numerically optimised parameter of the advection-diffusion equation (dash-dotted black line).

of the reactive transport system match rapidly while the numerically optimised parameter of the

advection-diffusion equation differs significantly. In order to verify the asymptotically matching of the parameters for the reactive transport system we plot in figure 4.16 the relative error

$$\frac{|p^*_{\text{num}}(nt) - p^*_{\text{asympt}}(nt)|}{p^*_{\text{num}}(nt)}$$

between the asymptotically and numerically optimised parameters. One observes that the rela-

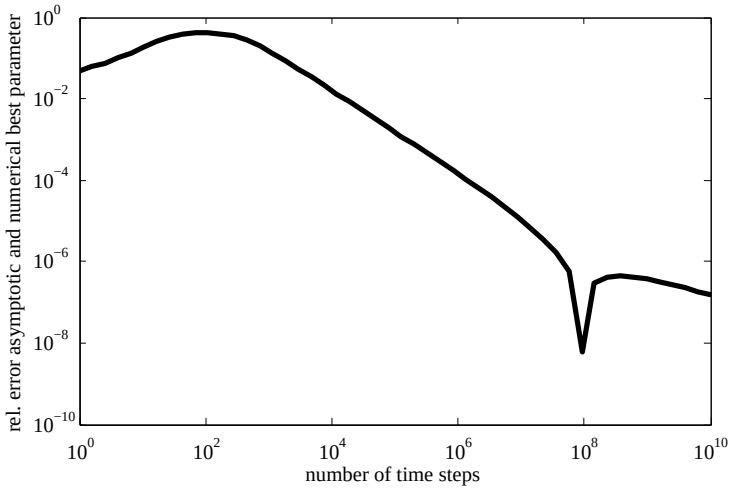


Figure 4.16: Asymptotic behaviour of the relative error between the numerically and asymptotically optimised parameters for the reactive transport system.

tive error decreases as the number of time steps is increased. The peak close to 10^8 time steps is due to the fact that at this point we change from $p^*_{\text{num}} < p^*_{\text{asympt}}$ to $p^*_{\text{num}} > p^*_{\text{asympt}}$ and hence the relative error is zero. Finally, we are interested in the real behaviour of the three parameters for a fixed time discretisation, i. e. 50 time steps. Therefore, we impose a random initial guess on the interface values and proceed a Schwarz waveform relaxation algorithm as in section 4.4.1 and plot in figure 4.17 the error of the interface values versus the number of iterations for the three different parameters. One observes that the algorithm using asymptotically and numerically optimised parameters for the reactive transport system needs only 13 and 19 iterations to reach the precision of 10^{-12} while using the numerically optimised parameters of the advection-diffusion equations needs 45 iterations.

4.4.4 Locally Optimised Transmission Conditions

An interesting application of the asymptotical solution of the best approximation problem in 1D is using it numerically in a local way on the interface: while in the classical approach in 2D and 3D the parameter for the transmission condition is equally chosen all over the interface and is determined by solving (analytically, asymptotically or numerically) the associated best approximation problem, one can choose also variable parameters. We apply this technique and

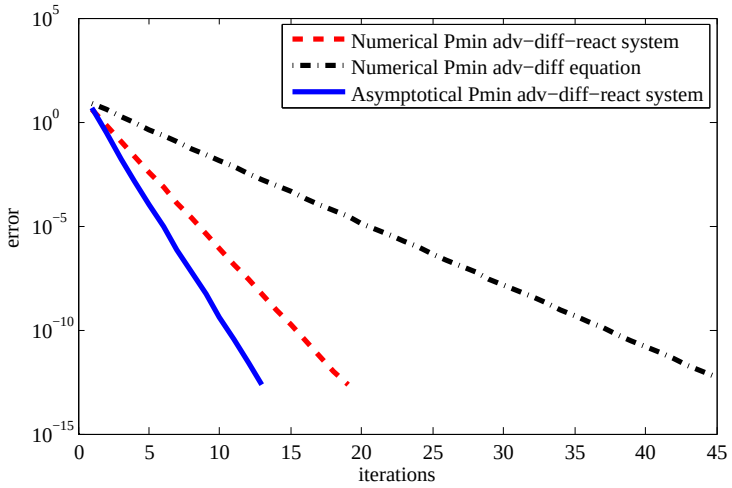


Figure 4.17: Error of the interface variables during the iterations using three different parameters.

set for every discrete interface face a parameter that we obtain by using the asymptotical solution for the one-dimensional problem (cf. theorem 4.12).
We fix the time period $t \in [0,1]$ and the global domain $\Omega = [0,1] \times [0,1] \times [0,1] \in \mathbb{R}^3$ which is decomposed into $\Omega_1 = [0,0.5] \times [0,1] \times [0,1]$ and $\Omega_2 = [0.5,1] \times [0,1] \times [0,1]$. Discrete steps are $\Delta t = 0.02$, $\Delta x = \Delta y = \Delta z = 0.05$. The diffusion parameter is $a = 1.5$, advection is $(b_x, b_y, b_z) = (5 \cdot 10^{-2}, 1 \cdot 10^{-3}, 1 \cdot 10^{-3})$. The reactive surface coefficient is set to $k = 100xyz$ and the equilibrium parameter to $c = 10$. Defining the function

$$\begin{aligned} f(x,y,z,t) = &(\sin(2\pi x)\cos(2\pi y)\cos(2\pi z)\cos(2\pi t) + \cos(2\pi x)\sin(2\pi y)\cos(2\pi z)\cos(2\pi t) \\ &+ \cos(2\pi x)\cos(2\pi y)\sin(2\pi z)\cos(2\pi t) + \cos(2\pi x)\cos(2\pi y)\cos(2\pi z)\sin(2\pi t) + 1)/2, \end{aligned}$$

we impose Dirichlet boundary conditions with values set to $u(x,y,z,t) = f(x,y,z,t)$ for $(x,y,z) \in \partial\Omega$. Initial values are set to $u_0 = 0.5$, $v_0 = 5.0$. The function f provides a heterogeneity in space and time so that we can ensure that the exact solution that we reconstruct numerically does not degenerate to a stationary problem. Moreover, the reactivity is set to a highly heterogeneous value in space.
In figure 4.18 we plot the interface error at the iterations of the Schwarz waveform relaxation method imposing a random initial guess, once with a globally optimised parameter in the 3D case and once with variable per face parameters obtained by the 1D asymptotical solution of the best approximation problem. One observes that both parameters lead to a considerable error reduction. While the algorithm using the 1D parameter is faster during the first 9 iterations the 3D parameter becomes more performing up from the 10th iteration. The 1D parameter attenuates quickly the low frequencies within two iterations and slows then down for high frequencies in the error. The 3D parameter attenuates both high and low frequencies in the same way since it is optimised over all frequencies (also in y and z direction) while the 1D parameter is only optimised over frequencies in t direction. Nevertheless, for practical use, the Schwarz waveform relaxation algorithm is processed until a certain error which is often chosen to be equal as or less

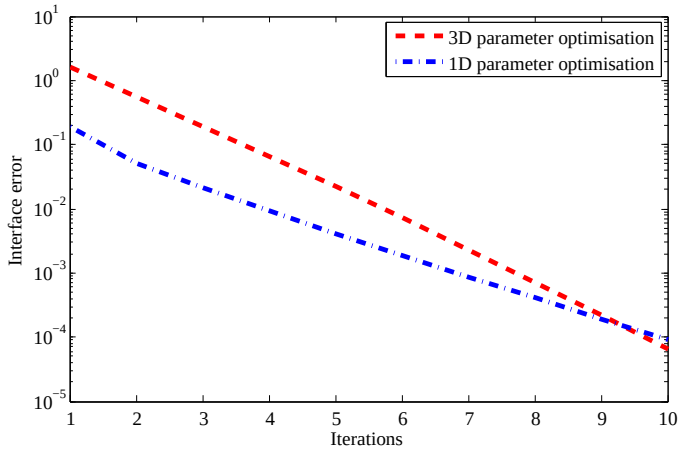


Figure 4.18: Interface error using a global optimised parameter for the 3D problem and a locally per face optimal parameter of the 1D problem

than the error of the discrete approximation. For this reason, the 1D parameter may be a good alternative to the 3D optimisation in practice.

The Schwarz waveform relaxation algorithm needs two iterations for every time window to converge.

Conclusion

In this chapter we have studied a Schwarz waveform relaxation algorithm applied to a coupled linear reactive transport system. After introducing the algorithm and different types of transmission conditions we stated their convergence factors. By using classical results and extending previous works we could state and prove well-posedness and convergence results. We then concentrated on the optimisation of transmission conditions on a numerical and analytical level. Finally, we have shown different numerical results.

By the work presented in this chapter we checked that the application of a Schwarz waveform relaxation algorithm in its various derivatives to the linear coupled reactive transport system is built on a solid theoretical foundation. This gives a mathematical sense to the numerical work of the prototype code presented in chapter 3.
Concerning the different transmission conditions, we obtained more general results that hold also for the scalar reactive transport equation. This issue has been studied in detail in the numerical section where we have emphasised several issues: first, the numerical approach of optimised transmission conditions is valid in practice and can be used in many cases as a powerful tool to predict or approach optimal transmission conditions. Then, we have shown, in prevision of chapter 5 where we study the Schwarz waveform relaxation for a nonlinear coupled reactive transport

system, that in some cases the coupling term can play a non negligible role when one wants to establish optimised transmission conditions. This is particularly interesting in the 1D case without overlap which justifies the work of the analytical solution of the best approximation problem carried out in this chapter. Moreover, we have proposed a technique that works quite well for more complex problems with variable coefficients along the interface where the analytical solution can be used locally.

Our nature hardly allows us to have enough of anything without having too much.

George Savile

5

A Nonlinear Coupled Two Species Reactive Transport System

Contents

Introduction

In this chapter we study a Schwarz waveform relaxation method on a nonlinear coupled two species reactive transport system. Having studied this system on a linear level in chapter 4, we are now interested in the change of behaviour of the domain decomposition algorithm with respect to the nonlinearity.

In the first part, we state the problem and apply a Schwarz waveform relaxation method. Then, we study the numerical approach and present three different approaches. We conclude finally with numerical results.

5.1 Problem Definition

In this chapter we consider the model problem of a coupled two species reactive transport system

$$\begin{aligned} \phi\partial_t u + \operatorname{div}(-a\vec{\nabla}u + \vec{b}u) - R(u,v) &= f_u \quad \text{on } \Omega \times (0,T), \\ \phi\partial_t v \qquad\qquad\qquad + R(u,v) &= f_v \quad \text{on } \Omega \times (0,T), \end{aligned} \tag{5.1}$$

on the spatial domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$ and the time period $(0, T)$. $\phi(x) > 0$ denotes the porosity. The mobile species u is subject to a linear transport operator $Lu := \operatorname{div}(-a\vec{\nabla}u + \vec{b}u)$ including diffusion described by a positive scalar diffusion coefficient $a > 0$ and advection described by a Darcy field vector $\vec{b} \in \mathbb{R}^d$. The fixed species v is coupled to the mobile species u by a nonlinear coupling term $R(u, v)$. Both species are subject to a right hand side term f_u, f_v. We impose an initial condition for the mobile and fixed species

$$u(x,0) = u_0(x), \qquad v(x,0) = v_0(x) \qquad \text{on } \Omega, \tag{5.2}$$

and a boundary condition for the mobile species u

$$\mathcal{J}u = g(x,t), \qquad \text{on } \partial\Omega \times (0,T), \tag{5.3}$$

where the linear boundary operator $\mathcal{J}$ can be of different types, e. g. Dirichlet, Neumann.

This problem arises as a subproblem of multispecies reactive transport systems like (1.17) defined in chapter 1: on the one hand, system (5.1) corresponds to equations (1.17d)–(1.17e). On the other hand, one can also condense system (5.1) to one equation which corresponds then to an equation of type (1.17c). This system appears to be the most challenging subsystem since mobile and fixed species are coupled by nonlinear reaction terms resulting of kinetic reactions.

5.2 Schwarz Waveform Relaxation Algorithm

We can approach problem (5.1)–(5.3) by a Schwarz waveform relaxation algorithm. Therefore, we decompose the spatial domain Ω in two possibly overlapping subdomains Ω_1, Ω_2 such that $\Omega = \Omega_1 \cup \Omega_2$. We call $\Gamma_1 = \partial\Omega_1 \setminus \partial\Omega$ and $\Gamma_2 = \partial\Omega_2 \setminus \partial\Omega$ the interfaces. In the case of nonoverlapping subdomains we have $\Gamma_1 = \Gamma_2 = \overline{\Omega}_1 \cap \overline{\Omega}_2$. We use linear transmission conditions $\mathcal{B}_1$ and $\mathcal{B}_2$ to transmit the information from one subdomain to another on the interfaces Γ_1, Γ_2. Providing an initial guess (u_2^0, v_2^0) on Γ_1, we can state the entire approach in algorithm 5.1. We have to

Algorithm 5.1 Alternating Schwarz waveform relaxation algorithm for the nonlinear coupled two species reactive transport system

$$
\begin{aligned}
&\phi\partial_t u_1^{k+1} + \operatorname{div}(-a\vec{\nabla}u_1^{k+1} + \vec{b}u_1^{k+1}) - R(u_1^{k+1}, v_1^{k+1}) = f_u && \text{on } \Omega_1 \times (0,T)\\
&\phi\partial_t v_1^{k+1} + R(u_1^{k+1}, v_1^{k+1}) = f_v && \text{on } \Omega_1 \times (0,T)\\
&\mathcal{J}u_1^{k+1} = g(x,t) && \text{on } (\partial\Omega_1 \setminus \Gamma_1) \times (0,T)\\
&(u_1^{k+1}(x,0), v_1^{k+1}(x,0)) = (u_0, v_0) && \text{on } \Omega_1\\
&\mathcal{B}_1 u_1^{k+1} = \mathcal{B}_1 u_2^{k} && \text{on } \Gamma_1 \times (0,T)
\end{aligned}
$$

$$
\begin{aligned}
&\phi\partial_t u_2^{k+1} + \operatorname{div}(-a\vec{\nabla}u_2^{k+1} + \vec{b}u_2^{k+1}) - R(u_2^{k+1}, v_2^{k+1}) = f_u && \text{on } \Omega_2 \times (0,T)\\
&\phi\partial_t v_2^{k+1} + R(u_2^{k+1}, v_2^{k+1}) = f_v && \text{on } \Omega_2 \times (0,T)\\
&\mathcal{J}u_2^{k+1} = g(x,t) && \text{on } (\partial\Omega_2 \setminus \Gamma_2) \times (0,T)\\
&(u_2^{k+1}(x,0), v_2^{k+1}(x,0)) = (u_0, v_0) && \text{on } \Omega_2\\
&\mathcal{B}_2 u_2^{k+1} = \mathcal{B}_2 u_1^{k+1} && \text{on } \Gamma_2 \times (0,T)
\end{aligned}
$$

emphasise several issues on this algorithm: first, in contrast to all so far presented Schwarz type methods in this work, it is the first time that the method is applied to a nonlinear problem. As the original problem itself is nonlinear, the subsidiary problems in the subdomain at every iteration of the Schwarz method are also nonlinear. Concerning the transmission conditions, we impose only linear transmission conditions. For this reason, the operators $\mathcal{B}_1$, $\mathcal{B}_2$ are linear. Moreover, we impose only transmission conditions of Dirichlet, Neumann or Robin type, all act only on the mobile species u and their definition is straightforward. In chapter 4.2.1, we have also developed transmission conditions of Ventcel type in the linear context. They include a tangential flux information on the interface which means that also temporal derivatives have to be included. In the linear case, the mobile species is coupled in a linear way to the fixed species and therefore this coupling term, namely the time derivative information of the fixed species v, is used in the Ventcel condition. The Ventcel condition in the linear case is developed with the help of Fourier transformation. In the nonlinear case, this approach is no longer valid and the definition of a Ventcel condition is not clear. Different strategies to establish a Ventcel type condition in the nonlinear case are imaginable:

1. One neglects the nonlinear coupling term and uses only a **linear Ventcel transmission condition** for the linear advection-diffusion problem. Under the theoretical assumptions of chapter 4, the corresponding transmission conditions are

$$\mathcal{B}_1(u,v) = \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_{\boldsymbol{y}} u + \vec{b}_{\boldsymbol{y}} \cdot \vec{\nabla}_{\boldsymbol{y}} u\right),$$
$$\mathcal{B}_2(u,v) = \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_{\boldsymbol{y}} u + \vec{b}_{\boldsymbol{y}} \cdot \vec{\nabla}_{\boldsymbol{y}} u\right).$$

   They are easy to use but their performance will not be optimal when the coupling nonlinearity becomes dominant compared to spatial tangential processes (advection or diffusion in tangential interface direction).

2. One linearises the nonlinear coupling term and uses the derivative information instead of the linear coefficients of the linear version. This results in **linearised Ventcel transmission conditions**. Suppose therefore the nonlinear coupling term to be linearised around a point $(\tilde{u}, \tilde{v})$, i. e.

$$R(u,v) \approx R(\tilde{u},\tilde{v}) + \left.\frac{\partial R}{\partial u}\right|_{(\tilde{u},\tilde{v})}(u-\tilde{u}) + \left.\frac{\partial R}{\partial v}\right|_{(\tilde{u},\tilde{v})}(v-\tilde{v}).$$

   Using then the relation between the linear form $R_{\text{lin}}(u,v) = k(v - cu)$ and the nonlinear form, one can set $k = \left.\frac{\partial R}{\partial u}\right|_{(\tilde{u},\tilde{v})}$ and $c = -\left(\left.\frac{\partial R}{\partial v}\right|_{(\tilde{u},\tilde{v})}\right)^{-1}\left.\frac{\partial R}{\partial u}\right|_{(\tilde{u},\tilde{v})}$ to obtain the linearised Ventcel conditions

$$\mathcal{B}_1(u,v) = \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_{\boldsymbol{y}} u + \vec{b}_{\boldsymbol{y}} \cdot \vec{\nabla}_{\boldsymbol{y}} u - \left.\frac{\partial R}{\partial u}\right|_{(\tilde{u},\tilde{v})} v - \left.\frac{\partial R}{\partial v}\right|_{(\tilde{u},\tilde{v})} u\right),$$
$$\mathcal{B}_2(u,v) = \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_{\boldsymbol{y}} u + \vec{b}_{\boldsymbol{y}} \cdot \vec{\nabla}_{\boldsymbol{y}} u - \left.\frac{\partial R}{\partial u}\right|_{(\tilde{u},\tilde{v})} v - \left.\frac{\partial R}{\partial v}\right|_{(\tilde{u},\tilde{v})} u\right).$$

   Caetano et al. propose in [12] this type of transmission conditions and call them "nonlinear transmission conditions". The conditions have shown to perform well in contrast to the first approach (linear Ventcel transmission conditions) in the context of a reaction diffusion equation with strong nonlinearities. Nevertheless, there are two crucial issues in order them not to loose performance: first, the use of the derivative itself in the transmission condition is only valid if the the point $(\tilde{u}, \tilde{v})$ is not "too far" from the actual solution, otherwise the linearisation is of bad quality. Then, concerning the optimised parameters, they also have to be updated regularly, otherwise the linearisation will only furnish poor information to the optimised parameter strategy and therefore the performance of the algorithm using this kind of transmission condition will deteriorate.

3. One uses directly the nonlinear coupling term in the transmission condition and creates in this way **nonlinear Ventcel transmission conditions**. According to the linear case, they have the form

$$\mathcal{B}_1(u,v) = \frac{\partial u}{\partial n_1} - \frac{b_x - p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_y u + \vec{b}_y \cdot \vec{\nabla}_y u - R(u,v)\right),$$
$$\mathcal{B}_2(u,v) = \frac{\partial u}{\partial n_2} + \frac{b_x + p}{2a}u + \frac{q}{2a}\left(\phi\partial_t u - a\Delta_y u + \vec{b}_y \cdot \vec{\nabla}_y u - R(u,v)\right).$$

The use of a nonlinear condition affects the character of the problem not adversely since the problem itself is already nonlinear and therefore it becomes not more complex when the transmission condition is also nonlinear. We suppose this transmission condition, theoretically, to have the most powerful performance under the condition that the parameters (p,q) can be adequately supplied. Note that the last point is quite critical: in the case of linear and linearised Ventcel transmission conditions, we can still use (with some precaution) the linear theory and develop analytically, asymptotically or numerically some kind of "optimised parameters". In the nonlinear context, this technique is no longer accessible analytically or asymptotically. Nevertheless, on a numerical level, it might be possible to obtain within discrete Fourier transformations (Fast Fourier Transformation for instance) an information of the error frequencies and deduce by this way parameters that are supposed to optimise the convergence factor. As this approach depends on the nonlinear function, it has to be performed frequently online during a numerical simulation and we doubt that the effort for this strategy is it worth to obtain a faster convergence than with linear or linearised Ventcel conditions or even Robin conditions together with accelerating ingredients like overlap and/or Krylov accelerators.

5.3 Numerical Approaches

It is possible to use algorithm 5.1 in its stated form directly in a numerical context. In the linear context, this is rarely done since other formulations are better suited to collaborate with other numerical ingredients like Krylov accelerators. In this section, we present three different approaches. The first one, the interface problem, closely follows the linear case and opens the door to two new approaches which we present afterwards. In order to make the readability easier and show the generality of the approaches, we reduce problem (5.1) to the following condensed general form:

$$\begin{aligned} \phi\partial_t w + \mathcal{L}w + \mathcal{F}(w) &= q && \text{on } \Omega\times(0,T), \\ w(x,0) &= w_0(x) && \text{on } \Omega, \\ \mathcal{G}w &= g(x,t) && \text{on } \partial\Omega\times(0,T). \end{aligned} \tag{5.4}$$

Note that by setting $w = (u,v)^t$, $\mathcal{F}(\cdot) = (-R(\cdot), R(\cdot))^t$, $q = (f_u, f_v)^t$, $w_0 = (u_0, v_0)^t$, $\mathcal{G} = (\mathcal{J}u, 0)^t$ and $\mathcal{L}w = (Lu, 0)^t$, one obtains problem (5.1), (5.2), (5.3) within formulation (5.4).

5.3.1 Classical Approach

The classical approach is to transform algorithm 5.1 such that the unknowns on the interfaces become the primary unknowns. In the linear case, one obtains then a linear system on the interface unknowns which is solved with a Krylov subspace solver like GMRES for instance. This approach is commonly known and illustrated for example in the books of Toselli and Widlund [77].
In the nonlinear context, one obtains a nonlinear function on the interface unknowns.

5.3.1.1 Reduction to Interface Variables and Fixed Point Algorithm on the Interface Problem

The reduction of algorithm 5.1 to the interface variables is similar in the linear and nonlinear context. First, one defines the subproblem solution operator for the subdomain i as

$$\mathcal{M}_i : \ (\lambda, f) \mapsto w_i \text{ solution of } \begin{cases} \phi\partial_t w_i + \mathcal{L}w_i + \mathcal{F}(w_i) &= q & \text{on } \Omega_i \times (0,T), \\ w_i(x,0) &= w_0(x) & \text{on } \Omega_i, \\ \mathcal{G}w_i &= g(x,t) & \text{on } (\partial\Omega_i \setminus \Gamma_i) \times (0,T), \\ \mathcal{B}_i w_i &= \lambda & \text{on } \Gamma_i \times (0,T), \end{cases}$$

where $f = (q, w_0, g)$ represents all "physical" source terms excepting the ones on the interface that are represented separately by λ. Define now the interface variable for the subproblem i at iteration k as

$$\lambda_i^k := \mathcal{B}_i w_i^k.$$

Note that the interface variable is, according to the Schwarz waveform relaxation approach, a time-space variable that lives on the interface Γ_i in space and on the entire time interval $(0,T)$. Owing to the transmission conditions of algorithm 5.1 on the interface, one obtains now the condensed interface relationship

$$\begin{aligned} \lambda_1^{k+1} &= \mathcal{B}_1\, \mathcal{M}_2(\lambda_2^k, f), \\ \lambda_2^{k+1} &= \mathcal{B}_2\, \mathcal{M}_1(\lambda_1^{k+1}, f), \end{aligned} \tag{5.5}$$

where the upper block lives on the interface Γ_1 and the lower block on the interface Γ_2. The alternating Schwarz waveform relaxation algorithm 5.1 is therefore a block-wise fixed point algorithm for the interface problem

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathcal{B}_1\, \mathcal{M}_2(\lambda_2, f) \\ \mathcal{B}_2\, \mathcal{M}_1(\lambda_1, f) \end{pmatrix}. \tag{5.6}$$

In the case of non overlapping subdomains and the use of Robin conditions, one can develop a simplified version of problem (5.5). The transmission operators write now

$$\mathcal{B}_i(w) = -\vec{f}(w)\vec{n_i} + pw,$$

with $\vec{n}_i$ the unit outward normal of Ω_i on Γ and $p \in \mathbb{R}$, $p > 0$ a constant and $\vec{f}$ the linear flux function associated to the linear transport operator $\mathcal{L}$. As both interfaces Γ_1 and Γ_2 coincide on one single interface Γ and the normals satisfy $\vec{n}_1 = -\vec{n}_2$, one can develop and simplify the interface relationships in the following way:

$$\lambda_i^{k+1} = \left(\vec{f}\vec{n}_i + p\right) w_i^{k+1} = \left(\vec{f}\vec{n}_i + p\right) w_{3-i}^k = -\left(\vec{f}\vec{n}_{3-i} + p\right) w_{3-i}^k + 2p w_{3-i}^k = -\lambda_{3-i}^k + 2p\, \mathcal{M}_{3-i}\left(\lambda_{3-i}^k, f\right).$$

Note that this statement is particularly interesting since we need no longer to reconstruct the flux information on the interface but only the trace information which is numerically much easier. The interface problem in this case writes now

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} -\lambda_2 + 2p\, \mathcal{M}_2\left(\lambda_2, f\right) \\ -\lambda_1 + 2p\, \mathcal{M}_1\left(\lambda_1, f\right) \end{pmatrix}.$$

5.3.1.2 Rediscovering the linear Case

In a linear context, suppose the operator $\mathcal{F}(\cdot)$ to be linear, the operator $\mathcal{M}_i$ is linear in both arguments. As we consider only linear transmission operators $\mathcal{B}_i$, one can rewrite problem (5.6) as a linear system

$$\begin{pmatrix} \mathrm{Id} & -\mathcal{B}_1\, \mathcal{M}_2(\cdot, 0) \\ -\mathcal{B}_2\, \mathcal{M}_1(\cdot, 0) & \mathrm{Id} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathcal{B}_1\, \mathcal{M}_2(0, f) \\ \mathcal{B}_2\, \mathcal{M}_1(0, f) \end{pmatrix},$$

where Id denotes the identity operator. Krylov subspace acceleration in this context is done by solving the linear interface problem by a Krylov type method like GMRES instead of using a splitting method like Jacobi or Gauß-Seidel. The resulting methods are called "Krylov-Schwarz" methods due to the combination of Krylov methods for linear systems and Schwarz methods for the domain decomposition approach. They can be applied to steady-state problems as well as to time-dependent problems. Brakkee and Wilders have studied in [9] the influence of interface conditions on the convergence of Krylov-Schwarz domain decomposition methods and showed that an application of a Krylov-type method on the interface problem has no overhead compared to a standard approach but accelerates significantly the convergence speed of the algorithm for all considered transmission condition types.

5.3.2 New Approaches

We extend the idea of a Krylov accelerator for the interface problem in the linear case to the nonlinear case in order to benefit from its accelerating properties. Similar approaches have been proposed for steady-state nonlinear problems. The class of Newton-Krylov-Schwarz methods (NKS) which is a combination of Newton-Krylov and Krylov-Schwarz methods is the most famous one. Those methods proceed the following strategy: If the nonlinear problem is time

dependent, it is discretised uniformly in time first and then one proceeds as for steady-state problems, i. e. the nonlinear problem is solved by a Newton method where the linear system at every iteration is solved by a Krylov-type method preconditioned by an algebraic Schwarz method (see [15] for the original paper of Cai et al.).
Here, we present a different strategy: We want to solve the original time-dependent nonlinear problem (5.4) using a Newton-type method to treat the nonlinearity and a Schwarz waveform relaxation method in order to benefit from the advantages of Schwarz type algorithms. Since we use Schwarz waveform relaxation methods for nonlinear time-dependent problems, we do not discretise the problem in time but keep the formulation continuous and global in time. There are now two key ideas. The first one is that the global and continuous formulation in time is also possible for a Newton algorithm. This means that there is nothing which would prevent us to state a global and continuous in time formulation of a Newton algorithm applied to problem (5.4). The second key idea is that there is *a priori* no given order to follow when one applies Schwarz waveform relaxation methods and Newton methods. We will develop therefore two new approaches, the first one will use first a Schwarz waveform relaxation algorithm, the resulting nonlinear interface problem (5.6) is then solved with a Newton-Krylov method. The second approach applies first a Newton method with a global and continuous formulation in time on the original problem (5.4). The resulting linear problem at every iteration of the Newton approach is then solved by a Krylov-Schwarz method as presented previously in the linear case.
The motivation of permuting domain decomposition methods and a Newton type method together with Krylov type methods is close to the works of Rey et al. given in [20] and [69] where nonlinearities are balanced upon subdomains by using the permutation of domain decomposition methods and Newton's method in combination with Krylov accelerators in the steady-state case.

5.3.2.1 Nested Iteration Approach

The first approach consists in treating system (5.6) by a Newton-Krylov approach. We seek the zeros of the nonlinear function

$$\Psi(\lambda) := \begin{pmatrix} \mathcal{B}_1\, \mathcal{M}_2(\lambda_2, f) - \lambda_1 \\ \mathcal{B}_2\, \mathcal{M}_1(\lambda_1, f) - \lambda_2 \end{pmatrix}.$$

One step $n \to n+1$ of Newton's method consists in solving the linear system

$$\Psi'(\lambda^n) \cdot (\lambda^{n+1} - \lambda^n) = -\Psi(\lambda^n),$$

where the Jacobian of Ψ is, due to the linearity of the operator $\mathcal{B}_i$, given by

$$\Psi'(\lambda) = \begin{pmatrix} -\operatorname{Id} & \mathcal{B}_1 \partial_\lambda\, \mathcal{M}_2(\lambda_2, f) \\ \mathcal{B}_2 \partial_\lambda\, \mathcal{M}_1(\lambda_1, f) & -\operatorname{Id} \end{pmatrix}.$$

Owing to the definition of a linearised operator

$$\mathcal{M}_1^{\text{lin}} : \ (A, h, f) \mapsto w_i$$

$$\text{solution of } \begin{cases} \phi\partial_t w_i + \mathcal{L}w_i + Aw_i &= q & \text{on } \Omega_i \times (0,T), \\ w_i(x,0) &= w_0(x) & \text{on } \Omega_i, \\ \mathcal{G}w_i &= g(x,t) & \text{on } (\partial\Omega_i \setminus \Gamma_i) \times (0,T), \\ \mathcal{B}_i w_i &= h & \text{on } \Gamma_i \times (0,T), \end{cases} \tag{5.7}$$

one can state the Jacobian

$$\Psi'(\lambda) = \begin{pmatrix} -\,\mathrm{Id} & \mathcal{B}_1\, \mathcal{M}_2^{\text{lin}}(\mathcal{F}'(\mathcal{M}_2(\lambda_2, f)), \cdot, 0) \\ \mathcal{B}_2\, \mathcal{M}_1^{\text{lin}}(\mathcal{F}'(\mathcal{M}_1(\lambda_1, f)), \cdot, 0) & -\,\mathrm{Id} \end{pmatrix}.$$

The entire procedure of the approach is then described by algorithm 5.2. The approach requires in every iteration of the outer loop (indices n) to set up a right hand side-vector that demands to solve two nonlinear problems in the subdomains. Therefore, a nested iterative procedure is necessary (Newton for instance), for this reason, we call this approach "Nested Iteration Approach" (NIA) due to the split iterative approaches of the nonlinear interface problem and the nonlinear subproblems. The name "Schwarz-Newton-Krylov" can be used in order to explain the order of application of the different tools: the global problem is first attacked by a Schwarz-type domain-decomposition method. The resulting nonlinear interface problem is attacked by a Newton-type method where, at every iteration, the resulting linear system is solved by a Krylov-type method. Unfortunately, the name "Newton-Krylov-Schwarz" has already been widely used for another type of methods and therefore "Schwarz-Newton-Krylov" may be confusing.

5.3.2.2 Common Iteration Approach

The second approach is not based on the nonlinear interface problem but attacks the global problem (5.4) up from the beginning. We apply Newton's method on this system and solve in every iteration $n \to n+1$ the linear system

$$\begin{aligned} (\phi\partial_t + \mathcal{L} + \mathcal{F}'(w^n))(w^{n+1} - w^n) &= -(\phi\partial_t w^n + \mathcal{L}w^n + \mathcal{F}(w^n) - q) && \text{on } \Omega \times (0,T), \\ w^{n+1}(x,0) &= w_0(x) && \text{on } \Omega, \\ \mathcal{G}(w^{n+1} - w^n) &= -\mathcal{G}w^n + g(x,t)) && \text{on } \partial\Omega \times (0,T), \end{aligned}$$

which can be reformulated to

$$\begin{aligned} (\partial_t\phi + \mathcal{L} + \mathcal{F}'(w^n))w^{n+1} &= \mathcal{F}'(w^n)w^n - \mathcal{F}(w^n) + q && \text{on } \Omega \times (0,T], \\ w^{n+1}(x, t=0) &= w_0(x) && \text{on } \Omega, \\ \mathcal{G}(w^{n+1}) &= g(x,t) && \text{on } \partial\Omega \times (0,T], \end{aligned}$$

owing to the linearity of the operators $\mathcal{L}$ and $\mathcal{G}$.

Algorithm 5.2 Nested Iteration Approach

Input: λ^0 (initial guess), ε (precision), $n_{\max}$ (maximum iterations)
Return: λ (solution)

$n = 0$
repeat
// Set up RHS
if $n = 0$ **then**
No previous iterate is available, use the previous time step as initial guess for the actual time step during evaluation of operators $\mathcal{M}_i$
else
Use the previous iterate globally in time as initial guess during the evaluation of operators $\mathcal{M}_i$
end if

$$-\Psi(\lambda^n) = \begin{pmatrix} \lambda_1^n - \mathcal{B}_1\, \mathcal{M}_2(\lambda_2^n, f) \\ \lambda_2^n - \mathcal{B}_2\, \mathcal{M}_1(\lambda_1^n, f) \end{pmatrix}$$

// Solve the linear system $\Psi' \delta_{\lambda^n} = -\Psi(\lambda^n)$ with a Krylov-type method
for every Krylov-iteration k **do**
realise a Matrix-vector multiplication by

$$\Psi' \delta_{\lambda^n}^k = \begin{pmatrix} -\delta_{\lambda_1^n}^k + \mathcal{B}_1\, \mathcal{M}_2^{\mathrm{lin}}(\mathcal{F}'(\mathcal{M}_2(\lambda_2^n, f)), \delta_{\lambda_2^n}^k, 0) \\ -\delta_{\lambda_2^n}^k + \mathcal{B}_2\, \mathcal{M}_1^{\mathrm{lin}}(\mathcal{F}'(\mathcal{M}_1(\lambda_1^n, f)), \delta_{\lambda_1^n}^k, 0) \end{pmatrix}$$

end for

// Update variables
$\lambda^{n+1} = \lambda^n + \delta_{\lambda^n}$
$n = n + 1$
until $n = n_{\max}$ **or** $\|\delta_{\lambda^n}\| < \varepsilon$ **or** $\|b\| < \epsilon$
$\lambda = \lambda^n$

We apply then the domain decomposition method on this linear system as we have done in section 5.3.1 and reduce the problem to the interface variables like in the linear case. Note that no further iteration index has to be introduced for the domain decomposition algorithm since no numerical method has yet been assigned to the resulting linear interface problem. The resulting linear problem for every iteration is then given by

$$\begin{pmatrix} \mathrm{Id} & -\mathcal{B}_1\, \mathcal{M}_2^{\mathrm{lin}}(\mathcal{F}'(w_2^n), \cdot, 0) \\ -\mathcal{B}_2\, \mathcal{M}_1^{\mathrm{lin}}(\mathcal{F}'(w_1^n), \cdot, 0) & \mathrm{Id} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1^{n+1} \\ \lambda_2^{n+1} \end{pmatrix} =$$

$$= \begin{pmatrix} \mathcal{B}_1\, \mathcal{M}_2^{\mathrm{lin}}(\mathcal{F}'(w_2^n), 0, (\mathcal{F}'(w_2^n) w_2^n - \mathcal{F}(w_2^n) + q, w_0, g)) \\ \mathcal{B}_2\, \mathcal{M}_1^{\mathrm{lin}}(\mathcal{F}'(w_1^n), 0, (\mathcal{F}'(w_1^n) w_1^n - \mathcal{F}(w_1^n) + q, w_0, g)) \end{pmatrix}.$$

Algorithm 5.3 Common Iteration Approach

Input: λ^0, (w_1^{-1}, w_2^{-1}) (initial guess), ε (precision), $n_{\max}$ (maximum iterations)
Return: λ (solution)

$n = 0$
repeat
// Update subdomain solutions
$w_1^n = \mathcal{M}_1^{\text{lin}}(\mathcal{F}'(w_1^{n-1}), 0, (\mathcal{F}'(w_1^{n-1})w_1^{n-1} - \mathcal{F}(w_1^{n-1}) + q, w_0, g)$
$w_2^n = \mathcal{M}_2^{\text{lin}}(\mathcal{F}'(w_2^{n-1}), 0, (\mathcal{F}'(w_2^{n-1})w_2^{n-1} - \mathcal{F}(w_2^{n-1}) + q, w_0, g)$

// Set up RHS
$$b = \begin{pmatrix} \mathcal{B}_1\, \mathcal{M}_2^{\text{lin}}(\mathcal{F}'(w_2^n), 0, (\mathcal{F}'(w_2^n)w_2^n - \mathcal{F}(w_2^n) + q, w_0, g)) \\ \mathcal{B}_2\, \mathcal{M}_1^{\text{lin}}(\mathcal{F}'(w_1^n), 0, (\mathcal{F}'(w_1^n)w_1^n - \mathcal{F}(w_1^n) + q, w_0, g)) \end{pmatrix}$$

// Solve the linear system $A\lambda^{n+1} = b$ with a Krylov-type method
for every Krylov-iteration k **do**
realise a Matrix-vector multiplication by
$$A\lambda^{n+1,k} = \begin{pmatrix} \lambda_1^{n+1,k} - \mathcal{B}_1\, \mathcal{M}_2^{\text{lin}}(\mathcal{F}'(w_2^n), \lambda_2^{n+1,k}, 0) \\ \lambda_2^{n+1,k} - \mathcal{B}_2\, \mathcal{M}_1^{\text{lin}}(\mathcal{F}'(w_1^n), \lambda_1^{n+1,k}, 0) \end{pmatrix}$$
end for

// Update variables
$\delta_{\lambda^n} = \lambda^{n+1} - \lambda^n$
$n = n + 1$
until $n = n_{\max}$ **or** $\|\delta_{\lambda^n}\| < \varepsilon$
$\lambda = \lambda^n$

Logically, the values for (w_1^n, w_2^n) that are needed to evaluate the right hand side and the matrix-vector multiplication at every iteration have to be provided by a nonlinear solution with operators $\mathcal{M}_i(\lambda_i^n, f)$. By giving an initial guess also for the first iterate (w_1^{-1}, w_2^{-1}), this procedure can be replaced by calculating the values only by the linearised operators

$$w_i^n = \mathcal{M}_i^{\text{lin}}(\mathcal{F}'(w_i^{n-1}), \lambda_i^n, (\mathcal{F}'(w_i^{n-1})w_i^{n-1} - \mathcal{F}(w_i^{n-1}) + q, w_0, g)),$$

because, suppose the solution has converged, the linearised operator gives the same solution as the nonlinear operator.

The entire procedure of the approach is then described by algorithm 5.3. The approach requires in every iteration of the outer loop (indices n) to set up a right hand side-vector that demands to solve two linear problems in the subdomains. Moreover, in the matrix-vector multiplication inside the Krylov-method, only linear problems in the subdomains are evaluated. No nested nonlinear iterative method is needed. For this reason and in contrast to the first approach, we call this approach "Common Iteration Approach" (CIA) due to the common iterative approach of

the nonlinear character of the monodomain problem. The name "Newton-Schwarz-Krylov" can be used in order to explain the order of application of the different numerical tools: the global problem is first attacked by a Newton-type method. At every iteration, the resulting linear problem is decomposed by a Schwarz-type algorithm where the problem is reduced to the interface variables. The resulting linear system is then solved by a Krylov-type method. As in the first case, we recommend not to use this name as the name "Newton-Krylov-Schwarz" has already been used for another type of methods.

5.3.2.3 Discussion on the New Approaches

Both methods make use of a Krylov-type method, GMRES for instance. In order both approaches to be competitive, we apply a precision strategy in the mood of inexact Newton methods, i. e. in the first iterations of the Newton method, we will not oversolve the linear system and can limit therefore the number of costly subdomain evaluations within a matrix-vector multiplication. The more we advance in the outer Newton iteration, the more precise the linear system has to be solved. A heuristic strategy for the precision of the solution of the appearing linear system at iteration n is solved up to a precision of

$$\max\left\{\min\left\{\frac{1}{1+n}, \|\Psi(\lambda^n)\|\right\}, \varepsilon\right\},$$

for the Nested Iteration Approach and

$$\max\left\{\alpha \cdot \min\left\{10^{-n}, \|\delta_{\lambda^n}\|\right\}, \varepsilon\right\},$$

for the Common Iteration Approach where ε is the user-supplied precision of the method and $\alpha > 0$ a real parameter. In practice $\alpha = 10^{-1}$ has shown to lead to a robust and performing strategy.

Note that the Common Iteration Approach needs to store the discretised values of a solution in both subdomains. This can be viewed as a huge drawback in high performance codes. The Nested Iteration Approach *a priori* does not suffer from this drawback. Anyway, it may be highly desirable to store the solution and use it as initial guess for the evaluation of the right hand side. In practice, using the solution of the previous iterate as initial guess reduces significantly the number of nested Newton iterations in the nonlinear subdomain solver.

Finally, concerning the stopping criterion of the outer Newton iteration, the Nested Iteration Approach can be classically controlled by both the residual and the step size norm. The Common Iteration Approach can though no longer be controlled by the residual norm since we have eliminated that term up from the beginning. Recalculating the residual term afterwards is possible but the cost for this may not be negligible since a nonlinear problem in global in time on every subdomain has to be calculated.

5.4 Numerical Results

We study now the numerical behaviour of the Schwarz waveform relaxation algorithm 5.1. In the first part, we concentrate on issues concerning the influence of the nonlinearity on the performance of the classical approach. In the second part, we compare the classical approach with the two new approaches presented in section 5.3.2.

5.4.1 Classical Approach

All numerical tests in this section have been performed using the classical approach as it is defined in algorithm 5.1.

5.4.1.1 Performance of Optimised Transmission Conditions

For the numerical results in this section we fix the time period $t \in [0, 1]$ and the global domain $\Omega = [0, 1] \times [0, 1] \in \mathbb{R}^2$. Discrete steps are $\Delta t = \Delta x = \Delta y = 2 \cdot 10^{-2}$. The diffusion parameter is $a = 1$, advection is $(b_x, b_y) = (1 \cdot 10^{-2}, 5 \cdot 10^{-2})$. Defining the function

$$f(x, y, t) = (\sin(\pi x)\cos(\pi y)\cos(\pi t) + \cos(\pi x)\sin(\pi y)\cos(\pi t) + \cos(\pi x)\cos(\pi y)\sin(\pi t) + 1)/2,$$

we can set the initial values to $u_0 = f(x, y, 0)$, $v_0 = f(x, y, 0)/c$ for $(x, y) \in \Omega$ and we impose Dirichlet boundary conditions with values set to $u_b(x, y, t) = f(x, y, t)$ for $(x, y) \in \partial\Omega$. The function f provides a heterogeneity in space and time so that we can ensure that the exact solution that we reconstruct numerically does not degenerate to a stationary problem.
We decompose Ω into non-overlapping subdomains $\Omega_1 = [0, 0.5] \times [0, 1]$ and $\Omega_2 = [0.5, 1] \times [0, 1]$. We impose a random initial guess on the interface Γ_1 in order to ensure the presence of a wide range of possible frequencies in the error.

As we have seen in the section 4.4.3, the optimised parameters in 2D for Robin transmission conditions do not show a significant sensitivity to the reaction coupling term k. As a consequence, one might think that, for the nonlinear case, one can use the optimised parameters of the single equation of advection-diffusion type for the nonlinear advection-diffusion-reaction system and the therefore obtained transmission condition behaves well. Therefore we study, as in the linear case, the convergence behaviour of the Schwarz waveform relaxation algorithm with Robin transmission conditions using different parameters p for the transmission condition. We proceed ten iterations of the algorithm and focus on the resulting error on the interface values which indicate us the numerical performance of the transmission condition with respect to the parameter p. We study different nonlinear coupling function. First, we consider an adsorption process that is modelled by a BET isotherm law:

$$\Psi(u) = \frac{Q_s K_L u}{(1 + K_L u - K_S u)(1 - K_S u)}.$$

BET theory is a rule for the physical adsorption of gas molecules on a solid surface and serves as the basis for an important analysis technique for the measurement of the specific surface area of a material (cf. Brunauer et al. [11]). This law is insofar mathematically interesting as it is neither convex nor concave (cf. figure 5.1). The coupling term is given by $R(u, v) = 100(v - \Psi(u))$ with

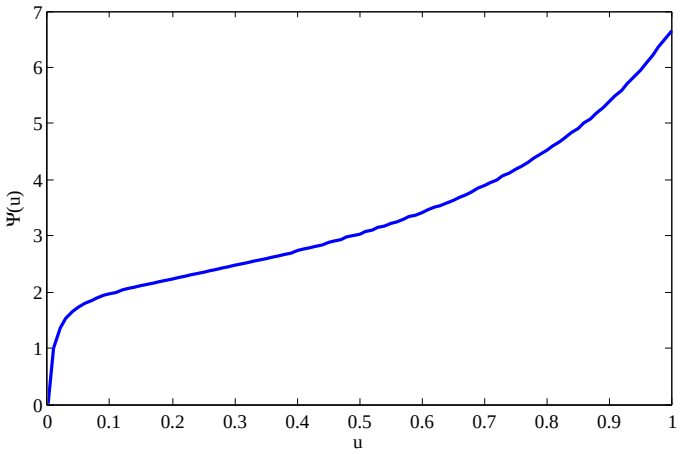


Figure 5.1: BET Isotherm with $Q_S = 2$, $K_S = 0.7$, $K_L = 100$

$Q_S = 2$, $K_S = 0.7$, $K_L = 100$. In figure 5.2 we plot the error at iteration 10 varying the parameter p of the Robin transmission condition. The square locates the numerically optimised parameter of the single equation of advection-diffusion type without reaction term.

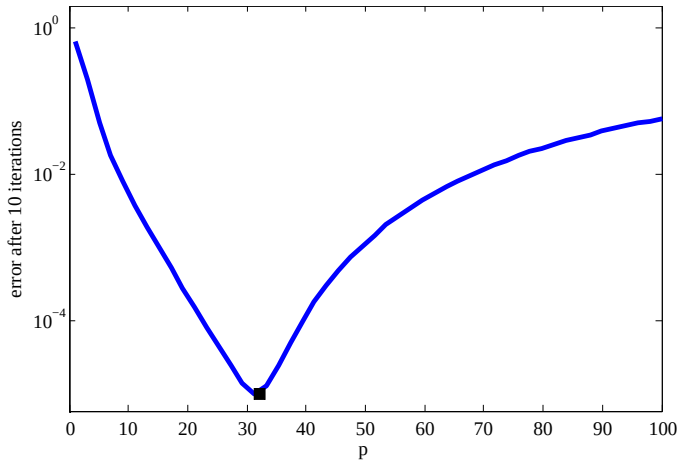


Figure 5.2: Variation of the parameter p of the Robin transmission condition versus the error of the 10th iterates. The asterisk locates the numerically optimised parameter of the advection-diffusion equation. Nonlinear function: BET isotherm law.

We study another nonlinear function that is given by an exponential equilibrium model

$$R(u, v) = \exp(10(2v - 3u)) - 1,$$

and plot in figure 5.3 again the error at the 10th iteration versus the parameter p of the Robin transmission condition.

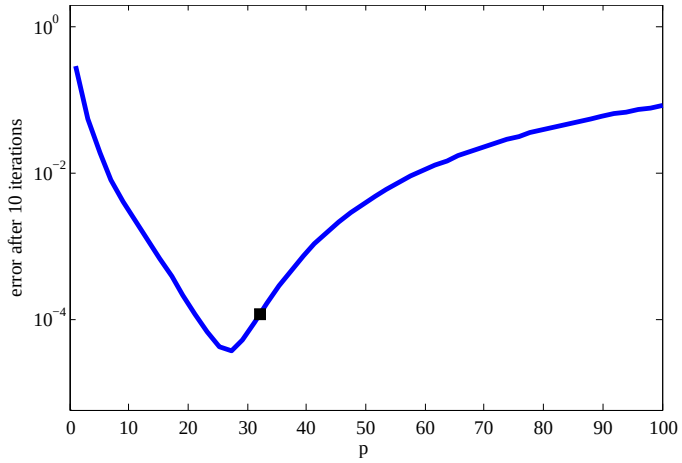


Figure 5.3: Variation of the parameter p of the Robin transmission condition versus the error of the 10th iterates. The asterisk locates the numerically optimised parameter of the advection-diffusion equation. Nonlinear function: exponential equilibrium model.

In both cases, we observe that the theoretically optimised parameter for the single equation of advection-diffusion type is close to the best performance for the nonlinear coupled advection-diffusion-reaction system. Tests with other nonlinear functions confirmed this behaviour.

5.4.1.2 Localising Time Step Constraints

In the linear case, domain decomposition methods are often used with many subdomains in order to distribute data and computational efforts to several processors. In the nonlinear context, domain decomposition methods can be used in order to localise time step constraints resulting from heterogeneity. In the context of CO_2 geological storage modelling, highly reactive moving fronts appear in the geochemical system. Those regions of strong chemical disequilibrium induce highly localised constraints on the time step using a global implicit approach. If the time step is chosen large, the number of Newton iterations for one time step increases drastically and if the time step is chosen too large, the standard Newton approach does no longer converge.
The Schwarz waveform relaxation approach provides the possibility to use different discretisations and numerical approaches in the subdomains. We exploit the possibility to choose different time grids in the subdomains. By this way, we can select a “reactive domain” with small time steps in order to keep the number of Newton iterations for the time steps small. In a “non reactive domain” we can chose much larger time steps and the number of Newton iterations stay acceptable. By this way we can limit the time step restrictions to the area where they appear instead of letting them influence the global time step of the whole simulation domain.

We exemplify this feature with the following test case: the time period is $[0, 1]$ and the global domain is $\Omega = [0, 1]\times[0, 1] \in \mathbb{R}^2$. Discrete steps are $\Delta x = \Delta y = 2 \cdot 10^{-2}$. The diffusion parameter is $\nu = 5 \cdot 10^{-2}$, advection is $(b_x,\ b_y) = (1.5,\ 1.0)$. The reaction term is realised by use of the BET isotherm. The initial values are set to $(u_0, v_0) = (0, 0)$ which is an equilibrium state. We model the entry of a reactive front by imposing a Dirichlet boundary condition on $x = 0$ with values $g(x = 0, y, t) = \sin(y\pi)$. All other boundaries are set to be of no diffusion type, i. e. we impose no concentration gradient on the boundary.
By the incoming reactive front, the chemical system is subject to a strong disequilibrium perturbation. In order the number of Newton iterations not to be too excessive (less than ten), one has to chose a time step of 10^{-1}, i. e. ten time steps. The global monodomain approach has a linear system of 5200 discrete unknowns to solve in every time step and every Newton iteration. The first time step needs 9 Newton iterations to reach convergence, the following time steps (2nd to 10th) need each 7 Newton iterations to reach convergence where we suppose the solution of the previous time step as initial guess of the Newton iteration. We measure the effort of the global approach by the effort of the inversion for one matrix multiplied by the number of matrix inversions since this is the most costly operation. The global effort is hence $(1\cdot9+9\cdot7)\cdot(5200)^3 = 1.01\cdot10^{13}$. In a domain decomposition approach, we can chose the reactive subdomain to be $\Omega_1 = [0,\ 0.4 + \Delta x]\times[0,\ 1]$ (2242 discrete unknowns) and the non reactive subdomain to be $\Omega_2 = [0.4,\ 1]\times[0,\ 1]$ (3160 unknowns discrete). Both subdomains have an overlap of one layer of cells. For the reactive subdomain we chose the same time step as in the global monodomain approach, i. e. ten time steps, while for the non reactive subdomain, we can use only one time step without the number of Newton iterations to become important. We impose the initial state as initial guess for the interface values in the Schwarz waveform relaxation iteration and impose the solution of the previous Schwarz waveform relaxation iteration in the Newton iterations. In the first iterate we proceed as in the global approach, i. e. we impose the solution of the previous time step as initial guess for the Newton iterates. In total, we need three Schwarz waveform relaxation iterations in order to reach convergence. In the first iteration, we need for the reactive subdomain 9 iterations in the first time step and 7 iterations for each following time step. In the non reactive subdomain we need 4 iterations. In the second Schwarz iteration, we need 2 iterations for the first 7 time steps and 3 iterations for the last 3 time steps in the reactive subdomain and 3 iterations in the non reactive subdomain. In the third iteration we need 2 iterations for all time steps in the reactive and non reactive subdomains. The total effort is hence $(1 \cdot 9 + 9 \cdot 7 + 17 \cdot 2 + 3 \cdot 3) \cdot (2242^3) + (4 + 3 + 2) \cdot (3160)^3 = 1.58 \cdot 10^{12}$. The effort for the domain decomposition solution is hence by a factor of 10 smaller than the effort for a global monodomain solution. Note that this estimation is quite pessimistic since using more performing linear solvers for the linear problems during the Newton iterations may reduce the gain of a domain decomposition approach compared to the global approach.

In figures 5.4 and 5.5 we plot the concentration of u and v at $t = 0.5$ and $t = 1.0$ at the third iteration of the domain decomposition algorithm.

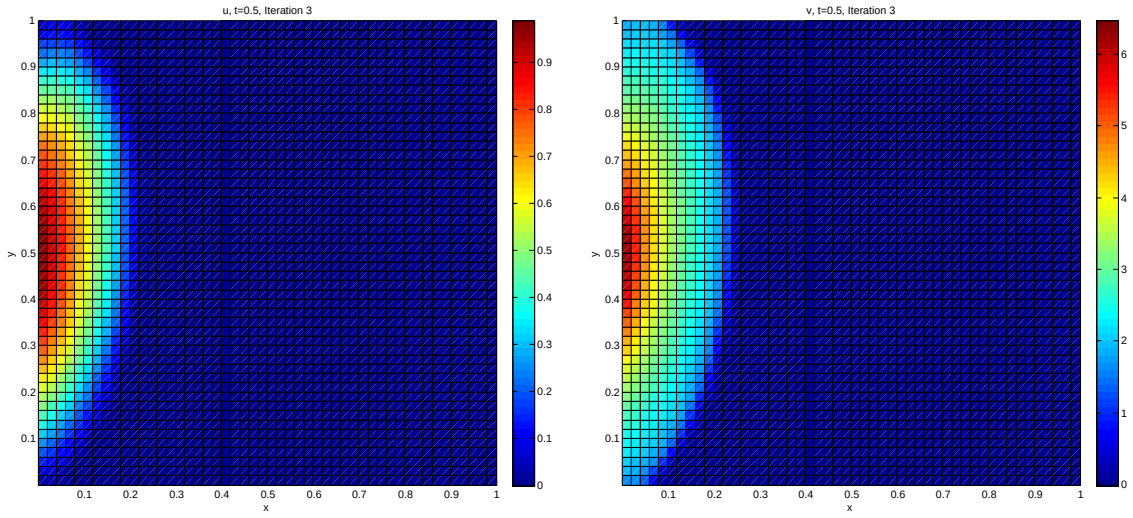


Figure 5.4: Concentration u and v at $t = 0.5$ using the BET isotherm with an incoming reactive front. Solution of the third domain decomposition iteration.

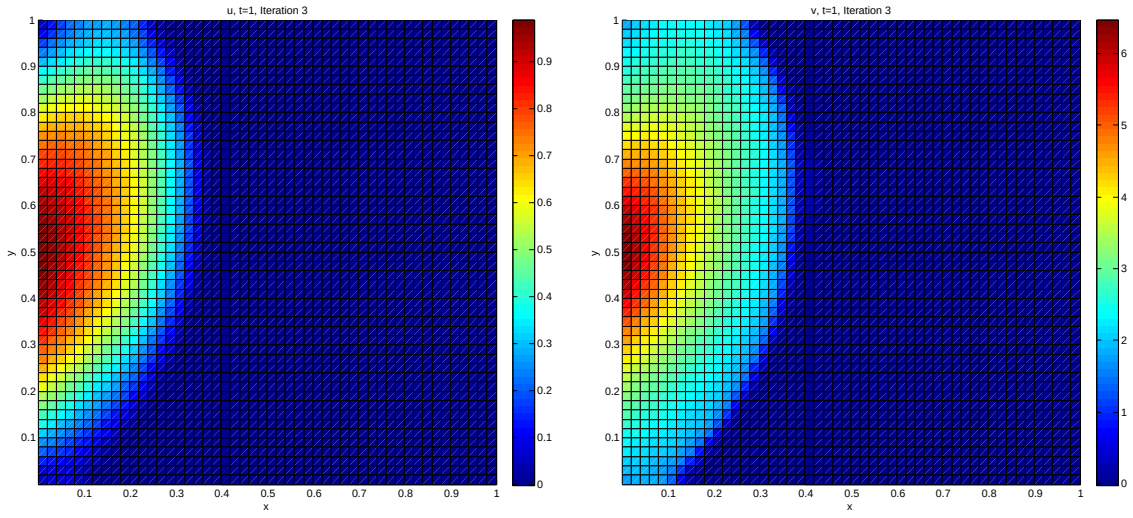


Figure 5.5: Concentration u and v at $t = 1.0$ using the BET isotherm with an incoming reactive front. Solution of the third domain decomposition iteration.

5.4.2 New Approaches

For the first tests in 2D, we set the simulation domain to $\Omega = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ with the subdomains $\Omega_1 = [0, 0.5] \times [0, 1]$ and $\Omega_2 = [0.5, 1] \times [0, 1]$. The considered time window is

$t \in [0, 1]$. Physical parameters are $\phi = 1$, $a = 1.5$, $(b_x, b_y) = (5 \cdot 10^{-2}, 1 \cdot 10^{-3})$. The nonlinear coupling term is defined by $R(u, v) = k(v - \Psi(u))$, where

$$\Psi(u) = \frac{Q_s K_L u}{(1 + K_L u - K_S u)(1 - K_S u)}$$

is the BET isotherm law. We set $k = 100$, $Q_S = 2$, $K_S = 0.7$ and $K_L = 100$. Initial values are set to $(u_0, v_0) = (\frac{1}{2}, \frac{1}{3})$. By defining the function

$$g(x, y, t) = (\sin(\pi x)\cos(\pi y)\cos(2\pi t) + \cos(\pi x)\sin(\pi y)\cos(2\pi t) + \cos(\pi x)\cos(\pi y)\sin(2\pi t) + 1)/2,$$

we impose Dirichlet boundary conditions with values set to $u_b(x, y, t) = g(x, y, t)$ for $(x, y) \in \partial\overline{\Omega}$.

5.4.2.1 Sensibility to the Parameter of the Robin Transmission Condition

In a first time, we are interested in the sensibility of the three approaches with respect to the parameter p of the Robin transmission condition. We discretise the numerical domain with $\Delta x = \Delta y = 1/40$ and $\Delta t = 1/10$ and impose a random initial guess on the interface for the first iteration. As both subdomains have the same size, the number of overall matrix inversions in the linear and nonlinear subdomain solvers for three approaches is a meaningful criterion to measure the numerical performance. We proceed the three approaches for different parameters p of the Robin transmission condition and plot in figure 5.6 the number of matrix inversions.

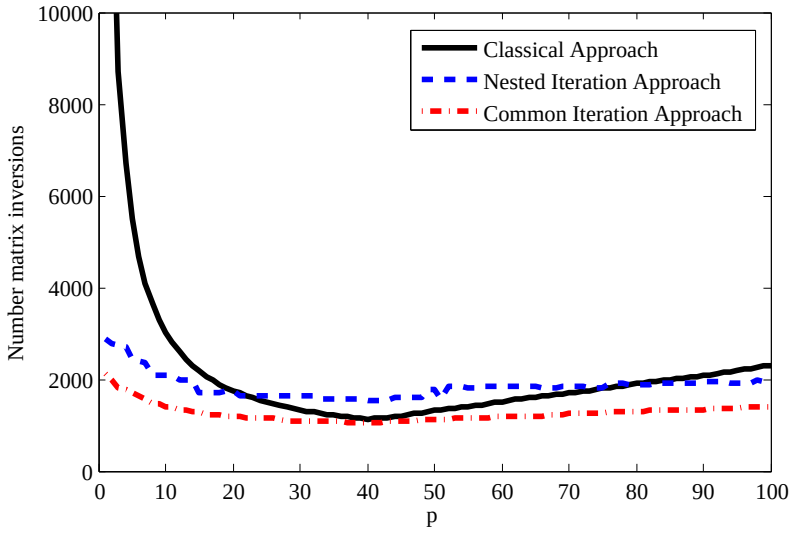


Figure 5.6: Number of matrix inversions versus parameter p of the Robin transmission conditions for the classical approach (fixed point on the nonlinear interface problem), Nested Iteration Approach and Common Iteration Approach

One observers first that the performance of the classical approach, i. e. a fixed point method on the nonlinear interface problem depends highly on the parameter p that one chooses for the Robin transmission condition. The best parameter is $p^* \approx 40$. The two new approaches, NIA and CIA also show the best performance at p^* but are much less sensitive to the choice of the parameter. Especially if, in realistic test cases where one has no idea of the best parameter, one underestimates the unknown parameter p^*, the new approaches do not suffer from the exponential loss of performance. Second, one observes that the CIA is always more performing than the classical and the NIA approach. The NIA is, in a wide range of parameters, more performing than the classical approach but the last one stays more performing in an important range of parameters around the optimal parameter p^*.

5.4.2.2 Performance within Mesh Refinement

It turned out that the two new methods, NIA and CIA, have a cost overhead that becomes non negligible if space discretisations are chosen too coarse. This overhead cost may be due to the fact that the subdomain evaluations in the classical and the new approaches are done with a different aim. While both classical and new approaches start from the same initial guess, the classical approach evaluates points that, during its iteration, lie more and more close to the exact solution. Therefore, the iterates move smoothly towards the physical solution and the guesses for the nonlinear subdomain solver may be very good, even if the convergence is slow. The new approaches use a Krylov method with tries in every iteration to minimise the residual of the linear system. Now, there are two points which are different: first, even if the global iterates of the two new methods are moving successively toward the physical solution, the inner iterates of the Krylov-type method do not, they follow only the aim to minimise the residual. The last evaluated point of the Krylov-type method is not the solution of the linear system but the point in the Krylov-subspace which minimises the residual. The physical meaning is lost. The second point lies in the Krylov-subspaces: at every iteration, they are reconstructed from scratch, all information on the Krylov-subspaces of the previous global iteration is lost. As a result, it might be interesting to develop a strategy where the information on the Krylov subspaces is kept.

Nevertheless, the cost overhead is only important, if the discretisations are chosen quite coarse. For this reason, we study the asymptotic behaviour within mesh refinement of the three approaches using for the new approaches always the optimal parameter of the classical approach. This is justified by the previous tests where all three approaches show their best performance at the same parameter. We refine the problem in space using always $\Delta x = \Delta y$. Note that we keep the time step constant at $\Delta t = 0.1$. Refining the discretisation also in time would lead to a problem that is quasi linear at every time step since we use a global implicit approach. The negligible nonlinearity would result in a minimal number of nested Newton iterations and the overhead cost would become more important. We measure again the overall number of matrix inversions in the three approaches. One observes that the overhead cost of the two new approaches compared to the classical approach becomes negligible up from a discretisation with about 150 grid points per dimension for the NIA and about 20 grid points per dimension for the SIA. For problems finer

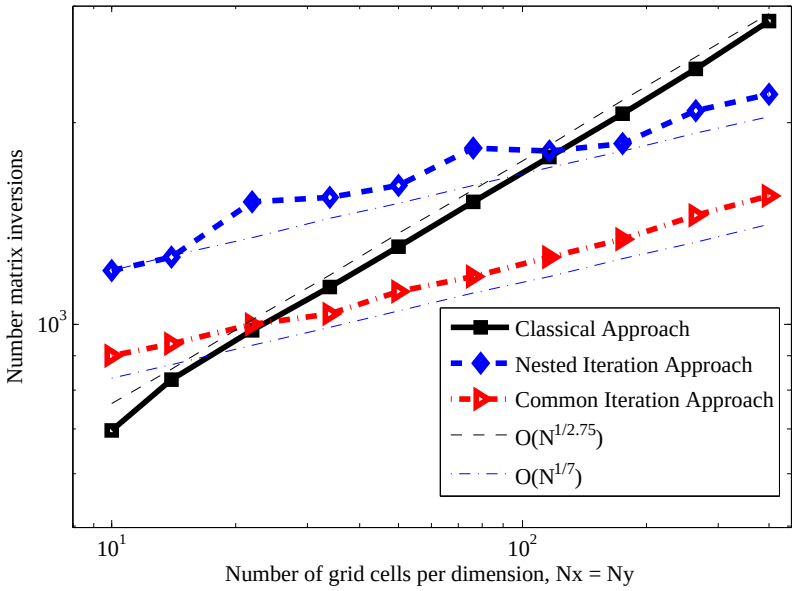


Figure 5.7: Number of matrix inversions versus number of discrete points per dimension ($N_x = N_y$) for the classical approach, Nested Iteration Approach and Common Iteration Approach

than the respective thresholds, the new approaches are always more performing compared to the classical approach with the best parameter for the transmission condition. Moreover, the finer the discretisation, the larger the problem, the more important the accelerating property of the two new approaches. Note that both new approaches have the same slope of $O(N^{1/7})$ in the asymptotic behaviour which is considerably less than the slope of the classical approach which behaves like $O(N^{1/2.75})$. The vertical translation of the curves for the two new approaches indicates their overhead cost.

5.4.2.3 Superlinear vs. Quadratic Convergence

In order to exemplify the accelerating property of the two new approaches, we perform a simulation with $N_x = N_y = 200$ points in each dimension keeping the number of time steps constant and compare the convergence behaviour of the stopping criteria of the three methods. In figure 5.8 we plot the convergence criterion versus the number of matrix inversions. Note that, for a better comparison, we set the residual norm of the nonlinear interface problem evaluated at the initial guess for all three methods at zero matrix inversions. One observes the quadratic convergence of the new approaches since they are Newton-based, the quadratic convergence is observed late in the history since the initial guess (randomly chosen) is far from the exact solution. The classical approach shows only a superlinear convergence, also in this case, the superlinear character appears late in the convergence history.

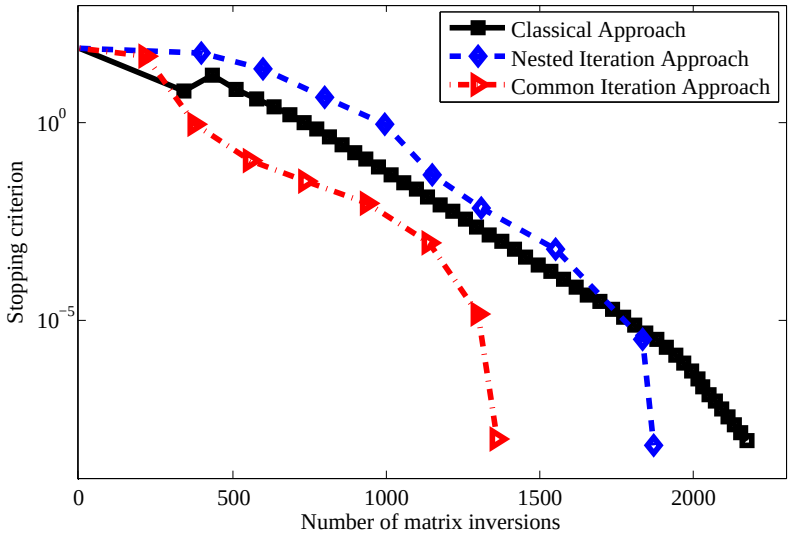


Figure 5.8: Convergence history with 200 points per space dimension for the classical approach, Nested Iteration Approach and Common Iteration Approach

5.4.2.4 Application to SHPCO2 Benchmark

In order to illustrate the performance of the new approaches, we compare the classical and new approaches on a benchmark test case in the context of CO_2 geological storage. The 3D test case is based on the benchmark for the SHPCO2 project (Simulation haute performance pour le stockage géologique du CO_2) which is described in [65] (see appendix A). The global domain is set to $\Omega = [0, 4750] \times [0, 3000] \times [-1100, -1000]$ with (38, 24, 8) grid cells in (x, y, z)-direction. The domain is decomposed into the two nonoverlapping subdomains $\Omega_1 = [1000, 2500] \times [0, 3000] \times [-1050, -1000]$ and $\Omega_2 = \Omega \setminus \Omega_1$. We call Ω_1 the reactive subdomain since in this subdomain an injection of the mobile species u is modelled by a source term. The initial state is zero for the mobile and immobile species. We consider again the BET isotherm law as nonlinear coupling term. The injected mobile species is partially absorbed by the reaction and partially transported by mainly advection.
Simulation time is [0, 100]. As the Schwarz waveform relaxation approach allows to use different discretisations in the subdomains, we chose to use ten time steps in the reactive subdomain Ω_1 and only five time steps in the subdomain Ω_2. This choice is insofar justified since the rapid injection in the reactive subdomain restricts the time step size by imposing a maximum number of Newton iterates of ten. As in the subdomain Ω_2, the mobile species appears only by transport processes on a slower time scale than the injection, one can chose a larger time step in order to respect the maximum number of Newton iterations.
Concerning the parameter of the Robin transmission condition, we use a low frequency approximation of the optimal parameter. The initial guess on the interface is zero for both subdomain interfaces.
In figure 5.9 we plot the convergence histogram, i. e. the stopping criterion in a logarithmic scale versus the CPU time (normalised to the CPU time of the classical approach). Note that, for a

better comparison, we set the residual norm of the nonlinear interface problem evaluated at the initial guess for all three methods at CPU-time zero. In this case both subdomains have a dif-

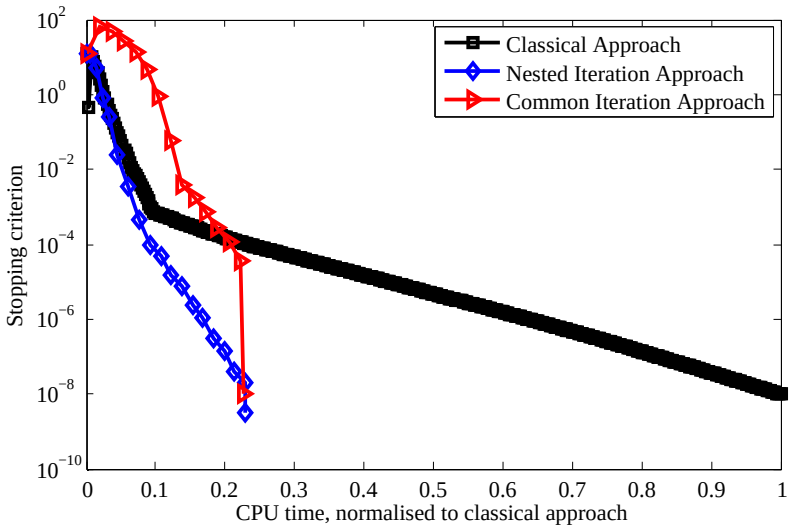


Figure 5.9: Convergence history of SHPCO2 benchmark case with (38, 24, 8) grid cells in (x, y, z)-direction, 10 time steps in Ω_1 and 5 time steps in Ω_2 for the classical approach, Nested Iteration Approach and Common Iteration Approach

ferent size of unknowns and therefore the number of matrix inversions, as used in the previous examples, is no longer a valid tool to measure the effort. One observes clearly that the classical approach converges only linearly while the two new approaches show a quadratic convergence. Moreover, the two new approaches need only about 20 percent of the CPU time of the classical approach.
Note that the stopping criterion of the Common Iteration Approach seems to crash down in the last iteration as one observes in figure 5.9. This behaviour is due to the fact that the GMRES solver provided in the last iteration the same solution as in the last but one iteration due to the precision strategy. As a consequence, the norm of the variation of the iterates is zero and this indicates that the algorithm has converged. To indicate this behaviour in the plot, we set the convergence criterion of the last iterate to the overall precision 10^{-8}.

5.4.2.5 Nested Iteration Approach vs. Common Iteration Approach

In the first test case using the BET isotherm one observes that the Common Iteration Approach is always more performing than the Nested Iteration Approach. However, in the SHPCO2 test case the opposite holds. Note that there is no *a priori* more performing method, all depends on the problem to treat. We performed several tests with different problems and we can state several points:
First, both new approaches, the Nested Iteration Approach and the Common Iteration Approach

show always much less sensibility to the parameter of the Robin transmission condition than the classical approach.
Second, both new approaches show always a more favourable slope in the case of asymptotic mesh refinement in space on the one hand and on the other hand they show always the same slope up to a vertical translation which indicates the different overhead cost.
Third, the question which of both new approaches is more performing on a problem is a challenge between the complexity to solve the problem with a domain decomposition approach and the difficulty to solve it with a nonlinear approach. Problems which are easy to solve with a domain decomposition approach since they are for example quite advective in normal interface direction and less diffusive have concentrated their main challenge in the nonlinearity. This is the case of the SHPCO2 test where the main difficulty lies in the massive injection of mobile species in the reactive domain which induces an elevated number of Newton iterations. The Nested Iteration approach showed to be more performing on this cases where the nonlinearity is the main challenge. On the other hand, for test cases which are difficult to solve with a domain decomposition approach, for example highly diffusive cases like the synthetic case in the previous part, the Common Iteration Approach seems to be more performing.
Altogether, we can summarise that both new approaches are more performing than the classical approach whenever the overhead cost due to coarse discretisations become negligible. Moreover, which one of the new approaches is more performing lies on the challenging character of the problem and can hardly be predicted in advance. Nevertheless, passing from one approach to another is very simple since they only inverse two loops.

Conclusion

In this chapter we have studied a Schwarz waveform relaxation algorithm applied to a coupled nonlinear reactive transport system. After introducing the algorithm and discussing several issues on higher order transmission conditions in the nonlinear case we concentrated on the numerical approach in the nonlinear case. We extended the classical idea of reducing the algorithm to its interface variables which is quite well-known in the linear case. By this technique we were able to propose two different approaches which allow the use of Krylov accelerators — a powerful tool to accelerate convergence speed without having to go the way of Fourier transform as we did in chapter 4 for the linear case. In the last part of this chapter, we could illustrate several issues on a numerical level: first, we have shown that the numerical approach of optimised transmission conditions still holds in the nonlinear case using the prediction techniques of the linear scalar case. Then, we have shown a very important issue which validates the main motivation of this thesis, namely the possibility to localise time step constraints in the nonlinear case using a Schwarz waveform relaxation algorithm. And finally, we showed the accelerating properties of the two new approaches for the interface problem.

There is no branch of mathematics, however abstract, which may not some day be applied to phenomena of the real world.

Nikolai Lobachevsky

6

Multispecies Nonlinear Reactive Transport and Optimised Schwarz Waveform Relaxation

Contents

Introduction

In this chapter, we join chapter 1 and chapter 2, i. e. we apply a Schwarz waveform relaxation algorithm on a multispecies reactive transport system. We use the experience gained throughout the prototyping phase, the techniques and methods are presented in chapter 3. The aim is to be able to simulate two different test cases in the context of CO_2 geological storage in a domain decomposition context and to use its local adaptivity in time.

We first recall the system used for the multispecies reactive transport problem and describe the numerical approach. Then, we state the domain decomposition approach and explain its realisation in the industrial development platform Arcane. Afterwards, we describe two different test cases, show numerical results and conclude.

6.1 Reactive Transport Problem

The reactive transport problem and its mathematical and numerical formulation used in this chapter is close to the one presented in chapter 1. Since chemical reactions with mineral species play an important role in the context of CO_2 geological storage, we extended the problem and used appropriate numerical approaches to treat the different behaviour of mineral reactions.

6.1.1 Problem Statement

In this chapter, we treat a multiphase, multispecies reactive transport system. The chemical species are associated to several phases:

- **Aqueous phase:** formed of the solvent H_2O and mobile dissolved species.
- **Sorbed phase:** formed of several fixed species.
- **Mineral phases:** every mineral phase contains only one fixed mineral and is pure.

The chemical species are denoted as follows:

- **Mobile species:** N_c primary species denoted c and N_x secondary species denoted x.
- **Sorbed species:** N_s primary species denoted s and N_y secondary species denoted y.
- **Mineral species:** N_q primary species denoted q and N_z secondary species denoted z.

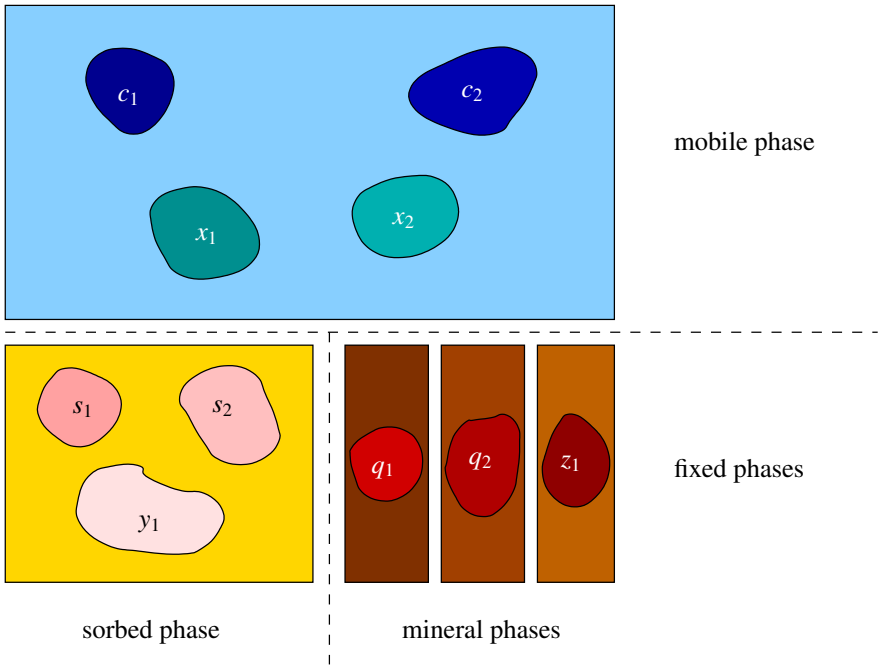


Figure 6.1: A multiphase chemical system with mobile and fixed phases.

We give a schematic example of the multiphase chemical system in figure 6.1.

We identify the chemical concentrations with their name. The concentration of chemical species is given with respect to its phase where we suppose that all mobile species form a phase, all sorbed species form a phase and every mineral is pure in its own phase. The associated amounts of the phases are noted θ_w for the mobile phase, θ_s for the sorbed phase, and θ_q, θ_w for the mineral phases.
We supposed ideal activities, i. e. the chemical activity for a species is equal to its concentration. The activity of a mineral species is always equal to one since they are pure in their phase and hence their concentration is always equal to one. Finally, we normalise the activity of mobile species with respect to the activity of water.
The chemical species are related by each other with kinetic reactions and by equilibrium reactions. The equilibrium system is described by the following matrix-valued Morel tableau

$$\begin{array}{c|ccc} & c & s & q \\ \hline c & \mathbf{Id} & & \\ s & & \mathbf{Id} & \\ q & & & \mathbf{Id} \\ \hdashline x & S_c^x & & \\ y & S_c^y & S_s^y & \\ z & S_c^z & & \end{array}, \tag{6.1}$$

where S_a^b are stoichiometric matrices, **Id** is the identity matrix, missing entries represent zero matrices. This system is an extension of the system presented in [5] in the sense that equilibrium reactions with minerals are taken into account. We do the following assumptions concerning the chemical equilibrium system:

- Mobile secondary species are formed only of mobile primary species.
- A mineral species is either in equilibrium with primary mobile species or it is considered as a primary mineral species which is not in equilibrium.

The second assumption ensures that there is at least a mobile species which interacts with the mineral species in an equilibrium reaction; a reaction between pure solid and mineral species is not allowed to be in equilibrium. The assumption is also used in order not to complicate the limitation of the validity of equilibrium conditions in the case when the mineral species is not present. As a consequence, the validity of an equilibrium condition is determined by the presence or absence of not more than one mineral species.
Besides the equilibrium reactions, we consider kinetic reactions. Kinetic reactions can react between mobile, sorbed and kinetic species (primary and secondary). We do the following assumptions concerning the kinetic reactions:

- Every mineral species (primary or secondary) appears in not more than one kinetic reaction as educt.
- If a mineral species appears in the same time as a product and as an educt, we suppose the reaction mechanisms to be decoupled.

The first assumption is the equivalent of the second assumption for equilibrium reactions and is due to the representation of the limitation of reaction rates which we explain later. The second assumption is to simplify the decoupling of the limitation of kinetic reaction rate.

The equilibrium state gives rise to the following equations which have to be satisfied. The first system describes the mass conservation of the total component amounts T, W, Q of the mobile, sorbed and mineral components:

$$\begin{aligned} \theta_w c + \theta_w (S_c^x)^t x + \theta_s (S_c^y)^t y + (S_c^z)^t (\theta_z \cdot z) &= T, \\ \theta_s s \qquad\qquad + \theta_s (S_s^y)^t y \qquad\qquad &= W, \\ \theta_q \cdot q \qquad\qquad\qquad\qquad\qquad &= Q. \end{aligned} \tag{6.2}$$

The second system describes the mass action laws of the equilibrium reactions where K_x, K_z, K_z are the equilibrium constants and a_i is the activity of species i:

$$\begin{aligned} \ln(a_x) - (S_c^x) \ln(a_c) \qquad\qquad &= \ln K_x, \\ \ln(a_y) - (S_c^y) \ln(a_c) - (S_s^y) \ln(a_s) &= \ln K_y, \\ \ln(a_z) - (S_c^z) \ln(a_c) \qquad\qquad &= \ln K_z. \end{aligned} \tag{6.3}$$

An equation of the last set of equilibrium conditions has only to be verified if the associated mineral phase is indeed present.

Finally, we impose a closure equation for the concentrations of every phase:

$$\begin{aligned}
\sum_{i=1}^{N_c} c_i + \sum_{i=1}^{N_x} x_i = 1, \quad & \text{(mobile phase)}, \\
\sum_{i=1}^{N_s} s_i + \sum_{i=1}^{N_y} y_i = 1, \quad & \text{(sorbed phase, if present)}, \\
q_i = 1, \quad & i = 1, \dots, N_q, \\
z_i = 1, \quad & i = 1, \dots, N_z.
\end{aligned} \tag{6.4}$$

The equations (6.2), (6.3) and (6.4) form together the chemical flash which has to be satisfied locally.

We state now the numerical formulation of the multispecies reactive transport problem as it is implemented. We consider the following extension of system (1.3):

$$\begin{aligned}
\partial_t(\phi C) + \partial_t(\phi F) + \mathcal{L}(C) + R_{T,\text{kin}} = 0 \qquad & (\#C), \\
\partial_t W + R_{W,\text{kin}} = 0 \qquad & (\#W), \\
\partial_t Q + R_{Q,\text{kin}} = 0 \qquad & (\#Q), \\
T - \phi C - \phi F = 0 \qquad & (\#T), \\
\phi F - \Psi(T, W, Q) = 0 \qquad & (\#F), \\
R_{T,\text{kin}} - \Theta(T, W, Q) = 0 \qquad & (\#R_{T,\text{kin}}), \\
R_{W,\text{kin}} - \Upsilon(T, W, Q) = 0 \qquad & (\#R_{W,\text{kin}}), \\
R_{Q,\text{kin}} - \Xi(T, W, Q) = 0 \qquad & (\#R_{Q,\text{kin}}).
\end{aligned} \tag{6.5}$$

For the sake of readability, we omit initial conditions and boundary conditions for physical boundaries.
The difference between system (1.3) and system (6.5) lies in the additional unknown Q which describes the total amount of mineral components. As a consequence, the chemical flash and hence the associated operator Ψ which describes the equilibrium state depend on all total components, mobile, sorbed and mineral. In the same way, the kinetic reaction rates $R_{T,\text{kin}}$, $R_{W,\text{kin}}$, $R_{Q,\text{kin}}$ depend on all total components. Another difference is, that we let appear correctly the porosity that we have neglected for the sake of simplicity in chapter 1.

6.1.2 Numerical Methods

System (6.5) is a nonlinear system of partial differential equations. We solve it by using a global implicit approach as it is described in section 1.3.2. The resulting nonlinear problem at every

time step is solved with a Newton method without linesearch. We apply an adaptive time stepping strategy that proceeds as follows:

- If the number of Newton iterations to reach convergence is lower than $N_{\min}$ (3 for instance), the actual integration step is accepted and the following time step is increased.
- If the number of Newton iterations to reach convergence is higher than $N_{\lim}$ (7 for instance) but lower than $N_{\max}$ (12 for instance), the actual time integration is accepted and the following time step is decreased.
- If the number of Newton iterations reaches $N_{\max}$ (12 for instance) and no convergence has been archived, then the actual time integration is rejected and repeated with a decreased time step.
- In all other cases (number of Newton iterations between $N_{\min}$ and $N_{\lim}$), the integration step is accepted and the following time integration uses the same time step.

The parameters $N_{\min}$, $N_{\lim}$, $N_{\max}$ as well as the increase and decrease multipliers are user defined parameters.
For solving the chemical flash, we use an extended version of the globalised Newton algorithm presented in [40] which now detects in some cases also non feasible total component concentrations: in this case the problem has no solution. Moreover, it has been made more robust by trying different initial guesses in the case of non convergence.
Both Newton approaches for the global problem and the local chemical problem are classically controlled by the residual norm in order to ensure convergence.
The linear systems appearing in the Newton iterations of the global system are preconditioned and solved by different iterative (or exact) solvers provided by the PetsC library (cf. [2]) or the hypre library (cf. [1]). The local linear systems appearing in the flash calculations are solved exactly with a LU-decomposition.
The spatial discretisation of the linear transport operator $\mathcal{L}$ is done using the hybrid finite volume scheme presented in chapter 3.2.2. The Darcy field is calculated globally by using standard two-point finite volume scheme as it is presented in chapter 3.2.1.

In order to prevent from negative concentrations of mineral species, we limit kinetic reaction rates where mineral species appear as educts by semismooth functions which are equivalent to complementary conditions. This technique has been proposed by Kräutle [52] and succesfully applied numerically by Hoffmann [48].
The underlying basic idea is the following: suppose a dissolution reaction M$\rightarrow$ A, where M is a mineral species and A is an aqueous species. The reaction speed $k \geq 0$ is independent of the concentration of M. Nevertheless, the reaction stops in a discontinuous way (i. e. without slowing down at the end if retarding factors like the reactive surface are neglected) when no more mineral is present. If no special care is taken in this situation, one would obtain a negative component amount when solving the equation $\partial_t \mathrm{M} + k = 0$. Classical approaches will cut the time

step to prevent from negative concentrations or use other techniques to detect the zero-crossing. Nevertheless, they are all quite heuristic and will not always lead to correct solutions. By using a semismooth function, for instance the minimum function, one can limit the reaction rate such that negative concentrations do not appear. The modified equation to solve is then

$$\frac{\mathrm{M}^{n+1} - \mathrm{M}^n}{\Delta t} + \min\left\{k, \frac{\mathrm{M}^n}{\Delta t}\right\} = 0. \tag{6.6}$$

As a consequence, if enough amount of M at the time t^n is available, the reaction rate k is not limited since the minimum is attained in the first amount. In the contrary case, the reaction rate is limited such that the amount of M at time t^{n+1} is zero. This interpretation of the complementary condition is different and extends the one presented by Hoffmann in [48] for kinetic reaction rates in that way that a sense is given to the limiting effect of the complementary condition.
Note that formulation (6.6) is a special reformulation of the following original logical condition

$$\Big((\partial_t \mathrm{M} + k = 0) \ \wedge \ (\mathrm{M} \geq 0)\Big) \vee \Big((\partial_t \mathrm{M} + k > 0) \ \wedge \ (\mathrm{M} = 0)\Big),$$

or of the following equivalent complementary condition

$$\Big((\partial_t \mathrm{M} + k) \cdot \mathrm{M} = 0\Big) \ \wedge \ \Big((\partial_t \mathrm{M} + k) \geq 0\Big) \ \wedge \ \Big(\mathrm{M} \geq 0\Big).$$

Both logical and complementary conditions can be reformulated in an equivalent way as the following algebraic condition

$$\min\left\{\frac{\mathrm{M}^{n+1} - \mathrm{M}^n}{\Delta t} + k, M^{n+1}\right\} = 0.$$

The interpretation of the limiting condition is the following: either the balance equation for M is verified and the amount of M^{n+1} is nonnegative (first case). Or the balance equation for M is not verified but the amount of M is zero (second case).

The formulation of complementary conditions with the help of minimum functions is insofar comfortable in the balance equations as the pure reaction rate k can be limited when calculating the reaction rate itself due to the additivity of the minimum function which lead to formulation (6.6).
In the case of equilibrium conditions for reactions with mineral species during the flash calculation, we use a different strategy: suppose $Q \geq 0$ to be the equilibrium condition for a reaction including a mineral species. The equilibrium condition is satisfied if $Q = 0$. Suppose that there is only one mineral species taking part in the equilibrium reaction. The amount of the mineral species is denoted by M. Now, we obtain the same complementary condition as before:

$$(Q \cdot \mathrm{M} = 0) \ \wedge \ (Q \geq 0) \ \wedge \ (\mathrm{M} \geq 0).$$

This complementary condition can be reformulated either by the algebraic formulation using the minimum function or using another function $\varphi(a, b)$ that has the property $\varphi(a, b) = 0 \Leftrightarrow$

$ab = 0 \wedge a \geq 0 \wedge b \geq 0$. A typical representative of φ besides the minimum function is the Fischer-Burmeister function (cf. [28])

$$\varphi_{\mathrm{FB}}(a, b) = a + b - \sqrt{a^2 + b^2}.$$

The advantage of the Fischer-Burmeister function in the context of the equilibrium condition calculation during the chemical flash is that its derivative contains the information of both a and b while the minimum function contains only information of the derivative of the case in which the minimum is actually attained. As in a dynamic context, one does not necessarily know in which case the guess for the solution in the following time step will lie, the numerical approach using the minimum function with a Newton method is problematic, singular Jacobian matrices appear and the standard method cannot be continued. This behaviour with the minimum can be explained in the following way: the equilibrium condition Q does not depend on the concentration of M. Hence, its derivative with respect to M vanishes. One can calculate the derivative of the minimum and the Fischer-Burmeister function in the four cases that may appear:

$\partial_M \min(Q, M)$

	$Q = 0$	$Q > 0$
$M = 0$	$\in [0, 1]$	1
$M > 0$	0	$\begin{cases} 1 & \text{if } M < Q, \\ 0 & \text{if } Q < M \end{cases}$

$\partial_M \varphi_{\mathrm{FB}}(Q, M)$

	$Q = 0$	$Q > 0$
$M = 0$	undefined	0
$M > 0$	0	$1 - \dfrac{M}{\sqrt{Q^2 + M^2}} > 0$

The problematic case is when the guess of the solution satisfies $Q > 0$, $M > 0$ and $Q < M$. Then the derivative of the minimum function is singular and cannot guide the Newton approach to the right case. By using the Fischer-Burmeister function, the derivative always includes information of both arguments and therefore, a Newton approach has a much higher chance to converge to the right case than in the minimum case.

By using the two algebraic versions of the complementary conditions, one can simply apply a Newton approach to the nonlinear problems. The method is called semismooth Newton approach since the derivatives of the nonlinear functions to solve are not smooth while the functions themselves are.

6.2 Domain Decomposition Approach

We now describe the domain decomposition approach used for the multispecies reactive transport problem (6.5).

6.2.1 Algorithm Statement

We apply a Schwarz waveform relaxation algorithm with two possibly overlapping subdomains and the windowing technique as it is presented in section 2.2. The global domain Ω is split into a non reactive domain Ω_N and a reactive domain Ω_R such that $\Omega = \Omega_N \cup \Omega_R$. The time integration period $[t_0, t_F]$ is split into n time-windows $[t_0, t_1]$, $[t_1, t_2]$, $\ldots$, $[t_{n-1}, t_F]$. The entire procedure is given in algorithm 6.1. The algorithm iterates over all time windows. The initial condition in the first iteration ($w = 0$) is given by the physical initial condition. For the following time windows, it is given by the solution of the previous time windows. Note that the previous time window furnishes a converged solution in the domain decomposition sense. Hence, the initial condition for the actual time windows is consistent and has a physical sense.
As initial guess for the SWR iteration, we set actually a constant value all over the time window that is equal to the reconstructed value of the initial state. This is a quite basic but useful information. In the general case, a richer information could be extrapolated values from the previous time window (when long time windows are used) or values obtained by a global coarse grid solution (when many subdomains are used). In our case with short time windows (for superlinear convergence) and only two subdomains, this seems to be a reasonable guess. The guess contains therefore only a small amplitude for low frequencies. Important amplitudes in high frequencies of the error due to the short time windows are quickly attenuated by the overlap that is used in practice.
As transmission conditions, we allow Dirichlet, Neumann and Robin transmission condition in the case of overlapping subdomains and Robin conditions in the case of non overlapping subdomains.

6.2.2 Numerical Realisation

The numerical realisation of algorithm 6.1 has been done on the platform Arcane (cf. [39]). Arcane is a software development platform originally developed by CEA-DAM (Commissariat à l'Énergie Atomique - Direction des Applications Militaires, Atomic Energie Commissariat - Military Applications Direction). Development is now done in a joint venture between IFP Energies nouvelles and CEA. Arcane is used at IFP Energies nouvelles as the numerical kernel for developing the next generation of industrial software in geoscience. The presented implementation is included in the Coores-Arcane project for CO_2 geological storage modelling.
Arcane is a C++ object oriented toolkit to develop high performance parallel simulators. It takes into account all architectural aspects for finite volumes/elements, e. g. grid management, time step management, parallelism, input data. Arcane has a modular structure so that different applications (modules) can use the same basic algorithms (services). Therefore, standard interfaces generalise the form of services and modules in order to ensure re-usability to be on a high level.

Arcane is restricted to codes that ensure the following two properties:

1. Execution advancement can be seen as an iteration of a sequence of codes.

Algorithm 6.1 Alternating Schwarz waveform relaxation method for the multispecies reactive transport system (6.5)

for $w = 0$ to n **do**

// treat time window $[t_w,\ t_{w+1}]$

$s = -1$ *// Index over SWR iterations*

$g_{\mathrm{N}}^{s} = \mathcal{B}_{\mathrm{N}} C(x, t_w)$ *// Set initial guess based on initial condition of time window*

repeat

// Do one SWR iteration

$s = s + 1$

// Non reactive domain

Solve problem (6.5) on Ω_{N}, with initial condition given by $c(x, t_w)$ with given physical boundary conditions and interface condition given by $\mathcal{B}_{\mathrm{N}} C_{\mathrm{N}}^{s} = g_{\mathrm{N}}^{s-1}$ on Γ_{N} transferred to the concerned time grid. Reconstruct $g_{\mathrm{R}}^{s} = \mathcal{B}_{\mathrm{R}} C_{\mathrm{N}}^{s}$ on Γ_{R}.

// Reactive domain

Solve problem (6.5) on Ω_{R}, with initial condition given by $c(x, t_w)$ with given physical boundary conditions and interface condition given by $\mathcal{B}_{\mathrm{R}} C_{\mathrm{N}}^{s} = g_{\mathrm{R}}^{s}$ on Γ_{R} transferred to the concerned time grid. Reconstruct $g_{\mathrm{N}}^{s} = \mathcal{B}_{\mathrm{N}} C_{\mathrm{R}}^{s}$ on Γ_{N}.

Set the actual solution $c^{s}(x,t) = \begin{cases} c_{\mathrm{N}}(x,t) & \text{for } x \in \Omega_{\mathrm{N}} \setminus \Omega_{\mathrm{R}} \\ c_{\mathrm{R}}(x,t) & \text{for } x \in \Omega_{\mathrm{R}} \end{cases}$, for $t \in [t_w,\ t_{w+1}]$.

// Check for convergence

if $s = 0$ **then**

cv is **false** *// Do at least two iterations*

else

if $\|c^{s}(x, t_{w+1}) - c^{s-1}(x, t_{w+1})\| < tol_s$ **then**

cv is **true**

else

cv is **false**

end if

end if

until cv is **true**

Set the global solution to $c(x,t) = c^{s}(x,t)$, for $t \in [t_w,\ t_{w+1}]$.

end for

2. Calculation domain is discretised in space over a collection of items. This collection is called grid and can be one-, two- or three-dimensional. Items of the mesh are nodes, edges, faces, cells. All manipulated values are based on one of those items.

Moreover, Arcane has its own terms of language for operations:

- A general operation/function is called entry point.
- The description of a sequence of operations is called time loop.
- Manipulated values are called variables.
- The collection of several entry points acting on the same variables is called module.
- Functions that provide tasks that are independent of the special application are ensured by services.

In our case, the code which solves a multispecies reactive transport problem on a given geometrical domain on a time window is implemented as a numerical model, which can be compared to a service. It has several entry points, the calculation of one given time step forms one entry point for instance. The variables of this numerical model are for instance the total concentrations amounts of all primary species.
The application which solves a time and space dependent problem described by partial differential equations with the help of a Schwarz waveform relaxation algorithm is implemented as module. Its time loop consists in initialising the service instances of the problem, realising a Schwarz waveform relaxation algorithm on a time window including the coupling between the interface variables and managing the solution on a global point of view.
The modular implementation of tasks under Arcane is useful in the sense that it can easily be reused by other algorithms. On a basic level, for instance, we created a service that calculates the transmissivity coefficients of the hybrid finite volume scheme presented in section 3.2.2. This service can now be used for other problems (e. g. for calculating a monophasic Darcy field) without modification in the code by a simple change in the input data file. On a higher level, the service for solving a reactive transport problem on a time window could be used to implement a parareal algorithm of Schwarz type for example.

In order to realise the Schwarz waveform relaxation algorithm with time windows on Arcane, we use a modular structure. The basic element is called a *Sequence* which explains the actions in an iterative algorithm to reach a final state up from an initial state. A sequence is divided into several categories:

- *Start*: all actions to initialise once the sequence.
- *Compute*: all actions to do an iterative step, the actions are divided in three categories:

  - *StartCompute*: all actions to initialise an iterative step.
  - *BaseCompute*: all actions to do the calculation of an iterative step itself.
  - *FinalizeCompute*: all actions to finish an iterative step. These actions are only executed if the *BaseCompute* sequence reports as successful execution.
- *Finalize*: all actions to finish once the sequence.

In order to do the calculations for a time windows $[t_w, t_{w+1}]$, first all actions of the *Start* category are executed, then as much as iterations on the *Compute* category are done as necessary to reach the final state of the time window, finally, the *Finalize* actions are processed.
An important tool in Arcane to manage the actions of the different categories is the concept of a *Collector*. It manages the actions to execute in the different moments of the sequence. Besides the actions which belong to the own service or module, it accepts also actions from other services or modules. This feature leads to a strict separation of actions to different implementations. We explain this feature with an example: the service for calculating a reactive transport problem should update its boundary condition values as an action in the *StartCompute* moment, this action is purely related to the reactive transport calculation service. Nevertheless, in a domain decomposition context, the Schwarz waveform relaxation module should transfer the interface condition values from one time grid to another in the *StartCompute* moment. This part of the code belongs purely to the coupling part and a reactive transport code should not know that he lives in a domain decomposition context. This emphasises the importance of a *Collector* to be able to execute actions from different services and modules.
In figure 6.2 we present a scheme of the collector for the actions to proceed and their affiliation to the modules/services in order to do the calculations for a time window $[t_w, t_{w+s}]$. In the *Start*, the SWR module sets first the initial state on which the reactive transport service can provide values of the trace and the flux on the interface Γ_b. Note that reconstruction action must belong to the reactive transport module since it is the only one to know which flux scheme is used for instance. Basing on this values, the SWR module can form the interface condition values for the complementary subdomain, this action is a pure SWR action since the reactive transport service does not need to know which coupling condition is used. Afterwards, the reactive transport service can proceed to its ordinary actions in the *Start* collection.
In the *StartInit* collection, i. e. in preparation of a time step, the SWR module has to project the transmission condition values from the time grid that has been used in the complementary subdomain at the previous iteration to the actual time t_a as it has been proposed by the adaptive time stepping strategy. The reactive transport module can then proceed its *StartCompute* and *BaseCompute* actions. If the *BaseCompute* was successful (i. e. the time step is accepted), then the reactive transport service can reconstruct trace and flux values which are needed by the SWR module to reconstruct the transmission condition values. The reactive transport service proceeds afterwards its *FinalizeCompute* actions. The *Compute* actions are repeated until the time window has been calculated.
Last but not least, in the *Finalize* collection, the SWR module sets the actual solution on Γ_a, and frees the memory of the no longer needed values on Ω_a saved in the previous Schwarz iteration. The reactive transport service proceeds then its *Finalize* actions.

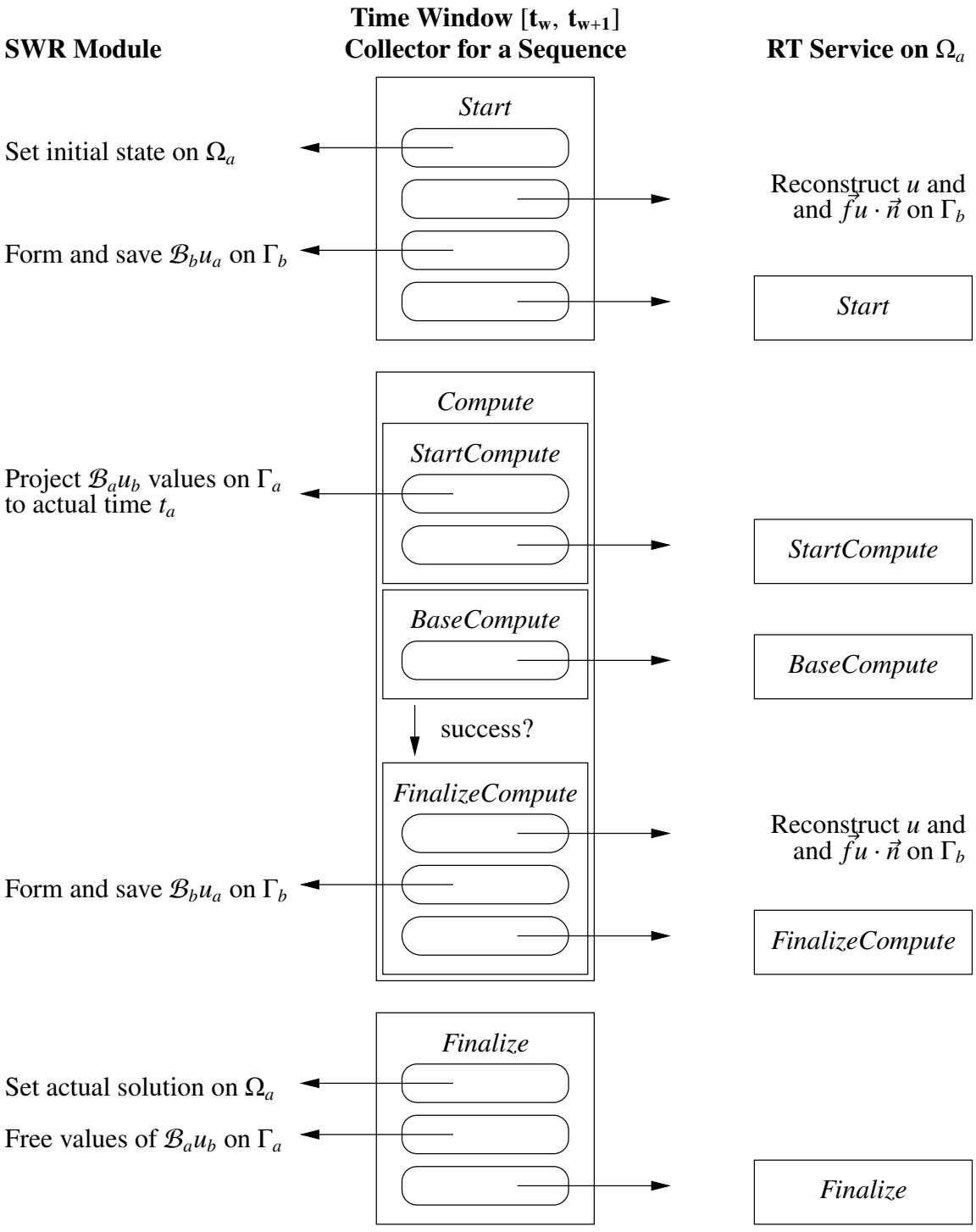


Figure 6.2: Collected operations for calculating the sequence on a time windows during the Schwarz waveform relaxation algorithm with $(a, b) \in \{(\mathrm{R}, \mathrm{N}), (\mathrm{N}, \mathrm{R})\}$.

6.3 Numerical Results

In this section, we present numerical results for two different test cases. The first test case describes the chemical attak of cement by CO_2 in an injection well casing and plug, a purely diffusive case with pure kinetic reactions. The second test case is the 2D benchmark case of the SHPCO2 project described in appendix A.

The numerical results concerning the domain decomposition approach are preliminary. At present, only a fixed decomposition into a reactive and a non reactive domain is realised numerically. The user chooses an initial partition of the domain supposed that he has a certain knowledge of the part of the domain which is highly reactive.

6.3.1 Cement Attack by CO_2 — Pure Kinetics

During CO_2 geological storage, liquid or supercritical CO_2 is injected in injection wells. A well is stabilised against breakdown by adding conical cases of cement. Once the injection has finished, the injection well is closed by a plug of cement. As a consequence, the injected CO_2 is in contact with cement and may attack its stabilisation and impermeability character (cf. figure 6.3).

The chemical system contains H_2O as solvent. Different mobile species are dissolved: a tracer, dissolved carbon dioxide CO_2(aq) and two dissolved minerals CaO(aq) and SiO_2(aq). Four fixed mineral species are present in the cement: Calcite ($Ca[CO_3]$), Wollastonite ($CaSiO_3$), Portlandite ($Ca(OH)_2$) and Silica (SiO_2).
Four different kinetic reactions are modelled:

- Portlandite Dissolution: Portlandite + CO_2(aq) $\longrightarrow$ Calcite
- Wollastonite Dissolution: Wollastonite $\xrightarrow{CO_2(aq)}$ CaO(aq) + Silica
- Calcite Precipitation: CaO(aq) + CO_2(aq) $\longrightarrow$ Calcite
- Silica Dissolution: Silica $\xrightarrow{CaO(aq)}$ SiO_2(aq)

The Wollastonite Dissolution reaction and the Silica Dissolution reaction need the catalysts CO_2(aq) and CaO(aq) in order to proceed, respectively. The reaction rate of both reactions is multiplied by the chemical activity of the catalyst while the catalyst itself is not influenced by the reaction.

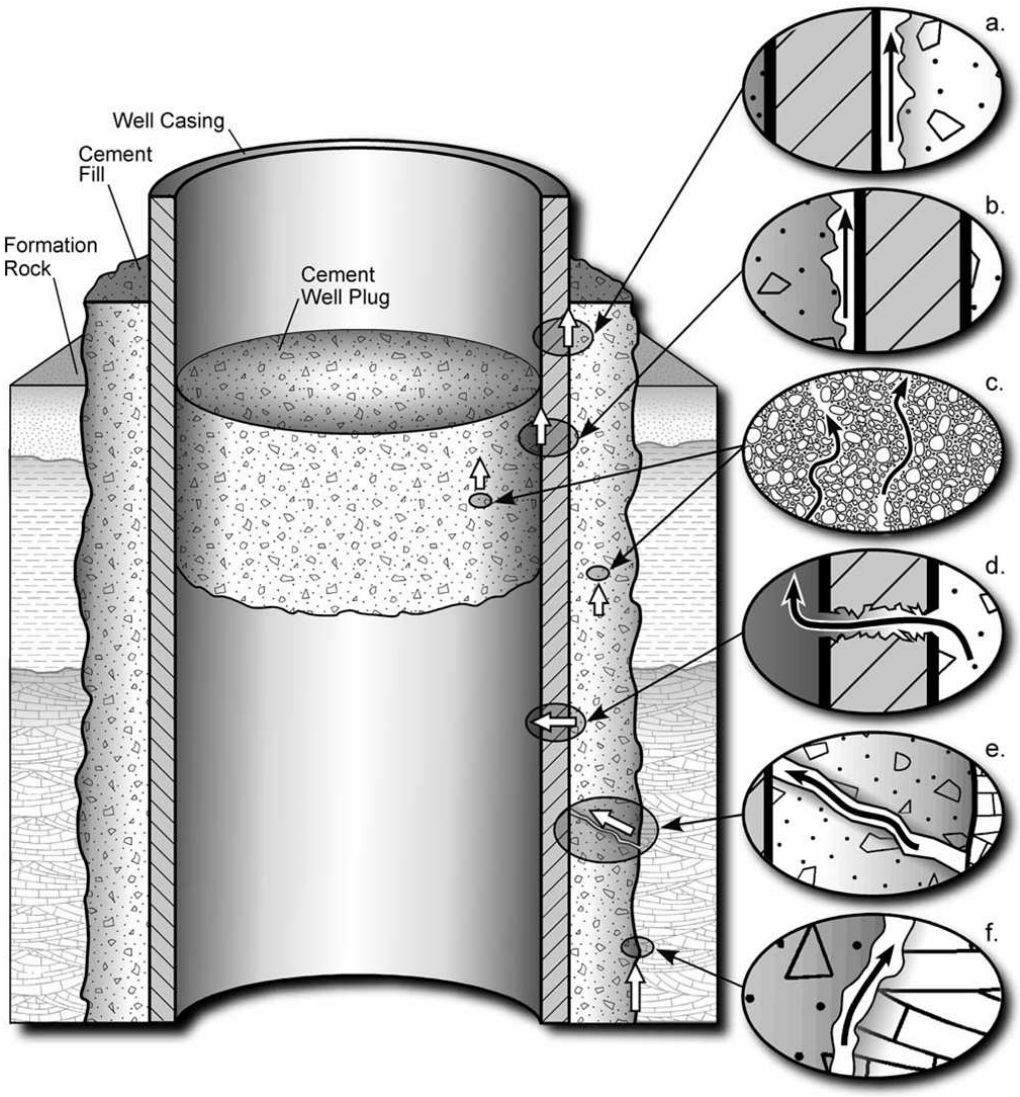


Figure 6.3: Schematic representation of an injection well with case and plug of cement and possible leakages. (source: Princeton University)

Based on the formulation of section 6.1.1, all chemical species are primary species since no equilibrium reactions exists. The following assignment is done:

$$c = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} \mathrm{H_2O} \\ \text{Tracer} \\ \mathrm{CO_2(aq)} \\ \mathrm{CaO(aq)} \\ \mathrm{SiO_2(aq)} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = \begin{pmatrix} \text{Calcite} \\ \text{Wollastonite} \\ \text{Portlandite} \\ \text{Silica} \end{pmatrix}.$$

The reaction rates are modelled using the first order kinetic formulation of equation (1.2). The chemical reactions are supposed to react only in the given direction, as a consequence, all backward constants k_j^b are zero. The forward reaction constants are set to

$$\begin{aligned} k^b_{\text{PortDiss}} &= 1, \\ k^b_{\text{WollDiss}} &= 10, \\ k^b_{\text{CalcPrec}} &= 0.1, \\ k^b_{\text{SiliDiss}} &= 20. \end{aligned}$$

In order to simplify the geometry, we consider here a one-dimensional tube of length 1 with porosity $\phi = 0.2$. The diffusion coefficient is $D_i = 0.02$, we impose no advective flow and hence no dispersion.
The initial state of the tube is

$$c(x,0) = \begin{pmatrix} 55 \\ 1.0 \cdot 10^{-8} \\ 1.0 \cdot 10^{-8} \\ 1.0 \cdot 10^{-8} \\ 1.0 \cdot 10^{-8} \end{pmatrix}, \qquad q(x,0) = \begin{pmatrix} 1.0 \cdot 10^{-8} \\ 4.0 \\ 4.0 \\ 1.0 \cdot 10^{-8} \end{pmatrix}.$$

Concerning the boundary, we impose a Dirichlet condition at $x = 1$ with values

$$c(1,t) = \begin{pmatrix} 55 \\ 1.0 \\ 1.0 \\ 1.0 \cdot 10^{-8} \\ 1.0 \cdot 10^{-8} \end{pmatrix},$$

for the mobile species. On $x = 0$ we impose a perfect impermeability modelled by a homogeneous Neumann condition, i. e. the tube is perfectly closed and the flux is zero.
Simulation time is $t \in [0, 20]$. The domain is decomposed in two overlapping subdomains $\Omega_{\text{NR}} = [0, 0.5]$ and $\Omega_{\text{R}} = [0.4, 1]$, two time windows $T_1 = [0, 10]$ and $T_2 = [10, 20]$ are used for the simulation. The initial time steps are 2.5 for the reactive and 5.0 for the non reactive domain, the global mesh uses a discretisation of 50 grid cells. We impose Robin transmission conditions on the interfaces with parameters $p = 0.02$ which consists in the best practical choice determined by try and error and start the alternating iteration with the reactive domain. The Schwarz waveform algorithm is iterated until the change of concentration at the end of a time window between two iterations is less than 10^{-8}.
In figures 6.4 and 6.5 we show the amount of the mineral species and the concentration of the mobile species at the initial state and at the end of the two time windows. The tracer and CO_2(aq) are entering on the right. CO_2(aq) is absorbed by the Portlandite dissolution reaction and as long as CO_2 is present as catalyst, the Wollastonite dissolution appears and produces CaO(aq) and

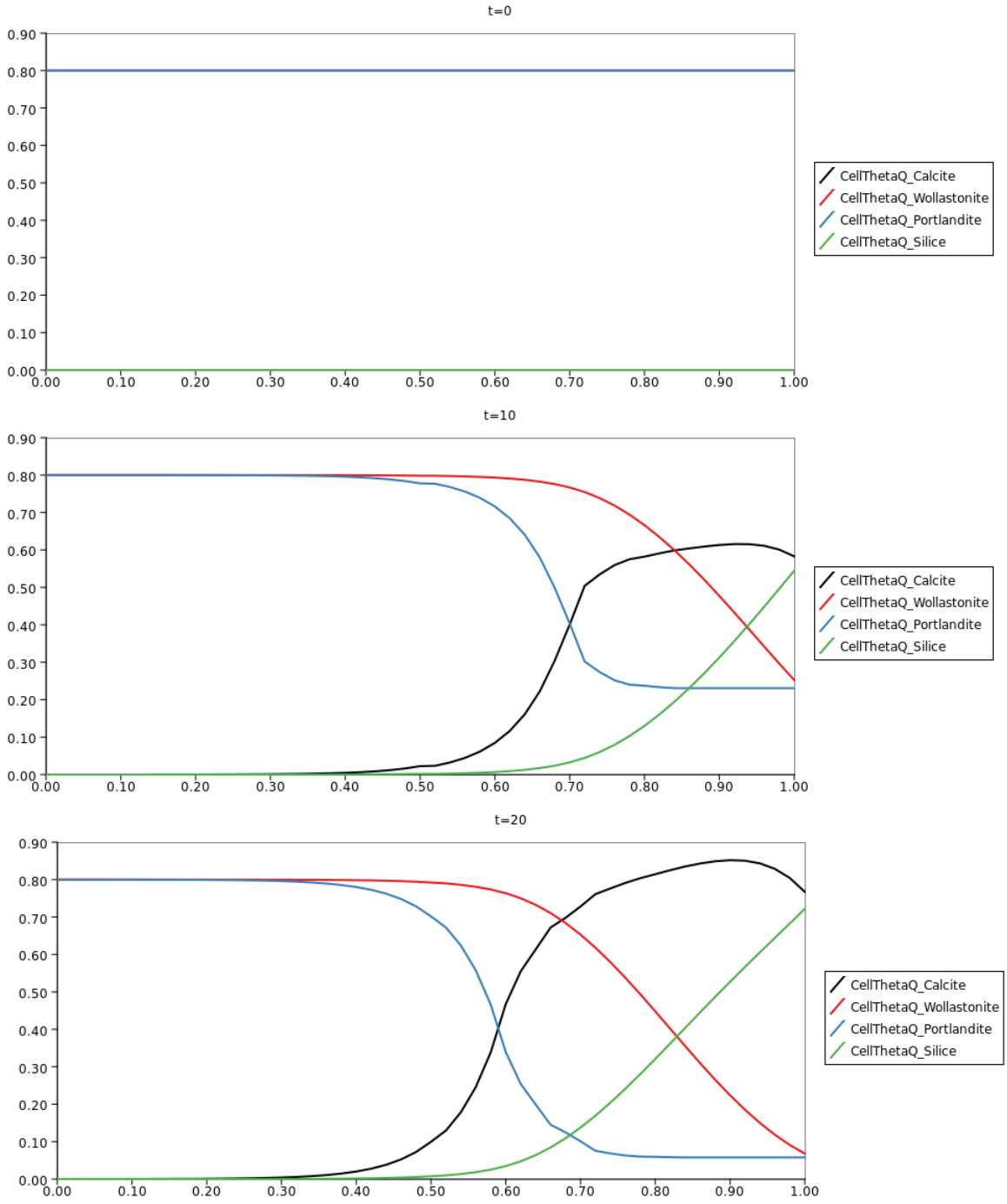


Figure 6.4: Domain decomposition solution of cement test case: amount of mineral species ($t = 0$, $t = 10$, $t = 20$ from up to down)

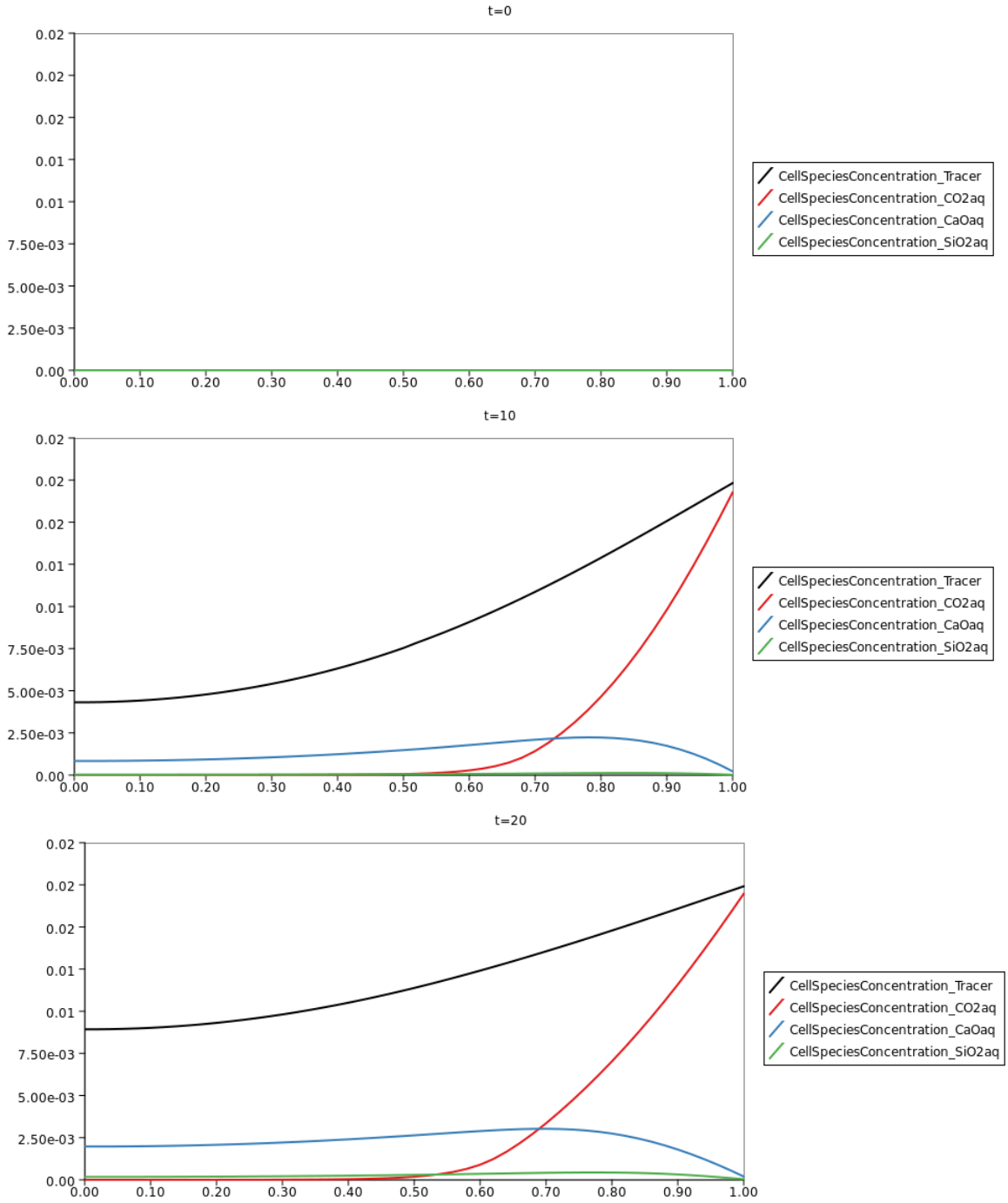


Figure 6.5: Domain decomposition solution of cement test case: concentration of mobile species ($t = 0$, $t = 10$, $t = 20$ from up to down)

Silica. CaO(aq) together with CO_2 forms Calcite and makes both the Portlandite and the Wollastonite dissolution slower since it consumes CO_2(aq). Silica dissolution is slowed down the missing catalyst CaO(aq)which is consumed in the Calcite precipitation.

The Schwarz solver needed 4 iterations for the first and 6 iterations for the second time window in order to reach convergence. The nonlinear solver in the subdomains proceeded 190 iterations in the reactive and 84 iterations in the non reactive subdomain.

6.3.2 SHPCO2 Test Case — Mixed Equilibrium and Kinetics

The SHPCO2 test case has been designed as a synthetic test case in the framework of CO_2 geological storage. It includes all major physical and chemical effects. The test case description can be found in appendix A.

For the numerical formulation of the chemical system, we chose the following distinction between primary species

$$c = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{pmatrix} = \begin{pmatrix} \mathrm{H_2O} \\ \text{Tracer} \\ \mathrm{CO_2(aq)} \\ \mathrm{Cl^-} \\ \mathrm{H^+} \\ \mathrm{Na^+} \\ \mathrm{Ca^{++}} \\ \mathrm{SiO_2(aq)} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \text{Calcite} \\ \text{Quartz} \end{pmatrix},$$

and secondary species

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mathrm{HCO_3^-} \\ \mathrm{OH^-} \end{pmatrix}, \qquad z = \begin{pmatrix} z_1 \end{pmatrix} = \begin{pmatrix} \mathrm{CO_2(solid)} \end{pmatrix}.$$

Note that CO_2(solid) models the gaseous CO_2 initially present and trapped in the geological dome.
The Morel tableau (cf. (6.1)) of the equilibrium system if given by

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	q_1	q_2
x_1	1		1		-1					
x_2	1				-1					
z_1			1							

,

where we omit the upper identity part for the sake of readability.

We simulate the test case on a 2D mesh with 76 cells in x-direction and 48 cells in y-direction. The initial subdomains are chosen as indicated in figure 6.6. We impose Robin transmission conditions with a local 1D approximation of 0th order of the optimal parameter (compare section 4.4.4 where this technique is applied with a local 1D optimisation of the parameter). The Schwarz waveform relaxation algorithm is iterated until the change of the concentration at the end of a time window is smaller than 10^{-4}. We perform a simulation of the first 95.13 years with three equally sized time windows and ten time steps per time window (i. e. $\Delta t = 3.17$ years). In order to show the localisation of numerical difficulties related to kinetic reactions appearing mainly in the reactive subdomain, we chose no adaptive time strategy and concentrate on the number of nonlinear steps needed to solve the subdomains. In figure 6.7 we show the pH at $t = 634$ years of the domain decomposition simulation ($t = 95.13$ year) who is an indicator of the chemical reactivity.

The Schwarz waveform relaxation algorithm needs 2 iterations per time window to converge. The overall number of global nonlinear steps are 421 in the reactive and 306 in the non reactive subdomains. The linear systems the during the nonlinear iteration are more difficult to solve in the reactive domain than in the non reactive domain. The iterative solver of GMRES type preconditioned with Hypre's Euclid preconditioner needed overall 10234 iterations in the reactive and and only 3806 iterations in the non reactive subdomain. This means, that in the reactive subdomain, the linear solver needed 24.3 iterations in average and in the non reactive domain it needed only 12.4 iterations.

In order to illustrate the global behaviour, we performed a global monodomain simulation until $t = 634$ years. In figures 6.8 and 6.9 we show the velocity magnitude of the Darcy field and the pressure field. In figures 6.10 and 6.11 we show the tracer at initial time and at $t = 634$ years. In figure 6.12 and 6.13 we show the aqueous CO_2 concentration and the Calcite amount at $t = 634$ years.

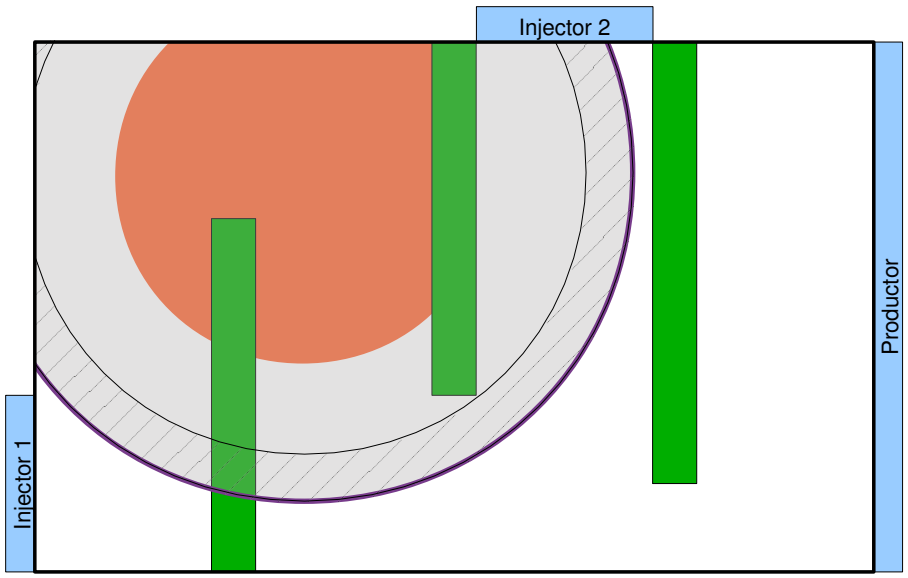


Figure 6.6: Choice of subdomains in the SHPCO2 2D test case. Orange: position of the initially present CO_2. Shaded grey: reactive subdomain. Violet solid line: Interface of the reactive subdomain. Hatched annulus: overlapping region between reactive and non reactive subdomain.

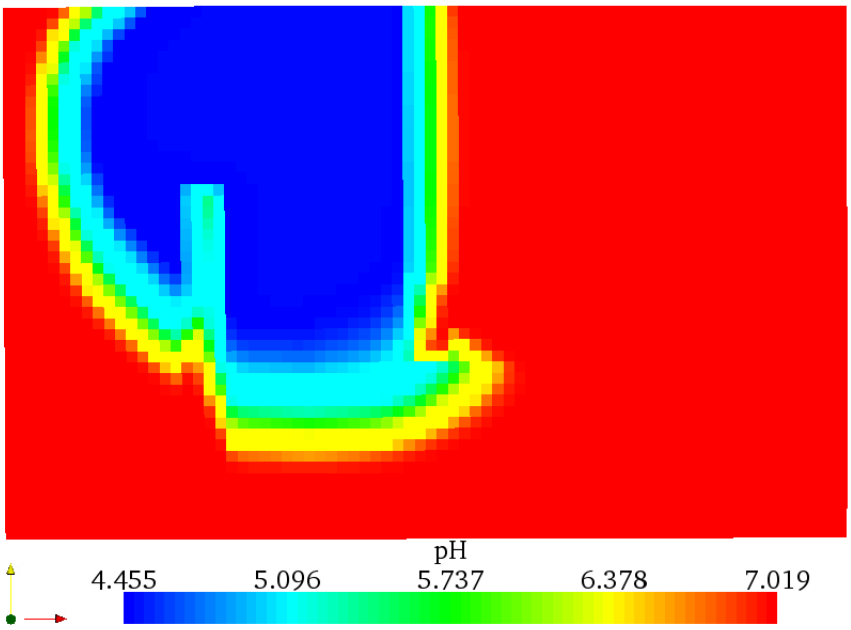


Figure 6.7: pH of the SHPCO2 test case at $t = 95.13$ years (end of the domain decomposition simulation)

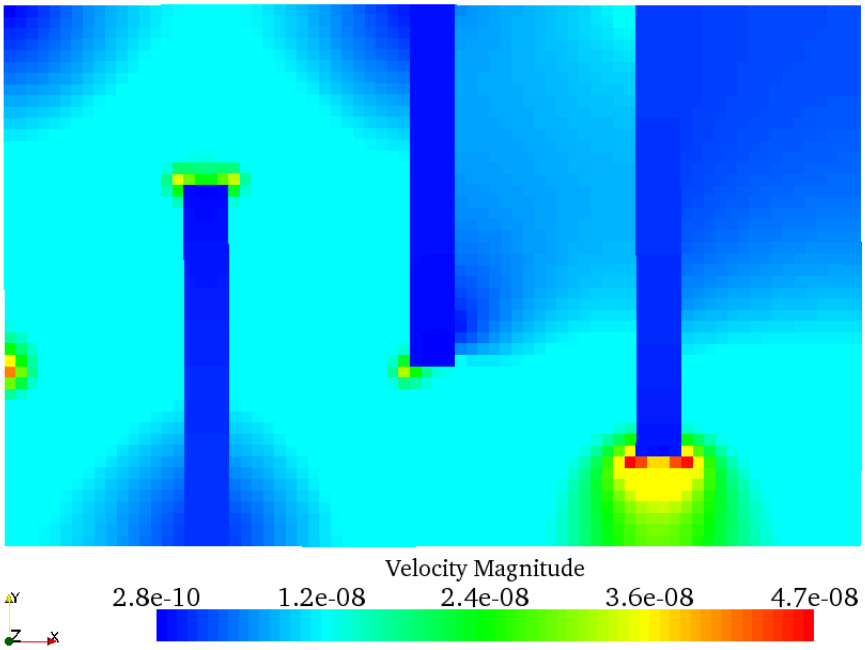


Figure 6.8: Velocity magnitude of the SHPCO2 test case

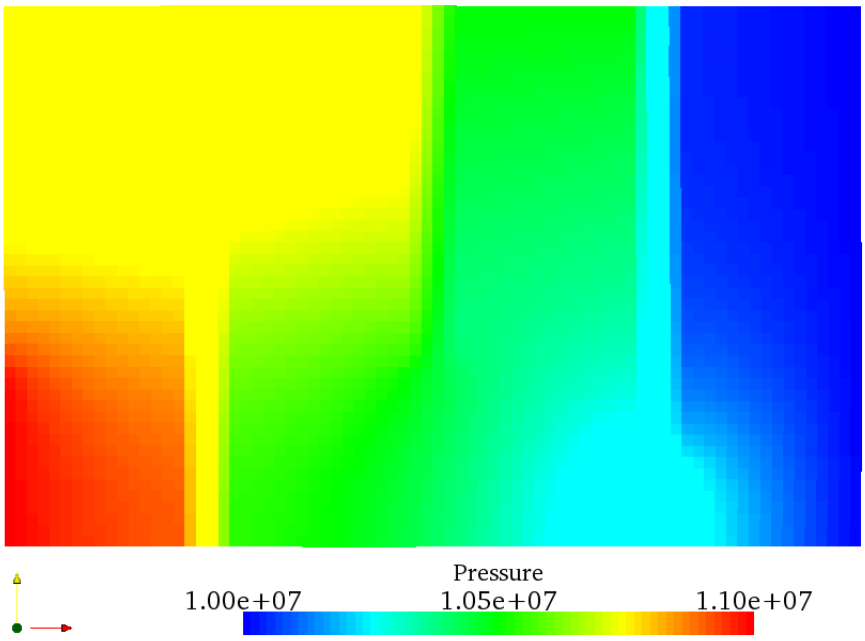


Figure 6.9: Pressure field of the SHPCO2 test case

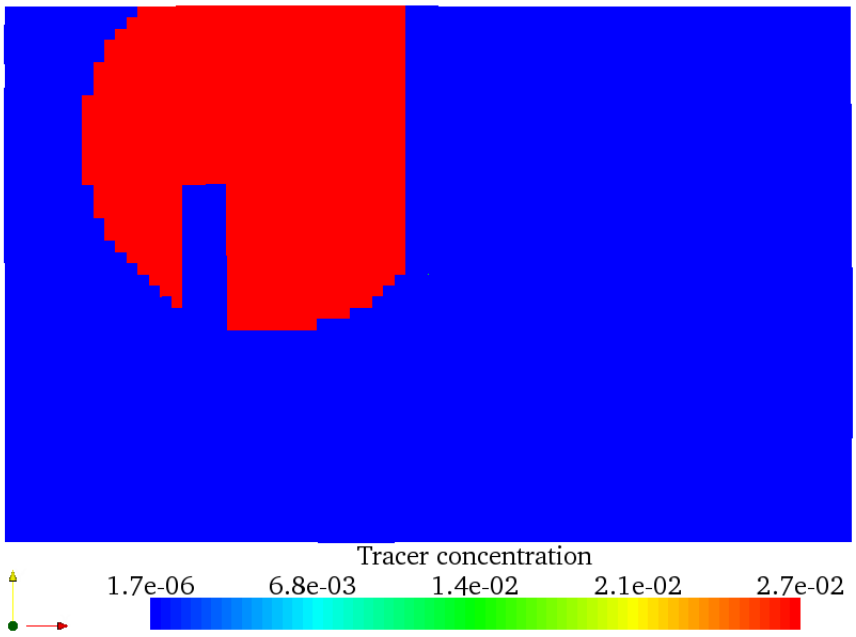


Figure 6.10: Tracer at initial time, monodomain solution

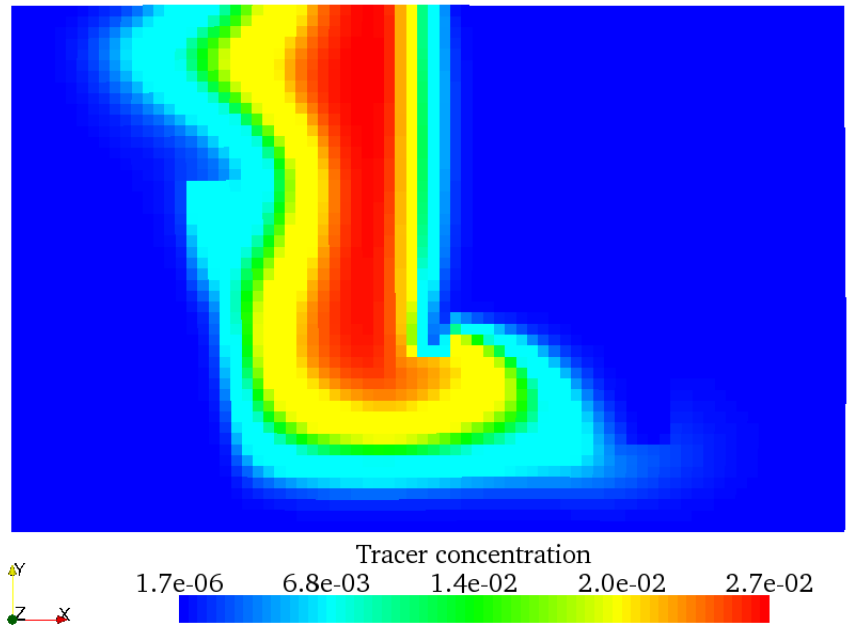


Figure 6.11: Tracer advancement at the $t = 634$ years, monodomain solution

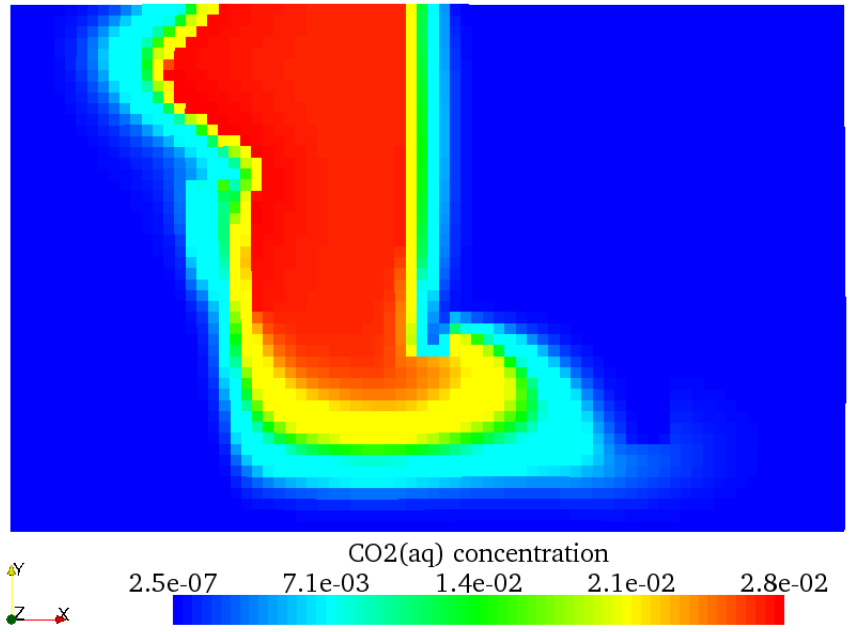


Figure 6.12: CO_2(aq) concentration at t = 634 years, monodomain solution

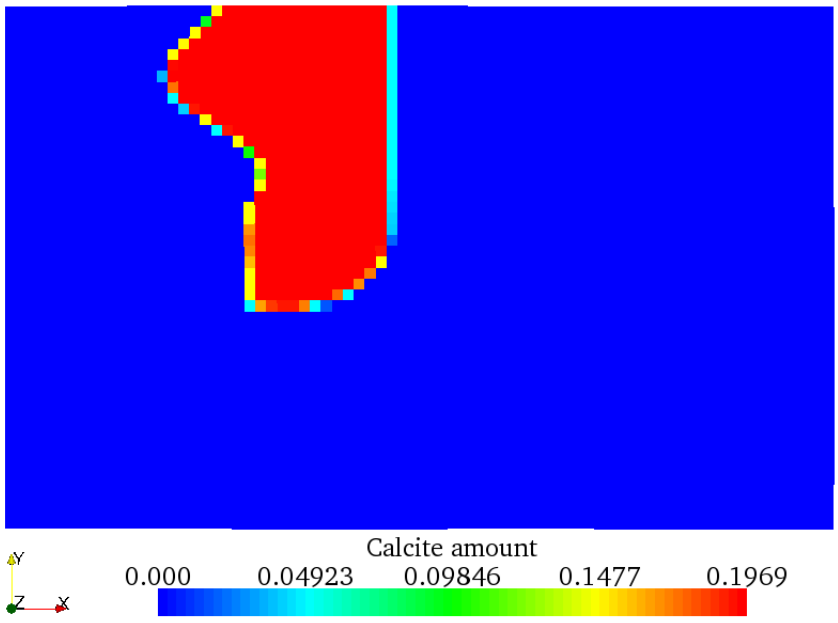


Figure 6.13: Calcite amount at t = 634 years, monodomain solution

Conclusion

In this chapter, we described the multispecies reactive transport model and the Schwarz waveform relaxation algorithm that we used for the numerical implementation on the Arcane platform. Then, we described the numerical tools and the structure of the realisation of the Schwarz algorithm with windowing technique. Finally, we described two test cases in the context of CO_2 geological storage and provided first numerical results.

The numerical results presented in this chapter allowed to give a first validation of the approach presented in this thesis. The first test case, attack of cement by CO_2, showed that it is possible to localise time step constraints in reactive subdomains. The second test case, the SHPCO2 benchmark, showed that it is possible to localise also nonlinear and linear numerical efforts in the reactive subdomain.
Nevertheless, we encountered robustness difficulties in the subdomain solvers, already on a global monodomain level, which need further investigations in order to provide a robust and black-box like solver. This is especially important in the context of a domain decomposition approach where the solver should be able to perform a resolution in every case.
Concerning the domain decomposition implementation further developments have to be done in order to realise automatically detected and moving subdomains.

Conclusion

The work presented in this manuscript has been carried out in order to apply a domain decomposition method on a reactive transport systems in the context of CO_2 geological storage simulation. The motivation lies in the localised time step restrictions due to strong nonlinear coupling terms related to chemical reactions in a global implicit approach. The time step restrictions are usually overcome by cutting the time step globally which leads to a loss of performance in a long-term simulation. The domain decomposition method used in this work allows a local time stepping strategy which makes global implicit approaches for reactive transport problems more performing.

Concerning the reactive transport problem, we have tried to develop a formulation and to use methods which are the state of the art. Nevertheless, solving a reactive transport problem in a global implicit approach stays a challenging task. Kinetic reactions and strong equilibrium reactions are still a difficult issue in the scientific community and more performing methods have to be developed.
Concerning our approach for discretising the transport operator, we limited our developments on rectangular meshes. We could establish a hybrid finite volume scheme which uses a two-point discretisation of the flux. Even if this two-point scheme is naturally suited for a domain decomposition approach and two-point schemes are in general widely used in industrial applications because of their simplicity, more performing and more general methods are available, especially for non-orthogonal grids and strong anisotropic diffusion. Multi Point Flux Approximation schemes (MPFA) have been developed in the last ten years which try to provide an efficient tool for this cases. It will be interesting to see how a domain decomposition approach can be realised using those schemes.

The heart of this work is the study of a domain decomposition method of Schwarz waveform relaxation type. Those methods offer on the one hand the possibility to treat subdomains with different numerical methods and discretisations and on the other hand they can easily be implemented and offer a fast convergence. A prototype study has been carried out in order to validate the approach on a reduced reactive transport problem. It came out that, in accordance with previous studies on related problem types, the convergence properties of the domain decomposition method depend on three factors:

1. **Transmission conditions**. While classical transmission conditions of Dirichlet and Neumann type can only be used when the subdomains do overlap and even then, the convergence is quite slow, advanced transmission conditions of Robin and Ventcel type are a more general and performing choice. We therefore developed Robin and Ventcel transmission conditions in the linear case for the coupled reactive transport system and studied the influence of the parameter choice on the convergence of the domain decomposition approach. The results are in accordance with previous studies on this topic, i. e. only optimised parameters offer a significant advantage compared to classical transmission conditions. Moreover, the assumptions done on the system in order to develop the strategy

to obtain optimised parameters have no significant influence on the real behaviour of the algorithm using the real system. Nevertheless, it appears a more difficult task to develop a strategy to obtain optimised parameters in reality when these assumptions cannot be done anymore. While, in the linear case, first results for the case of discontinuous or variable coefficients are available, no strategy for the nonlinear case exists yet. Even the definition of advanced transmission conditions of Ventcel type in the nonlinear case is far from being clear.

2. **Overlap**. For both the linear and the nonlinear case, overlap is an interesting possibility to obtain faster convergence in our application. While for multiple subdomains with the aim of parallelism overlap is a penalty factor since the amount of doubled data and work rises with the number of subdomains, this is less crucial in the case of only two subdomains (if the subdomains are connected and do not have a degenerated shape like meanders for example). In our case, it has turned out, that an overlap of one layer of grid cells can enhance the convergence speed tremendously, independently of the choice of the parameters for the transmission condition. Nevertheless, there may be applications where overlap is not possible.

3. **Krylov accelerators**. Basing on the linear case, we could develop two new approaches in the nonlinear case that apply Krylov accelerators on the interface problem rising from the domain decomposition method. We studied intensively their properties, always in comparison to the classical approach, and could show that they do have a significant accelerating property. It has been shown that this property becomes especially interesting in the case when the parameters for the transmission conditions are badly estimated. Moreover, the accelerating property becomes more important when the problem size increases. Nevertheless, we have shown that the new approaches have an important drawback compared to the linear case: while in the linear cased, Krylov accelerators do accelerate the convergence speed independently of the type of the transmission condition without overhead cost, in the nonlinear case, they have an overhead cost that is not negligible for coarse discretisations.

As to our application on a multispecies problem, based on this three points, we decided to use the following strategy. First, we decided to use Robin transmission conditions since they are as easy to use as Neumann and Dirichlet conditions but offer a significant gain in convergence speed. Together with a localised strategy for the parameter choice which consists in a low order approximation of the optimised parameter, we could obtain fast convergence in practice. Second, overlap is in our case an important factor in accelerating the convergence speed since in the case of two subdomains the amount of doubled data is negligible compared to the gain of convergence speed. Finally, we decided not to use a Krylov accelerator as the studies and developments concerning the overhead cost have not yet finished and therefore the overhead cost may have a too strong influence on the overall performance.

Meanwhile, there are open questions concerning a high performing domain decomposition approach.

The first issue concerns the optimised parameters. It was shown, that they play an important role in the convergence speed of the domain decomposition method. The strategy for obtaining optimised parameters in the linear case is under some assumptions affordable and has confirmed its validity. For nonlinear cases with variable coefficients on interfaces with arbitrary shape and a time discretisation that is not known in advance since it is a result of an adaptive time stepping strategy, no approach for obtaining optimised parameters is available yet. It will be interesting to see how the strategy for obtaining optimised transmission conditions changes when only few information of the time discretisation are available, especially in the case where chemistry introduces strong coupling terms which have an influence on the temporal variation on the interface and hence only poor initial guesses are available. A study of the sensibility of optimised parameters with respect to the information of the time discretisation has to be done and then the influence of the real behaviour of the algorithm has to be tested.
Besides the missing information on the time discretisation, variable coefficients are an interesting detail. Since in theory, only constant parameters along the interface are considered until now, it will be interesting to allow also variable parameters on the interface and compare them to a constant choice. We tested a first strategy with a localised 1D optimisation at every face which is a comfortable and performing choice when the problem is advection dominant. Nevertheless, as soon as diffusion becomes dominant, this strategy could lead to poor convergence and other approaches have to be developed.
Finally, if overlap is acceptable, it can be used, with the cost of doubled resolution amount, to accelerate convergence of the domain decomposition algorithm. But when, as in the case of many subdomains, overlap becomes a penalty factor, only optimised parameters and Krylov accelerators can be used. The two new approaches for the interface problem using Krylov accelerators in the nonlinear case have shown interesting properties. Nevertheless, further effort has to be put in this methods in order to attenuate or eliminate the overhead cost for coarse problems. We proposed a linesearch strategy for the Newton method and motivated to develop a strategy for keeping the information on the Krylov subspaces during the global iterations. Finally, the new approaches have to be tested in a framework with many subdomains.

Last, but not least, we want to emphasise an issue concerning the combination of a performing domain decomposition approach and the reactive transport model. We have seen that for performance reasons advanced transmission conditions like Robin or Ventcel conditions have to be used. The associated operators are in general not monotone, i. e. they do not respect the maximum principle. As a consequence, even if the initial state and the converged domain decomposition solution of the problem is physically meaningfull, the iterates of the domain decomposition approach may not be within a physical range, negative concentrations may appear.
Often, it is not possible to limit the influence of a non-monotone operator by cutting the time step and hence the domain decomposition approach fails since it is not possible to solve the subproblems. We observed that changing the parameters of the transmission conditions such that the influence of the monotone part of the operator becomes more important helped. For Robin transmission conditions it suffices to increase the parameter, hence, the Dirichlet part becomes more important and the operator violates less the maximum principle. Now, the subproblems are solvable but the performance of the domain decomposition approach deteriorates. Nevertheless,

this technique consists in a try-and-error approach and has no mathematical or numerical basis. A different approach has to be developed and may appear naturally during a wellposedness proof of the nonlinear multispecies reactive transport problem with advanced boundary conditions.

A

SHPCO2 Benchmark

A.1 Introduction

The synthetic test case which is presented in this document has initially been defined in the context of the ANR-SHPCO$_2$ project in order to validate the resolution methods for reactive transport. We hope that, in a later time, it can be used as benchmark for other working groups which are interested in modelling reactive transport for CO$_2$ geological storage.

The problem definition is widely inspired by the GDR MoMaS reactive transport benchmark. It also uses items from other benchmarks concerning CO$_2$ storage. Its relative simplicity allows to consider it for numerical convergence tests and for performance measurements on a series of grids.

We have chosen to use simple models in order to describe all parameters in this note. We hope that in such a way strict numerical comparisons of results can take place. Nevertheless, we also hope to construct a representative example of chemical interactions appearing in real CO$_2$ geological storage cases in such a way that numerical results can be discussed from a physical viewpoint with the modelling comunity. We also tried to respect as much as possible the order of the considered physical variables.

A.2 General Simulation Context

We want to model the modifications of the natural environment on a period of several thousands of years due to a process of CO$_2$ geological storage in a saline aquifer. The target aquifer is heterogeneous and located at 1,000 metres of depth. The studied zone is 4,750 metres in est-west direction and 3,000 metres in north-south direction. The considered geological layer has a thickness of 100 metres. One can represent this simulation configuration in the figure given here below.

The spatial dimensions of the studied zone justify entirely the use of a 2D geometry if one neglects gravity effects.

We suppose that the geological layers limiting the zone of the aquifer in its lower and upper part are entirely impermeable. We also suppose that this zone is mainly formed of sandstone with high permeability and that there are several barriers with low permeability. The aquifer is opened for a flow on three lateral surfaces which are called *Injector1*, *Injector2* and *Productor*, respectively. We suppose that one can impose a uniform pressure which is different on the three surfaces boundaries. The spatial configuration of the boundary which is associated to the positioning of boundary conditions defines a certain hydrogen network. The permanent flow generated in the aquifer is globally orientated from the west to the east but is highly subject to the surrounding of the barriers.

The volume of transported water will flush the gas storage zone situated initially in the north-western part of the domain. This will induce the creation of a gas dissolution front and of associated reaction fronts within the water and afterwards between water and the rock matrix. On a first level, one will observe characteristic reactions in a carbonated aquifer: dissociation of aqueous CO_2, acidification of water and dissolution of calcite. Other more complex phenomena may interfere with this basic processes depending on the imposed boundary conditions or on the complexity of the chemical system taken into account.

A.3 Modelling Hypotheses

Water flow in porous media will be calculated with the help of Darcy's law. We suppose that the permeability is not affected by the change of porosity. We also can reasonably decouple in this first state the flow calculation from reactive transport calculation.

The gas saturation is below the critical value up from which the relative permeability of gas is non-zero. Moreover, since it is a pure phase, there is no diffusion in the gas.

As a result, only species in the aqueous solution are mobile. They can be transported by advection, diffusion and dispersion.

Remark. The hypothesis of immobile gas allows to concentrate oneself on reactive transport and not to treat readily questions concerning multiphase flow. This is indeed a limit situation but it appears effectively in storage zones where the gas is trapped by capillary forces and hence this is not an entirely artificial hypothesis. On the other hand, up to certain details, this is the environment used for the reactive transport phase in the coupling algorithms implemented in ToughReact, PFlotran or Coores: gas transport is taken in charge by the polyphasic flow model. This test case allows so to test especially this phase of the calculation.

A.4 Expected Results

One simulates the evolution of the system for a period of 10,000 years, i. e. 365*10,000 = 3,650,000 days, one day is equivalent to 86,400 seconds.

One saves intermediate results of the simulation every 1,000 years in order to measure the evolution of the variables of the problem during the time. In total, this represents eleven intermediate results including the initial state.

One can also save the evolution curves of the associated balances of gas quantity, mineral phases quantity and porosity.

A.5 Geometric Domain and Mesh

By inspiring ourselves in the proposed tests in the MoMaS benchmark, we have chosen to propose different geometries for the simulation domain. The 1D and 2D geometries are destined to compare the results of different groups and to proceed numerical convergence studies. The use of a 3D geometry should facilitate the diffusion of the results of the benchmark in a larger community.

The reference geometry of the benchmark is the 2D geometry. The use of a planar 2D geometry allows to avoid problems related to the mesh deformation. In this way, one also ensures that the test case stays accessible to all codes, academic as well as industrial ones.

A 1D geometry is also proposed in order to facilitate the comparison between solutions by superposition of curves. By this, one can also measure problems related to instability or precision of a resolution method or the used numerical schemes.

In order to place the test in its initial context, we finally propose a 3D geometry. Now, the aim is no longer to quantify the results but to obtain more a global estimation of the process. Adding a vertical dimension can also allow to simulate and discuss finally the effects due to gravity.

We propose a series of four reference meshes for every type of geometry:

- “Extra Small” mesh, can be used for setting up the code.
- “Small” and “Medium” meshes are the aims of the simulation on which one can compare the solutions of different groups.
- “Large” mesh is more challenging, it can be used to measure to robustness and the performances of the application.

A.5.1 1D Geometry

The 1D problem is obtained by measuring the driven distance of a fluid from the boundary condition situated at the entrance at the south-east side of the domain to the boundary condition at the exit at the east side of the domain. The presence of barriers induces a tortuosity which doubles the effective length between the two boundary conditions. The depth is reduced to L = 1,000 metres, corresponding to the mean section used by the flow. The total volume is close to the one in the 2D case suppressing the barrier zones and the flow channel which is transverse to the principal flow.

The boundary conditions *Injector1* and *Productor* are directly imposed at the entrance and at the exit of the 1D domain. As the barriers are not directly transverse to the flow (cf. the 2D geometry), it is not useful to consider them in 1D. In return, it is necessary to transcribe the boundary condition *Injector2* situated at the north of the domain since it allows the mixing of different waters.

If one uses a 2D or a 3D code to solve the 1D problem, one can directly impose this boundary condition on the faces located at the north of the injection zone between x = 7,000 metres and x = 8,000 metres. If one uses a pure 1D code, one can no longer proceed in this way. In order to replace the boundary fluxes, we propose to use a volumic source term localised at the interaction zone. The formula which allow to calculate those source terms as well as a technical justification based on a finite volume approximation are furnished in the appendix.

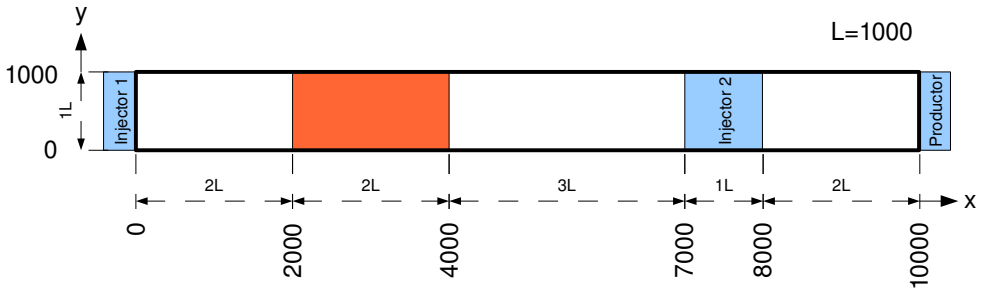


Figure A.1: 1D Geometry

A.5.2 2D Geometry

It consists in an approximation of the 3D geometry of the aquifer. The z-dimension which is not shown in the figure is 100 metres. It is not mandatory to integrate this parameter for the simulation if one uses a real 2D code but 3D codes use this dimension with representative volumes of the considered problem.

Mesh	XS	S	M	L
Dx	250	50	10	5
Dy	1000	1000	1000	1000
Dz	100	100	100	100
Nx	40	200	1000	2000
NCell	40	200	1000	2000

Table A.1: 1D mesh parameters

Mesh	XS	S	M	L
Dx	250	50	10	5
Dy	250	50	10	5
Dz	100	100	100	100
Nx	19	95	475	950
Ny	12	60	300	600
NCell	228	5700	142500	570000

Table A.2: 2D mesh parameters

The dimensions of the problem are given as a function of the characteristic length L = 1,000 metres. The three boundary conditions named *Injector1*, *Injector2* and *Productor*, respectively, are illustrated in blue. They are located in the south-east, in the north and in the east of the domain. The barriers illustrated in green are disposed by three rectangular areas whose dimensions and positions are given in the figure here below.

The storage zone of gas in which gaseous CO_2 is initially present is illustrated in orange. It consists of a disc with radius $\frac{3\sqrt{2}}{4}L$, whose centre point lies in $(1.5\,L, 2.25\,L)$ and which is limited by the barriers illustrated in green. This shape is natural for a storage of gas in a anticlinal geological structure. The rectangular zone limited by the yellow contour is entirely artificial. It occupies approximately the same surface area. It can be used during a set up of data but the final tests have to integrate the orange circular zone. In the appendix, a filter function to select the inner points of the circle is given.

A.5.3 3D Geometry

In order to test the application in a context which is as close as possible to real world use and to discuss eventually the aspects of gravity, we propose to define a 3D geometry of the simulation

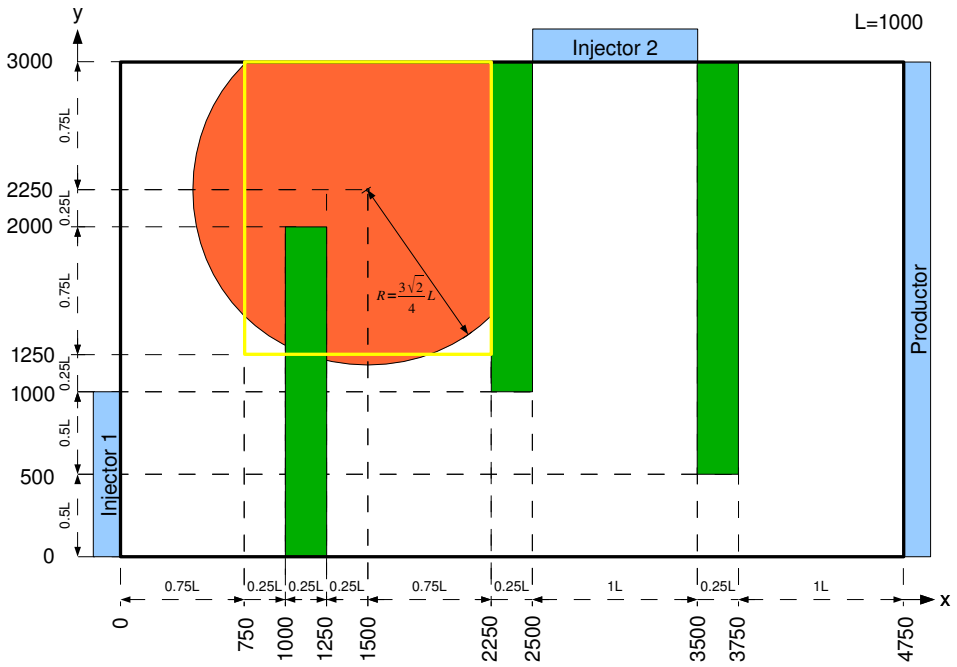


Figure A.2: 2D Geometry

Mesh	XS	S	M	L
Dx	250	62.5	25	25
Dy	250	62.5	25	25
Dz	100	10	10	1
Nx	19	76	190	190
Ny	12	48	120	120
Nz	1	5	10	100
NCell	228	18240	228000	2280000

Table A.3: 3D Mesh parameters

domain. We impose a roof topography of the structure by an analytical deformation. In the appendix, a Matlab function which allows to calculate the upper limit of the roof of the domain at every point in (x, y) is given.

The deformation of the roof is such that the centre of the zone which contains initially CO_2 corresponds to a climax of the geological structure. The positioning of the gas zone can be interpreted as a result of vertical injection at this climax.

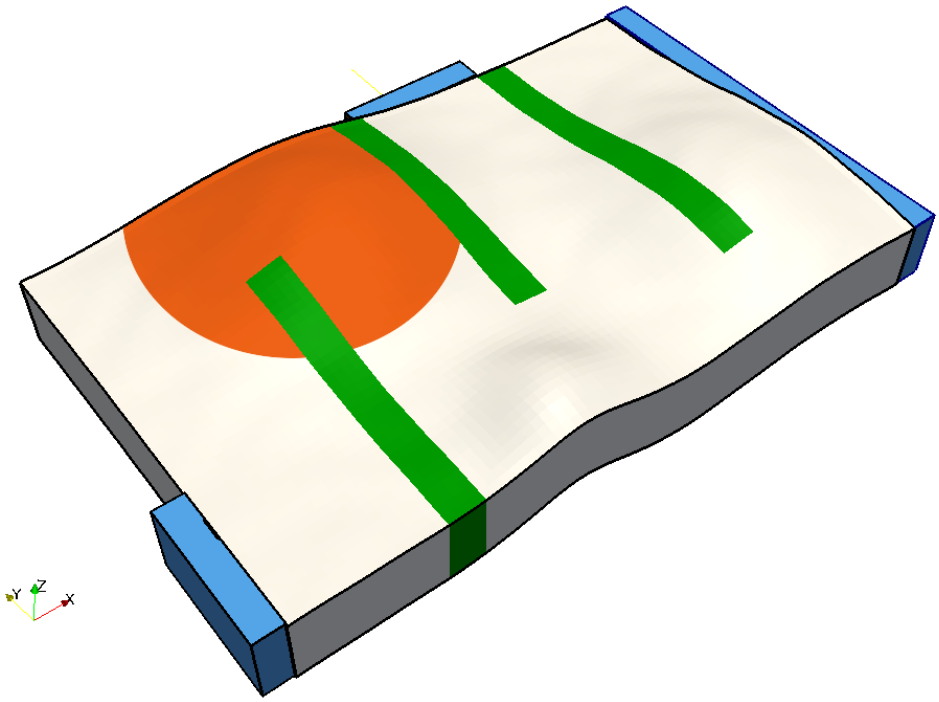


Figure A.3: 3D Geometry

A.6 Compositional System

A.6.1 Phases and Species

We consider a compositional system formed of 12 species divided into 4 phases:

- Phase 1. Gas (or supercritical) : $CO_2(g)$
- Phase 2. Aqueous solution: H_2O, H^+, $CO_2(aq)$, Cl^-, Na^+, Ca^{+2}, $SiO_2(aq)$, HCO_3^-, OH^-
- Phase 3. Calcite mineral: Calcite
- Phase 4. Quartz mineral: Quartz

A.6.2 Equilibrium Reactions

We consider the following three equilibrium reactions:

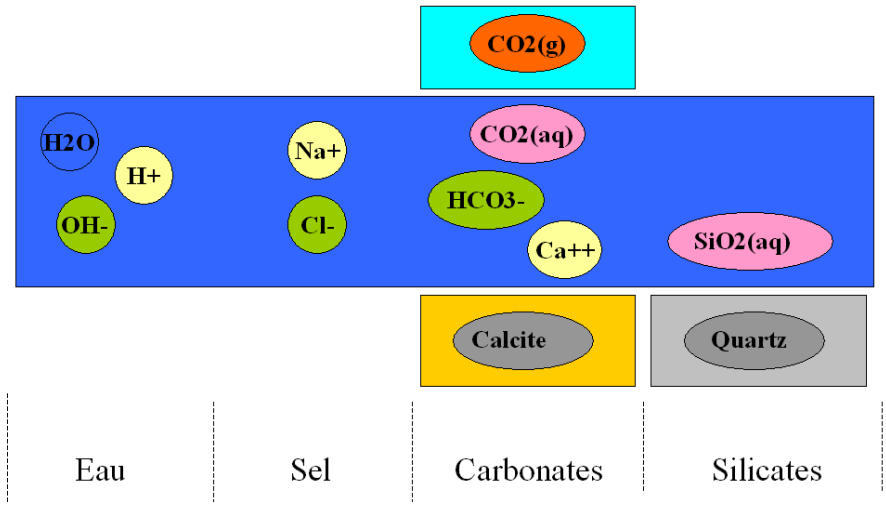


Figure A.4: Structure of the compositional system

Req 1. Hydrolysis of water
$H_2O \Longleftrightarrow H^+ + OH^-$

Req 2. Dissolution of CO_2(g) in water
$CO_2(g) \Longleftrightarrow CO_2$

Req 3. Dissociation of CO_2(aq)
$H_2O + CO_2 \Longleftrightarrow HCO_3^- + H^+$

A.6.3 Kinetic Reactions

We model dissolution-precipitation reactions by kinetics:

Rkin 1. Dissolution of Calcite
$Calcite + H^+ \longrightarrow Ca^{++} + HCO_3^-$

Rkin 2. Precipitation of Calcite
$Calcite + H^+ \longleftarrow Ca^{++} + HCO_3^-$

Rkin 3. Dissolution of Quartz
$Quartz \longrightarrow SiO_2(aq)$

Rkin 4. Precipitation of Quartz
$Quartz \longleftarrow SiO_2(aq)$

A.6.4 Thermodynamic Properties

A.6.4.1 Activity Models

The activity of a chemical species is linked to its chemical potential by the following relation:

$$\mu_i = \mu_i^0(T, P) + RT \ln(a_i)$$

We propose to use ideal activities for every phase:

- Aqueous solution phase, solvent species $i = H_2O$: $a_i = x_i$
- Aqueous solution phase, dissolved species $i \neq H_2O$: $a_i = m_i$
- Calcite mineral phase, mineral species i = Calcite: $a_i = 1$
- Quartz mineral phase, mineral species i = Quartz: $a_i = 1$
- Gas phase, gaseous species $i = CO_2(g)$: $a_i = P_i$

where

- x_i is the molar fraction of the considered species in its phase
- $P_i = P \times x_i$ is the partial pressure of the considered gaseous species
- m_i represents the amount of a species expressed in mol per kilogram of solvent.

A.6.4.2 Elements

A.6.4.3 Species

Remark: in order to obtain a reasonable solubility of CO_2 in water we directly integrated a correction term in the chemical reference potential of the species $CO_2(g)$.

A.6.4.4 Chemical Reactions

The equilibrium reaction constants are obtained by combining the chemical reference potentials of the species which interact in the reaction.

Element	Name	Molar mass [$g.mol^{-1}$]
Na	Sodium	22.990
Ca	Calcium	40.078
Si	Silicium	28.085
Cl	Chlorine	35.453
C	Carbon	12.0105
H	Hydrogen	1.008
O	Oxygen	15.9995
+	Charge	0.000

Table A.4: Characteristics of elements

Phase	Species	Log10(K)	Formula	Molar mass(g)
Aqueous	H_2O	0	H(2)O(1)	18.0155
Aqueous	H^+	0	H(1)+(1)	1.008
Aqueous	CO_2(aq)	0	C(1)O(2)	44.0095
Aqueous	Cl^-	0	Cl(1)+(-1)	35.453
Aqueous	Na^+	0	Na(1)+(1)	22.990
Aqueous	Ca^{+2}	0	Ca(1)+(2)	40.078
Aqueous	SiO_2(aq)	0	Si(1)O(2)	60.0840
Aqueous	HCO_3^-	-6.2206340	H(1)C(1)O(3)+(-1)	61.0170
Aqueous	OH^-	-13.235362	O(1)H(1)+(-1)	17.0075
Mineral Calcite	Calcite	-7.7454139	Ca(1)C(1)O(3)	100.087
Mineral Quartz	Quartz	3.5862160	Si(1)O(2)	60.0840
Gas	CO_2(g)	2.0861861	C(1)O(2)	44.0095

Table A.5: Thermodynamic parameters of the species

Reaction	Type	Name	Log10(K)
Req 1	Aqu	Hydrolysis of water	-13.235362
Req 2	Gas-Aqu	Dissolution of CO_2(g) in water	-2.0861861
Req 3	Aqu	Dissociation of CO_2(aq)	-6.220634
Rkin 1	Min-Aqu	Dissolution of Calcite	1.5247799
Rkin 2	Min-Aqu	Precipitation of Calcite	-1.5247799
Rkin 3	Min-Aqu	Dissolution of Quartz	-3.5862160
Rkin 4	Min-Aqu	Precipitation of Quartz	3.5862160

Table A.6: Thermodynamic parameters of chemical reactions

	$T_0[C]$	$k(T_0)[mol.s^{-1}.m^{-2}]$	$Ea[J]$	$S\,r_M[m^2.kg-1]$
Calcite	25	1.6e-09	87.5	100
Quartz	25	1.2e-14	41.87	1000

Table A.7: Kinetic reactions parameters

A.6.5 Dissolution-Precipitation Kinetics

We use the following kinetic model:

$$V_d^M = k_d(T) \times S\,r_M \times (1 - Q/K),$$

$$V_p^M = k_p(T) \times S\,r_M \times (Q/K - 1),$$

where Q is the activity product of the considered reaction, K its equilibrium constant and $S\,r_M$ its reactive surface. The variables k_d, k_p are reaction speeds normalised to a surface unit. They are modelled by a formula of the following type:

$$k(T) = k(T_0) \times \exp(Ea/RT_0) \times \exp(-Ea/RT)$$

The parameters are given here below, we suppose that the reactive surface of minerals is constant.

A.7 Boundary Conditions

Concerning the fluid flow, we impose boundary conditions with uniform pressure at the boundary surfaces *Injector1*, *Injector2* and *Productor*. At the rest of the boundary of the domain, we impose boundary conditions with zero flux.

Concerning the transport, we impose on the three boundary surfaces a neutral water composition which is identical to the initial state outside the gas region. This conditions are *a priori* subject to both advective and diffusive flux. With respect to the used code of the groups, one can also take into account a pure advective flux on the boundary. On the rest of the boundary of the domain, we consider a closed boundary which can be translated by a zero flux.

	Pressure [Pa]	Composition
Injector 1	100.e+05	Neutral water
Injector 2	105.e+05	Neutral water
Productor	110.e+05	Neutral water

Table A.8: Limit condition parameters

Temperature [K]	Pressure [Pa]	Reference depth [m]
323.15	$100.e+05 \leq P \leq 110.e+05$	-1050

Table A.9: Pressure and temperature parameters for the initial state

A.8 Initial Conditions

A.8.1 Pressure and Temperature

Temperature is supposed to be uniform and constant during all the simulation. The initial pressure is calculated with respect to the imposed limit conditions on *Injector1*, *Injector2* and *Productor*. As a first approach, one can use a uniform pressure.

A.8.2 PetrophysicalProperties

The domain is partitioned in two zones. The first zone, called "Barriers", is formed by the three geological units coloured in green. The second zone, called "Drain", is formed of the rest of the domain. In the inner of a zone, the petrophysical parameters and the transport parameters are constant and uniform. Meanwhile, the parameters of the two zones are different. We suppose that the porosity and the initial chemical composition of the rock are equal everywhere in the domain.

Remark. We suppose that the mobility of water is not affected by the presence of gas which is lower than the critical saturation and we neglect the capillarity effect on this interval. For a diphasic model, this can be translated by the constraint $kr_w(S\,g) = 1$ and $Pc^{g,w}(S\,g) = 0$ on the interval $0 \leq S\,g \leq S\,gc$. The initial saturation of gas is lower than $S\,gc$ and the system evolutes in the sens of the dissolution of gas. One does not leave this saturation level all over the simulation.

	Barrier zone	Drain zone
Porosity [−]	0.2	0.2
Permeability [m^2]	1.e-15	100.e-15
Critical saturation of gas $S\,gc$ [−]	0.25	0.25
Relative permeability kr_w in $S\,gc$ [−]	1	1
Capillary pressure $Pc^{g,w}$ in $S\,gc$ [−]	0	0
Longitudinal dispersion β_L [m]	50	50
Transversal dispersion β_T [m]	10	10
Diffusion coefficient [$m^2.s-1$]	1.e-09	1.e-09
Volumic fraction of solid Calcite	0.2	0.2
Volumic fraction of solid Quartz	0.8	0.8

Table A.10: Petrophysical Parameters

	Gas zone	Outside gas zone
Gas saturation	0.2	0
Water composition	Acid water	Neutral water

Table A.11: Initial state of the fluids

A.8.3 Fluid Parameters

We distinguish the central circular zone coloured in orange named “gas zone” from the rest of the domain. Initially, we suppose that in the gas zone the saturation is not zero but smaller than the critical saturation. By this, the gas stays fixed during all the simulation.

The composition of water in the gas zone is obtained by imposing an equilibrium in neutral water with the rock matrix and an excess of CO_2(g), Calcite and Quartz. One obtains finally a more acid water which is in equilibrium with both gas and the rock matrix. Neutral water is used to fill the rest of the domain. The initial state is therefore locally in equilibrium, it is only disturbed by the transport.

A.8.4 Numerical Parameters

It is difficult to impose absolute convergence criteria since those parameters are interpreted in a specific way for different formulations or numerical methods used. One can interpret the following criteria individually but every group may indicate the used criteria to determine convergence of implemented solvers in their code.

	Volumic mass $[kg.m^{-3}]$	Viscosity $[Pa.s]$
Aqueous solution phase	1000	0.571e-03
Gas phase	470	0.0285e-03
Calcite phase	2710	0
Quartz phase	2643	0

Table A.12: Physical properties of fluids

	Tolerance
Relative Error - Balance Equations	1.e-08
Relative Error - Equilibrium Equations	1.e-08

Table A.13: Numerical convergence criteria

A.9 Appendix 1. Topography of the Roof Structure in a 3D Geometry

The here given Matlab function allows to calculate the vertical height of the roof structure for a given point (x, y) of the considered domain.

```
% ------------------------------------------------------
% Calculation of the roof topography for TestSHPCO2 3D
%------------------------------------------------------
function z = topSHPCO2(x,y)

% Mean value
ztop = -1000.;

% Interpolated topographie by interpolation
A1 = 36   ; B1 = 1; x1 = 1750  ; y1 = 2200  ; r1 = 1000  ;
A2 = 24   ; B2 = 1; x2 = 2000  ; y2 = 300   ; r2 = 500   ;
A3 = 34   ; B3 = 1; x3 = 4000  ; y3 = 1000  ; r3 = 1000  ;
A4 = -10  ; B4 = 1; x4 = 3200  ; y4 = 700   ; r4 = 2000  ;

zk =  A1*exp(-B1*((x-x1).^2 + (y-y1).^2)/(r1.^2)) ...
    + A2*exp(-B2*((x-x2).^2 + (y-y2).^2)/(r2.^2)) ...
    + A3*exp(-B3*((x-x3).^2 + (y-y3).^2)/(r3.^2)) ...
    + A4*exp(-B4*((x-x4).^2 + (y-y4).^2)/(r4.^2));
```

```
% Local pseudo-periodic perturbation
zp = 2*sin(x*0.005 + sin(y)).*sin(y*0.01 + cos(x));

% Large distance deformation
zc = - 0.000001* ((x-2000).^2 + (y-2000).^2);

% Final surface
z = ztop + zk + zp + zc ;
% ------------------------------------------
```

A.10 Appendix 2. Volumic Calculation of the Gas Zone

The cylindrical zone in orange can be indicated by the following filter function in Matlab.

```
% ------------------------------------------------
% Filter selection of the gas zone
%------------------------------------------------
function f = GasZoneFilter(x,y)

L       = 1000;

% Free circular zone
x0  = 1.5*L;
y0  = 2.25*L;
r   = 2*sqrt(2)/4*L;
fcircle = ( (x-x0).*(x-x0) + (y-y0).*(y-y0) ) < r*r ;

% Barriers and Limits
fbar1      = (x>L) && (x<1.25*L) && (y<2*L);
flimbar1   = ~fbar1;
flimbar2   = (x<2.25*L)

% Final circular zone
f = fcircle && flimbar1 && flimbar2

% ------------------------------------------
```

For a given 2D mesh, by applying the presence filter for all cell centres, one can determine the global parameters of the gas zone. The given values in the table here below have to be verified by all groups with their codes in order to ensure that the initial gas zone is interpreted in a correct way.

Geometry	Dx	Dy	Dz	Volume [m^3]	Gas volume [m^3]	Mass of gas [kg]
2D Mesh - XS	250	250	100	2.562500e+08	1.02500e+07	4.817500e+09
2D Mesh - S	50	50	100	2.687500e+08	1.07500e+07	5.052500e+09
2D Mesh - M	10	10	100	2.706400e+08	1.08256e+07	5.088032e+09
2D Mesh - L	5	5	100	2.706875e+08	1.08275e+07	5.088925e+09

Table A.14: Global parameters of the gas zone in its initial state

A.11 Appendix 3. Realisation of the Injector2 Boundary in a 1D Code

If one simulates the test case with a 1D geometry using a pure 1D code, lateral boundaries are not represented in the mesh. To transcribe the entry of water due to the boundary condition *Injector2*, we propose to transform the limit condition into a volumic source term. The here proposed formula are equivalent for a cell-centred finite volume discretisation in 2D with only one layer of cells in y-direction. The very easy formula of the source term should not pose any particular problem under the condition that adding source terms is allowed in the code.

We consider our 1D domain as a rectangular domain in 2D of length $10L$ and of depth $dy = L$. We consider an elementary control volume denoted k with length dx and depth dy which is centred in the point $(x, dy/2)$. The faces which we are interested in are situated at half the distance of the centre of cell in y-direction. The area of the face σ located at the north is $|\sigma| = dx$ and the distance between the cell centre and the face centre is $d_{k,\sigma} = dy/2$. The flux between the centre of the cell k and the boundary face σ can therefore be approached by the following classical formula:

$$Q_{Darcy} = T_{k_\sigma} * (P_k - P_\sigma),$$

where

$$T_{k,\sigma} = K/\mu_w * |\sigma|/d_{k,\sigma} = K/\mu_w * dx/(dy/2) = K/\mu_w * 2 * dx/dy$$

By dividing the results through the cell volume $|k| = dx * dy$, one obtains the following formula for the equivalent source term which has to be added in the pressure equation:

$$q_{Darcy}(x) = K/\mu_w * 2/(dy)^2 * (P(x) - P_{Inj2}) * \chi([7000, 8000]).$$

Concerning the transport equation, a simple upwind approximation gives the following formula:

$$q_C(x) = (q_{Darcy})^{(+)} * C(x) - (q_{Darcy})^{(-)} * C_{Inj2}$$

Species	Neutral water	Acid water	Molar mass (g)	Charge
Molality H_2O	55.509	55.509	18.0155	0
Molality H^+	0.10000E-06	0.25973E-04	1.008	1
Molality CO_2(aq)	0.14715E-04	0.82000	44.0095	0
Molality Cl^-	1.0784	1.0784	35.453	-1
Molality Na^+	1.0000	1.0002	22.990	1
Molality Ca^{+2}	0.39256E-01	0.48222E-01	40.078	2
Molality SiO_2(aq)	0.25929E-03	0.25929E-03	60.0840	0
Molality HCO_3^-	0.85286E-04	0.18032E-01	61.0170	-1
Molality OH^-	0.56024E-06	0.21258E-08	17.0075	-1
Mass fraction H_2O	0.9409	0.9088		
Molar fraction H_2O	0.9632	0.9493		
pH	7.0000	4.5855		
Charge balance	2.6254e-05	2.3797e-04		

Table A.15: Composition of initial water

The term q_{Darcy} is expressed in $[m^3.s^{-1}]$ per $[m^3]$, or in $[s^{-1}]$ while q_C is expressed in $[mol.s^{-1}]$ per $[m^3]$, i. e. by $[mol.s^{-1}.m^{-3}]$. Remark that the two added terms have a restricted support in the interval located between x = 7,000 m and x = 8,000 m. On the other hand, one has also to observe that those terms are affine with respect to the pressure and concentration unknowns. It is hence enough to integrate them properly in the system of equations.

A.12 Appendix 4. Chemical Composition for the Initial Water

A.12.1 Detailled Composition per Species

The composition of specified water corresponds to give an explicit value of the initial composition state by the form of the effective concentrations of the species which are in the solution. It is frequently obtained by the calculation of an equilibrium state under constraints. The indicated parameters in the lower part of the synthetic table are deduced from the indicated values in the upper part of the table. The charge balance is a simple verification of the coherence of the indicated results.

Recall of the calculation of concentrations

The unit concentration which is used for the aqueous solution is the molality $[mol.kg^{-1}_{H_2O}]$, denoted m_i. It is the fraction of the number of moles of a considered species and the mass of

solvent in the solution. In particular, the molality of the solvent, denoted m_{H_2O}, is a constant which depends only of its molar mass, denoted $MolWt(H_2O)$. The molality does not depend on the volumic mass of the aqueous phase.

The mass fraction of H_2O, denoted c_{H_2O}, makes use of the molar mass of the species which are in solution, denoted $MolWt(i)$. It allows to reference the measured quantities to one kilogram of phase solution instead of one kilogram of solvent species. To pass to molarities, i. e. to concentrations in $[mol.l^{-1}]$ denoted ζ_i, one has to multiply the result by the volumic mass of the aqueous solution.

$$\begin{aligned} m_i &= Molal(i) = \frac{n_i}{n_{H_2O} \times MolWt(H_2O)} \\ m_{H_2O} &= Molal(H_2O) = \frac{1}{MolWt(H_2O)} \simeq 55.509 \\ c_{H_2O} &= MassFraction(H_2O) = \frac{1}{\sum_i m_i \times MolWt(i)} \\ \zeta_i &= Molar(i) = \rho_w \times m_i \times c_{H_2O} \end{aligned}$$

A.12.2 Total Composition Deduced

By projecting the compositions into a basis, one can also deduce total equivalent compositions. Attention, the total of a basis species can be negative (cf. total H^+) and has to be used separately of the other totals. The charge balance is calculated up from numerical values given in the table. Up to rounding errors, one has to found the charge balances which have been calculated in the table of species compositions.

A.12.3 Initial Equilibrium State

Based on the initial compositions of the specified water, one can recalculate the log(Q/K) of the reactions in order to verify the level of internal disequilibrium in the water and between the water and other phases. By this, one can verify if the compositions are compatible with the imposed equilibrium hypothesis.

Remarks.

The cited activity models will interact in the this calculations. The activities of dissolved species is equal to its molality, the molar fraction is used for the solvent, the partial pressure for the gas and mineral species have an activity equal to 1.

Constraint	Neutral water	Acid water	Charge
Molality Total H_2O	5.5509e+01	5.5527e+01	0
Molality Total H^+	-8.5746e-05	-1.8006e-02	1
Molality Total CO_2(aq)	1.0000e-04	8.3803e-01	0
Molality Total Cl^-	1.0784e+00	1.0784e+00	-1
Molality Total Na^+	1.0000e+00	1.0002e+00	1
Molality Total Ca^{+2}	3.9256e-02	4.8222e-02	2
Molality Total SiO_2(aq)	2.5929e-04	2.5929e-04	0
Charge balances	2.6254e-05	2.3800e-04	

Table A.16: Composition of the initial water

Reaction	Type	Name	Neutral water	Acid water	Log10(K)
Req 1	Aqu	Water hydrolysis	-1.1587e-06	6.9524e-06	-13.235362
Req 2	Gas-Aqu	Dissolution of CO_2(g)	**-4.7461e+00**	-4.7616e-08	-2.0861861
Req 3	Aqu	Dissociation of CO_2(aq)	1.4208e-05	-1.3404e-05	-6.2206340
Rkin 1	Min-Aqu	Dissolution of Calcite	3.8910e-06	-1.2898e-05	1.5247799
Rkin 2	Min-Aqu	Precipitation of Calcite	-3.8910e-06	1.2898e-05	-1.5247799
Rkin 3	Min-Aqu	Dissolution of Quartz	1.7677e-06	1.7677e-06	-3.5862160
Rkin 4	Min-Aqu	Precipitation of Quartz	-1.7677e-06	-1.7677e-06	3.5862160

Table A.17: Initial log(Q/K) of the reactions

The initial neutral water is undersaturated with respect to CO_2 with an index of -4.7 which consists in a partial pressure of 0.002 bar of CO_2 compared to the 100 bar imposed in the gas zone. It is quite logic, that the gas is absent in the outer gas zone. Meanwhile, this is less important than situations which can be observed where water is in contact with the atmosphere. This explains also why the initial concentration of Calcium is relatively elevated in order to ensure an equilibrium with Calcite.

A.13 Bibliography

[1] Karsten Pruess, Chin-Fu Tsang, David Law, and Curt Oldenburg. *Intercomparison of simulation models for CO_2 disposal in underground storage reservoirs* (January 1, 2001). Lawrence Berkeley National Laboratory. Paper LBNL-47353. DOE Report 775181.
http://repositories.cdlib.org/lbnl/LBNL-47353

[2] Class, H., Ebigbo, A., Helmig, R., Dahle, H.K., Nordbotten, J.M., Celia, M.A., Audigane, P., Darcis, M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S.E., Jin, M., Krug, S., Labregere, D., Naderi Beni, A., Pawar, R.J., Sbai, A., Thomas, S.G., Trenty, L. and Wei, L. *A benchmark study on problems related to CO_2 storage in geologic formations: Summary and discussion of the results*. Journal of Comp. Geosc. 13(4), 409-434, 2009.
`http://www.hydrosys.uni-stuttgart.de/\ce{CO2}-workshop`

[3] A. Bourgeat, S. Bryant, J. Carrayrou, A. Dimier, H. C. J. Van Duijn, M. Kern, P. Knabner, and N. Leterrier. *GDR MoMaS Benchmark Reactive Transport* (2008).
`http://www.gdrmomas.org/Ex_qualif`

B

Carbon Footprint

Transportation and housing are, besides industry and power plants, the two major ejectors of greenhouse gases into atmosphere. During this Ph. D. thesis, I had the chance to attend several workshops, summer and winter schools, national and international scientific conferences. The transportation to reach the locations are the most weighty factor of the carbon footprint of this Ph. D. thesis. We give here an overview of the different journeys and their corresponding CO_2 emission. All emission values count the entire round trip and are calculated according to the websites `www.actioncarbone.org` for long distance itineraries and `www.ratp.fr` for public transport in the Parisan region. We omit local transportation between home and airports/train stations since they can be neglected (less than 500 g per itinerary by using public transport) except for the first entry which consists of my journeys at University Paris XIII. We omit in this carbon footprint all housing factors like electricity for light and computers, fossil and electric power for heating and cooling since they are difficult to measure in an isolated context.

Itinerary	**Means of travel**	**CO_2 emission**
Rueil-Malmaison — Villetaneuse	Public Transport (Suburban train and bus)	42 journeys at 338 g = 14 kg
Paris — Amsterdam	Train	44 kg
Paris — Marseille	Train	48 kg
Paris — Nice	Plane	204 kg
Paris — Pau	Plane	236 kg
Paris — Barcelona	Plane	268 kg
Paris — Ajaccio	Plane	351 kg
Paris — San Diego	Plane	2455 kg
		Total: 3620 kg

An average tree removes about 35 kg of CO_2 from the atmosphere during 3 years of a Ph. D. thesis. In order to neutralise the impact of the greenhouse gas CO_2 that has been ejected into atmosphere

by travelling during the Ph. D. thesis, about 104 trees had to “work on earth for my emissions” in the last three years. Alternatively, by planting one tree at the end of the thesis, one has to wait until the year 2322 (311 years up from now) before cutting it down until it will have consumed the ejected CO_2. The average maximal age of ash, black poplar, European beech or larch trees is about 300 years.

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] *High performance preconditioners*, https://computation.llnl.gov/casc/linear_solvers/sls_hypre.html.

[2] *Portable, Extensible Toolkit for Scientific Computation*, webpage http://www.mcs.anl.gov/petsc/petsc-as/.

[3] *BP Statistical Review of World Energy June 2011*, Tech. report, BP, June 2011, available online http://www.bp.com/statisticalreview.

[4] Y. Achdou, C. Japhet, Y. Maday, and F. Nataf, *A new cement to glue non-conforming grids with Robin interface conditions: The finite volume case*, Numerische Mathematik **92** (2002), no. 4, 593–620.

[5] L. Amir and M. Kern, *A global method for coupling transport with chemistry in heterogeneous porous media*, Computational Geosciences **14** (2010), 465–481.

[6] D. Bennequin, M. Gander, and L. Halpern, *A Homographic Best Approximation Problem with Application to Optimized Schwarz Waveform Relaxation*, Math. Comp. **78** (2009), no. 265, 185–223.

[7] E. Blayo, L. Halpern, and C. Japhet, *Optimized Schwarz waveform relaxation algorithms with nonconforming time discretization for coupling convection-diffusion problems with discontinuous coefficients*, Domain Decomposition Methods in Science and Engineering XVI (2007), 267–274.

[8] A. Bourgeat, S. Bryant, J. Carrayrou, A. Dimier, H. C. J. Van Duijn, M. Kern, P. Knabner, and N. Leterrier, *GDR MoMaS Benchmark Reactive Transport*, published online at http://www.gdrmomas.org/Ex_qualif/Geochimie/Documents/Benchmark-MoMAS.pdf, 2008.

[9] E. Brakkee and P. Wilders, *The Influence of Interface Conditions on Convergence of Krylov-Schwarz Domain Decomposition for the Advection-Diffusion Equation*, Journal of Scientific Computing **12** (1997), 11–30.

[10] F. Brezzi, K. Lipnikov, and M. Shashkov, *Convergence of mimetic finite difference method for diffusion problems on polyhedral meshes*, CONVERGENCE **43**, no. 5, 1872–1896.

[11] S. Brunauer, P. H. Emett, and E. Teller, *Adsorption of Gases in Multimolecular Layers*, Journal American Chemical Society **60** (1938), no. 2, 309–319.

[12] F. Caetano, L. Halpern, M. J. Gander, and J. Szeftel, *Schwarz Waveform Relaxation Algorithms with Nonlinear Transmission Conditions for Reaction-Diffusion Equations*, DD19-19th International Conference on Domain Decomposition Methods, Zhangjiajie, China, 2009.

[13] X. C. Cai, *Additive Schwarz algorithms for parabolic convection-diffusion equations*, Numerische Mathematik **60** (1991), no. 1, 41–61.

[14] ———, *Multiplicative Schwarz methods for parabolic problems*, SIAM Journal on Scientific Computing **15** (1994), no. 3, 587–603.

[15] X. C. Cai, W. D. Gropp, D. E. Keyes, and M. D. Tidriri, *Parallel implicit methods for aerodynamics*, In Keyes, Citeseer, 1994.

[16] X. C. Cai and D. E. Keyes, *Nonlinearly Preconditioned Inexact Newton Algorithms*, SIAM Journal on Scientific Computing **24** (2000), 183–200.

[17] J. Carrayrou, R. Mosé, and P. Behra, *New efficient algorithm for solving thermodynamic chemistry*, AIChE journal **48** (2002), no. 4, 894–904.

[18] C. Chniti, F. Nataf, and F. Nier, *Improved interface conditions for 2D domain decomposition with corners: a theoretical determination*, Calcolo **45** (2008), no. 2, 111–147.

[19] ———, *Improved Interface Conditions for 2 D Domain Decomposition with Corners: Numerical Applications*, Journal of Scientific Computing **38** (2009), no. 2, 207–228.

[20] P. Cresta, O. Allix, C. Rey, and S. Guinard, *Nonlinear localization strategies for domain decomposition methods: application to post-buckling analyses*, Computer Methods in Applied Mechanics and Engineering **196** (2007), no. 8, 1436–1446.

[21] L. B. da Veiga, J. Droniou, and G. Manzini, *A unified approach to handle convection terms in Finite Volumes and Mimetic Discretization Methods for elliptic problems*, (2010).

[22] B. Després, *A domain decomposition method for the helmholtz problem*, (1991).

[23] J. Droniou and R. Eymard, *A mixed finite volume scheme for anisotropic diffusion problems on any grid*, Numerische Mathematik **105** (2006), no. 1, 35–71.

[24] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, Math. Models Methods Appl. Sci **20** (2010), no. 2.

[25] G. Enchéry, R. Eymard, and A. Michel, *Numerical approximation of a two-phase flow problem in a porous medium with discontinuous capillary forces*, SIAM Journal on Numerical Analysis **43** (2006), no. 6, 2402–2422.

[26] B. Engquist and A. Majda, *Absorbing boundary conditions for numerical simulation of waves*, Proceedings of the National Academy of Sciences of the United States of America **74** (1977), no. 5, 1765.

[27] R. Eymard, T. Gallouët, and R. Herbin, *Discretisation of heterogeneous and anisotropic diffusion problems on general non-conforming meshes, sushi: a scheme using stabilisation and hybrid interfaces*, IMAJNA, see also `http://hal.archives-ouvertes.fr/docs/00/21/18/28/PDF/suchi.pdf` (2008).

[28] A. Fischer, *A special Newton-type optimization method*, Optimization **24** (1992), no. 3, 269–284.

[29] M. Gander, L. Halpern, and M. Kern, *A Schwarz Waveform Relaxation Method for Advection-Diffusion-Reaction Problems with Discontinuous Coefficients and non-Matching Grids*, DD16, New Work, USA, January 12–15, 2005, 2007.

[30] M. Gander, L. Halpern, and F. Nataf, *Optimal Schwarz Waveform Relaxation for the one dimensional Wave Equation*, SIAM Journal on Numerical Analysis **41** (2003), no. 5, 1643–1681.

[31] M. G. Gander and C. Japhet, *An algorithm for non-matching grid projections with linear complexity*, Domain Decomposition Methods in Science and Engineering XVIII (2009), 185–192.

[32] M. J. Gander, *Overlapping Schwarz for parabolic problems*, Ninth International Conference on Domain Decomposition Methods (Petter E. Bjørstad, Magne Espedal, and David Keyes, eds.), ddm. org, 1997, pp. 97–104.

[33] _____, *Schwarz methods over the course of time*, Electronic Transactions on Numerical Analysis **31** (2008), 228–255.

[34] M. J. Gander and L. Halpern, *Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems*, SIAM Journal on Numerical Analysis **45** (2008), no. 2, 666–697.

[35] M. J. Gander and F. Kwok, *Best Robin parameters for optimized Schwarz methods at cross points*.

[36] M. J. Gander and A. M. Stuart, *Space-time continuous analysis of waveform relaxation for the heat equation*, SIAM Journal on Scientific Computing **19** (1998), no. 6, 2014–2031.

[37] J. W. Gibbs, *On the Equilibrium of Heterogeneous Substances: The Collected Works*, Yale University Press, New Haven, CT **1** (1906), 63,96,332.

[38] E. Giladi and H. B. Keller, *Space-time domain decomposition for parabolic problems*, Numerische Mathematik **93** (2002), no. 2, 279–313.

[39] G. Grospellier and B. Lelandais, *The Arcane development framework*, Proceedings of the 8th workshop on Parallel/High-Performance Object-Oriented Scientific Computing, ACM, 2009, pp. 1–11.

[40] F. Haeberlein, *Reactive Transport Applied To CO_2 Geological Storage Modelling*, Master's thesis, University of Bayreuth, 2008.

[41] L. Halpern, *Artificial boundary conditions for the linear advection diffusion equation*, Mathematics of Computation **46** (1986), no. 174, 425–438.

[42] ———, *Optimized Schwarz Waveform Relaxation: Roots, Blossoms and Fruits*, Proceedings of the Eighteenth International Conference of Domain Decomposition Methods, Springer, 2009, pp. 225–232.

[43] L. Halpern and F. Hubert, *Optimized Schwarz algorithms in the classical finite volume framework*, unpublished (2009).

[44] L. Halpern and C. Japhet, *Discontinuous galerkin and nonconforming in time optimized schwarz waveform relaxation for heterogeneous problems*, Domain Decomposition Methods in Science and Engineering XVII (2008), 211–219.

[45] L. Halpern, C. Japhet, and J. Szeftel, *Discontinuous galerkin and nonconforming in time optimized schwarz waveform relaxation*, Domain Decomposition Methods in Science and Engineering XIX (2011), 133–140.

[46] ———, *Space-time nonconforming optimized schwarz waveform relaxation for heterogeneous problems and general geometries*, Domain Decomposition Methods in Science and Engineering XIX (2011), 75–86.

[47] L. Halpern, J. Szeftel, and C. Japhet, *Optimized Schwarz waveform relaxation and discontinuous Galerkin time stepping for heterogeneous problems*, Arxiv preprint arXiv:1006.2601 (2010).

[48] J. Hoffmann, *Reactive Transport and Mineral Dissolution/Precipitation in Porous Media: Efficient Solution Algorithms, Benchmark Computations and Existence of Global Solutions*, Ph. d. thesis, University of Erlangen-Nuremberg, 2010.

[49] C. Japhet, *Optimized Krylov-Ventcell method. Application to convection-diffusion problems*, Proceedings of the 9th International Conference on Domain Decomposition Methods, P. E. Bjørstad, M. S. Espedal, and D. E. Keyes, 1998, pp. 382–389.

[50] C. Japhet, F. Nataf, and F.-X. Roux, *The Optimized Order 2 Method with a coarse grid preconditioner. Application to convection-diffusion problems*, Ninth International Conference on Domain Decompositon Methods in Science and Engineering (P. Bjørstad, M. Espedal, and David Keyes, eds.), John Wiley & Sons, New York, 1999, pp. 382–389.

[51] D. A. Knoll and D. E. Keyes, *Jacobian-free Newton-Krylov methods: a survey of approaches and applications*, Journal of Computational Physics **193** (2004), 357–397.

[52] S. Kräutle, *General Mutli-Species Reactive Transport Problems in Porous Media: Efficient Numerical Approaches and Existence of Global Solutions*, Habiliation thesis, University of Erlangen-Nuremberg, March 2008.

[53] S. Kräutle and P. Knabner, *A reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions*, Water Resources Research **43** (2007), no. 3.

[54] E. Lelarasmee, A. E. Ruehli, and A. L. Sangiovanni-Vincentelli, *The waveform relaxation method for time-domain analysis of large scale integrated circuits*, IEEE Trans. CAD IC Systems, 1, 1982.

[55] J. L. Lions, Y. Maday, and G. Turinici, *Résolution d'EDP par un schéma en temps Ğpararéel ğ: A Ğparareal ğ in time discretization of PDE's*, Comptes Rendus de l'Académie des Sciences-Series I-Mathematics **332** (2001), no. 7, 661–668.

[56] J. L. Lions and E. Magenes, *Problemes aux limites non homogènes et applications*, vol. 3, Dunod, 1970.

[57] P.-L. Lions, *On the Schwarz alternating method. I*, First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Philadelphia) (R. Glowinski, G. H. Golub, G. A. Meurant, and J. Périaux, eds.), SIAM, 1988, p. 1–42.

[58] ———, *On the Schwarz alternating method. II: Stochastic Interpretation and Orders Properties*, Second International Symposium on Domain Decomposition Methods for Partial Differential Equations (Philadelphia) (T. F. Chan, R. Glowinski, J. Périaux, and O. B. Widlund, eds.), SIAM, 1989, p. 47–70.

[59] ———, *On the Schwarz alternating method. III: A variant for nonoverlapping subdomains*, Third International Symposium on Domain Decomposition Methods for Partial Differential Equations (Philadelphia) (T. F. Chan, R. Glowinski, J. Périaux, and O. B. Widlund, eds.), SIAM, 1990.

[60] V. Martin, *Méthodes de décomposition de domaine de type relaxation d'ondes pour des équations de l'océanographie*, PhD thesis, Université Paris 13, 2003.

[61] ———, *An optimized Schwarz waveform relaxation method for unsteady convection diffusion equation*, Appl. Numer. Math., no. 52, 2005, p. 401–428.

[62] ———, *Schwarz waveform relaxation method for the viscous shallow water equations*, Domain Decomposition Methods in Science and Engineering (2005), 653–660.

[63] K. U. Mayer and K. T. B. MacQuarrie, *MoMaS 1D and 2D advective benchmark problems – MIN3P implementation*, published online at `http://www-imfs.u-strasbg.fr/colloques/mrtpm2008/papers/Bench_Mayer_Quarrie.pdf`, 2008.

[64] G. A. Meurant, *Numerical experiments with a domain decomposition method for parabolic problems on parallel computers*, Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations, Philadelphia, PA, 1991.

[65] A. Michel, F. Haeberlein, and L. Trenty, *Cas Tests SHPCO2 n°3 — Test synthétique pour le transport réactif dans le cadre du stockage géologique de CO2*, (2009).

[66] S. Molins, J. Carrera, C. Ayora, and M. W. Saaltink, *A formulation for decoupling components in reactive transport problems*, Water Resources Research **40** (2004), no. 10, W10301.

[67] F. Morel and J. G. Hering, *Principles and applications of aquatic chemistry*, Wiley-Interscience, 1993.

[68] K. A. Morin, *Simplified explanations and examples of computerized methods for calculating chemical equilibrium in water*, Computers & Geosciences **11** (1985), no. 4, 409–416.

[69] J. Pebrel, C. Rey, and P. Gosselet, *A nonlinear dual domain decomposition method: application to structural problems with damage*, international journal of multiscale computational engineering **6** (2008), no. 3, 251–262.

[70] A. Quarteroni and A. Valli, *Domain Decomposition Methods for Partial Differential Equations*, Oxford Science Publications, London, 1999.

[71] M. W. Saaltink, C. Ayora, and J. Carrera, *A mathematical formulation for reactive transport that eliminates mineral concentrations*, Water Resour. Res. **34** (1998), 1649–1656.

[72] M. W. Saaltink, J. Carrera, and C. Ayora, *A comparison of two approaches for reactive transport modelling*, Journal of Geochemical Exploration **69–70** (2000), 97–101.

[73] _____, *A comparison of two approaches for reactive transport modelling*, Journal of Geochemical Exploration **69** (2000), 97–101.

[74] _____, *On the behavior of approaches to simulate reactive transport*, Journal of Contaminant Hydrology **48** (2001), 213–235.

[75] H. A. Schwarz, *Über einen Grenzübergang durch alternierendes Verfahren*, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, 15 ed., 1870, p. 272–286.

[76] B. F. Smith, P. E. Bjørstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, 1996.

[77] A. Toselli and O. B. Widlund, *Domain decomposition methods–algorithms and theory*, Springer Verlag, 2005.

[78] E. Weltin, *Are the equilibrium concentrations for a chemical reaction always uniquely determined by the initial concentrations?*, Journal of Chemical Education **67** (1990), no. 7, 548.

[79] G. T. Yeh and V. S. Tripathi, *A critical evaluation of recent developments in hydrogeochemical transport models of reactive multichemical components*, Water Resources Research **25** (1989), no. 1, 93–108.

[80] H. Zhao and M. J. Gander, *Overlapping Schwarz Waveform Relaxation for Parabolic Problems in Higher Dimensions*, (1997).

Résumé
Les modèles de transport réactif sont un outil basique pour la modélisation de l'interaction entre les réactions chimiques et l'écoulement du fluide dans un milieu poreux. Nous présentons un modèle de transport réactif multi-espèces totalement réduit incluant des réactions cinétiques et en équilibre. Une formulation structurée ainsi que différentes approches numériques sont proposées. Les méthodes de décomposition de domaine offrent la possibilité de diviser des problèmes de grande taille dans des problèmes plus petits dont la solution se fait en parallèle. Partant d'un point de vue géométrique, nous présentons la classe des méthodes de Schwarz ayant prouvé une haute performance dans de nombreuses applications. Des questions quant à la réalisation d'une décomposition de domaine et des conditions de transmission au niveau discret sont traitées dans le contexte des volumes finis. Nous proposons et validons numériquement un schéma de volumes finis hybrides pour l'opérateur d'advection-diffusion étant particulièrement adapté à l'utilisation dans le contexte d'une décomposition de domaine. Nous étudions théoriquement et numériquement des méthodes de Schwarz relaxation d'ondes en détail pour un système de deux espèces couplées de type transport réactif avec des termes de couplage linéaire et non-linéaire. Des résultats qualifiant le problème comme bien posé ainsi que la convergence des méthodes de décomposition de domaine sont développés et la sensibilité du comportement de convergence de l'algorithme de Schwarz par rapport au terme de couplage est étudiée. Finalement, nous appliquons une méthode de Schwarz relaxation d'ondes au modèle de transport réactif multi-espèces présenté.

Mots clefs : *transport réactif multi-espèces, réactions cinétiques, modèles couplés, approche globalement implicite, décomposition de domaine espace-temps, méthodes de Schwarz, Schwarz relaxation d'ondes optimisées, conditions de transmission optimisées, volumes finis hybrides, stockage géologique du CO_2, calcul haute-performance.*

Summary
Reactive transport modelling is a basic tool to model chemical reactions and flow processes in porous media. A totally reduced multi-species reactive transport model including kinetic and equilibrium reactions is presented. A structured numerical formulation is developed and different numerical approaches are proposed. Domain decomposition methods offer the possibility to split large problems into smaller subproblems that can be treated in parallel. The class of Schwarz-type domain decomposition methods that have proved to be high-performing algorithms in many fields of applications is presented with a special emphasis on the geometrical viewpoint. Numerical issues for the realisation of geometrical domain decomposition methods and transmission conditions in the context of finite volumes are discussed. We propose and validate numerically a hybrid finite volume scheme for advection-diffusion processes that is particularly well-suited for the use in a domain decomposition context. Optimised Schwarz waveform relaxation methods are studied in detail on a theoretical and numerical level for a two species coupled reactive transport system with linear and nonlinear coupling terms. Wellposedness and convergence results are developed and the influence of the coupling term on the convergence behaviour of the Schwarz algorithm is studied. Finally, we apply a Schwarz waveform relaxation method on the presented multi-species reactive transport system.

Keywords: *multi-species reactive transport, kinetic reactions, coupled models, global implicit approach, time-space domain decomposition, Schwarz methods, optimised Schwarz Waveform Relaxation, optimised transmission conditions, hybrid finite volumes, CO_2 geological storage, high performance computing.*