Reactive Transport and Mineral Dissolution/Precipitation in Porous Media: Efficient Solution Algorithms, Benchmark Computations and Existence of Global Solutions

Der Naturwissenschaftlichen Fakultät
der Friedrich–Alexander–Universität Erlangen–Nürnberg
zur
Erlangung des Doktorgrades Dr. rer. nat.

vorgelegt von
Joachim Hoffmann

aus Erlangen

Als Dissertation genehmigt von der Naturwissenschaftlichen Fakultät der Universität Erlangen–Nürnberg

Tag der mündlichen Prüfung:	4. März 2010
Vorsitzender der Promotionskommission:	Prof. Dr. Eberhard Bänsch
Erstberichterstatter:	Prof. Dr. Peter Knabner
Zweitberichterstatter:	Prof. Dr. Christian Wieners
Drittberichterstatter:	Dr. Michel Kern

Acknowledgements

First of all I am grateful to my supervisor Prof. Dr. Peter Knabner for making it possible to work on such a fascinating subject in the framework of my PhD thesis. Especially I would like to thank PD Dr. Serge Kräutle. I could come to him with questions at any time. This thesis would not have been possible without his preliminary work on this field. I am also indebted to PD Dr. Markus Bause who supports me especially at the beginning of my PhD studies. Next I am glad to express my gratitude to Prof. Dr. Christian Wieners, who made his software M++ available to me. Furthermore I would like to thank all my colleagues at the Chair of Applied Mathematics I for the excellent working atmosphere. Last but not least I thank my parents for their continuous support.

Erlangen, December 2009
Joachim Hoffmann

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are numerous examples of abandoned polluted areas caused by old industrial plants. It is not possible to remediate all of those areas. So it is necessary to get an assessment which polluted area is most dangerous. For example a nearby river can be at risk of contamination. Such an assessment can be given by a numerical simulation of the pollutant distribution in the soil. With help of the numerical results one can see if a pollutant reaches the river or will be degraded before. The needed mathematical models for reactive transport (see e.g. [SM96a], [SM96b], [Bet96]) are well known. So if all model parameters are known sufficiently accurate reliable predictions are possible.

1.1 Current State of the Research

One method for solving reactive transport problems, which is widely used, is operator splitting. Here the problem is split into a chemical and a transport problem. The advantage is that the problem decomposes in two subproblems and so this approach is easier to implement. Mainly there are two types of splitting schemes, the sequential non-iterative approach (SNIA) and the sequential iterative approach (SIA). For example the software SPECY uses a non-iterative operator splitting scheme (see [Car01]). But using non-iterative operator splitting schemes there is the problem that these methods lead to an operator splitting error (see e.g. [VM92], [BMCB97], [CMB04]). To circumvent this problem an iterative operator splitting scheme can be used. For example the software HYTEC uses an iterative operator splitting scheme (see [vdLWLG03]). But an iterative operator splitting scheme needs many iteration steps and requires small time step sizes to converge in chemically difficult cases and so it is not very effective (see e.g. [SCA00]).

The other method besides operator splitting for solving reactive transport problems is the global implicit approach (GIA). One method following this approach, which is known for several years, is the direct substitutional approach (DSA). For example the code MIN3P (see [MFB02], [May99]) uses this method. One disadvantage of DSA is that it leads to a nonlinear system which is difficult to solve numerically. Another disadvantage is that there is no decomposition in chemical problem and transport problem and so it is more difficult to implement. A comparison of DSA and SIA can be found in [SCA00].

Twenty years ago [YT89] concludes that operator splitting is preferable. But in the recent time the global implicit approach gets more and more popular and new global methods were developed. One is to use a differential algebraic equation (DAE) solver (see [dD08], [dDEK09]). Here the transport equations, the mass balance equations and the equations describing the chemical equilibrium are solved in one very large system of equations. The disadvantage is that this approach leads to excessive computation times. The other one is the global method out of [AK09], which uses a resolution function to handle the chemical problem. Such a resolution function is also applied in the reduction scheme which will be used in this work.

Solving reactive transport problems, the handling of the equilibrium conditions of equilibrium mineral reactions is a crucial point for the effectiveness of the code. In the literature different approaches for the handling of equilibrium minerals can be found. The first one is a swap procedure (see [Bet96], [CMB02]). If the solution is unphysical (negative mineral concentration or supersaturated mineral) the "most undersaturated mineral" is removed from the basis or, if there is no undersaturated mineral, the "most supersaturated mineral" is added to the basis and then a new solution is computed. This procedure is repeated until one gets a physical solution. The second possibility is to rewrite the reactive transport problem as a moving boundary problem with a generalized Rankine–Hugoniot condition (see [Lic85], [Lic96]). The third possibility is to rewrite the chemical subproblem as a minimization problem with constraints and to solve this problem with an optimization method, for example an interior point method (see [Saa96]).

In this work a new handling of equilibrium minerals suggested in [Krä08, Chap. 4] is used. The equilibrium condition is rewritten as a complementarity problem. Then the complementarity condition is replaced by an equivalent equation and the resulting problem is solved with a semismooth Newton method. The advantages are that only one Newton iteration is necessary (contrary to swap procedure), the same equations are valid on the whole domain (contrary to moving boundary) and no operator splitting between chemistry and transport is needed (contrary to minimization formulation).

Reaction networks with many species lead to coupled systems with many equations and so simulating such a problem requires much CPU time. So it is desirable to find strategies to reduce the amount of CPU time. In the literature one can find different reformulation techniques that try to reduce the number of equations in the coupled nonlinear system. The first one is the elimination of constant activity species (minerals, water) in [SAC98]. A more enhanced method, which leads to the decoupling of components in certain situations, can be found in [MCAS04]. For a comparison with the reduction scheme, which will be used in this work, see [KK07, Sec. 5.2, 5.3]. For the example considered there the number of coupled nonlinear differential equations by use of the reduction scheme is half the number as by use of the method out of [MCAS04]. In [Fri91], [FR92] a linear variable transformation is described causing the decoupling of some equations. A discussion of some problems arising when this method is used as a GIA method can be found in [KK05, Sec. 3].

This work is based on the newly developed reduction scheme which is proposed in [KK05], [KK07], [Krä08]. There a linear transformation of the equations and variables is performed such that some linear differential equations decouple form the nonlinear system. The key point is that the transformation is performed separately for the variables that correspond to mobile species and that ones corresponding to immobile species. In addition a resolution function eliminating the local equations (equilibrium conditions, ordinary differential equations for immobile species) is employed. The number of equations in the resulting coupled nonlinear system is always smaller or equal than by use of the Morel formulation (see [HKK09]). In [Hof05] simple problems were solved successfully with this new reduction scheme.

Regarding existence results for reactive transport problems, a proof of a global solution in the case of homogeneous kinetic reactions according to law of mass action can be found in [Krä08, Chap. 3]. Also in [Krä08, Chap. 3] there is a existence proof for heterogeneous kinetic reactions under certain restrictions on the exchange reactions.

1.2 Objective of this Work

The goal of this work is to modify the reduction scheme which is presented in [KK05], [KK07], [Krä08] and implemented in [Hof05] in such a way that it is possible to apply the reduction scheme also to realistic problems. Hence it is necessary that the modified reduction scheme can handle the numerical difficulties arising from concentration values varying over many orders of magnitude and from large reaction constants. To show that the modified reduction scheme can really be applied to realistic problems the MoMaS benchmark (see [BBC$^+$], [CKK09]), a numerically very challenging reactive transport benchmark, should be computed successfully. Furthermore the existence of a global solution for the kinetic mineral problem should be proven.

1.3 Overview of the Work

The used reactive transport model is described in Chapter 2. The model includes kinetic reactions according to law of mass action, equilibrium reactions according to law of mass action and mineral reactions in equilibrium. For Monod reactions it is referred to [KK05, Sec. 5] and kinetic mineral reactions are subject of Chapter 5.

The reduction scheme presented in Chapter 3 is an extension to that one in [KK05], [KK07], [Hof05], [Krä08]. In Section 3.1 the equations of the reduction scheme are derived by taking linear combinations of the original equations and performing a linear variable transformation. Doing so some linear partial differential equations decouple from the nonlinear system. In Section 3.2 the number of equations in the coupled nonlinear system is diminished even more with help of a resolution function. Some equations, that (after space discretization) depend only on the values of one nodal point, are solved for certain variables and are plugged into the other ones. The existence of such a resolution function is proven in two ways. In Section 3.3 the used discretization techniques are explained. As no explicit formula of the resolution function is known it is necessary to use a Newton iteration for the evaluation of the resolution function (Sec. 3.3.2). The evaluation of the resolution function is called local problem while solving the remaining coupled nonlinear system is called global problem.

In Section 3.4 the special numerical treatment due to the numerical difficulties of realistic problems is described. It is necessary to use the logarithms of the concentrations and a special solver for the linear system in the local problems because of concentration values varying over many orders of magnitude (Sec. 3.4.1). As the logarithms are used it is essential that the concentration values are positive. To ensure this it is necessary to modify the starting value of the global Newton iteration (Sec. 3.4.2) and to cut off the global Newton steps (Sec. 3.4.3). For convection dominated problems a stabilization is needed (Sec. 3.4.4). Also an anisotropic diffusion tensor can lead to negative concentration values. To avoid this an adapted grid is used (Sec. 3.4.5).

In Section 3.5 it is analyzed why this method has good convergence properties.

It is shown that the derivative of the resolution function, which appears in the global Jacobian matrix, is bounded (Sec. 3.5.1). For a representative example it is shown that for $\Delta t = 0$ the condition number of the global Jacobian matrix is bounded by a fixed number (Sec. 3.5.2) and that for the other limit case, a very large time step size, the problem decomposes in well conditioned subproblems (Sec. 3.5.3).

In Chapter 3.6 variants of the used formulation are mentioned. The three variants in the Sections 3.6.1-3.6.3 have less equations in the global problem but it turns out that all these variants are not applicable to realistic problems. In 3.6.1 the original formulation out of [KK07], [Krä08] is considered. Here the derivatives of the resolution function are not bounded and so this method is not convergent for realistic problems. In Section 3.6.2 some variables are eliminated to get a smaller coupled system. This variant has a ill-conditioned Jacobian for large time step sizes. So it converges only for very small time step sizes. In Section 3.6.3 some other variables are eliminated. It can be shown that the resulting method is equivalent to that one in Section 3.6.2. Instead of using a resolution function it is possible to eliminate the local equations on the linear level (Sec. 3.6.4). But it turns out that using a resolution function is much more efficient.

In Section 3.7 the implementation developed in the framework of this thesis is described. In Section 3.8 the connections between the reduction scheme and the widely used Morel formulation are shown: The variables used in the Morel formulation are linear combinations of the variables used in the reduction scheme. In absence of kinetic reactions the local problem of the reduction scheme is equivalent to the chemical subproblem of the Morel formulation and the equations in the transport problem of the Morel formulation are linear combinations of the equations of the global problem of the reduction scheme. So a chemical solver (Morel formulation) can be used to solve the local problem. In this sense a modular implementation of the reduction scheme is possible. In Section 3.9 a generalization of the reduction scheme is presented. The generalization is constructed in such a way that the reduction scheme out of the Sections 3.1-3.4 and the Morel formulation are special cases of the generalized formulation.

The MoMaS–benchmark [BBC$^+$] is a numerically very challenging reactive transport benchmark. Using the implementation of the reduction scheme computations of this benchmark were carried out. In Section 4.1 a short problem formulation is given. In Section 4.2 the reduction scheme is applied to the MoMaS–benchmark. The results of the computations can be found in Section 4.3. In Section 4.4 this results are compared with that ones of other benchmark participants. In Section 4.5 the implementation of the generalized formulation of the reduction scheme (see Sec. 3.9) is used to compare the reduction scheme with the global ODE approach and with iterative splitting (SIA). The different methods used by the benchmark participants are briefly presented in Section 4.6. Despite of the different methods used to solve the transport the results are very similar. So in Section 4.7 a suggestion for a second version of the benchmark is given, in which more differences are expected.

Chapter 5 handles the kinetic mineral problem. For this problem there are three different mathematical formulations. Concerning weak solutions these formulations are equivalent (Sec. 5.1). In Section 5.2 the three different formulations are compared regarding algorithmic aspects. In Section 5.3 the reduction scheme is applied to the kinetic mineral problem and with the resulting method travelling waves are computed. Last the existence of a global solution of the kinetic mineral problem is proven (Sec. 5.4).

Chapter 2

Mathematical Model

2.1 Mass Transport

The following physical quantities are required for the modelling:

- **concentration vector $\boldsymbol{c} = (c_1, \ldots, c_I)^T$:** amount of substance of all mobile species per volume water
- **water content θ:** volume water per total volume
- **Darcy flow $\boldsymbol{q}$:** volume water per time and cross-sectional area
- **concentration vector $\bar{\boldsymbol{c}} = (\bar{c}_{I+1}, \ldots, \bar{c}_{I+\bar{I}})^T$:** amount of substance of all immobile species per volume water

The number of the mobile species is denoted with I and the number of the immobile species with $\bar{I}$.

In this work the concentrations of the immobile species are given in amount of substance per volume water like in [YT89]. Another possibility for the unit of the immobile concentrations would be amount of substance per mass earth. The advantage of the choice taken here is that it is possible to add mobile and immobile concentrations without the need of multiplying one of them with a conversion factor.

Mass balance for the i-th mobile species leads to

$$\partial_t(\theta c_i) + \nabla \cdot (\boldsymbol{q} c_i) + \nabla \cdot \boldsymbol{j}_i = f_i\,, \qquad i = 1, \ldots, I$$

with the diffusive mass flow $\boldsymbol{j}_i$ and the source/sink term f_i. The mass flow $\boldsymbol{j}_i$ is caused by two different physical phenomena. The first one is mechanic dispersion

$$\boldsymbol{j}_{i,1} = -\theta \boldsymbol{D}_{mech} \nabla c_i$$

with the symmetric positive definite mechanic dispersion matrix $\boldsymbol{D}_{mech}$, which depends on $\boldsymbol{q}/\theta$. The second one is the molecular diffusion according to Fick's law

$$\boldsymbol{j}_{i,2} = -\theta d_{diff,i} \nabla c_i \,,$$

where $d_{diff,i}$ is the diffusion coefficient of the i-th species.

In the following the Scheidegger diffusion/dispersion tensor $\boldsymbol{D}_i$ (see [Sch61]) is used to describe these two phenomena

$$\boldsymbol{j}_i = -\Bigg(\underbrace{(\theta d_{diff,i} + \beta_t |\boldsymbol{q}|)\boldsymbol{I} + (\beta_l - \beta_t) \frac{\boldsymbol{q} \otimes \boldsymbol{q}}{|\boldsymbol{q}|}}_{=:\, \boldsymbol{D}_i} \Bigg) \nabla c_i \tag{2.1}$$

with the notation $\boldsymbol{I}$ for the identity matrix and the two parameters β_l and β_t, the longitudinal and the transversal dispersion coefficients, with $\beta_l > \beta_t$.

Altogether for every mobile species the partial differential equation

$$\partial_t(\theta c_i) - \nabla \cdot (\boldsymbol{D}_i \nabla c_i - \boldsymbol{q} c_i) = f_i \,, \qquad i = 1, \dots, I \tag{2.2}$$

is obtained. For every immobile species mass balance leads to the ordinary differential equation

$$\partial_t(\theta \bar{c}_i) = f_i \,, \qquad i = I+1, \dots, I+\bar{I} \,. \tag{2.3}$$

2.2 Chemical Reactions

The chemical reactions are given by the stoichiometric matrix $\boldsymbol{S}$. Each column of $\boldsymbol{S}$ corresponds to one chemical reaction. The number of chemical reactions is named J. The entries s_{ij}, called stoichiometric coefficients, specify if and on which scale a species takes part in a chemical reaction. A negative sign of the stoichiometric coefficient denotes that the species is an educt and a positive sign stands for a product.

The reaction rate vector $\boldsymbol{r} = (r_1, \dots, r_J)^T$ specifies how fast the chemical reactions proceed, i.e., how many moles per volume and time are reacting. These reaction rates appear in the source/sink term f_i. As all chemical reactions considered here can only take place in aqueous solution there is a factor θ in front of r_j. So the source/sink term is the following sum of the reaction rates

$$f_i = \sum_{j=1}^{J} \theta s_{ij} r_j \,.$$

2.2.1 Kinetic Reactions According to Law of Mass Action

Concerning kinetic reactions according to the law of mass action the rate term is given by the difference of the forward and the backward reaction rate (see e.g. [Bet96])

$$r_{kin,j}(\boldsymbol{c}, \bar{\boldsymbol{c}}) = k_{f,j} \prod_{\substack{i=1 \\ s_{ij}<0}}^{I+\bar{I}} c_i^{-s_{ij}} - k_{b,j} \prod_{\substack{i=1 \\ s_{ij}>0}}^{I+\bar{I}} c_i^{+s_{ij}} \,. \tag{2.4}$$

Here $k_{f,j}$ denotes the forward coefficient and $k_{b,j}$ the backward coefficient of the chemical reaction. For the sake of clarity the bars over the c_i regarding immobile species are left out. The number of reactions of this type is denoted with J_{kin}.

Another kind of kinetic reactions are biodegradation reactions, that can be described with help of the Monod model. A presentation of the model and how to apply the reduction mechanism (see Chap. 3) to this kind of reactions can be found in [KK05, Sec. 5].

2.2.2 Equilibrium Reactions According to Law of Mass Action

Reactions that are fast in comparison to the flow and dispersion/diffusion processes can be assumed to adopt an stationary state at all times, i.e., on every point an equilibrium condition holds. If the j-th equilibrium reaction can be described by the law of mass action the j-th equilibrium condition reads

$$\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}) := -\ln(K_j) + \sum_{i=1}^{I+\bar{I}} s_{ij} \ln(c_i) = 0 \tag{2.5}$$

with the equilibrium constant K_j.

In case of equilibrium reactions the reaction rate $r_{eq,j}$ is not known. So $r_{eq,j}$ gets an additional unknown and the equilibrium condition (2.5) is added to the system of equations (2.2), (2.3) as an additional equation.

2.2.3 Equilibrium Minerals

In the mineral case the sum analogous to that one in (2.5) does not depend on the mineral concentration. Furthermore we assume that no other immobile species take part in the mineral reaction. So if the j-th equilibrium reaction is a mineral reaction we define

$$\psi_j(\boldsymbol{c}) := -\ln(K_j) + \sum_{i=1}^{I} s_{ij} \ln(c_i) \tag{2.6}$$

where it is assumed that the stoichiometric coefficient of the mineral is positive. In this case the equilibrium condition consisting of equations and inequalities reads

$$\Big(\psi_j(\boldsymbol{c}) = 0 \wedge \bar{c}_{min,j} \geq 0\Big) \vee \Big(\psi_j(\boldsymbol{c}) > 0 \wedge \bar{c}_{min,j} = 0\Big)$$

where $\bar{c}_{min,j}$ denotes the concentration of that mineral taking part in the j-th equilibrium reaction. The first case, where the mineral is present, is called saturated and the second one, where no mineral is existing without saturation, is called undersaturated. It is possible to rewrite this condition as a complementarity condition (see [PHKK06] or [Krä08])

$$\psi_j(\boldsymbol{c})\, \bar{c}_{min,j} = 0 \wedge \psi_j(\boldsymbol{c}) \geq 0 \wedge \bar{c}_{min,j} \geq 0\,.$$

This complementarity condition can be replaced by the equivalent algebraic equation (see e.g. [Kan04])

$$\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) := \min\left\{\psi_j(\boldsymbol{c}),\, \bar{c}_{min,j}\right\} = 0\,. \tag{2.7}$$

This equation can be added to the system (2.2), (2.3) like in the case of equilibrium according to law of mass action.

2.3 Reactive Transport Model

The columns of the stoichiometric matrix $\boldsymbol{S}$ are sorted in the following way. First we take the columns associated with equilibrium reactions, then the columns associated with kinetic reactions. The submatrices with the index 1 contain the stoichiometric coefficients of the mobile species and the ones with index 2 the coefficients of the immobile species. So we get the block structure

$$\boldsymbol{S} = \begin{pmatrix}\boldsymbol{S}_{eq} & \boldsymbol{S}_{kin}\end{pmatrix} = \begin{pmatrix}\boldsymbol{S}_1 \\ \boldsymbol{S}_2\end{pmatrix} = \begin{pmatrix}\boldsymbol{S}_{1,eq} & \boldsymbol{S}_{1,kin} \\ \boldsymbol{S}_{2,eq} & \boldsymbol{S}_{2,kin}\end{pmatrix}\,. \tag{2.8}$$

The number of all equilibrium reactions is denoted J_{eq} .

In the present of kinetic and equilibrium reactions the source/sink terms f_i are

$$f_i = \sum_{j=1}^{J_{kin}} \theta s_{kin,ij} r_{kin,j}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \sum_{j=1}^{J_{eq}} \theta s_{eq,ij} r_{eq,j}\,.$$

Plugging this in (2.2), (2.3) and adding the equilibrium conditions (2.5), (2.7) leads to

$$\begin{aligned}
\partial_t(\theta c_i) + L_i c_i &= \sum_{j=1}^{J_{kin}} \theta s_{kin,ij} r_{kin,j}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \sum_{j=1}^{J_{eq}} \theta s_{eq,ij} r_{eq,j} && i = 1, \dots, I \\
\partial_t(\theta \bar{c}_i) &= \sum_{j=1}^{J_{kin}} \theta s_{kin,ij} r_{kin,j}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \sum_{j=1}^{J_{eq}} \theta s_{eq,ij} r_{eq,j} && i = I+1, \dots, I+\bar{I} \\
\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}) &= 0 && j = 1, \dots, J_{eq}
\end{aligned}$$

with the linear transport operator $L_i u := -\nabla \cdot (\boldsymbol{D}_i \nabla u - \boldsymbol{q} u)$. This system of equations is widely used for modelling reactive transport (compare e.g. [SAC98, eq. (19),(1)]). In matrix notation this system reads

$$\begin{aligned}
\partial_t(\theta \boldsymbol{c}) + L\boldsymbol{c} &= \theta \boldsymbol{S}_{1,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \theta \boldsymbol{S}_{1,eq} \boldsymbol{r}_{eq} \\
\partial_t(\theta \bar{\boldsymbol{c}}) &= \theta \boldsymbol{S}_{2,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \theta \boldsymbol{S}_{2,eq} \boldsymbol{r}_{eq} \qquad (2.9) \\
\boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) &= \boldsymbol{0}\,.
\end{aligned}$$

Chapter 3

The Reduction Scheme

The reduction scheme presented in this chapter is an extension to the reduction scheme described in [KK05], [KK07], [Hof05] and [Krä08]. To apply the reduction scheme the assumption that the diffusion coefficient $d_{diff,i}$ is the same for all species is needed. This assumption is justified because the molecular diffusion is small compared to the mechanic dispersion.

3.1 Transformation of the System of Equations

We sort the equilibrium reactions in the following order. First we take the reactions in that only mobile species take part, then the equilibrium sorption reactions, i.e., heterogeneous without mineral species, and last the equilibrium mineral reactions:

$$\boldsymbol{S}_{1,eq} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}, \qquad \boldsymbol{S}_{2,eq} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{pmatrix} \tag{3.1}$$

$\boldsymbol{I}_{J_{min}}$ denotes the identity matrix of the size J_{min}. The number of the reactions of each type is denoted with J_{mob}, J_{sorp} and J_{min}, respectively. Also the immobile species are sorted. Here we take the nonminerals first and then the minerals:

$$\bar{\boldsymbol{c}} = \begin{pmatrix} \bar{\boldsymbol{c}}_{nmin} \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix}$$

The number of nonminerals[1] is named $\bar{I}_{nmin}$ and the number of minerals $\bar{I}_{min}$.

We assume that we can write $\boldsymbol{S}_{1,sorp}$ in the form

$$\boldsymbol{S}_{1,sorp} = \begin{pmatrix} \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min}\boldsymbol{A}_{ld} \end{pmatrix} \tag{3.2}$$

[1] In this work nonminerals always denotes all *immobile* species that are not a mineral

with a coefficient matrix $\boldsymbol{A}_{ld}$ such that the columns of $\begin{pmatrix}\boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min}\end{pmatrix}$ are linear independent[2]. The number of columns of the matrix $\boldsymbol{S}_{1,sorp,li}$ is denoted $J_{sorp,li}$. Hence the coefficient matrix $\boldsymbol{A}_{ld}$ has the size $J_{min} \times (J_{sorp} - J_{sorp,li})$. When the linear independence condition is not fulfilled because some columns of $\begin{pmatrix}\boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp,li}\end{pmatrix}$ are linear dependent it is possible to apply preprocessing steps described in [KK07, Chap. 4] to make also in this situation the use of the reduction scheme possible. Furthermore we assume that the columns of $\boldsymbol{S}_{2,sorp}$ are linear independent.

Then the matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, containing all entries of $\boldsymbol{S}$ connected to mobile and immobile species, respectively, are of the form:

$$\begin{aligned}\boldsymbol{S}_1 &= \begin{pmatrix}\boldsymbol{S}_{1,eq} & \boldsymbol{S}_{1,kin}\end{pmatrix} = \begin{pmatrix}\boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min}\boldsymbol{A}_{ld} & \boldsymbol{S}_{1,min} & \boldsymbol{S}_{1,kin}\end{pmatrix}\\ \boldsymbol{S}_2 &= \begin{pmatrix}\boldsymbol{S}_{2,eq} & \boldsymbol{S}_{2,kin}\end{pmatrix} = \begin{pmatrix}\boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} & \tilde{\boldsymbol{S}}_{2,kin}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0}\end{pmatrix}\end{aligned} \tag{3.3}$$

It is not allowed that an equilibrium mineral participates in a kinetic reaction. So the stoichiometric coefficients in $\boldsymbol{S}_{2,kin}$ connected to minerals are zero. Hence it is possible to write $\boldsymbol{S}_{2,kin}$ as $\begin{pmatrix}\tilde{\boldsymbol{S}}_{2,kin}\\ \boldsymbol{0}\end{pmatrix}$.

With (3.3) we can rewrite the system (2.9) as

$$\partial_t(\theta \boldsymbol{c}) + L\boldsymbol{c} = \theta \boldsymbol{S}_1 \begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\end{pmatrix} \tag{3.4}$$

$$\partial_t(\theta \bar{\boldsymbol{c}}) = \theta \boldsymbol{S}_2 \begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\end{pmatrix} \tag{3.5}$$

$$\boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) = \boldsymbol{0}\,. \tag{3.6}$$

Now we define the matrices $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$ that contain a maximal system of linear independent columns of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively. Because of the linear independence assumption (see (3.2)) it is always possible to choose $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$ such that the matrices have the form

$$\boldsymbol{S}_1^* = \begin{pmatrix}\boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min} & \boldsymbol{S}_{1,kin}^*\end{pmatrix} \tag{3.7}$$

$$\boldsymbol{S}_2^* = \begin{pmatrix}\boldsymbol{S}_{2,sorp} & \boldsymbol{0} & \boldsymbol{S}_{2,kin}^*\\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0}\end{pmatrix} \tag{3.8}$$

where $\boldsymbol{S}_{1,kin}^*$ consists of some of the columns of $\boldsymbol{S}_{1,kin}$ and $\boldsymbol{S}_{2,kin}^*$ consists of some of the columns of $\tilde{\boldsymbol{S}}_{2,kin}$. For clarity we do not put a tilde on $\boldsymbol{S}_{2,kin}^*$. The number

[2] In [Krä08, Chap. 4] it is assumed that the columns of $\begin{pmatrix}\boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min}\end{pmatrix}$ are linear independent.

of columns of $\boldsymbol{S}^*_{1,kin}$ is denoted $J^*_{1,kin}$ and the number of columns of $\boldsymbol{S}^*_{2,kin}$ is denoted $J^*_{2,kin}$.

There are always matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ such that

$$\boldsymbol{S}_i = \boldsymbol{S}^*_i \boldsymbol{A}_i \qquad i = 1, 2\,. \tag{3.9}$$

With the block structure from (3.3) for $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and the block structure from (3.7), (3.8) for $\boldsymbol{S}^*_1$, $\boldsymbol{S}^*_2$, respectively, we get for $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ the block structure

$$\begin{aligned} \boldsymbol{A}_1 &= \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{1,mob} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{sorp,li}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{1,sorp} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{ld} & \boldsymbol{I}_{J_{min}} & \boldsymbol{A}_{1,min} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{1,kin} \end{pmatrix} \\ \boldsymbol{A}_2 &= \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} & \boldsymbol{A}_{2,sorp} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{2,kin} \end{pmatrix}\,. \end{aligned} \tag{3.10}$$

Then we construct the matrices $\boldsymbol{S}^{\perp}_1$ and $\boldsymbol{S}^{\perp}_2$. These matrices consist of a maximal system of linear independent vectors that are orthogonal to all columns of $\boldsymbol{S}^*_1$ and $\boldsymbol{S}^*_2$, respectively. So $\boldsymbol{S}^{\perp}_1$ has $I - J_{mob} - J_{sorp,li} - J_{min} - J^*_{1,kin}$ columns and $\boldsymbol{S}^{\perp}_2$ has $\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin}$ columns. We choose matrices $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ being of the same size as $\boldsymbol{S}^*_1$, $\boldsymbol{S}^*_2$, respectively, that fulfil the condition that the columns of $\boldsymbol{B}_i$, $\boldsymbol{S}^{\perp}_i$ form a basis of the whole space. Furthermore $\boldsymbol{B}_2$ should be of the form

$$\begin{pmatrix} * & \boldsymbol{0} & * \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix} \tag{3.11}$$

like it is $\boldsymbol{S}^*_2$. The simplest choice fulfilling these conditions is $\boldsymbol{B}_1 = \boldsymbol{S}^*_1$, $\boldsymbol{B}_2 = \boldsymbol{S}^*_2$. In [KK05], [KK07], [Krä08] the reduction scheme is only formulated for this choice. Analogously to $\boldsymbol{S}^{\perp}_1$, $\boldsymbol{S}^{\perp}_2$ matrices $\boldsymbol{B}^{\perp}_1$, $\boldsymbol{B}^{\perp}_2$ are constructed from $\boldsymbol{B}_1$, $\boldsymbol{B}_2$. There hold the orthogonality relations

$${\boldsymbol{S}^{\perp}_i}^T \boldsymbol{S}^*_i = \boldsymbol{0}, \qquad \boldsymbol{B}^T_i \boldsymbol{B}^{\perp}_i = \boldsymbol{0}\,. \tag{3.12}$$

With the condition that the columns of $\boldsymbol{B}_i$, $\boldsymbol{S}^{\perp}_i$ form a basis of the whole space it is possible to represent $\boldsymbol{S}^*_i$, $\boldsymbol{B}^{\perp}_i$ as

$$\boldsymbol{S}^*_i = \boldsymbol{B}_i \boldsymbol{N}_i + \boldsymbol{S}^{\perp}_i \boldsymbol{M}_i\,, \qquad\qquad \boldsymbol{B}^{\perp}_i = \boldsymbol{B}_i \boldsymbol{U}_i + \boldsymbol{S}^{\perp}_i \boldsymbol{V}_i\,.$$

with the quadratic coefficient matrices $\boldsymbol{N}_i$, $\boldsymbol{V}_i$ and the rectangular coefficient matrices $\boldsymbol{M}_i$, $\boldsymbol{U}_i$. Multiplication from left with $\boldsymbol{S}^{*T}_i$ and ${\boldsymbol{B}^{\perp}_i}^T$, respectively, leads to

$$\boldsymbol{S}^{*T}_i \boldsymbol{S}^*_i = \boldsymbol{S}^{*T}_i \boldsymbol{B}_i \boldsymbol{N}_i\,, \qquad\qquad {\boldsymbol{B}^{\perp}_i}^T \boldsymbol{B}^{\perp}_i = {\boldsymbol{B}^{\perp}_i}^T \boldsymbol{S}^{\perp}_i \boldsymbol{V}_i\,.$$

Since the columns of $\boldsymbol{S}_i^*$ and $\boldsymbol{B}_i^\perp$ are linear independent according to the construction of the matrices, the matrices on the left hand side are invertible. It follows that also the quadratic matrices $\boldsymbol{S}_i^{*T}\boldsymbol{B}_i$ and $\boldsymbol{B}_i^{\perp^T}\boldsymbol{S}_i^\perp$, respectively, on the right hand side as well as the transposed of these matrices are invertible.

Multiplying from left block (3.4) with

$$\left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T} \text{ and } (\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T$$

and block (3.5) with

$$\left(\boldsymbol{S}_2^{\perp^T}\boldsymbol{B}_2^\perp\right)^{-1}\boldsymbol{S}_2^{\perp^T} \text{ and } (\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T$$

and plugging in (3.9) leads to

$$\begin{aligned}
\left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T}\left(\partial_t(\theta\boldsymbol{c}) + L\boldsymbol{c}\right) &= \theta\left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T}\boldsymbol{S}_1^*\boldsymbol{A}_1\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\left(\partial_t(\theta\boldsymbol{c}) + L\boldsymbol{c}\right) &= \theta(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{S}_1^*\boldsymbol{A}_1\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
\left(\boldsymbol{S}_2^{\perp^T}\boldsymbol{B}_2^\perp\right)^{-1}\boldsymbol{S}_2^{\perp^T}\partial_t(\theta\bar{\boldsymbol{c}}) &= \theta\left(\boldsymbol{S}_2^{\perp^T}\boldsymbol{B}_2^\perp\right)^{-1}\boldsymbol{S}_2^{\perp^T}\boldsymbol{S}_2^*\boldsymbol{A}_2\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\partial_t(\theta\bar{\boldsymbol{c}}) &= \theta(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\boldsymbol{S}_2^*\boldsymbol{A}_2\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
\boldsymbol{\phi}(\boldsymbol{c},\bar{\boldsymbol{c}}) &= \boldsymbol{0}\,.
\end{aligned}$$

Since the matrices, the equations are multiplied with, are constant in time and space the matrix multiplication commutates with the differential operators. Using this and the orthogonality relations (3.12) we get

$$\begin{aligned}
\partial_t\left(\theta\left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T}\boldsymbol{c}\right) + L\left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T}\boldsymbol{c} &= \boldsymbol{0}\\
\partial_t\left(\theta(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c}\right) + L(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c} &= \theta\boldsymbol{A}_1\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
\partial_t\left(\theta\left(\boldsymbol{S}_2^{\perp^T}\boldsymbol{B}_2^\perp\right)^{-1}\boldsymbol{S}_2^{\perp^T}\bar{\boldsymbol{c}}\right) &= \boldsymbol{0}\\
\partial_t\left(\theta(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}\right) &= \theta\boldsymbol{A}_2\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}\\
\boldsymbol{\phi}(\boldsymbol{c},\bar{\boldsymbol{c}}) &= \boldsymbol{0}\,.
\end{aligned}$$

This implies the following definition of the new variables:

$$\begin{aligned}
\boldsymbol{\eta} &:= \left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^\perp\right)^{-1}\boldsymbol{S}_1^{\perp^T}\boldsymbol{c}, \quad \boldsymbol{\xi} := (\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c}\\
\bar{\boldsymbol{\eta}} &:= \left(\boldsymbol{S}_2^{\perp^T}\boldsymbol{B}_2^\perp\right)^{-1}\boldsymbol{S}_2^{\perp^T}\bar{\boldsymbol{c}}, \quad \bar{\boldsymbol{\xi}} := (\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}
\end{aligned} \tag{3.13}$$

The vector $\boldsymbol{\xi}$ has $J_{mob} + J_{sorp,li} + J_{min} + J^*_{1,kin}$ and the vector $\boldsymbol{\eta}$ has $I - J_{mob} - J_{sorp,li} - J_{min} - J^*_{1,kin}$ entries. So $\begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix}$ is an representation of the vector $\boldsymbol{c}$ regarding to another basis of the $\mathbb{R}^I$. Analogously $\bar{\boldsymbol{\xi}}$ has $J_{sorp} + J_{min} + J^*_{2,kin}$ and $\bar{\boldsymbol{\eta}}$ has $\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin}$ entries and the vector $\begin{pmatrix} \bar{\boldsymbol{\xi}} \\ \bar{\boldsymbol{\eta}} \end{pmatrix}$ is an representation of the vector $\bar{\boldsymbol{c}}$ regarding to another basis of the $\mathbb{R}^{\bar{I}}$. The variables $\boldsymbol{\xi}$ and $\bar{\boldsymbol{\xi}}$ are partitioned analogously to the partitioning of the columns of $\boldsymbol{S}^*_1$ and $\boldsymbol{S}^*_2$, respectively, in (3.7), (3.8) in

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix}, \qquad \bar{\boldsymbol{\xi}} = \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \\ \bar{\boldsymbol{\xi}}_{kin} \end{pmatrix}. \tag{3.14}$$

As retransformation we get

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{S}^*_1 \boldsymbol{\xi} + \boldsymbol{B}^\perp_1 \boldsymbol{\eta} \\ \bar{\boldsymbol{c}} &= \boldsymbol{S}^*_2 \bar{\boldsymbol{\xi}} + \boldsymbol{B}^\perp_2 \bar{\boldsymbol{\eta}}\,. \end{aligned}$$

That this is the retransformation of (3.13) can easily be seen by plugging the retransformation in (3.13) together with the orthogonality relations (3.12).

As the matrices $\boldsymbol{S}^*_2$ and $\boldsymbol{B}_2$ are of the form (3.11) the transformed variables $\bar{\boldsymbol{\xi}}_{sorp}$ and $\bar{\boldsymbol{\xi}}_{kin}$ depend only on $\bar{\boldsymbol{c}}_{nmin}$ and not on $\bar{\boldsymbol{c}}_{min}$. Furthermore as the second column block of (3.11) consists of unit vectors the last J_{min} entries in every column of $\boldsymbol{S}^\perp_2$ and $\boldsymbol{B}^\perp_2$ are always zero. Hence also $\bar{\boldsymbol{\eta}}$ depends only on $\bar{\boldsymbol{c}}_{nmin}$ and not on $\bar{\boldsymbol{c}}_{min}$. So we can write $\boldsymbol{B}^\perp_2$ as $\begin{pmatrix} \tilde{\boldsymbol{B}}^\perp_2 \\ \boldsymbol{0} \end{pmatrix}$. Using this and the partitionings (3.7), (3.8), (3.14) we can rewrite the retransformation as

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{S}_{1,mob} \boldsymbol{\xi}_{mob} + \boldsymbol{S}_{1,sorp,li} \boldsymbol{\xi}_{sorp} + \boldsymbol{S}_{1,min} \boldsymbol{\xi}_{min} + \boldsymbol{S}^*_{1,kin} \boldsymbol{\xi}_{kin} + \boldsymbol{B}^\perp_1 \boldsymbol{\eta} \\ \bar{\boldsymbol{c}} &= \begin{pmatrix} \boldsymbol{S}_{2,sorp} \bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin} \bar{\boldsymbol{\xi}}_{kin} + \tilde{\boldsymbol{B}}^\perp_2 \bar{\boldsymbol{\eta}} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix}. \end{aligned} \tag{3.15}$$

Then the equilibrium reaction rates $\boldsymbol{r}_{eq}$ and the equilibrium conditions $\boldsymbol{\phi}$ are partitioned analogously to $\boldsymbol{S}_{1,eq}$ in (3.1) in $\boldsymbol{r}_{eq} = \begin{pmatrix} \boldsymbol{r}_{mob} \\ \boldsymbol{r}_{sorp} \\ \boldsymbol{r}_{min} \end{pmatrix}$ and $\boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_{mob} \\ \boldsymbol{\phi}_{sorp} \\ \boldsymbol{\phi}_{min} \end{pmatrix}$, respectively. Furthermore the vector $\boldsymbol{r}_{sorp}$ is partitioned analogously to $\boldsymbol{S}_{1,sorp}$ in (3.2) in $\boldsymbol{r}_{sorp} = \begin{pmatrix} \boldsymbol{r}_{sorp,li} \\ \boldsymbol{r}_{sorp,ld} \end{pmatrix}$. Using this, the block structure of $\boldsymbol{A}_i$ (3.10), the definition of the transformed variables (3.13) and the partitioning of $\boldsymbol{\xi}$, $\bar{\boldsymbol{\xi}}$ (3.14) we get

$$\begin{aligned}
\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} &= \mathbf{0} \\
\partial_t\left(\theta\begin{pmatrix}\boldsymbol{\xi}_{mob}\\ \boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\\ \boldsymbol{\xi}_{kin}\end{pmatrix}\right) + L\begin{pmatrix}\boldsymbol{\xi}_{mob}\\ \boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\\ \boldsymbol{\xi}_{kin}\end{pmatrix} &= \theta\begin{pmatrix}\boldsymbol{I}_{J_{mob}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{1,mob}\\ \mathbf{0} & \boldsymbol{I}_{J_{sorp,li}} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{1,sorp}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{ld} & \boldsymbol{I}_{J_{min}} & \boldsymbol{A}_{1,min}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{1,kin}\end{pmatrix}\begin{pmatrix}\boldsymbol{r}_{mob}\\ \boldsymbol{r}_{sorp,li}\\ \boldsymbol{r}_{sorp,ld}\\ \boldsymbol{r}_{min}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix} \\
\partial_t(\theta\bar{\boldsymbol{\eta}}) &= \mathbf{0} \\
\partial_t\left(\theta\begin{pmatrix}\bar{\boldsymbol{\xi}}_{sorp}\\ \bar{\boldsymbol{\xi}}_{min}\\ \bar{\boldsymbol{\xi}}_{kin}\end{pmatrix}\right) &= \theta\begin{pmatrix}\mathbf{0} & \boldsymbol{I}_{J_{sorp}} & \mathbf{0} & \boldsymbol{A}_{2,sorp}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{J_{min}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{2,kin}\end{pmatrix}\begin{pmatrix}\boldsymbol{r}_{mob}\\ \boldsymbol{r}_{sorp}\\ \boldsymbol{r}_{min}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix} \\
\boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \mathbf{0} \\
\boldsymbol{\phi}_{sorp}(\boldsymbol{c},\bar{\boldsymbol{c}}_{nmin}) &= \mathbf{0} \\
\boldsymbol{\phi}_{min}(\boldsymbol{c},\bar{\boldsymbol{c}}_{min}) &= \mathbf{0}\,.
\end{aligned}$$

Analogously to $\boldsymbol{S}_{1,sorp}$ in (3.2) the vector $\bar{\boldsymbol{\xi}}_{sorp}$ is split in

$$\bar{\boldsymbol{\xi}}_{sorp} = \begin{pmatrix}\bar{\boldsymbol{\xi}}_{sorp,li}\\ \bar{\boldsymbol{\xi}}_{sorp,ld}\end{pmatrix}\,. \tag{3.16}$$

Expanding leads to

$$\begin{aligned}
\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} &= \mathbf{0} && (3.17)\\
\partial_t(\theta\boldsymbol{\xi}_{mob}) + L\boldsymbol{\xi}_{mob} &= \theta(\boldsymbol{r}_{mob} + \boldsymbol{A}_{1,mob}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})) && (3.18)\\
\partial_t(\theta\boldsymbol{\xi}_{sorp}) + L\boldsymbol{\xi}_{sorp} &= \theta(\boldsymbol{r}_{sorp,li} + \boldsymbol{A}_{1,sorp}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})) && (3.19)\\
\partial_t(\theta\boldsymbol{\xi}_{min}) + L\boldsymbol{\xi}_{min} &= \theta(\boldsymbol{r}_{min} + \boldsymbol{A}_{ld}\boldsymbol{r}_{sorp,ld} + \boldsymbol{A}_{1,min}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})) && (3.20)\\
\partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) && (3.21)\\
\partial_t(\theta\bar{\boldsymbol{\eta}}) &= \mathbf{0} && (3.22)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp,li}) &= \theta(\boldsymbol{r}_{sorp,li} + \boldsymbol{A}_{2,sorp,li}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})) && (3.23)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp,ld}) &= \theta(\boldsymbol{r}_{sorp,ld} + \boldsymbol{A}_{2,sorp,ld}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})) && (3.24)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{min}) &= \theta\boldsymbol{r}_{min} && (3.25)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{kin}) &= \theta\boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) && (3.26)\\
\boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \mathbf{0} && (3.27)\\
\boldsymbol{\phi}_{sorp}(\boldsymbol{c},\bar{\boldsymbol{c}}_{nmin}) &= \mathbf{0} && (3.28)\\
\boldsymbol{\phi}_{min}(\boldsymbol{c},\bar{\boldsymbol{c}}_{min}) &= \mathbf{0}\,. && (3.29)
\end{aligned}$$

Here $\boldsymbol{A}_{2,sorp,li}$ denotes the submatrix of $\boldsymbol{A}_{2,sorp}$ which contains the first $J_{sorp,li}$ rows of $\boldsymbol{A}_{2,sorp}$ and $\boldsymbol{A}_{2,sorp,ld}$ denotes the submatrix of $\boldsymbol{A}_{2,sorp}$ which contains the last $(J_{sorp} - J_{sorp,li})$ rows of $\boldsymbol{A}_{2,sorp}$.

To compute the equilibrium reaction rates $\boldsymbol{r}_{eq} = (\boldsymbol{r}_{mob}, \boldsymbol{r}_{sorp}, \boldsymbol{r}_{min})^T$ is not of interest. So when an equilibrium reaction rate appears in only one equation this equation can be left out. This is the case for the rate $\boldsymbol{r}_{mob}$ in block (3.18). By subtracting block (3.23) from block (3.19) and block (3.24) multiplied with $\boldsymbol{A}_{ld}$ from block (3.20) it can be achieved that $\boldsymbol{r}_{sorp,li}$ and $\boldsymbol{r}_{sorp,ld}$ appear only in block (3.23) and block (3.24), respectively. Likewise by subtracting block (3.25) from block (3.20) it can be achieved that $\boldsymbol{r}_{min}$ appears only in block (3.25). Then also the blocks (3.23)-(3.25) can be left out. Doing so we get

$$\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} = \boldsymbol{0} \tag{3.30}$$

$$\partial_t(\theta\boldsymbol{\xi}_{sorp}) + L\boldsymbol{\xi}_{sorp} = \partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp,li}) + \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp,li})\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \tag{3.31}$$

$$\begin{aligned}\partial_t(\theta\boldsymbol{\xi}_{min}) + L\boldsymbol{\xi}_{min} &= \partial_t(\theta\bar{\boldsymbol{\xi}}_{min}) + \boldsymbol{A}_{ld}\partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp,ld}) \\ &\quad + \theta(\boldsymbol{A}_{1,min} - \boldsymbol{A}_{ld}\boldsymbol{A}_{2,sorp,ld})\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\end{aligned} \tag{3.32}$$

$$\partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} = \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \tag{3.33}$$

$$\partial_t(\theta\bar{\boldsymbol{\eta}}) = \boldsymbol{0} \tag{3.34}$$

$$\partial_t(\theta\bar{\boldsymbol{\xi}}_{kin}) = \theta\boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \tag{3.35}$$

$$\boldsymbol{\phi}_{mob}(\boldsymbol{c}) = \boldsymbol{0} \tag{3.36}$$

$$\boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) = \boldsymbol{0} \tag{3.37}$$

$$\boldsymbol{\phi}_{min}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) = \boldsymbol{0}\,. \tag{3.38}$$

The blocks (3.30) and (3.34) decouple from the rest of the system. So these equations can be solved independent from the other ones. (3.30) consists of linear PDEs for that a numerical solution can be computed with standard methods. For θ constant in time (3.34) says that $\bar{\boldsymbol{\eta}}$ is always equal to its initial value. In numerical computations the decoupled equations (3.30) and (3.34) are solved at the beginning of every time step and afterwards the remaining system is solved. In the following we assume that the solution of the linear PDEs (3.30) and the ODEs (3.34) are known and focus on solving the remaining system.

The blocks (3.35)-(3.38) do not contain any space derivative. These equations are called local equations, because after space discretization these equations do only depend on the values of the variables at one point. In the next section we will see that (after time discretization of (3.35) with the implicit Euler method) there is a resolution function solving these blocks for certain variables. Then we can plug this resolution function in the remaining equations. So the system we have to solve consists only of the blocks (3.31)-(3.33).

3.2 Resolution Function

First we introduce the additional variables[3]

$$\tilde{\boldsymbol{\xi}} := \begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\xi}_{sorp} - \bar{\boldsymbol{\xi}}_{sorp,li} \\ \boldsymbol{\xi}_{min} - \bar{\boldsymbol{\xi}}_{min} - \boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} \end{pmatrix} . \tag{3.39}$$

The motivation for this definition is that the new variables $\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}$ are reaction invariant regarding to the equilibrium reactions. This holds because in case of no kinetic reactions and no transport (3.31) and (3.32) degenerate to $(\theta\tilde{\boldsymbol{\xi}}_{sorp}) = \mathbf{0}$, $(\theta\tilde{\boldsymbol{\xi}}_{min}) = \mathbf{0}$, respectively. Due to the additional variables there is a need for additional equations. Therefore the defining equations of $\tilde{\boldsymbol{\xi}}_{sorp}$ and $\tilde{\boldsymbol{\xi}}_{min}$ (3.39) are added to the system (3.30)-(3.38).

Because of the additional variables there is a freedom of choice in the formula for the calculation of the concentrations (3.15). Here the following possibility is chosen: The defining equations for $\tilde{\boldsymbol{\xi}}_{sorp}$ and $\tilde{\boldsymbol{\xi}}_{min}$ (3.39) are solved for $\boldsymbol{\xi}_{sorp}$ and $\boldsymbol{\xi}_{min}$ and plugged in the retransformation (3.15). So we get the new retransformation ((3.2) and (3.16) are used to simplify the new retransformation)

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{S}_{1,sorp,li}\tilde{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,min}(\tilde{\boldsymbol{\xi}}_{min} + \bar{\boldsymbol{\xi}}_{min}) \\ &\quad + \boldsymbol{S}^*_{1,kin}\boldsymbol{\xi}_{kin} + \boldsymbol{B}_1^{\perp}\boldsymbol{\eta} \\ \bar{\boldsymbol{c}} &= \begin{pmatrix} \boldsymbol{S}_{2,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \tilde{\boldsymbol{B}}_2^{\perp}\bar{\boldsymbol{\eta}} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix} . \end{aligned} \tag{3.40}$$

3.2.1 Existence of the Resolution Function

Then the unknowns are split in so-called local unknowns

$$\boldsymbol{\xi}_{loc} := \begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \\ \bar{\boldsymbol{\xi}}_{kin} \end{pmatrix} \tag{3.41}$$

and global unknowns

$$\boldsymbol{\xi}_{glob} := \begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \\ \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix} . \tag{3.42}$$

[3] In [KK05], [KK07], [Krä08] no additional variables are used. See Section 3.6.1 for the proceeding without additional variables.

Now we want to prove the existence of a resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ defined by the equations (3.35)-(3.38) (after time discretization of (3.35) with the implicit Euler method) and $(\boldsymbol{c}, \bar{\boldsymbol{c}})$ given by (3.40). Note that the resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ does not depend on the variables $\boldsymbol{\xi}_{sorp}$ and $\boldsymbol{\xi}_{min}$ because these variables do not appear in the equations (3.35)-(3.38) and in the retransformation (3.40). In a first step we show the existence of a resolution function

$$(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}, \bar{\boldsymbol{\xi}}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})$$

for the blocks (3.36)-(3.38) under the assumption that a positive bound for the concentrations exists. For that purpose the equilibrium mineral reactions are split in inactive (index $\mathcal{I}$) and active (index $\mathcal{A}$) reactions:

$$\boldsymbol{S}_{1,min} = \begin{pmatrix} \boldsymbol{S}_{1,min,\mathcal{I}} & \boldsymbol{S}_{1,min,\mathcal{A}} \end{pmatrix}$$

A reaction is called inactive when in $\min\{-\ln(K_j) + \sum_{i=1}^{I} s_{min,ij} \ln(c_i), \bar{c}_{min,j}\}$ the minimum is attained in the first argument and otherwise called active.

The following proof is adapted from [KK07, Appendix]. To apply the implicit function theorem we have to check that the matrix $\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})}$ is invertible. We get

$$\begin{aligned}
&\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} \\
&= \begin{pmatrix}
\frac{\partial \boldsymbol{\phi}_{mob}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,mob} & \frac{\partial \boldsymbol{\phi}_{mob}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,sorp} & \frac{\partial \boldsymbol{\phi}_{mob}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,min} \\
\frac{\partial \boldsymbol{\phi}_{sorp}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,mob} & \frac{\partial \boldsymbol{\phi}_{sorp}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,sorp} + \frac{\partial \boldsymbol{\phi}_{sorp}}{\partial \bar{\boldsymbol{c}}_{nmin}} \boldsymbol{S}_{2,sorp} & \frac{\partial \boldsymbol{\phi}_{sorp}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,min} \\
\frac{\partial \boldsymbol{\phi}_{min}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,mob} & \frac{\partial \boldsymbol{\phi}_{min}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,sorp} & \frac{\partial \boldsymbol{\phi}_{min}}{\partial \boldsymbol{c}} \boldsymbol{S}_{1,min} + \frac{\partial \boldsymbol{\phi}_{min}}{\partial \bar{\boldsymbol{c}}_{min}}
\end{pmatrix} \\
&= \begin{pmatrix}
\boldsymbol{S}_{1,mob}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,mob}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,mob}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,min} \\
\boldsymbol{S}_{1,sorp}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,sorp} + \boldsymbol{S}_{2,sorp}^T \bar{\boldsymbol{\Lambda}}_{nmin} \boldsymbol{S}_{2,sorp} & \boldsymbol{S}_{1,sorp}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,min} \\
\boldsymbol{S}_{1,min,\mathcal{I}}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,min,\mathcal{I}}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{min,\mathcal{I}}^T \boldsymbol{\Lambda} \boldsymbol{S}_{1,min} \\
\boldsymbol{0} & \boldsymbol{0} & \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{\mathcal{A}} \end{pmatrix}
\end{pmatrix}
\end{aligned}$$

with the diagonal matrices

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1/c_1 & & 0 \\ & \ddots & \\ 0 & & 1/c_I \end{pmatrix}, \qquad \bar{\boldsymbol{\Lambda}}_{nmin} = \begin{pmatrix} 1/\bar{c}_{I+1} & & 0 \\ & \ddots & \\ 0 & & 1/\bar{c}_{I+\bar{I}_{nmin}} \end{pmatrix}.$$

We can rewrite the matrix as

$$\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} = \begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{0} & \boldsymbol{I}_{\mathcal{A}} \end{pmatrix}$$

with

$$\boldsymbol{M}_1 = \begin{pmatrix} \boldsymbol{S}_{1,mob}^T & \boldsymbol{0} \\ \boldsymbol{S}_{1,sorp}^T & \boldsymbol{S}_{2,sorp}^T \\ \boldsymbol{S}_{1,min,\mathcal{I}}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{\Lambda}}_{nmin} \end{pmatrix} \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min,\mathcal{I}} \\ \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}$$

$$\boldsymbol{M}_2 = \begin{pmatrix} \boldsymbol{S}_{1,mob}^T \\ \boldsymbol{S}_{1,sorp}^T \\ \boldsymbol{S}_{1,min,\mathcal{I}}^T \end{pmatrix} \boldsymbol{\Lambda} \boldsymbol{S}_{1,min,\mathcal{A}} \, .$$

Due to the linear independence assumption (see after (3.2)) the columns of the matrices $\begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,min} \end{pmatrix}$ and $\boldsymbol{S}_{2,sorp}$ are linear independent. Therewith also the columns of the matrix $\begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min,\mathcal{I}} \\ \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}$ are linear independent. It follows that the matrix $\boldsymbol{M}_1$ is symmetric positive definite because the matrix $\begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{\Lambda}}_{nmin} \end{pmatrix}$ is a diagonal matrix with only positive entries. So all diagonal blocks of the upper triangular matrix $\begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{0} & \boldsymbol{I}_{\mathcal{A}} \end{pmatrix}$ are invertible and therefore the whole matrix is invertible.

The next steps to show the existence of the resolution function

$$(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})$$

for sufficiently small Δt are exactly the same as in the proof of the existence of a resolution function $(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{kin}, \bar{\boldsymbol{\xi}}_{immo})$, which is done in [KK07].

Using this resolution function and assuming that the η-equations (3.30) and (3.34) are solved, (3.30)-(3.38) together with (3.39) can be reduced to

$$\tilde{\boldsymbol{\xi}}_{sorp} = \boldsymbol{\xi}_{sorp} - \bar{\boldsymbol{\xi}}_{sorp,li}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \tag{3.43}$$

$$\begin{aligned} \tilde{\boldsymbol{\xi}}_{min} = \boldsymbol{\xi}_{min} &- \bar{\boldsymbol{\xi}}_{min}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \\ &- \boldsymbol{A}_{ld} \bar{\boldsymbol{\xi}}_{sorp,ld}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \end{aligned} \tag{3.44}$$

$$\partial_t(\theta \tilde{\boldsymbol{\xi}}_{sorp}) + L\boldsymbol{\xi}_{sorp} = \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp,li})\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \tag{3.45}$$

$$\partial_t(\theta \tilde{\boldsymbol{\xi}}_{min}) + L\boldsymbol{\xi}_{min} = \theta(\boldsymbol{A}_{1,min} - \boldsymbol{A}_{ld}\boldsymbol{A}_{2,sorp,ld})\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \tag{3.46}$$

$$\partial_t(\theta \boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} = \theta \boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \tag{3.47}$$

The reaction rates $\boldsymbol{r}_{kin}$ are written as a function of $\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}$. This can be achieved by (3.40) and the resolution function. This nonlinear system will be used for the numerical computations.

3.2.2 As Minimum Problem

When there are no variables $\bar{\boldsymbol{\xi}}_{kin}$ it is possible to rewrite the local problem as minimization problem. In [Krä08, Sec. 2.4.4] this is done for the case that there are no equilibrium minerals and without using the variables $\tilde{\boldsymbol{\xi}}$.

First we define the functional

$$G(\boldsymbol{c}, \bar{\boldsymbol{c}}) := \sum_{i=1}^{I} \mu_i(c_i) c_i + \sum_{i=I+1}^{I+\bar{I}} \bar{\mu}_i(\bar{c}_i) \bar{c}_i$$

where

$$\mu_i(c_i) := \mu_{0,i} - 1 + \ln(c_i)$$

$$\bar{\mu}_i(\bar{c}_i) := \begin{cases} \bar{\mu}_{0,i} - 1 + \ln(\bar{c}_{nmin,i}) & \text{for } i = I+1, \dots, I+\bar{I}_{nmin} \\ \bar{\mu}_{0,i} & \text{for } i = I+\bar{I}_{nmin}+1, \dots, I+\bar{I}\,. \end{cases}$$

Thereby the vector $\begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_0 \end{pmatrix} \in \mathbb{R}^{I+\bar{I}}$ is a solution of the linear system

$$\boldsymbol{S}_{eq}^T \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_0 \end{pmatrix} = -\ln(\boldsymbol{K})\,. \tag{3.48}$$

We calculate

$$\nabla G(\boldsymbol{c}, \bar{\boldsymbol{c}}) = \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_{0,nmin} \\ \bar{\boldsymbol{\mu}}_{0,min} \end{pmatrix} + \begin{pmatrix} \ln(\boldsymbol{c}) \\ \ln(\bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{0} \end{pmatrix}.$$

We see that ∇G is monotonically increasing for positive concentration values. So it follows that G as a function of $(\boldsymbol{c}, \bar{\boldsymbol{c}})$ is a convex functional.

With help of the retransformation (3.40) and for given values for the variables $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$, $\boldsymbol{\xi}_{kin}$, $\bar{\boldsymbol{\eta}}$ we can write G as a function of the variables $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{min}$. Note that in this subsection we assume that there are no variables $\bar{\boldsymbol{\xi}}_{kin}$. Also as a function of $(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})$ G is a convex functional. This can be seen as follows. Using the fact that $\frac{\partial(\boldsymbol{c},\bar{\boldsymbol{c}})}{\partial(\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min})} = \boldsymbol{S}_{eq}$, which follows immediately from (3.40), we compute

$$\nabla_{(\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min})} G = \boldsymbol{S}_{eq}^T \nabla_{(\boldsymbol{c},\bar{\boldsymbol{c}})} G = \boldsymbol{S}_{eq}^T \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_{0,nmin} \\ \bar{\boldsymbol{\mu}}_{0,min} \end{pmatrix} + \boldsymbol{S}_{eq}^T \begin{pmatrix} \ln(\boldsymbol{c}) \\ \ln(\bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{0} \end{pmatrix} \tag{3.49}$$

Using again the fact that $\frac{\partial(\boldsymbol{c},\bar{\boldsymbol{c}})}{\partial(\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min})} = \boldsymbol{S}_{eq}$ we get for the second derivatives of G

$$D^2_{(\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min})} G = \boldsymbol{S}_{eq}^T \boldsymbol{\Lambda} \boldsymbol{S}_{eq}$$

with $\boldsymbol{\Lambda} = \operatorname{diag}(1/c_1, \dots, 1/c_I, 1/\bar{c}_{I+1}, \dots, 1/\bar{c}_{I+\bar{I}_{nmin}}, 0, \dots, 0)$. We know that the matrix $\boldsymbol{S}_{eq}^T \boldsymbol{\Lambda} \boldsymbol{S}_{eq}$ is positive semidefinite because $\boldsymbol{\Lambda}$ is a diagonal matrix with nonnegative entries. Hence G as a function of $(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})$ is a convex functional.

Now we consider the minimization problem

$$\begin{aligned} &\min \; G(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ &s.t. \quad \bar{\boldsymbol{\xi}}_{min} \geq \boldsymbol{0}\,. \end{aligned} \tag{3.50}$$

The Lagrange functional of this minimization problem reads

$$\mathcal{L}(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \boldsymbol{\nu}) = G(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) - \bar{\boldsymbol{\xi}}_{min} \cdot \boldsymbol{\nu}\,.$$

Using (3.49) and (3.48) we get for the associated KKT system

$$\begin{aligned} 0 &= \nabla_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} \mathcal{L} \\ &= \boldsymbol{S}_{eq}^T \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_{0,nmin} \\ \bar{\boldsymbol{\mu}}_{0,min} \end{pmatrix} + \boldsymbol{S}_{eq}^T \begin{pmatrix} \ln(\boldsymbol{c}) \\ \ln(\bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{\nu} \end{pmatrix} \\ &= -\ln(\boldsymbol{K}) + \boldsymbol{S}_{eq}^T \begin{pmatrix} \ln(\boldsymbol{c}) \\ \ln(\bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{\nu} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\phi}_{mob}(\boldsymbol{c}) \\ \boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{\psi}_{min}(\boldsymbol{c}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{\nu} \end{pmatrix}, \\ \nu_j \, \bar{\xi}_{min,j} &= 0, \bar{\xi}_{min,j} \geq 0, \nu_j \geq 0 \qquad j = 1, \dots, J_{min} \end{aligned}$$

where the entries of $\boldsymbol{\psi}_{min}$ are defined according to (2.6). Hence the minimization problem (3.50) is equivalent to solving the equilibrium conditions (3.36)-(3.38). So the resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ of the previous subsection can also be defined as the solution of the minimization problem (3.50) when there are no variables $\bar{\boldsymbol{\xi}}_{kin}$. The difference to [Krä08, Sec. 2.4.4], where the additional variables $\tilde{\boldsymbol{\xi}}$ are not used, is that we can write the local problem as *one* minimization problem while there two coupled minimization problems are needed.

3.3 Discretization

The linear partial differential equation (3.30) and the system of equations (3.43)-(3.47) should be solved numerically on the space-time-cylinder $\Omega \times (0, T)$ with

$T > 0$ and $\Omega \subset \mathbb{R}^2$ bounded Lipschitz domain. We need boundary conditions for the variables $\boldsymbol{\eta}$, $\boldsymbol{\xi}_{sorp}$, $\boldsymbol{\xi}_{min}$ and $\boldsymbol{\xi}_{kin}$. As boundary conditions Dirichlet- and homogeneous Neumann boundary conditions are considered, so e.g. for $\boldsymbol{\eta}$ we have

$$\begin{aligned} \eta_i &= \eta_{D,i} && \text{on } \Gamma_D \times (0,T) \\ \boldsymbol{D}\nabla\eta_i \cdot \boldsymbol{\nu} &= 0 && \text{on } \Gamma_N \times (0,T) \end{aligned}$$

with $\boldsymbol{\nu}$ the outer normal of the domain. The boundary parts Γ_D, Γ_N form a disjunct partitioning of $\partial\Omega$

$$\partial\Omega = \Gamma_D \,\dot{\cup}\, \Gamma_N$$

where Γ_D is closed.

Dirichlet boundary conditions are used at the inflow boundary ($\boldsymbol{q}\cdot\boldsymbol{\nu} < 0$) while homogeneous Neumann boundary conditions are used at the rest of the boundary. At that parts of the boundary where the flow is tangential to the boundary ($\boldsymbol{q}\cdot\boldsymbol{\nu} = 0$) homogeneous Neumann boundary conditions describe an impermeable barrier. At the outflow boundary ($\boldsymbol{q}\cdot\boldsymbol{\nu} > 0$) homogeneous Neumann boundary conditions imply that the solute can leave the domain driven by advection but not driven by dispersion or diffusion.

When Dirichlet values for the concentrations $\boldsymbol{c}$ are given the needed values for the variables $\boldsymbol{\eta}$, $\boldsymbol{\xi}_{sorp}$, $\boldsymbol{\xi}_{min}$ and $\boldsymbol{\xi}_{kin}$ can be calculated with the definition of these variables (see (3.13), (3.14)).

Additionally we need initial values for the variables $\boldsymbol{\eta}$, $\bar{\boldsymbol{\eta}}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$, $\boldsymbol{\xi}_{kin}$ and $\bar{\boldsymbol{\xi}}_{kin}$. Because of equation (3.34) $\bar{\boldsymbol{\eta}}$ is always equal to its initial value for θ constant in time. $\bar{\boldsymbol{\xi}}_{kin}$ is computed in the local problem with help of the ordinary differential equation (3.35). Again for given initial values for the concentrations $(\boldsymbol{c}, \bar{\boldsymbol{c}})$ the needed values for the transformed variables can be calculated with their definitions (see (3.13), (3.14), (3.39)).

3.3.1 Space and Time Discretization

The used discretization for the partial differential equations should be illustrated with the model equation

$$\partial_t(\theta u) - \nabla \cdot (\boldsymbol{D}\nabla u - \boldsymbol{q}u) = G(u).$$

This equation contains all types of terms that appear in the partial differential equations (3.30), (3.45)-(3.47) (terms with time derivative, second space derivative, first space derivative and a nonlinear term not depending on any derivative).

First we need a variational formulation of this equation. We define the function space

$$H^1_{0,D}(\Omega) := \left\{ v \in H^1(\Omega) | \gamma_0(v) = 0 \text{ on } \Gamma_D \right\}$$

where γ_0 denotes the trace operator. First it is assumed that the given Dirichlet values on Γ_D are zero. We multiply the partial differential equation with a test function $v \in H^1_{0,D}(\Omega)$ and integrate over the domain Ω:

$$\int_\Omega \partial_t(\theta u)v\, d\boldsymbol{x} - \int_\Omega \nabla \cdot (\boldsymbol{D}\nabla u)v\, d\boldsymbol{x} + \int_\Omega \nabla \cdot (\boldsymbol{q}u)v d\boldsymbol{x} = \int_\Omega G(u)v\, d\boldsymbol{x}$$

Partial integration of the diffusive term gives:

$$\int_\Omega \partial_t(\theta u)v\, d\boldsymbol{x} + \int_\Omega \boldsymbol{D}\nabla u \cdot \nabla v\, d\boldsymbol{x} - \int_{\partial\Omega} \boldsymbol{D}\nabla u \cdot \boldsymbol{\nu} v\, d\sigma + \int_\Omega \nabla \cdot (\boldsymbol{q}u)v d\boldsymbol{x} = \int_\Omega G(u)v\, d\boldsymbol{x}$$

The boundary integral vanishes because $v = 0$ on Γ_D and $\boldsymbol{D}\nabla u \cdot \boldsymbol{\nu} = 0$ on Γ_N. So we get the variational formulation for homogeneous Dirichlet boundary conditions:
Find $u \in L^2((0,T), H^1_{0,D}(\Omega))$ *with* $u' \in L^2((0,T), L^2(\Omega))$ *such that for almost every* $t \in (0,T)$

$$\begin{aligned}\int_\Omega \partial_t(\theta(\cdot,t)u(t))v\, d\boldsymbol{x} + \int_\Omega \boldsymbol{D}(\cdot,t)\nabla u(t) \cdot \nabla v\, d\boldsymbol{x} + \int_\Omega \nabla \cdot (\boldsymbol{q}(\cdot,t)u(t))v d\boldsymbol{x} \\ = \int_\Omega G(u(t))v\, d\boldsymbol{x} \qquad \forall v \in H^1_{0,D}(\Omega)\end{aligned}$$

and

$$u(0) = u_0$$

The case inhomogeneous Dirichlet boundary conditions can be reduced to the homogeneous case by replacing u through $\tilde{u} + w$ with $\tilde{u} \in L^2((0,T), H^1_{0,D}(\Omega))$ and $w \in H^1(\Omega)$ where w takes the given boundary values, i.e., $\gamma_0(w) = u_D$ on Γ_D. Therewith we get the variational formulation:
Find $\tilde{u} \in L^2((0,T), H^1_{0,D}(\Omega))$ *with* $\tilde{u}' \in L^2((0,T), L^2(\Omega))$ *such that for almost every* $t \in (0,T)$

$$\begin{aligned}\int_\Omega \partial_t(\theta(\cdot,t)\tilde{u}(t))v\, d\boldsymbol{x} + \int_\Omega \boldsymbol{D}(\cdot,t)\nabla \tilde{u}(t) \cdot \nabla v\, d\boldsymbol{x} + \int_\Omega \nabla \cdot (\boldsymbol{q}(\cdot,t)\tilde{u}(t))v d\boldsymbol{x} \\ = -\int_\Omega \partial_t(\theta(\cdot,t)w)v\, d\boldsymbol{x} - \int_\Omega \boldsymbol{D}(\cdot,t)\nabla w \cdot \nabla v\, d\boldsymbol{x} - \int_\Omega \nabla \cdot (\boldsymbol{q}(\cdot,t)w)v d\boldsymbol{x} \\ + \int_\Omega G(\tilde{u}(t) + w)v\, d\boldsymbol{x} \qquad \forall v \in H^1_{0,D}(\Omega)\end{aligned}$$

and

$$\tilde{u}(0) = u_0 - w$$

For the space discretization conform Finite Elements are used. Therefore a triangulation $\mathcal{T}_h$ of the domain Ω is needed. Only triangulations consisting of triangles or parallelograms are considered. Furthermore only Lagrange elements are considered, i.e., all degrees of freedom are function values at certain points of the element $T \in \mathcal{T}_h$. In this work these points are called nodes. The number of the nodes is denoted with M. Further let the nodes $\boldsymbol{a}_i$ be numerated in such a way that

$$\begin{aligned} \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{M_1} &\in \Omega \cup \Gamma_N \\ \boldsymbol{a}_{M_1+1}, \ldots, \boldsymbol{a}_M &\in \Gamma_D \,. \end{aligned}$$

Let the basis functions φ_i be polynomial on each element $T \in \mathcal{T}_h$ and let them fulfil

$$\varphi_i(\boldsymbol{a}_j) = \delta_{ij} \,.$$

Using the Finite Element method the basic space $H^1_{0,D}(\Omega)$ is replace by the finite dimensional space

$$V_h := \operatorname{span}\{\varphi_1, \ldots, \varphi_{M_1}\} \,.$$

Thus a function $u_h(t) = \sum_{i=1}^{M} d_i(t)\varphi_i$ is searched which fulfills the variational formulation for all test functions $v_h \in V_h$ for almost every t, the Dirichlet condition for every Dirichlet node $\boldsymbol{a}_j$ $(j = M_1 + 1, \ldots, M)$ and the initial condition for every node $\boldsymbol{a}_j$ $(j = 1, \ldots, M)$:
Find u_h of the form $u_h(t) = \sum_{i=1}^{M} d_i(t)\varphi_i$ such that for almost every $t \in (0, T)$

$$\begin{aligned} \int_\Omega \partial_t(\theta(\cdot,t)u_h(t))v_h \, d\boldsymbol{x} + \int_\Omega \boldsymbol{D}(\cdot,t)\nabla u_h(t) \cdot \nabla v_h \, d\boldsymbol{x} + \int_\Omega \nabla \cdot (\boldsymbol{q}(\cdot,t)u_h(t))v_h \, d\boldsymbol{x} \\ = \int_\Omega G(u_h(t))v_h \, d\boldsymbol{x} \qquad\qquad \forall v_h \in V_h \\ u_h(t)(\boldsymbol{a}_j) = u_D(\boldsymbol{a}_j) \qquad\qquad j = M_1 + 1, \ldots, M \end{aligned}$$

and

$$u_h(0)(\boldsymbol{a}_j) = u_0(\boldsymbol{a}_j) \qquad\qquad j = 1, \ldots, M$$

For the time discretization the implicit Euler method is used. The time interval $(0, T)$ is divided in N subintervals (t_{n-1}, t_n) $(n = 1, \ldots, N)$. In the implicit Euler method the time derivative is replaced by the backward difference quotient

$$\frac{u_h - u_{h,old}}{\Delta t}$$

with the time step size $\Delta t = t_n - t_{n-1}$. For the sake of clarity the quantities do not get an index specifying the point in time. Only the values at the old point in time t_{n-1} get a subscript *old*. All other quantities are evaluated at t_n. Also the time step size Δt does not get an index specifying the number of the time step although the time step size Δt may differ from time step to time step.

Using the implicit Euler method the time discrete problem reads:
For every point in time t_n $(n = 1, \ldots, N)$ find $u_h \in \operatorname{span}\{\varphi_1, \ldots, \varphi_M\}$ such that

$$\begin{aligned}\int_\Omega \frac{\theta u_h - (\theta u_h)_{old}}{\Delta t} v_h \, d\boldsymbol{x} + \int_\Omega \boldsymbol{D}\nabla u_h \cdot \nabla v_h \, d\boldsymbol{x} + \int_\Omega \nabla \cdot (\boldsymbol{q} u_h) v_h \, d\boldsymbol{x}& \\ = \int_\Omega G(u_h) v_h \, d\boldsymbol{x} \qquad\qquad & \forall v_h \in V_h \\ u_h(\boldsymbol{a}_j) = u_D(\boldsymbol{a}_j) \qquad\qquad & j = M_1 + 1, \ldots, M\end{aligned}$$

where $u_h(\boldsymbol{a}_j) = u_0(\boldsymbol{a}_j)$ $(j = 1, \ldots, M)$ at the time t_0

With the representation

$$u_h(x) = \sum_{i=1}^{M} u_i \varphi_i(x)$$

and using the test functions $v_h = \varphi_j$ $(j = 1, \ldots, M_1)$ we get

$$\begin{aligned}\sum_{i=1}^{M} \left(\frac{\theta u_i - (\theta u_i)_{old}}{\Delta t} \int_\Omega \varphi_i \varphi_j \, d\boldsymbol{x} + u_i \int_\Omega \boldsymbol{D}\nabla \varphi_i \cdot \nabla \varphi_j \, d\boldsymbol{x} + u_i \int_\Omega \nabla \cdot (\boldsymbol{q} \varphi_i) \varphi_j \, d\boldsymbol{x} \right)& \\ = \int_\Omega G(u_h) \varphi_j \, d\boldsymbol{x} \qquad\qquad & j = 1, \ldots, M_1 \\ u_j = u_D(\boldsymbol{a}_j) \qquad\qquad & j = M_1 + 1, \ldots, M\end{aligned}$$

where $u_j = u_0(\boldsymbol{a}_j)$ $(j = 1, \ldots, M)$ at the time t_0. This is the fully discrete formulation used for the numerical computations.

Mass Lumping

The integrals $\int_\Omega \varphi_i \varphi_j \, d\boldsymbol{x}$ and $\int_\Omega G(u_h) \varphi_j \, d\boldsymbol{x}$ are approximated by a node orientated quadrature rule, i.e., a quadrature rule of the form

$$I(f) = \sum_{i=1}^{M} \omega_i f(\boldsymbol{a}_i)$$

which is exact for all basis functions.

This is not possible for the standard quadratic element P_2 on triangles. For the standard quadratic element P_2 the only node orientated quadrature rule which is exact for polynomials of degree 2 has the weights $\omega_s = 0$ (s for summit) at the corner points. So using this quadrature rule the mass matrix is not invertible. From the theory it is not clear if this leads to a convergent method. So it is not possible to do mass lumping for the element P_2.

Instead of that the modified element $\tilde{P}_2$ (see [CJRT01]) can be used. It has the barycenter as additional degree of freedom and the bubble function as additional ansatz function (Written in barycentric coordinates the bubble function b on one triangle is $b = \lambda_1\lambda_2\lambda_3$).

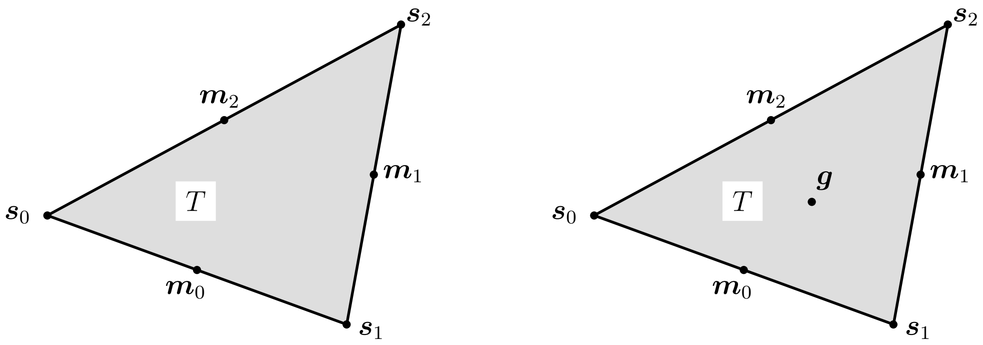


Figure 3.1: Degrees of freedom of P_2 (left) and $\tilde{P}_2$ (right) on one triangle T

The shape functions of this element are

$$\begin{aligned}
&\lambda_i(2\lambda_i - 1) + 3\lambda_1\lambda_2\lambda_3 && i = 1, 2, 3\\
&4\lambda_i\lambda_j - 12\lambda_1\lambda_2\lambda_3 && i, j = 1, 2, 3,\ i < j\\
&27\lambda_1\lambda_2\lambda_3\,.
\end{aligned}$$

For this element there is a node orientated quadrature rule with strictly positive weights. On one triangle this quadrature rule has the weights

$$\omega_s = \frac{1}{20}\,, \qquad \omega_m = \frac{2}{15}\,, \qquad \omega_g = \frac{9}{20}$$

for the corner points, the midpoints and the center of gravity, respectively. So it is possible to do mass lumping for quadratic elements on triangles without any problems.

3.3.2 Implicit Elimination

For the local problem it is only known that a resolution function exists. There is no explicit representation for this function which can be plugged in the system of coupled partial differential equations. But such a representation is not necessary to solve the system of nonlinear equations, which results from the discretization of the coupled partial differential equations (3.43)-(3.47), with Newton's method.

We consider the system

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{u},\boldsymbol{v}) &= \mathbf{0}\\ \boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) &= \mathbf{0}\end{aligned}$$

of nonlinear equations. Let the number of equations in the first block correspond with the length of the vector $\boldsymbol{u}$ and the number of equations in the second block with the length of the vector $\boldsymbol{v}$. Furthermore let it be known that for the second block a resolution function $\boldsymbol{v}(\boldsymbol{u})$ exists, i.e., the function $\boldsymbol{v}(\boldsymbol{u})$ is defined by

$$\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}(\boldsymbol{u})) = \mathbf{0} \text{ for all } \boldsymbol{u}\,.$$

Therewith we can rewrite the system of nonlinear equations as

$$\boldsymbol{f}(\boldsymbol{u},\boldsymbol{v}(\boldsymbol{u})) = \mathbf{0}\,.$$

To solve this system with Newton's method the evaluation of $\boldsymbol{v}(\boldsymbol{u})$ and the Jacobian matrix are needed. The evaluation can be done by solving $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = \mathbf{0}$ with Newton's method for fixed $\boldsymbol{u}$. The Jacobian matrix is

$$\boldsymbol{J} = \partial_1\boldsymbol{f} + \partial_2\boldsymbol{f}\boldsymbol{v}' \tag{3.51}$$

where $\partial_{1/2}$ denotes the partial derivative with respect to the first/second block of variables. The derivative $\boldsymbol{v}'$ is obtained by differentiating the defining equation of the function $\boldsymbol{v}$ with respect to $\boldsymbol{u}$. This yields the linear system

$$\partial_2\boldsymbol{g}\boldsymbol{v}' = -\partial_1\boldsymbol{g}\,. \tag{3.52}$$

The matrix $\partial_2\boldsymbol{g}$ is invertible because for $\boldsymbol{g}$ a resolution function with respect to the second variable exists. So this linear system can be used to calculate $\boldsymbol{v}'$.

Calculation of $\boldsymbol{v}'$ for the reduction scheme

We want to apply this approach to our problem. The variables $\boldsymbol{u}$ correspond to the global unknowns $\boldsymbol{\xi}_{glob}$ defined in (3.42) and the variables $\boldsymbol{v}$ to the local unknowns $\boldsymbol{\xi}_{loc}$ defined in (3.41). So $\boldsymbol{v}'$ contains the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$. Furthermore the equations $\boldsymbol{f}(\boldsymbol{u},\boldsymbol{v})$ correspond to the global problem consisting of the discretization of the equations (3.43)-(3.47) and the equations $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v})$ correspond to the local problem consisting of the equations (3.35)-(3.38) (after time discretization of (3.35)).

First we make the approximation $D_{\boldsymbol{\xi}_{glob}}\bar{\boldsymbol{\xi}}_{kin} \approx \mathbf{0}$. $\bar{\boldsymbol{\xi}}_{kin}$ describes kinetic reactions while the other local variables $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$ and $\bar{\boldsymbol{\xi}}_{min}$ describe equilibrium reactions. Kinetic reactions proceed much more slowly than equilibrium reactions. So only small changes of $\bar{\boldsymbol{\xi}}_{kin}$ are due and therewith the approximation $D_{\boldsymbol{\xi}_{glob}}\bar{\boldsymbol{\xi}}_{kin} \approx \mathbf{0}$ is justified. Using this approximation we have to solve a smaller linear system to calculate $\boldsymbol{v}'$. Now in (3.52) $\boldsymbol{v}$ corresponds to $(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})$ and $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$ corresponds to (3.36)-(3.38).

The equations (3.36) − (3.38) do not depend on $\boldsymbol{\xi}_{sorp}$ and $\boldsymbol{\xi}_{min}$. Hence the derivatives $D_{\boldsymbol{\xi}_{sorp}}\boldsymbol{\xi}_{loc}$ and $D_{\boldsymbol{\xi}_{min}}\boldsymbol{\xi}_{loc}$ vanishes. So it is not necessary to consider the whole vector $\boldsymbol{\xi}_{glob}$, as the vector corresponding to $\boldsymbol{u}$, we can restrict us to the subvector $(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin})$.

To set up the linear system for $\boldsymbol{v}'$ we need the matrix

$$\partial_2 \boldsymbol{g} = \frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})}\,.$$

With the computations of Section 3.2.1 we can write this matrix as

$$\partial_2 \boldsymbol{g} = \begin{pmatrix} \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B} & * \\ \mathbf{0} & \boldsymbol{I}_{\mathcal{A}} \end{pmatrix}$$

with

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min,\mathcal{I}} \\ \mathbf{0} & \boldsymbol{S}_{2,sorp} & \mathbf{0} \end{pmatrix}, \qquad \tilde{\boldsymbol{\Lambda}} = \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{\Lambda}}_{nmin} \end{pmatrix}$$

and $\boldsymbol{\Lambda}$, $\bar{\boldsymbol{\Lambda}}_{nmin}$ defined as in Section 3.2.1. Analogously we can calculate that

$$-\partial_1 \boldsymbol{g} = -\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min,\mathcal{I}}, \boldsymbol{\phi}_{min,\mathcal{A}})}{\partial(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin})} = \begin{pmatrix} \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{C} \\ \mathbf{0} \end{pmatrix}$$

with

$$\boldsymbol{C} = \begin{pmatrix} -\boldsymbol{S}_{1,sorp,li} & -\boldsymbol{S}_{1,min} & -\boldsymbol{S}^*_{1,kin} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Because of the structure of $\partial_1 \boldsymbol{g}$ and $\partial_2 \boldsymbol{g}$ we get that the lower block of $\boldsymbol{v}'$, by a partitioning of $\boldsymbol{v}'$ analogously to $\partial_1 \boldsymbol{g}$, is zero. This lower block consists of the derivatives $D_{(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin})}\bar{\boldsymbol{\xi}}_{min,\mathcal{A}}$. We can write $\boldsymbol{v}'$ as $\begin{pmatrix} \boldsymbol{X} \\ \mathbf{0} \end{pmatrix}$ where $\boldsymbol{X}$ is the solution of the linear systems

$$\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B}\boldsymbol{X} = \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{C}\,. \tag{3.53}$$

$\boldsymbol{X}$ contains the derivatives $D_{(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin})}(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min,\mathcal{I}})$. The matrix of the linear systems $\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B}$ is symmetric and positive definite and hence it is always invertible.

To assemble the global Jacobi matrix with (3.51) the terms $\partial_1 \boldsymbol{f}$ and $\partial_2 \boldsymbol{f}$ are needed. By differentiation of the discretization of (3.43)-(3.47) with respect to $\boldsymbol{\xi}_{glob}$ (treating $\boldsymbol{\xi}_{loc}$ as a variable and not as a function of $\boldsymbol{\xi}_{glob}$) one gets

$$\partial_1 \boldsymbol{f} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0} \\ \theta\boldsymbol{I} & \boldsymbol{0} & \Delta t \boldsymbol{L}_h & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \theta\boldsymbol{I} & \boldsymbol{0} & \Delta t \boldsymbol{L}_h & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \theta\boldsymbol{I} + \Delta t \boldsymbol{L}_h \end{pmatrix}$$

$$- \Delta t\theta \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{A}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{sorp} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \\ \boldsymbol{A}_{min} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{min} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{min} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \\ \boldsymbol{A}_{1,kin} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{1,kin} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{1,kin} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \end{pmatrix}$$

with $\boldsymbol{A}_{sorp} := \boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp,li}$, $\boldsymbol{A}_{min} := \boldsymbol{A}_{1,min} - \boldsymbol{A}_{ld}\boldsymbol{A}_{2,sorp,ld}$ and $\boldsymbol{L}_h$ the discretization of the transport operator L. The derivatives of $\boldsymbol{r}_{kin}$ can be obtained by plugging in (3.40) and using the chain rule.

By differentiation of the discretization of (3.43)-(3.47) with respect to the variables $(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})$ ($\bar{\boldsymbol{\xi}}_{sorp}$ is split like in (3.16)) one gets

$$\partial_2 \boldsymbol{f} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{ld} & \boldsymbol{I} \\ \multicolumn{4}{c}{-\Delta t\theta \boldsymbol{A}_{sorp} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}} \\ \multicolumn{4}{c}{-\Delta t\theta \boldsymbol{A}_{min} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}} \\ \multicolumn{4}{c}{-\Delta t\theta \boldsymbol{A}_{1,kin} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}} \end{pmatrix}.$$

Performing the matrix multiplication with $\boldsymbol{v}'$ gives

$$\partial_2 \boldsymbol{f} \boldsymbol{v}' =$$

$$\begin{pmatrix} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp,li} & D_{\tilde{\boldsymbol{\xi}}_{min}} \bar{\boldsymbol{\xi}}_{sorp,li} & \boldsymbol{0} & \boldsymbol{0} & D_{\boldsymbol{\xi}_{kin}} \bar{\boldsymbol{\xi}}_{sorp,li} \\ D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{min} + \boldsymbol{A}_{ld} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp,ld} & D_{\tilde{\boldsymbol{\xi}}_{min}} \bar{\boldsymbol{\xi}}_{min} + \boldsymbol{A}_{ld} D_{\tilde{\boldsymbol{\xi}}_{min}} \bar{\boldsymbol{\xi}}_{sorp,ld} & \boldsymbol{0} & \boldsymbol{0} & D_{\boldsymbol{\xi}_{kin}} \bar{\boldsymbol{\xi}}_{min} + \boldsymbol{A}_{ld} D_{\boldsymbol{\xi}_{kin}} \bar{\boldsymbol{\xi}}_{sorp,ld} \\ \boldsymbol{R}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{sorp} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ \boldsymbol{R}_{min} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{min} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{min} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ \boldsymbol{R}_{kin} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{kin} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{kin} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \end{pmatrix}$$

with $\boldsymbol{R}_{sorp} := -\Delta t\theta \boldsymbol{A}_{sorp} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}$,
$\boldsymbol{R}_{min} := -\Delta t\theta \boldsymbol{A}_{min} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}$ and
$\boldsymbol{R}_{kin} := -\Delta t\theta \boldsymbol{A}_{1,kin} D_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp,li}, \bar{\boldsymbol{\xi}}_{sorp,ld}, \bar{\boldsymbol{\xi}}_{min})} \boldsymbol{r}_{kin}$.

Altogether we get for the global Jacobian matrix

$$\boldsymbol{J}_{glob} = \boldsymbol{J}_{nkin} + \boldsymbol{J}_{kin} \tag{3.54}$$

where

$$\boldsymbol{J}_{nkin} = \begin{pmatrix} \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp,li} & D_{\tilde{\boldsymbol{\xi}}_{min}} \bar{\boldsymbol{\xi}}_{sorp,li} & -\boldsymbol{I} & \boldsymbol{0} & D_{\boldsymbol{\xi}_{kin}} \bar{\boldsymbol{\xi}}_{sorp,li} \\ D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{min,ld} & \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{min}} \bar{\boldsymbol{\xi}}_{min,ld} & \boldsymbol{0} & -\boldsymbol{I} & D_{\boldsymbol{\xi}_{kin}} \bar{\boldsymbol{\xi}}_{min,ld} \\ \theta \boldsymbol{I} & \boldsymbol{0} & \Delta t \boldsymbol{L}_h & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \theta \boldsymbol{I} & \boldsymbol{0} & \Delta t \boldsymbol{L}_h & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \theta \boldsymbol{I} + \Delta t \boldsymbol{L}_h \end{pmatrix}$$

with $\bar{\boldsymbol{\xi}}_{min,ld} := \bar{\boldsymbol{\xi}}_{min} + \boldsymbol{A}_{ld} \bar{\boldsymbol{\xi}}_{sorp,ld}$ and

$$\boldsymbol{J}_{kin} = -\Delta t \theta \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{A}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{sorp} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \\ \boldsymbol{A}_{min} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{min} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{min} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \\ \boldsymbol{A}_{1,kin} D_{\tilde{\boldsymbol{\xi}}_{sorp}} \boldsymbol{r}_{kin} & \boldsymbol{A}_{1,kin} D_{\tilde{\boldsymbol{\xi}}_{min}} \boldsymbol{r}_{kin} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{1,kin} D_{\boldsymbol{\xi}_{kin}} \boldsymbol{r}_{kin} \end{pmatrix} +$$

$$\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{R}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{sorp} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{sorp} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ \boldsymbol{R}_{min} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{min} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{min} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ \boldsymbol{R}_{kin} D_{\tilde{\boldsymbol{\xi}}_{sorp}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{R}_{kin} D_{\tilde{\boldsymbol{\xi}}_{min}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_{kin} D_{\boldsymbol{\xi}_{kin}} (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \end{pmatrix}.$$

Using the implicit elimination strategy we get the following algorithm for one time step of the reduction scheme:

One time step of the reduction scheme

```
Solve $\eta$-problem
Solve local problem
Calculate defect $\boldsymbol{d}$ of the global problem
Stopping criteria for global problem not fulfilled
    Assemble Jacobi matrix $\boldsymbol{J}$ of the global problem
    Solve linear system $\boldsymbol{J}\Delta\boldsymbol{\xi}_{glob} = \boldsymbol{d}$
    Update $\boldsymbol{\xi}_{glob} -= \Delta\boldsymbol{\xi}_{glob}$
    Solve local problem
    Calculate defect $\boldsymbol{d}$ of the global problem
```

3.4 Special Numerical Treatment

3.4.1 Local Problem

In the local problem the variables $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{min}$ and $\bar{\boldsymbol{\xi}}_{kin}$ are calculated with the equations

$$\begin{aligned}\boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \boldsymbol{0}\\ \boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) &= \boldsymbol{0}\\ \boldsymbol{\phi}_{min}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) &= \boldsymbol{0}\\ \frac{\theta\bar{\boldsymbol{\xi}}_{kin} - (\theta\bar{\boldsymbol{\xi}}_{kin})_{old}}{\Delta t} &= \theta \boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\end{aligned}$$

where $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$ are defined by (3.40). Solving this system of nonlinear equations with Newton's method can lead to very ill conditioned linear systems. For example in a test computation with the chemistry of the MoMaS–benchmark (see Chap. 4) the condition number of the Jacobian matrix is $\kappa(\boldsymbol{J}) \approx 10^{70}$. It is not possible to solve such a linear system numerically. In this computation the order of magnitude of the smallest concentration value is 10^{-25}, the one of the largest concentration value is 1. Because of this large range of magnitudes the Jacobian matrix is so ill conditioned.

Therefore it is necessary to use the logarithms of the variables as unknowns of the numerical computation. But in the formulation above this is not possible because in the equilibrium conditions there is the logarithm of a sum, e.g.:

$$\boldsymbol{k}_{mob} + \boldsymbol{S}_{1,mob}^T \ln(\boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \cdots + \boldsymbol{S}_{1,kin}^*\boldsymbol{\xi}_{kin} + \boldsymbol{B}_1^\perp \boldsymbol{\eta}) = \boldsymbol{0}$$

The reason for this is that we have plugged in (3.40) for the concentrations. So it is not allowed to plug in the retransformation. Instead of that we have to use the concentrations as variables. So we replace the variables $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$ and $\bar{\boldsymbol{\xi}}_{min}$ by $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$. The number of the concentrations is $I + \bar{I}$ and is larger than the number of the replaced transformed variables, which is $J_{mob} + J_{sorp} + J_{min}$. So we need $I + \bar{I} - J_{mob} - J_{sorp} - J_{min}$ additional equations so that the number of unknowns is equal to the number of equations again. Hence the defining equations of the transformed variables $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$, $\boldsymbol{\xi}_{kin}$, $\bar{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\xi}}_{kin}$ are added as additional equations. Indeed the number of these defining equations is

$$\begin{aligned}&(I - J_{mob} - J_{sorp,li} - J_{min} - J_{1,kin}^*) + J_{sorp,li} + J_{min} + J_{1,kin}^*\\ &\quad + (\bar{I} - J_{sorp} - J_{min} - J_{2,kin}^*) + J_{2,kin}^*\\ &= I + \bar{I} - J_{mob} - J_{sorp} - J_{min}\,.\end{aligned}$$

Then in the local problem we have to compute the variables $\boldsymbol{c}$, $\bar{\boldsymbol{c}}$ and $\bar{\boldsymbol{\xi}}_{kin}$ with help of the equations

$$\begin{aligned}
\boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \boldsymbol{0} \\
\boldsymbol{\eta} &= \left({\boldsymbol{S}_1^{\perp}}^T \boldsymbol{B}_1^{\perp}\right)^{-1} {\boldsymbol{S}_1^{\perp}}^T \boldsymbol{c} \\
\tilde{\boldsymbol{\xi}}_{sorp} &= \left((\boldsymbol{B}_1^T \boldsymbol{S}_1^*)^{-1} \boldsymbol{B}_1^T \boldsymbol{c}\right)_{i=J_{mob}+1,\dots,J_{mob}+J_{sorp,li}} \\
&\quad - \left((\boldsymbol{B}_2^T \boldsymbol{S}_2^*)^{-1} \boldsymbol{B}_2^T \bar{\boldsymbol{c}}\right)_{i=1,\dots,J_{sorp,li}} \\
\tilde{\boldsymbol{\xi}}_{min} &= \left((\boldsymbol{B}_1^T \boldsymbol{S}_1^*)^{-1} \boldsymbol{B}_1^T \boldsymbol{c}\right)_{i=J_{mob}+J_{sorp,li}+1,\dots,J_{eq,li}} - \bar{\boldsymbol{c}}_{min} \\
&\quad - \boldsymbol{A}_{ld} \left((\boldsymbol{B}_2^T \boldsymbol{S}_2^*)^{-1} \boldsymbol{B}_2^T \bar{\boldsymbol{c}}\right)_{i=J_{sorp,li}+1,\dots,J_{sorp}} \\
\boldsymbol{\xi}_{kin} &= \left((\boldsymbol{B}_1^T \boldsymbol{S}_1^*)^{-1} \boldsymbol{B}_1^T \boldsymbol{c}\right)_{i=J_{eq,li}+1,\dots,J_{eq,li}+J^*_{1,kin}} \\
\boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) &= \boldsymbol{0} \\
\boldsymbol{\phi}_{min}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) &= \boldsymbol{0} \\
\bar{\boldsymbol{\eta}} &= \left({\boldsymbol{S}_2^{\perp}}^T \boldsymbol{B}_2^{\perp}\right)^{-1} {\boldsymbol{S}_2^{\perp}}^T \bar{\boldsymbol{c}} \\
\bar{\boldsymbol{\xi}}_{kin} &= \left((\boldsymbol{B}_2^T \boldsymbol{S}_2^*)^{-1} \boldsymbol{B}_2^T \bar{\boldsymbol{c}}\right)_{i=J_{sorp}+J_{min}+1,\dots,J_{sorp}+J_{min}+J^*_{2,kin}} \\
\frac{\theta \bar{\boldsymbol{\xi}}_{kin} - (\theta \bar{\boldsymbol{\xi}}_{kin})_{old}}{\Delta t} &= \theta \boldsymbol{A}_{2,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})
\end{aligned}$$

with $J_{eq,li} = J_{mob} + J_{sorp,li} + J_{min}$. Remember that because of the special structure (3.11) of $\boldsymbol{S}_2^*$ and $\boldsymbol{B}_2$ the transformed variables $\bar{\boldsymbol{\eta}}$, $\bar{\boldsymbol{\xi}}_{kin}$ and the last summands in the defining equations for $\tilde{\boldsymbol{\xi}}_{sorp}$ and $\tilde{\boldsymbol{\xi}}_{min}$ do not depend on $\bar{\boldsymbol{c}}_{min}$ (see Sec. 3.1).

Now the equilibrium conditions are, e.g.:

$$\boldsymbol{k}_{mob} + \boldsymbol{S}_{1,mob}^T \ln(\boldsymbol{c}) = \boldsymbol{0}$$

So it possible to use the logarithms of the concentrations as unknowns. It turns out that is useful to use the logarithms of the concentrations of the mobile species and the nonminerals but not of the mineral concentrations. The reason why the mineral concentrations are not logarithmized is that because of the mineral equilibrium $\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) = \min\{\psi_j(\boldsymbol{c}), \bar{c}_{min,j}\} = 0$ it is necessary that $\bar{\boldsymbol{c}}_{min}$ can have the value zero which would not be possible if we replaced $\bar{\boldsymbol{c}}_{min}$ by its logarithm. Written as a root-finding problem with the logarithms of the mobile concentrations $\boldsymbol{l}$ and the logarithms of the nonminerals $\bar{\boldsymbol{l}}_{nmin}$ instead of $\boldsymbol{c}$ and

$\bar{\boldsymbol{c}}_{nmin}$, respectively, the local problem reads

$$\boldsymbol{\phi}_{mob}(\boldsymbol{l}) = \boldsymbol{0} \tag{3.55}$$

$$-\boldsymbol{\eta} + \left(\boldsymbol{S}_1^{\perp T} \boldsymbol{B}_1^{\perp}\right)^{-1} \boldsymbol{S}_1^{\perp T} \exp(\boldsymbol{l}) = \boldsymbol{0} \tag{3.56}$$

$$-\tilde{\boldsymbol{\xi}}_{sorp} + \left(\left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \exp(\boldsymbol{l}) \right)_{i=J_{mob}+1,\dots,J_{mob}+J_{sorp,li}} \tag{3.57}$$

$$- \left(\left(\boldsymbol{B}_2^T \boldsymbol{S}_2^*\right)^{-1} \boldsymbol{B}_2^T \begin{pmatrix} \exp(\bar{\boldsymbol{l}}_{nmin}) \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix} \right)_{i=1,\dots,J_{sorp,li}} = \boldsymbol{0} \tag{3.58}$$

$$\begin{aligned} -\tilde{\boldsymbol{\xi}}_{min} + \left(\left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \exp(\boldsymbol{l}) \right)_{i=J_{mob}+J_{sorp,li}+1,\dots,J_{eq,li}} - \bar{\boldsymbol{c}}_{min} & \\ - \boldsymbol{A}_{ld} \left(\left(\boldsymbol{B}_2^T \boldsymbol{S}_2^*\right)^{-1} \boldsymbol{B}_2^T \begin{pmatrix} \exp(\bar{\boldsymbol{l}}_{nmin}) \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix} \right)_{i=J_{sorp,li}+1,\dots,J_{sorp}} &= \boldsymbol{0} \end{aligned} \tag{3.59}$$

$$-\boldsymbol{\xi}_{kin} + \left(\left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \exp(\boldsymbol{l}) \right)_{i=J_{eq,li}+1,\dots,J_{eq,li}+J^*_{1,kin}} = \boldsymbol{0} \tag{3.60}$$

$$\boldsymbol{\phi}_{sorp}(\boldsymbol{l}, \bar{\boldsymbol{l}}_{nmin}) = \boldsymbol{0} \tag{3.61}$$

$$\boldsymbol{\phi}_{min}(\boldsymbol{l}, \bar{\boldsymbol{c}}_{min}) = \boldsymbol{0} \tag{3.62}$$

$$-\bar{\boldsymbol{\eta}} + \left(\boldsymbol{S}_2^{\perp T} \boldsymbol{B}_2^{\perp}\right)^{-1} \boldsymbol{S}_2^{\perp T} \begin{pmatrix} \exp(\bar{\boldsymbol{l}}_{nmin}) \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix} = \boldsymbol{0} \tag{3.63}$$

$$-\bar{\boldsymbol{\xi}}_{kin} + \left(\left(\boldsymbol{B}_2^T \boldsymbol{S}_2^*\right)^{-1} \boldsymbol{B}_2^T \begin{pmatrix} \exp(\bar{\boldsymbol{l}}_{nmin}) \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix} \right)_{i=J_{sorp}+J_{min}+1,\dots,J_{sorp}+J_{min}+J^*_{2,kin}} = \boldsymbol{0} \tag{3.64}$$

$$\frac{\theta \bar{\boldsymbol{\xi}}_{kin} - (\theta \bar{\boldsymbol{\xi}}_{kin})_{old}}{\Delta t} - \theta \boldsymbol{A}_{2,kin} \boldsymbol{r}_{kin}(\exp(\boldsymbol{l}), \exp(\bar{\boldsymbol{l}}_{nmin}), \bar{\boldsymbol{c}}_{min}) = \boldsymbol{0}. \tag{3.65}$$

The transformed variables $\boldsymbol{\xi}_{mob}$ and $\bar{\boldsymbol{\xi}}_{sorp}$ are calculated after solving the local problem with their definitions

$$\boldsymbol{\xi}_{mob} = \left(\left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \boldsymbol{c} \right)_{i=1,\dots,J_{mob}}, \qquad \bar{\boldsymbol{\xi}}_{sorp} = \left(\left(\boldsymbol{B}_2^T \boldsymbol{S}_2^*\right)^{-1} \boldsymbol{B}_2^T \bar{\boldsymbol{c}} \right)_{i=1,\dots,J_{sorp}}$$

and the transformed variable $\bar{\boldsymbol{\xi}}_{min}$ is equal to $\bar{\boldsymbol{c}}_{min}$.

If we neglect the dependency of $\boldsymbol{r}_{kin}$ on $\bar{\boldsymbol{c}}_{min}$ (in most cases $\boldsymbol{r}_{kin}$ is indeed independent of $\bar{\boldsymbol{c}}_{min}$) each mineral concentration $\bar{c}_{min,k}$ appears only once in the system (3.55)-(3.65). When the k-th mineral is present $\bar{c}_{min,k}$ appears only in (3.59) because in this case the k-th equation of (3.62) gets $\psi_k(\boldsymbol{l}) = 0$ and so is independent of $\bar{c}_{min,k}$. If the k-th mineral is not present $\bar{c}_{min,k}$ appears only in (3.62) because in this case the k-th mineral concentration is identical to zero and so the term $-\bar{c}_{min,k}$ in (3.59) can be left out. So it is possible to use a smaller system and to calculate the mineral concentrations $\bar{\boldsymbol{c}}_{min}$ afterwards.

In the implementation done in the framework of this thesis this is done in the following way: For each mineral reaction it is checked if the minimum in the k-th equation of (3.62) is attained in the first or in the second argument (definition of $\boldsymbol{\phi}_{min}$ see (2.7)) and the result is stored in a vector $\boldsymbol{AI}$. A '1' stands for the minimum is attained in the first argument that means that the mineral is present. A '0' stands for the minimum is attained in the second argument that means that the mineral is not present and the mineral concentration $\bar{c}_{min,k}$ is set to zero. This is done before the computation of the defect of the local problem.

The computation of the mineral concentrations $\bar{\boldsymbol{c}}_{min}$ is also done before the computation of the local defect. Here it is important that the mineral concentration is computed before it is checked if the minimum is attained in the first or in the second argument. Furthermore if the value of AI_k has changed it is necessary to compute $\bar{c}_{min,k}$ again. Altogether the following algorithm is used:

Detailed algorithm for calculating the local defect

```
For each mineral reaction $k$
    $AI_k$
        1:
            $\bar{c}_{min,k} = -\tilde{\xi}_{min,k} + \left(\left(\boldsymbol{B}_1^T\boldsymbol{S}_1^*\right)^{-1}\boldsymbol{B}_1^T\exp(\boldsymbol{l})\right)_{J_{sorp,li}+k} - \left(\boldsymbol{A}_{ld}\left(\left(\boldsymbol{B}_2^T\boldsymbol{S}_2^*\right)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}_{nmin}\right)_{i=J_{sorp,li}+1,\dots,J_{sorp}}\right)_k$
        0:
            $\varnothing$
    $\psi_k(\boldsymbol{l}) \quad > \quad \bar{c}_{min,k}$
        TRUE:
            $AI_k = 0$
            $\bar{c}_{min,k} = 0$
            $\varnothing$
        FALSE:
            $AI_k$
                0:
                    $\bar{c}_{min,k} = -\tilde{\xi}_{min,k} + \left(\left(\boldsymbol{B}_1^T\boldsymbol{S}_1^*\right)^{-1}\boldsymbol{B}_1^T\exp(\boldsymbol{l})\right)_{J_{sorp,li}+k} - \left(\boldsymbol{A}_{ld}\left(\left(\boldsymbol{B}_2^T\boldsymbol{S}_2^*\right)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}_{nmin}\right)_{i=J_{sorp,li}+1,\dots,J_{sorp}}\right)_k$
                1:
                    $\varnothing$
            $AI_k = 1$
Assemble local defect (3.55)-(3.65) without (3.62)
For each mineral reaction $k$
    $AI_k$
        1:
            $defect_{J_{mob}+N_\eta+J_{sorp,li}+k} = \psi_k(\boldsymbol{l})$
        0:
            $\varnothing$
```

It is important to use a damped Newton's method for solving the local problem because otherwise for realistic examples the Newton's method does not converge. In the implementation done in the framework of this thesis line search is used.

Newton's method with line search

```
$l$ = Maximal number of line search steps
Calculate defect $\boldsymbol{b}$
$d = \|\boldsymbol{b}\|$
Stopping criteria not fulfilled
    Calculate Jacobi matrix $\boldsymbol{J}$
    Solve linear system $\boldsymbol{J}\boldsymbol{c} = \boldsymbol{b}$
    $\boldsymbol{x} -= \boldsymbol{c}$
    Calculate defect $\boldsymbol{b}$
    $j = 0$
    $j < l$
        $d_1 = \|\boldsymbol{b}\|$
        $d_1 < d$
            TRUE
                break
            FALSE
                $\varnothing$
        $\boldsymbol{c} *= 0.5$, $\boldsymbol{x} += \boldsymbol{c}$
        Calculate defect $\boldsymbol{b}$
        $j = j + 1$
    $d = d_1$
```

In the test computation mentioned at the beginning of this section the order of magnitude of the condition number is 10^3 (instead of 10^{70}) when solving the enlarged problem using the logarithms of the unknowns. Such a linear system can be solved numerically without problems.

However it can happen that in one row of the Jacobian matrix which corresponds to a defining equation of a transformed variable all entries are very close to zero and so the Jacobian matrix gets numerically singular. This is the case when one of the global variables is close to zero and this variable is defined by a linear combination of concentrations in which all coefficients have the same sign, e.g.:

$$\tilde{\xi}_{sorp,1} = -c_2 - c_4 - 3c_5 - c_8 - \bar{c}_{11}$$

If $\tilde{\xi}_{sorp,1}$ is close to zero it holds for every positive solution that all occurring concentrations values are close to zero. Because of the use of the logarithms as variables the corresponding row of the Jacobian matrix is

$$\begin{pmatrix} 0 & -c_2 & 0 & -c_4 & -3c_5 & 0 & 0 & -c_8 & 0 & 0 & -\bar{c}_{11} & 0 \end{pmatrix}.$$

In this row all entries gets close to zero as soon as the Newton iterate is situated near the solution. Then the Jacobian matrix gets numerically singular. Numerical tests showed that multiplying the row in which all entries are close to zero with a large number, such that at least one entry is close to one, does not improve the condition number of the Jacobian matrix. Obviously the resulting row is numerically linear dependent.

Therefore instead of solving the linear system $\boldsymbol{Jx} = \boldsymbol{d}$ the substitution problem

$$\begin{aligned} &\text{Minimize } \|\boldsymbol{x}\|^2 \text{ on} \\ &L(\boldsymbol{d}) := \left\{ \boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \|\boldsymbol{Jx} - \boldsymbol{d}\|^2 = \min\left\{ \|\boldsymbol{Jy} - \boldsymbol{d}\|^2 \,|\, \boldsymbol{y} \in \mathbb{R}^n \right\} \right\} \end{aligned}$$

is solved. This problem always has a unique solution. To compute the solution a QR-factorization for matrices with numerical rang smaller than n and with column pivot search $\boldsymbol{JP} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$ can be used (see [Kna02]). With help of this factorization we get the solution of the minimization problem in the following way (The indices 1,2 denote the partitioning in two blocks analogously to the partitioning of the rows in the right triangular matrix of the factorization):

Calculation of the smallest-norm minimum

Calculate QR-factorization for matrices with numerical rang smaller than n and with column pivot search $\boldsymbol{JP} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$
Calculate $\boldsymbol{V}$ by solving $\boldsymbol{RV} = \boldsymbol{B}$
Calculate $\boldsymbol{z}$ by solving $\boldsymbol{Rz} = (\boldsymbol{Q}^T\boldsymbol{d})_1$
Calculate $\boldsymbol{y}_2$ by solving $(\boldsymbol{I} + \boldsymbol{V}^T\boldsymbol{V})\boldsymbol{y}_2 = \boldsymbol{V}^T\boldsymbol{z}$
Set $\boldsymbol{y}_1 = \boldsymbol{z} - \boldsymbol{V}\boldsymbol{y}_2$ and $\boldsymbol{x} = \boldsymbol{Py}$

When solving the substitution problem all equations of the local problem are solved exactly except of defining equations of transformed variables in that the order of magnitude of the transformed variable is the machine precision. So the only error, which is obtain by solving the substitution problem instead of the linear system, is a mass balance error whose order of magnitude is the machine precision.

To enlarge the robustness of the method it is also useful to replace the linear system (3.53) by a minimization problem and to solve this problem with the algorithm above. Especially when many concentration values are close to zero solving the linear system can fail.

3.4.2 Starting Value Search After η–Problem

The obvious choice of the starting value of the global Newton method is the value at the old time step. This choice is here not possible. It can be seen with the following example consisting of two equilibrium reactions

$$\begin{aligned} \mathrm{B} &\leftrightarrow 2\mathrm{A} \\ \mathrm{B}+\mathrm{C} &\leftrightarrow \overline{\mathrm{D}}\,. \end{aligned}$$

The choice of the transformation matrix

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

leads to the transformed variables

$$\begin{aligned}
\begin{pmatrix} \xi_{mob} \\ \xi_{sorp} \end{pmatrix} &= \left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \boldsymbol{c} \\
&= \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & -1 \\ 0 & -1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c(\mathrm{A}) \\ c(\mathrm{B}) \\ c(\mathrm{C}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}c(\mathrm{A}) \\ -c(\mathrm{C}) \end{pmatrix} \\
\eta &= \left(\boldsymbol{S}_1^{\perp\,T} \boldsymbol{B}_1^{\perp}\right)^{-1} \boldsymbol{S}_1^{\perp\,T} \boldsymbol{c} \\
&= \left(\begin{pmatrix} 1 & 2 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 2 & -2 \end{pmatrix} \begin{pmatrix} c(\mathrm{A}) \\ c(\mathrm{B}) \\ c(\mathrm{C}) \end{pmatrix} \\
&= \frac{1}{2}c(\mathrm{A}) + c(\mathrm{B}) - c(\mathrm{C}) \\
\bar{\xi}_{sorp} &= c(\overline{\mathrm{D}}) \\
\tilde{\xi}_{sorp} &= \xi_{sorp} - \bar{\xi}_{sorp} = -c(\mathrm{C}) - c(\overline{\mathrm{D}})\,.
\end{aligned}$$

The retransformation is

$$\begin{aligned}
c(\mathrm{A}) &= 2\xi_{mob}\\
c(\mathrm{B}) &= -\xi_{mob} - \tilde{\xi}_{sorp} - \bar{\xi}_{sorp} + \eta\\
c(\mathrm{C}) &= -\tilde{\xi}_{sorp} - \bar{\xi}_{sorp}\\
c(\overline{\mathrm{D}}) &= \bar{\xi}_{sorp}\,.
\end{aligned}$$

Let the value at the old time step be

$$c(\mathrm{A}) = c(\mathrm{B}) = c(\mathrm{C}) = c(\overline{\mathrm{D}}) = 0\,.$$

As the variable transformation is linear all transformed variables are also zero at the old time step.

After solving the η-problem let $\eta = -0.5$ at one node. This matches an inflow of C. With this value for η and the values at the old time step for the other transformed variables we get the concentrations

$$c(\mathrm{A}) = c(\mathrm{B}) = c(\overline{\mathrm{D}}) = 0, \quad c(\mathrm{B}) = -0.5.$$

The concentration of B is negative. With the values of $\eta = -0.5$ and $\tilde{\xi}_{sorp} = 0$ there is no nonnegative solution of the local problem. We see this in the following way: Plugging $\xi_{mob} = \frac{1}{2}c(\mathrm{A})$ and $\bar{\xi}_{sorp} = c(\overline{\mathrm{D}})$ in $c(\mathrm{B}) = -\xi_{mob} - \tilde{\xi}_{sorp} - \bar{\xi}_{sorp} + \eta$ leads to

$$c(\mathrm{B}) = -\frac{1}{2}c(\mathrm{A}) - c(\overline{\mathrm{D}}) - 0.5$$

and we see immediately that there exists no nonnegative solution.

Hence it is necessary to modify the starting value of the global Newton method when it corresponds to negative concentration values because it can happen that there is no nonnegative solution of the local problem. As the logarithms of the concentrations are used as unknowns the solver of the local problem will not converge when there is no nonnegative solution. So we have to change the transformed variables in such a way that they correspond to nonnegative concentration values.

Under the assumption that there is a positive bound for the concentrations it exists a resolution function $(\tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})$. Therefore it is secured that there is a positive solution of the local problem when the values of the transformed variables correspond to positive concentrations.

It is sufficient to modify the variables $\boldsymbol{\xi}_{mob}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$ and $\boldsymbol{\xi}_{kin}$

$$\boldsymbol{\xi}_{mob} \mathrel{+}= \Delta\boldsymbol{\xi}_{mob}\,, \quad \tilde{\boldsymbol{\xi}}_{sorp} \mathrel{+}= \Delta\tilde{\boldsymbol{\xi}}_{sorp}\,, \quad \tilde{\boldsymbol{\xi}}_{min} \mathrel{+}= \Delta\tilde{\boldsymbol{\xi}}_{min}\,, \quad \boldsymbol{\xi}_{kin} \mathrel{+}= \Delta\boldsymbol{\xi}_{kin}\,.$$

As $\boldsymbol{\eta}$ only affects the mobile species it is not possible that the concentration of an immobile species gets negative during the η-step. Hence the transformed variables $\bar{\boldsymbol{\eta}}$, $\bar{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{min}$, $\bar{\boldsymbol{\xi}}_{kin}$ that correspond to immobile species need not to be considered.

Under the assumption that the discrete solution is positive it always exists an appropriate starting value where we need not modify $\boldsymbol{\eta}$ because $\boldsymbol{\eta}$ is already computed and because of the linearity of the η-equations equal to the $\boldsymbol{\eta}$-value of the discrete solution. There always exists an appropriate starting value because the values of the transformed variables, corresponding to the discrete solution, are one possible starting value.

During the starting value search the values of the transformed variables should be altered as less as possible. This leads to a linear minimization problem with constraints:

$$\begin{aligned} \min\ & \epsilon|\Delta\boldsymbol{\xi}_{mob}| + |\Delta\tilde{\boldsymbol{\xi}}_{sorp}| + |\Delta\tilde{\boldsymbol{\xi}}_{min}| + |\Delta\boldsymbol{\xi}_{kin}| \\ s.t.\ & \boldsymbol{S}_{1,mob}\Delta\boldsymbol{\xi}_{mob} + \boldsymbol{S}_{1,sorp}\Delta\tilde{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,min}\Delta\tilde{\boldsymbol{\xi}}_{min} \\ & \quad + \boldsymbol{S}^*_{1,kin}\Delta\boldsymbol{\xi}_{kin} \geq -\boldsymbol{c} + \mathcal{E}c_{max}\boldsymbol{e} \end{aligned} \tag{3.66}$$

with $\boldsymbol{e} = (1,\dots,1)^T$, $c_{max} = \max_{i=1,\dots,I} c_i$ and the parameters, e.g., $\epsilon = 0.1$, $\mathcal{E} = 2\cdot 10^{-16}$. The term $\mathcal{E}c_{max}\boldsymbol{e}$ ensures that despite of numerical rounding errors it holds $c_i \geq 0$. The parameter ϵ effects that first an starting value is search at which only the local variables are modified. When there is no such starting value then the global variables are modified.

3.4.3 Cutting-off of Global Newton Step

Also after a global Newton step it can happen that the transformed variables correspond to negative concentrations and so it is possible that the local problem has no positive solution. Therefore it is necessary to check after each global Newton step if the Newton iterate has to be modified.

For modifying the current Newton iterate there are two different approaches. First the whole Newton step can be cut-off by taking only a part of the Newton update $\Delta\boldsymbol{\xi}_{glob}$ instead of the whole Newton update, i.e., the update of the global variables

$$\boldsymbol{\xi}_{glob} -= \Delta\boldsymbol{\xi}_{glob}$$

is replaced by

$$\boldsymbol{\xi}_{glob} -= \tau\Delta\boldsymbol{\xi}_{glob}$$

with $\tau \in (0,1)$. This has consequences for all nodes.

The other approach is to modify the Newton iterate only locally. Thereby only the values of nodes at that the Newton iterate corresponds to negative concentrations are modified. Numerical tests showed that it is most efficient to cut-off the whole Newton step only when a large modification of the current Newton iterate is necessary and otherwise to modify the Newton iterate only locally.

This adaptive approach can be performed by solving the minimization problem (3.66) two times at each node. First the values of $|\Delta\tilde{\boldsymbol{\xi}}_{sorp}|$, $|\Delta\tilde{\boldsymbol{\xi}}_{min}|$ and $|\Delta\boldsymbol{\xi}_{kin}|$ are determined by solving (3.66) on each node but no modification of the variables $\boldsymbol{\xi}_{mob}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$ and $\boldsymbol{\xi}_{kin}$ is performed. If there are values of $|\Delta\tilde{\boldsymbol{\xi}}_{sorp}|$, $|\Delta\tilde{\boldsymbol{\xi}}_{min}|$ or $|\Delta\boldsymbol{\xi}_{kin}|$ that are larger than a given bound S_{max} the whole Newton step is cut-off. After that the minimization problem (3.66) is solved again at each node. But this time the modification of the variables $\boldsymbol{\xi}_{mob}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$ and $\boldsymbol{\xi}_{kin}$ is performed.

Simplification for linear combinations with same sign

In the case that all global variables are linear combination of concentrations where all coefficients have the same sign there is a much simpler approach. The MoMaS–benchmark (see Chap. 4) is an example where this simplification can be applied. In the “easy test case” of this benchmark we have the four global variables

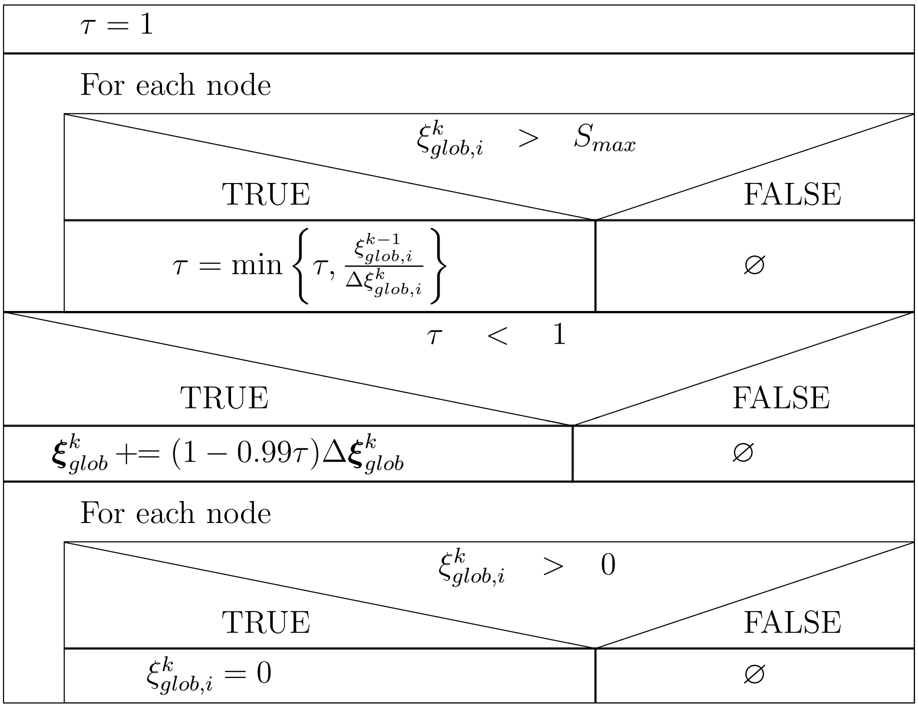
$$\begin{aligned}
\tilde{\xi}_{sorp,1} &= -\mathrm{C}_2 - \mathrm{C}_4 - 3\mathrm{C}_5 - \mathrm{X}_3 - \mathrm{CS}_1\\
\tilde{\xi}_{sorp,2} &= -\mathrm{C}_3 - 3\mathrm{C}_4 - \mathrm{C}_5 - \mathrm{X}_4 - \mathrm{CS}_2\\
\xi_{sorp,1} &= -\mathrm{C}_2 - \mathrm{C}_4 - 3\mathrm{C}_5 - \mathrm{X}_3\\
\xi_{sorp,2} &= -\mathrm{C}_3 - 3\mathrm{C}_4 - \mathrm{C}_5 - \mathrm{X}_4.
\end{aligned}$$

Note that this is only possible because of the decoupling of the variables $\boldsymbol{\eta}$. The variable η_1 is a linear combination of the concentrations with positive and negative coefficients:

$$\eta_1 = -\mathrm{C}_1 - 2\mathrm{C}_2 + 2\mathrm{C}_3 + 2\mathrm{C}_4 - 2\mathrm{C}_5 + \mathrm{X}_2 - 3\mathrm{X}_3 + 3\mathrm{X}_4$$

In this case it is sufficient to guarantee that the global variables are nonpositive. With the help of the resolution function it is secured that a solution of the local problem with positive concentration values exists. Like in the general case the whole Newton step should be cut off for large modifications of the Newton iterate and the Newton iterate should be modified only locally for small modifications. This can be done with the following algorithm (The upper index denotes the Newton step. Note the sign of the Newton update: $\boldsymbol{\xi}^k_{glob} = \boldsymbol{\xi}^{k-1}_{glob} - \Delta\boldsymbol{\xi}^k_{glob}$):

Cutting-off of global Newton step — simplified case



```
$\tau = 1$
For each node
    $\xi^k_{glob,i} > S_{max}$
        TRUE:  $\tau = \min\left\{\tau, \frac{\xi^{k-1}_{glob,i}}{\Delta\xi^k_{glob,i}}\right\}$
        FALSE: $\varnothing$
$\tau < 1$
    TRUE:  $\boldsymbol{\xi}^k_{glob} \mathrel{+}= (1 - 0.99\tau)\Delta\boldsymbol{\xi}^k_{glob}$
    FALSE: $\varnothing$
For each node
    $\xi^k_{glob,i} > 0$
        TRUE:  $\xi^k_{glob,i} = 0$
        FALSE: $\varnothing$
```

The factor 0.99 ensures that despite of numerical rounding errors it holds that $\xi^k_{glob,i}$ is nonpositive.

Update of concentration values

As two sets of variables (the transformed variables and the concentrations) are used it is necessary to update the concentrations when the transformed variables have changed. So after solving the η-problem the concentrations must be updated by

$$\boldsymbol{c} \mathrel{+}= \boldsymbol{B}_1^{\perp}(\boldsymbol{\eta} - \boldsymbol{\eta}_{old})\,.$$

Also after solving the global problem the concentrations must be updated. Here it must be considered that we use a resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ to get a smaller global system. So we have $\Delta\boldsymbol{\xi}_{loc} = D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}\Delta\boldsymbol{\xi}_{glob}$. Hence if we write the retransformation (3.40) in the form $\boldsymbol{c} = \boldsymbol{S}_{glob}\boldsymbol{\xi}_{glob} + \boldsymbol{S}_{loc}\boldsymbol{\xi}_{loc}$ we have to update the concentrations by

$$\boldsymbol{c} \mathrel{-}= \boldsymbol{S}_{glob}\Delta\boldsymbol{\xi}_{glob} + \boldsymbol{S}_{loc}D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}\Delta\boldsymbol{\xi}_{glob}\,.$$

Like in Section 3.3.2 the derivatives $D_{\boldsymbol{\xi}_{glob}}\bar{\boldsymbol{\xi}}_{kin}$ are neglected. Written in detail it reads

$$\begin{aligned}
\boldsymbol{c} \mathrel{-}= {} & \boldsymbol{S}_{1,mob} D_{\boldsymbol{\xi}_{glob}} \boldsymbol{\xi}_{mob} \Delta\boldsymbol{\xi}_{glob} + \boldsymbol{S}_{1,sorp,li} \Delta\tilde{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,sorp} D_{\boldsymbol{\xi}_{glob}} \bar{\boldsymbol{\xi}}_{sorp} \Delta\boldsymbol{\xi}_{glob} \\
& + \boldsymbol{S}_{1,min} (\Delta\tilde{\boldsymbol{\xi}}_{min} + D_{\boldsymbol{\xi}_{glob}} \bar{\boldsymbol{\xi}}_{min} \Delta\boldsymbol{\xi}_{glob}) + \boldsymbol{S}^{*}_{1,kin} \Delta\boldsymbol{\xi}_{kin} \\
\bar{\boldsymbol{c}}_{nmin} \mathrel{-}= {} & \boldsymbol{S}_{2,sorp} D_{\boldsymbol{\xi}_{glob}} \bar{\boldsymbol{\xi}}_{sorp} \Delta\boldsymbol{\xi}_{glob} \\
\bar{\boldsymbol{c}}_{min} \mathrel{-}= {} & D_{\boldsymbol{\xi}_{glob}} \bar{\boldsymbol{\xi}}_{min} \Delta\boldsymbol{\xi}_{glob} \,.
\end{aligned}$$

Incorporating the strategies of this section and the section before we get the following algorithm for one step of the reduction scheme:

One time step of the reduction scheme

```
Solve $\eta$-problem
Update concentrations $\boldsymbol{c} \mathrel{+}= \boldsymbol{B}_1^{\perp}(\boldsymbol{\eta} - \boldsymbol{\eta}_{old})$
If $c_i < 0$ starting value search
Solve local problem (unknowns: $\boldsymbol{c}, \bar{\boldsymbol{c}}, \bar{\boldsymbol{\xi}}_{kin}$)
Calculate $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{min}$ by their definitions
Calculate defect $\boldsymbol{d}$ of the global problem
Stopping criteria for global problem not fulfilled
    Assemble Jacobi matrix $\boldsymbol{J}$ of the global problem
    Solve linear system $\boldsymbol{J}\Delta\boldsymbol{\xi}_{glob} = \boldsymbol{d}$
    Update $\boldsymbol{\xi}_{glob} \mathrel{-}= \Delta\boldsymbol{\xi}_{glob}$
    If $\xi_{glob,i} > 0$ cutting-off of global Newton step
    Update concentrations
    $\boldsymbol{c} \mathrel{-}= \boldsymbol{S}_{glob}\Delta\boldsymbol{\xi}_{glob} + \boldsymbol{S}_{loc} D_{\boldsymbol{\xi}_{glob}} \boldsymbol{\xi}_{loc} \Delta\boldsymbol{\xi}_{glob}$
    Solve local problem (unknowns: $\boldsymbol{c}, \bar{\boldsymbol{c}}, \bar{\boldsymbol{\xi}}_{kin}$)
    Calculate $\boldsymbol{\xi}_{mob}$, $\bar{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{min}$ by their definitions
    Calculate defect $\boldsymbol{d}$ of the global problem
```

3.4.4 FV–Stabilization for Convection Dominated Problems

The stabilization technique given in this section is only applicable for linear elements on triangles. To a given triangulation $\mathcal{T}_h$ consisting of triangles T a dual grid is generated by connecting all edge midpoints of one triangle with the barycenter of the triangle. Thus to each node $\boldsymbol{a}_k$ a control volume Ω_k is generated (see Fig. 3.2). Such a family of control volumes is called *Donald–Diagram*.

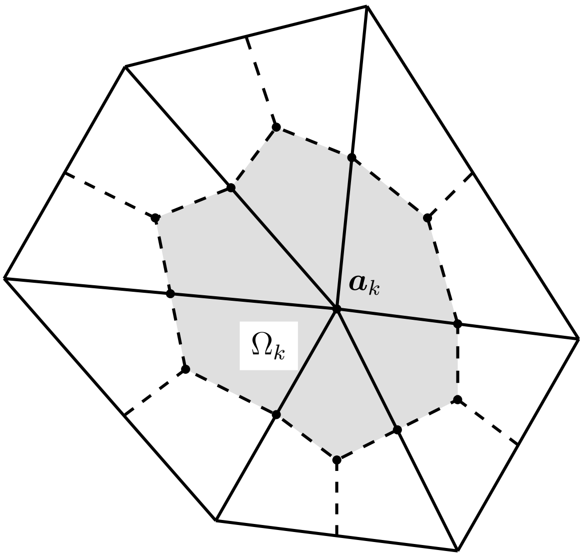


Figure 3.2: Control volume Ω_k associated to the node $\boldsymbol{a}_k$

The integral appearing at the discretization of the advective term with linear Finite Elements can be approximated by

$$\int_\Omega \nabla \cdot (\boldsymbol{q} c_{h,i}) \varphi_k \, d\boldsymbol{x} \approx \int_{\Omega_k} \nabla \cdot (\boldsymbol{q} c_{h,i}) \, d\boldsymbol{x} = \int_{\partial\Omega_k} (\boldsymbol{q} c_{h,i}) \cdot \boldsymbol{\nu} \, d\sigma$$

where φ_k is the basis function associated to the node $\boldsymbol{a}_k$. This boundary integral allows the treatment of the advective term with the Finite Volume method. The approximation of the integral is the same approximation which is applied to terms without spatial derivatives by using mass lumping. Hence the approximation is justified.

Another point of view is that we carry out a Finite Volume discretization for the whole partial differential equation. Let us consider the Finite Volume scheme treated in [KA03, chapter 6.2] for the case of the Donald Diagram. In this scheme the diffusive term is discretized as in the Finite Element method and the discretization of the terms without any space derivative is the same as using linear Finite Elements and mass lumping, because for the Donald Diagram it holds $|\Omega_i \cap T| = |T|/3$. So in this FV scheme the discretization of all terms except of the advective one coincides with that one of the linear Finite Element Method using mass lumping. Hence we only have to replace the advective term in the linear Finite Element discretization to get a Finite Volume scheme.

The integral

$$\int_{\partial\Omega_k} (\boldsymbol{q}c_{h,i}) \cdot \boldsymbol{\nu}\, d\sigma = \sum_{T\in\mathcal{T}_h} \int_{\partial\Omega_k\cap T} (\boldsymbol{q}c_{h,i}) \cdot \boldsymbol{\nu}\, d\sigma$$

is discretized in the following way (see Fig. 3.3 for the used notation):

$$\begin{aligned}\int_{\partial\Omega_k\cap T} (\boldsymbol{q}c_{h,i}) \cdot \boldsymbol{\nu}\, d\sigma \approx \boldsymbol{q}((\bar{\boldsymbol{a}}_0 + \boldsymbol{a}_c)/2) \cdot \boldsymbol{\nu}_{[\bar{\boldsymbol{a}}_0,\boldsymbol{a}_c]}(r_{01}c_{h,i}(\boldsymbol{a}_0) + (1-r_{01})c_{h,i}(\boldsymbol{a}_1)) \\ + \boldsymbol{q}((\boldsymbol{a}_c + \bar{\boldsymbol{a}}_2)/2) \cdot \boldsymbol{\nu}_{[\boldsymbol{a}_c,\bar{\boldsymbol{a}}_2]}(r_{02}c_{h,i}(\boldsymbol{a}_0) + (1-r_{02})c_{h,i}(\boldsymbol{a}_2))\end{aligned}$$

where the normal vectors $\boldsymbol{\nu}_{[\bar{\boldsymbol{a}}_0,\boldsymbol{a}_c]}$ and $\boldsymbol{\nu}_{[\boldsymbol{a}_c,\bar{\boldsymbol{a}}_2]}$ have the length $|\bar{\boldsymbol{a}}_0 - \boldsymbol{a}_c|$ and $|\boldsymbol{a}_c - \bar{\boldsymbol{a}}_2|$, respectively. For the choice of the parameters $r_{01}, r_{02} \in [0,1]$ there are two possibilities.

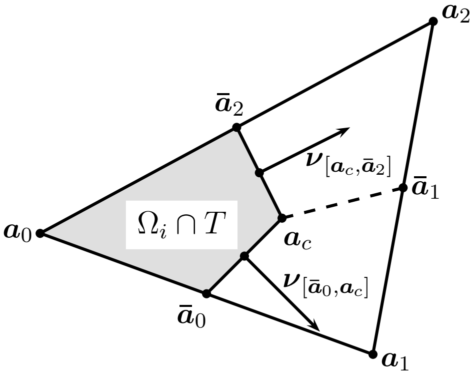


Figure 3.3: Intersection of a control volume with a triangle

The first one is full upwinding. Here the parameters r_{01}, r_{02} are chosen in the following way:

$$\begin{aligned} r_{01} &= \begin{cases} 1 & \text{for } \boldsymbol{q}((\bar{\boldsymbol{a}}_0 + \boldsymbol{a}_c)/2) \cdot \boldsymbol{\nu}_{[\bar{\boldsymbol{a}}_0,\boldsymbol{a}_c]} \geq 0 \\ 0 & \text{else} \end{cases} \\ r_{02} &= \begin{cases} 1 & \text{for } \boldsymbol{q}((\boldsymbol{a}_c + \bar{\boldsymbol{a}}_2)/2) \cdot \boldsymbol{\nu}_{[\boldsymbol{a}_c,\bar{\boldsymbol{a}}_2]} \geq 0 \\ 0 & \text{else} \end{cases} \end{aligned} \tag{3.67}$$

This discretization is suitable for convection dominated problems.

But using full upwinding the PDE is stabilized on the whole domain. When the PDE is convection dominated only on a part of the domain this is not necessary and leads to an imprecise solution in that part of the domain which is not convection dominated. In this case it is preferable to use exponential upwinding instead.

Using exponential upwinding the parameters r_{ij} are chosen in the following way

$$r_{ij} = 1 - \frac{1}{z_{ij}} \left(1 - \frac{z_{ij}}{\exp(z_{ij}) - 1} \right)$$

with the local Péclet numbers z_{ij}

$$\begin{aligned} z_{01} &= \frac{\boldsymbol{q}((\bar{\boldsymbol{a}}_0 + \boldsymbol{a}_c)/2) \cdot \boldsymbol{\nu}_{[\bar{\boldsymbol{a}}_0, \boldsymbol{a}_c]}}{-|T| \boldsymbol{D} \nabla \varphi_1 \cdot \nabla \varphi_0} \\ z_{02} &= \frac{\boldsymbol{q}((\boldsymbol{a}_c + \bar{\boldsymbol{a}}_2)/2) \cdot \boldsymbol{\nu}_{[\boldsymbol{a}_c, \bar{\boldsymbol{a}}_2]}}{-|T| \boldsymbol{D} \nabla \varphi_2 \cdot \nabla \varphi_0} \end{aligned}$$

where φ_i is the linear Finite Element basis function to the node $\boldsymbol{a}_i$ on the triangle T. Note that $\nabla \varphi_i$ is constant on the triangle T. Using for $\boldsymbol{D}$ the Scheidegger tensor (2.1) $\boldsymbol{D}$ depends on $\boldsymbol{q}$. In this case $\boldsymbol{q}$ is evaluated at the barycenter $\boldsymbol{a}_c$ of the triangle T, i.e.

$$\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{q}(\boldsymbol{a}_c)) \,.$$

When the denominator $-|T| \boldsymbol{D} \nabla \varphi_2 \cdot \nabla \varphi_0$ is not positive (which can happen because $\boldsymbol{D}$ is a full tensor) the parameters r_{ij} are computed with (3.67).

The calculation of the local Péclet numbers z_{ij} is slightly different to that one used in [KA03, chapter 6.2]. There the quantities on the two adjacent triangles, which have the vertices $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ in common, are used to calculate z_{ij}. Here because of programming restrictions only the quantities on one triangle can be used.

At the Neumann boundary additional boundary integrals on $\partial\Omega \cap \partial\Omega_k$ appear. Here we only handle homogeneous Neumann boundary conditions. Let $[\boldsymbol{a}_0, \bar{\boldsymbol{a}}_0]$ be a half edge on the Neumann boundary. Then the additional boundary integral is discretized in the following way:

$$\int_{[\boldsymbol{a}_0, \bar{\boldsymbol{a}}_0]} (\boldsymbol{q} c_{h,i}) \cdot \boldsymbol{\nu} \; d\sigma \approx \boldsymbol{q}((\boldsymbol{a}_0 + \bar{\boldsymbol{a}}_0)/2) \cdot \boldsymbol{\nu}_{[\boldsymbol{a}_0, \bar{\boldsymbol{a}}_0]} c_{h,i}(\boldsymbol{a}_0)$$

The advantage of the Finite Volume method is the *inverse monotonicity.* For full upwinding, a triangulation $\mathcal{T}_h$ consisting of solely nonobtuse triangles and a scalar diffusion coefficient, which is constant on the single elements of $\mathcal{T}_h$, one can proof that the resulting discretization is inverse monotone (see [KA03, Theorem 6.19]). Numerical experiments showed that in most cases also for exponential upwinding this FV method is inverse monotone.

In the transport reaction problems considered in this work the equilibrium conditions are formulated with the logarithms of the concentrations. So it is important that the discrete solution is nonnegative and hence an inverse monotone method is needed.

3.4.5 Anisotropic Diffusion Tensors

Another fact which can cause negative values in the discrete solution is the anisotropic diffusion tensor. In this section we want to examine under which conditions we can guarantee that the discretization of the diffusive term leads to an M-matrix despite of the anisotropic tensor. We restrict the examination to linear elements on regular triangular grids. By use of the Scheidegger dispersion tensor (2.1) we have to study the terms

$$\int_\Omega \left((\beta_l - \beta_t) \frac{\boldsymbol{q} \otimes \boldsymbol{q}}{|\boldsymbol{q}|} \nabla\varphi_i \cdot \nabla\varphi_j + \beta_t |\boldsymbol{q}| \nabla\varphi_i \cdot \nabla\varphi_j \right) \, d\boldsymbol{x}$$

where φ_i and φ_j are basis functions of linear Finite Elements. First we assume that we have a constant flow $\boldsymbol{q} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$. Then we can rewrite the integral as

$$|\boldsymbol{q}| \left((\beta_l - \beta_t) \int_\Omega \frac{\boldsymbol{q} \otimes \boldsymbol{q}}{|\boldsymbol{q}|^2} \nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x} + \beta_t \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x} \right) . \tag{3.68}$$

For a Friedrichs–Keller triangulation with side length of the squares h we get the following contributions of the term $\int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x}$. For simplicity we write the contributions as a stencil:

$$\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix} \tag{3.69}$$

In a Friedrichs–Keller triangulation with diagonals from the lower right to the upper left (compare Fig. 3.4 left) we get from the term $\int_\Omega \frac{\boldsymbol{q} \otimes \boldsymbol{q}}{|\boldsymbol{q}|^2} \nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x}$

$$\frac{1}{q_1^2 + q_2^2} \begin{bmatrix} q_1 q_2 & -q_1 q_2 - q_2^2 & \\ -q_1 q_2 - q_1^2 & (q_1^2 + q_2^2) + (q_1 - q_2)^2 & -q_1 q_2 - q_1^2 \\ & -q_1 q_2 - q_2^2 & q_1 q_2 \end{bmatrix} .$$

We see immediately that for a flow parallel to a coordinate axis the matrix fulfills that the diagonal entries are positive and the nondiagonal entries are nonpositive because in this case we have $q_1 q_2 = 0$. We also see that it is not possible to get an M-matrix when $q_1 q_2 > 0$ because then the upper left and lower right entry are positive and there are no entries in (3.69) that can cancel these positive entries.

In a Friedrichs–Keller triangulation with diagonals from the lower left to the upper right (compare Fig. 3.4 right) we get from the term $\int_\Omega \frac{\boldsymbol{q}\otimes\boldsymbol{q}}{|\boldsymbol{q}|^2}\nabla\varphi_i \cdot \nabla\varphi_j \, d\boldsymbol{x}$

$$\frac{1}{q_1^2+q_2^2}\begin{bmatrix} & q_1q_2 - q_2^2 & -q_1q_2 \\ q_1q_2 - q_1^2 & (q_1^2+q_2^2)+(q_1-q_2)^2 & q_1q_2-q_1^2 \\ -q_1q_2 & q_1q_2-q_2^2 & \end{bmatrix}.$$

For this choice of the diagonals it holds that for $q_1q_2 < 0$ it is not possible to get an M-matrix.

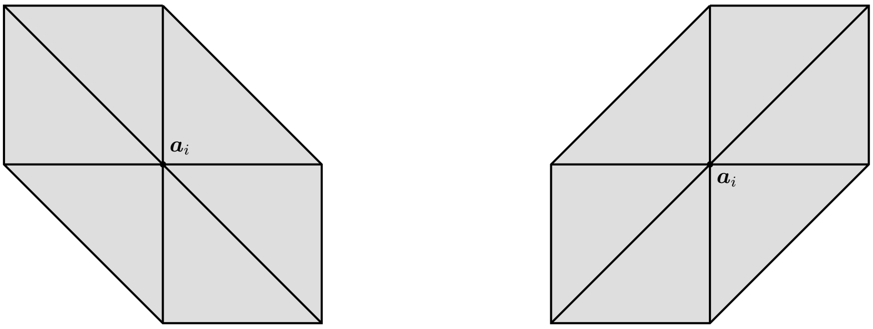


Figure 3.4: Support of one basis function

So we choose the direction of the diagonals depending on the flow. For $q_1q_2 > 0$ we take the diagonals from the lower left to the upper right and for $q_1q_2 < 0$ we take the diagonals from the lower right to the upper left. That means we choose the diagonals as parallel as possible to the flow direction.

To guarantee that (3.68) is nonpositive for $i \neq j$ we have to check that

$$(\beta_l - \beta_t)\frac{-q_1q_2 - q_i^2}{q_1^2+q_2^2} + \beta_t(-1) \leq 0 \qquad \text{for } q_1q_2 < 0,\ i = 1,2 \tag{3.70}$$

$$(\beta_l - \beta_t)\frac{q_1q_2 - q_i^2}{q_1^2+q_2^2} + \beta_t(-1) \leq 0 \qquad \text{for } q_1q_2 > 0,\ i = 1,2\,. \tag{3.71}$$

For this purpose we determine the maximum of $\frac{-q_1q_2-q_i^2}{q_1^2+q_2^2}$, $\frac{q_1q_2-q_i^2}{q_1^2+q_2^2}$, respectively. By differentiating and some simple calculations we get that these maxima are $\frac{\sqrt{2}-1}{2}$. So with this choice of the diagonals it is secured for

$$(\beta_l - \beta_t)\frac{\sqrt{2}-1}{2} - \beta_t \leq 0 \qquad \Leftrightarrow \frac{\beta_l}{\beta_t} \leq 3 + 2\sqrt{2}\ (\approx 5.528)$$

that all nondiagonal entries are nonpositive.

For a nonconstant flow this motivates the following proceeding. We start with a grid of squares with side length h. Let c_i denote the center of the i-th square. Then for each cell we check if $q_1(c_i)q_2(c_i)$ is positive or negative. For $q_1(c_i)q_2(c_i) < 0$ we take the diagonal from the lower right to the upper left and for $q_1(c_i)q_2(c_i) > 0$ we take the diagonal from the lower left to the upper right. In Fig. 3.5 a nonconstant flow and an appropriate grid constructed with this strategy is shown.

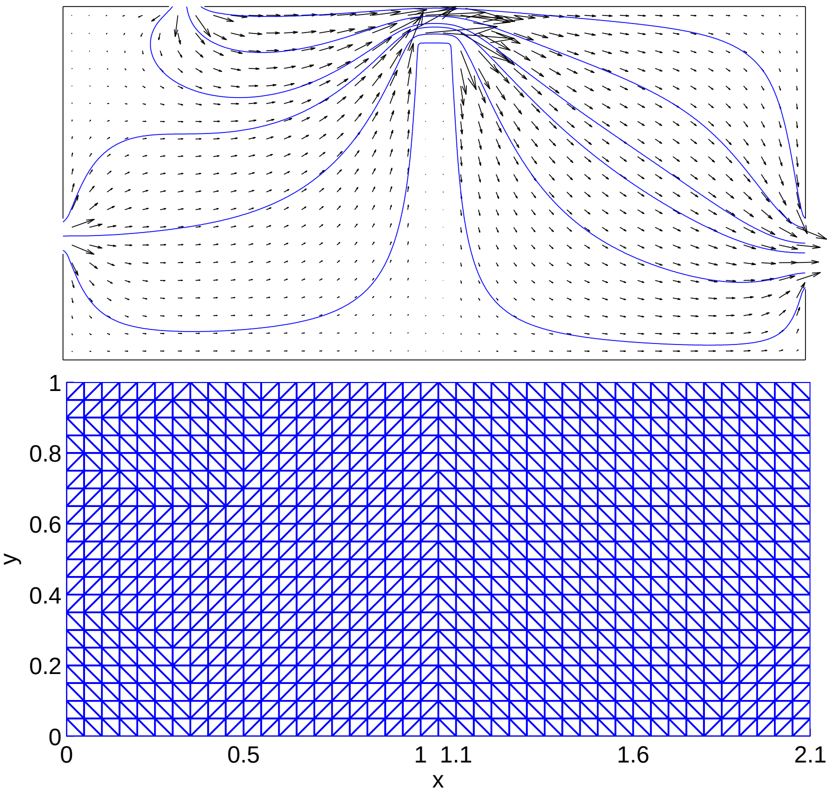


Figure 3.5: Nonconstant flow and appropriate grid

Even for ratios β_l/β_t that are greater than $3 + 2\sqrt{2}$ it is advantageous to use this alignment of the diagonals. Numerical tests with pure transport problems and $\beta_l/\beta_t = 10$ have shown that the number of points on which the discrete solution is negative can be decreased significantly and that the absolute value of the negative values is much smaller.

There is a second possibility to derive the condition for the choice of the diagonals and the conditions (3.70) and (3.71). First we define the linear transformation

$$\boldsymbol{F}(\tilde{\boldsymbol{x}}) = \boldsymbol{B}\tilde{\boldsymbol{x}} + \boldsymbol{d}$$

and the transformed triangle

$$\tilde{T} := \boldsymbol{F}^{-1}(T)\,.$$

Using the transformation formula (see e.g. [KA03, Sec. 3.5.2])

$$\int_T \boldsymbol{D}\nabla_{\boldsymbol{x}}\varphi_i(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\varphi_j(\boldsymbol{x})\; d\boldsymbol{x} = \int_{\tilde{T}} \boldsymbol{D}\boldsymbol{B}^{-T}\nabla_{\tilde{\boldsymbol{x}}}\tilde{\varphi}_i(\tilde{\boldsymbol{x}}) \cdot \boldsymbol{B}^{-T}\nabla_{\tilde{\boldsymbol{x}}}\tilde{\varphi}_j(\tilde{\boldsymbol{x}})\; d\tilde{\boldsymbol{x}} |\det(\boldsymbol{B})|$$

with $\boldsymbol{B} = \boldsymbol{D}^{1/2}$ ($\boldsymbol{D}^{1/2}$ exists because $\boldsymbol{D}$ is symmetric) we get

$$\int_\Omega \boldsymbol{D}\nabla_{\boldsymbol{x}}\varphi_i(\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}}\varphi_j(\boldsymbol{x})\; d\boldsymbol{x} = \int_{\tilde{\Omega}} \nabla_{\tilde{\boldsymbol{x}}}\tilde{\varphi}_i(\tilde{\boldsymbol{x}}) \cdot \nabla_{\tilde{\boldsymbol{x}}}\tilde{\varphi}_j(\tilde{\boldsymbol{x}})\; d\tilde{\boldsymbol{x}} |\det(\boldsymbol{B})|\,.$$

Remember that to derive the conditions (3.70) and (3.71) we have assumed that the flow $\boldsymbol{q}$ is constant and so the diffusion tensor $\boldsymbol{D}$ is constant. For the integral on the right hand side it is known that the angle condition must be fulfilled to get an M-matrix (see [KA03, Sec. 3.9]). The angle condition says that for any two triangles of $\mathcal{T}_h$ with a common edge the sum of the interior angles opposite to this edge does not exceed the value π.

In a Friedrichs–Keller triangulation two angles opposite to one edge are equal. This remains true for the transformed triangulation because we use a linear transformation with a constant matrix $\boldsymbol{B}$. So here the angle condition is equivalent to all angles are not obtuse. So we have to check under which conditions all angles of the transformed triangles $\tilde{T}$ are not obtuse.

This can be done by examining the sign of the scalar product of two sides of a transformed triangle $\tilde{T}$. The vertices of the transformed triangle $\tilde{T}$ are denoted with $\tilde{\boldsymbol{a}}_{\boldsymbol{1}}, \tilde{\boldsymbol{a}}_{\boldsymbol{2}}, \tilde{\boldsymbol{a}}_{\boldsymbol{3}}$ and the vertices of the original triangle T with $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$. So we get for the scalar product of two sides of a transformed triangle $\tilde{T}$

$$\begin{aligned}
&(\tilde{\boldsymbol{a}}_{\boldsymbol{3}} - \tilde{\boldsymbol{a}}_{\boldsymbol{1}}) \cdot (\tilde{\boldsymbol{a}}_{\boldsymbol{2}} - \tilde{\boldsymbol{a}}_{\boldsymbol{1}}) \\
&= \big(\boldsymbol{B}^{-1}\boldsymbol{a}_3 + \boldsymbol{B}^{-1}\boldsymbol{d} - (\boldsymbol{B}^{-1}\boldsymbol{a}_1 + \boldsymbol{B}^{-1}\boldsymbol{d})\big) \cdot \big(\boldsymbol{B}^{-1}\boldsymbol{a}_2 + \boldsymbol{B}^{-1}\boldsymbol{d} - (\boldsymbol{B}^{-1}\boldsymbol{a}_1 + \boldsymbol{B}^{-1}\boldsymbol{d})\big) \\
&= (\boldsymbol{a}_3 - \boldsymbol{a}_1)^T \boldsymbol{D}^{-1} (\boldsymbol{a}_2 - \boldsymbol{a}_1)\,.
\end{aligned}$$

We calculate that (see (2.1) for the definition of $\boldsymbol{D}$)

$$\boldsymbol{D}^{-1} = \frac{1}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \begin{pmatrix} q_1^2\beta_t + q_2^2\beta_l & -q_1q_2(\beta_l - \beta_t) \\ -q_1q_2(\beta_l - \beta_t) & q_1^2\beta_l + q_2^2\beta_t \end{pmatrix}.$$

For a triangle T of a Friedrichs–Keller triangulation with the diagonal from the lower left to the upper right we get the three conditions

$$\begin{pmatrix}-h & 0\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}0\\ h\end{pmatrix} = h^2 \frac{q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0 \tag{3.72}$$

$$\begin{pmatrix}0 & -h\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}-h\\ -h\end{pmatrix} = h^2 \frac{q_1^2\beta_l + q_2^2\beta_t - q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0 \tag{3.73}$$

$$\begin{pmatrix}h & h\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}h\\ 0\end{pmatrix} = h^2 \frac{q_1^2\beta_t + q_2^2\beta_l - q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0\,. \tag{3.74}$$

Condition (3.72) is equivalent to $q_1 q_2 \geq 0$, i.e., the direction of the diagonal from the lower left to the upper right can only be chosen for $q_1 q_2 \geq 0$, condition (3.73) is equivalent to (3.71) with $i = 1$ and condition (3.74) is equivalent to (3.71) with $i = 2$.

For a triangle T of a Friedrichs–Keller triangulation with the diagonal from the lower right to the upper left we get the three conditions

$$\begin{pmatrix}0 & h\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}h\\ 0\end{pmatrix} = -h^2 \frac{q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0 \tag{3.75}$$

$$\begin{pmatrix}h & -h\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}0\\ -h\end{pmatrix} = h^2 \frac{q_1^2\beta_l + q_2^2\beta_t + q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0 \tag{3.76}$$

$$\begin{pmatrix}-h & 0\end{pmatrix} \boldsymbol{D}^{-1} \begin{pmatrix}-h\\ h\end{pmatrix} = h^2 \frac{q_1^2\beta_t + q_2^2\beta_l + q_1 q_2(\beta_l - \beta_t)}{(q_1^2 + q_2^2)^{(3/2)}\beta_l\beta_t} \geq 0\,. \tag{3.77}$$

Condition (3.75) is equivalent to $q_1 q_2 \leq 0$, i.e., this direction of the diagonal can only be chosen for $q_1 q_2 \leq 0$, condition (3.76) is equivalent to (3.70) with $i = 1$ and condition (3.77) is equivalent to (3.70) with $i = 2$.

3.5 Analysis of the Method

3.5.1 Boundedness of the Derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$

The reason, why the method with the additional variables $\tilde{\boldsymbol{\xi}}$ has been developed, is that by use of the additional variables the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ has small absolute values. In the first formulation of the reduction scheme out of [KK07], [Krä08] it is not the case that the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ has small absolute values (see Sec. 3.6.1). This is probably the reason why the method with the additional variables has much better convergence properties.

We look at one example with the chemistry of the "easy test case" of the MoMaS–benchmark (see Chap. 4). The matrix of the linear system to calculate $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ (see Sec. 3.3.2 how to compute $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$) is in this example

```
    1.00      -0.00       0.00       0.00      -0.00      -0.00       0.00
   -0.00       1.00      -0.00       0.21       0.62       0.21      -0.00
    0.16      -0.16       0.31       1.00      -0.52      -0.48       0.60
    0.00       0.33       0.00       0.69       1.00       0.33       0.00
   -0.00       0.00      -0.00       0.00       1.00       0.00      -0.00
   -0.00       0.33      -0.00       0.33       1.00       0.33      -0.00
    0.21      -0.21       0.26       1.00      -0.79      -0.54       0.93
```

and the right hand side is:

```
    0.00      -0.00
   -0.21       0.00
    0.48      -0.60
   -0.33      -0.00
   -0.00       0.00
   -0.33       0.00
    0.63      -0.68
```

This yields for $D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$:

```
 -1.00      -0.00
  0.22      -0.42
```

In this example the derivatives $D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$ have small absolute values. In the following a proof will be given that using the additional variables $\tilde{\boldsymbol{\xi}}$ the absolute values of the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ are bounded with a bound independent of the concentration values and the reaction constants.

The linear system of equations for calculating $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ has the form (see (3.53))

$$\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B}\boldsymbol{X} = \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{C}$$

where $\boldsymbol{B}$, $\boldsymbol{C}$ consist only of stoichiometric coefficients. We can rewrite the linear system as $\boldsymbol{B}\boldsymbol{X} = \boldsymbol{P}\boldsymbol{C}$ with

$$\boldsymbol{P} := \boldsymbol{B}(\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B})^{-1}\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\,. \tag{3.78}$$

$\boldsymbol{P}$ is a projection because $\boldsymbol{P}^2 = \boldsymbol{P}$. Obviously $\mathrm{im}(\boldsymbol{P}) \subset \boldsymbol{B}$ and $\boldsymbol{P}\boldsymbol{B} = \boldsymbol{B}$. So $\boldsymbol{P}$ projects onto $\boldsymbol{B}$.

The linear system can also be interpreted as the normal equations of the linear least squares problem

$$\min \|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{B}\boldsymbol{X} - \boldsymbol{C})\|_2\,. \tag{3.79}$$

This formulation will be used to prove that $\boldsymbol{X}$ is bounded with a bound only depending on the entries of $\boldsymbol{B}$ and $\boldsymbol{C}$. So the bound depends only on the stoichiometric coefficients.

For $\boldsymbol{B}$ consisting of one column $\boldsymbol{b}$, $\boldsymbol{b} \neq \boldsymbol{0}$, $\boldsymbol{C}$ consisting of one column $\boldsymbol{c}$ and $\boldsymbol{X}$ a scalar x it holds

$$\min_{b_i \neq 0} \frac{c_i}{b_i} \leq x \leq \max_{b_i \neq 0} \frac{c_i}{b_i} \,. \tag{3.80}$$

This should be proven by contradiction. If the second relation were not be true we would have

$$x > \max_{b_i \neq 0} \frac{c_i}{b_i} \,.$$

Then choose $\varepsilon > 0$ sufficiently small such that

$$x - \varepsilon > \max_{b_i \neq 0} \frac{c_i}{b_i} \,.$$

It follows

$$b_i x > b_i(x - \varepsilon) > c_i \qquad \text{for all } b_i > 0 \tag{3.81}$$

$$b_i x < b_i(x - \varepsilon) < c_i \qquad \text{for all } b_i < 0 \,. \tag{3.82}$$

1. case: $b_i > 0$
Because of $\varepsilon > 0$ we always have

$$\begin{aligned} & 0 > -b_i\varepsilon \\ \Leftrightarrow \quad & b_i x - c_i > b_i(x - \varepsilon) - c_i \,. \end{aligned}$$

Because of (3.81) it follows

$$|b_i x - c_i| > |b_i(x - \varepsilon) - c_i| \,.$$

2. case: $b_i < 0$
Because of $\varepsilon > 0$ we always have

$$\begin{aligned} & 0 > b_i\varepsilon \\ \Leftrightarrow \quad & -b_i x + c_i > -b_i(x - \varepsilon) + c_i \,. \end{aligned}$$

Because of (3.82) it follows

$$|b_i x - c_i| > |b_i(x - \varepsilon) - c_i| \,.$$

Altogether we have

$$|b_i x - c_i| > |b_i(x - \varepsilon) - c_i| \qquad \text{for all } b_i \neq 0.$$

So the absolute value of each component of $\boldsymbol{b}x(1-\varepsilon)-\boldsymbol{c}$ is smaller or equal to that one of $\boldsymbol{b}x-\boldsymbol{c}$, where it is strictly smaller for $b_i \neq 0$. So if $\boldsymbol{b} \neq \boldsymbol{0}$ it is not possible that x is the minimum of

$$\|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{b}x-\boldsymbol{c})\|_2$$

because $\tilde{\boldsymbol{\Lambda}}$ is a diagonal matrix with positive entries and so we have $\|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{b}x(1-\varepsilon)-\boldsymbol{c})\|_2 < \|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{b}x-\boldsymbol{c})\|_2$. So we have proven the second relation of (3.80). The first relation of (3.80) can be proven completely analogously.

In the case $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^2$ this proceeding can be interpreted in a graphical way: The point $\boldsymbol{c}$ is projected on the straight line span$\{\boldsymbol{b}\}$ such that the distance between the point and the projection of the point is minimal in the norm $\|\tilde{\boldsymbol{\Lambda}}^{1/2}\cdot\|_2$. With the contradiction argument used above one shows that the projection of the point must be in the red part of the straight line in Fig. 3.6.

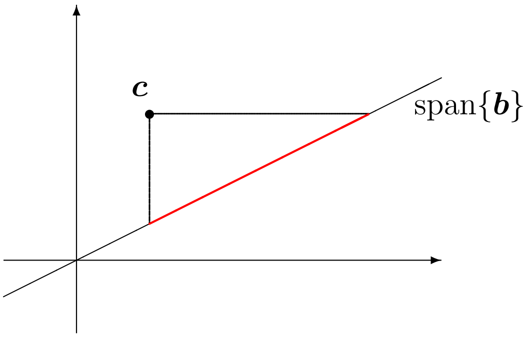


Figure 3.6: Graphical interpretation

Now let us consider the case where $\boldsymbol{B}$ is a matrix and $\boldsymbol{C}$ consists of one column $\boldsymbol{c}$. As (3.79) is a linear problem it is sufficient to examine the case where $\boldsymbol{c}$ is a unit vector $\boldsymbol{e}_k$. We can assume that the matrix $\boldsymbol{B}$ does not contain a row in which all entries are zero. If the matrix $\boldsymbol{B}$ contains a row in which all entries are zero this row and the corresponding entry of $\boldsymbol{c}$ can be left out without changing the result of (3.79). Let $\boldsymbol{b}_i$ denote the i-th *row* of $\boldsymbol{B}$.

Assertion:

$$0 \leq \boldsymbol{b}_k \cdot \boldsymbol{x} \leq 1 \tag{3.83}$$

If the second relation were not be true we would have

$$\boldsymbol{b}_k \cdot \boldsymbol{x} > 1.$$

Now choose $\varepsilon > 0$ sufficiently small such that

$$\boldsymbol{b}_k \cdot \boldsymbol{x}(1-\varepsilon) > 1. \tag{3.84}$$

Let us consider

$$\boldsymbol{Bx}(1-\varepsilon) - \boldsymbol{e}_k\,.$$

For the absolute value of the i-th component of this vector it holds

$$\begin{aligned} |\boldsymbol{b}_i \cdot \boldsymbol{x}(1-\varepsilon)| &\le |\boldsymbol{b}_i \cdot \boldsymbol{x}| && \text{for } i \neq k \\ |\boldsymbol{b}_i \cdot \boldsymbol{x}(1-\varepsilon) - 1| &< |\boldsymbol{b}_i \cdot \boldsymbol{x} - 1| && \text{for } i = k\,. \end{aligned}$$

The second relation is true because of (3.84).

Altogether it follows that the absolute value of each component of $\boldsymbol{Bx}(1-\varepsilon) - \boldsymbol{e}_k$ is smaller or equal to that one of $\boldsymbol{Bx} - \boldsymbol{e}_k$ where the k-th component is strictly smaller. So it is not possible that $\boldsymbol{x}$ is the minimum of

$$\|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{Bx} - \boldsymbol{e}_k)\|_2$$

because $\tilde{\boldsymbol{\Lambda}}$ is a diagonal matrix with positive entries. So we have proven the second relation of assertion (3.83). The first relation can be proven analogously.

We write $\boldsymbol{x}$ as $\alpha\boldsymbol{b}_k + \boldsymbol{a}$ with a vector $\boldsymbol{a}$ fulfilling $\boldsymbol{a} \cdot \boldsymbol{b}_k = 0$. Taking the scalar product with $\boldsymbol{b}_k$ of the equation $\boldsymbol{x} = \alpha\boldsymbol{b}_k + \boldsymbol{a}$ and using (3.83) gives

$$\begin{aligned} \alpha &= \frac{\boldsymbol{b}_k \cdot \boldsymbol{x}}{|\boldsymbol{b}_k|^2} \\ \Rightarrow \quad 0 \le \alpha &\le \frac{1}{|\boldsymbol{b}_k|^2}\,. \end{aligned}$$

Plugging the representation $\boldsymbol{x} = \alpha\boldsymbol{b}_k + \boldsymbol{a}$ in (3.79) gives

$$\min \left\|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{Bx} - \boldsymbol{c})\right\| = \min \left\|\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{Ba} - (\boldsymbol{c} - \alpha\boldsymbol{Bb}_k))\right\|.$$

Because of $\boldsymbol{b}_k \cdot \boldsymbol{a} = 0$ (notice that $\boldsymbol{b}_k$ is the k-th row of $\boldsymbol{B}$) we can leave out the k-th component of the vector $\tilde{\boldsymbol{\Lambda}}^{1/2}(\boldsymbol{Ba} - (\boldsymbol{c} - \alpha\boldsymbol{Bb}_k))$. This does not affect the value of $\boldsymbol{a}$ at the minimum. As $\boldsymbol{a}$ fulfills the condition $\boldsymbol{b}_k \cdot \boldsymbol{a} = 0$ we can write $\boldsymbol{a}$ as $\boldsymbol{B}_k^{\perp}\hat{\boldsymbol{a}}$ with $\boldsymbol{B}_k^{\perp}$ the orthogonal complement of $\boldsymbol{b}_k$. One possible choice for $\boldsymbol{B}_k^{\perp}$ is (w.l.o.g. we assume that $b_{k1} \neq 0$, this is possible because of the assumption that the matrix $\boldsymbol{B}$ does not contain a row in which all entries are zero)

$$\boldsymbol{B}_k^{\perp} = \begin{pmatrix} -b_{k2} & -b_{k3} & \dots & -b_{kn} \\ b_{k1} & 0 & \dots & 0 \\ 0 & b_{k1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{k1} \end{pmatrix}. \tag{3.85}$$

So we get the minimization problem

$$\min \left\| \hat{\boldsymbol{\Lambda}}^{1/2} (\hat{\boldsymbol{B}} \hat{\boldsymbol{a}} - \hat{\boldsymbol{c}}) \right\|$$

with

$$\begin{aligned}
\hat{\boldsymbol{B}} &= (\boldsymbol{B} \text{ without } k\text{-th row}) \boldsymbol{B}_k^{\perp} \\
\hat{\boldsymbol{c}} &= (\boldsymbol{c} \text{ without } k\text{-th row}) - \alpha (\boldsymbol{B} \text{ without } k\text{-th row}) \boldsymbol{b}_k \\
\hat{\boldsymbol{\Lambda}} &= (\tilde{\boldsymbol{\Lambda}} \text{ without } k\text{-th row and } k\text{-th column}) \,.
\end{aligned}$$

Then we apply the assertion (3.83) to this new problem. The matrix $\hat{\boldsymbol{B}}$ has one column less than the matrix $\boldsymbol{B}$. We iterate this procedure until the matrix $\hat{\boldsymbol{B}}$ consists of one column. Then we can apply (3.80). So we get a bound for $\boldsymbol{x}$. In the case $\boldsymbol{C}$ is a matrix we can treat every column of $\boldsymbol{C}$ separately.

The matrices $\boldsymbol{B}$, $\boldsymbol{C}$ contain stoichiometric coefficients. Typically stoichiometric coefficients are integer and have a small absolute value. Hence we know that:

- For $\boldsymbol{B}$ integer and the choice (3.85) for $\boldsymbol{B}_k^{\perp}$ the matrix $\boldsymbol{B}_k^{\perp}$ has only integer entries.

- Then the entries of $\hat{\boldsymbol{B}}$ are also integer and they depend only on the entries of $\boldsymbol{B}$. Especially dividing by $|\hat{\boldsymbol{b}}_k|^2$ does not lead to large numbers.

- It holds $|\hat{c}_i| \leq |c_i| + \frac{1}{|\boldsymbol{b}_k|^2} |(\boldsymbol{B}\boldsymbol{b}_k)_i|$. So the entries of $\hat{\boldsymbol{c}}$ are bounded with a bound only depending on $\boldsymbol{B}$ and $\boldsymbol{c}$.

So we get a bound for $D_{\boldsymbol{\xi}_{glob}} \boldsymbol{\xi}_{loc}$ which depends only on the stoichiometric coefficients of the matrices $\boldsymbol{B}$, $\boldsymbol{C}$. Particularly the bound is independent of the concentration values and the reaction constants.

The estimate (3.80) is sharp. This should be shown with help of the following example. Let us consider one mobile species, two immobile species and one equilibrium sorption reaction, namely

$$\mathrm{A} + \overline{\mathrm{B}} \leftrightarrow 2\overline{\mathrm{C}}$$

with the equilibrium condition

$$K c(\mathrm{A}) \bar{c}(\overline{\mathrm{B}}) = \bar{c}(\overline{\mathrm{C}})^2 \,.$$

Applying the reduction scheme with the transformation matrices

$$\boldsymbol{S}_2^{\perp} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} , \qquad \boldsymbol{B}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \qquad \boldsymbol{B}_2^{\perp} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

leads to the new variables

$$\xi_{sorp} = -c(\mathrm{A})\,,\quad \bar{\xi}_{sorp} = \frac{1}{2}\bar{c}(\overline{\mathrm{C}})\,,\quad \bar{\eta} = \bar{c}(\overline{\mathrm{B}}) + \frac{1}{2}\bar{c}(\overline{\mathrm{C}})\,,\quad \tilde{\xi}_{sorp} = -c(\mathrm{A}) - \frac{1}{2}\bar{c}(\overline{\mathrm{C}})\,.$$

For this example the resolution function $\bar{\xi}_{sorp}(\tilde{\xi}_{sorp})$ can be calculated explicitly. One gets

$$\bar{\xi}_{sorp}(\tilde{\xi}_{sorp}) = \frac{-K(\tilde{\xi}_{sorp} - \bar{\eta}) - \sqrt{K^2(\tilde{\xi}_{sorp}^2 - 2\tilde{\xi}_{sorp}\bar{\eta} + \bar{\eta}^2) + 4K(K-4)\tilde{\xi}_{sorp}\bar{\eta}}}{2(K-4)}\,.$$

So we get for the derivative of the resolution function

$$\bar{\xi}'_{sorp}(\tilde{\xi}_{sorp}) = \frac{-K}{2K-8} - \frac{K^2\tilde{\xi}_{sorp} + K^2\bar{\eta} - 8K\bar{\eta}}{(2K-8)\sqrt{K^2\tilde{\xi}_{sorp}^2 + 2K^2\tilde{\xi}_{sorp}\bar{\eta} + K^2\bar{\eta}^2 - 16K\tilde{\xi}_{sorp}\bar{\eta}}}\,.$$

As $\tilde{\xi}_{sorp}$ is a linear combination of concentrations with negative coefficients $\tilde{\xi}_{sorp}$ is nonpositive and so $\sqrt{\tilde{\xi}_{sorp}^2} = -\tilde{\xi}_{sorp}$. Using this one gets for $\bar{\eta} = 0$

$$\bar{\xi}'_{sorp}(\tilde{\xi}_{sorp}) = \frac{-K^2}{K(2K-8)} - \frac{K^2}{(2K-8)(-K)} = 0\,.$$

As $\bar{\eta}$ is a linear combination of concentrations with positive coefficients $\bar{\eta}$ is nonnegative. So one gets for $\tilde{\xi}_{sorp} = 0$

$$\bar{\xi}'_{sorp}(0) = \frac{-K^2}{K(2K-8)} - \frac{K^2 - 8K}{(2K-8)K} = -1\,.$$

In this example the vectors $\boldsymbol{b}$, $\boldsymbol{c}$ in estimate (3.80) are $\boldsymbol{b} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$, $\boldsymbol{c} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. So (3.80) yields

$$-1 \le \bar{\xi}'_{sorp}(\tilde{\xi}_{sorp}) \le 0\,.$$

Hence one can see that the estimate is sharp.

3.5.2 Condition Number of the Global Jacobian Matrix for $\Delta t = 0$

For $\Delta t = 0$ the global Jacobian matrix (see (3.54)) simplifies to

$$\boldsymbol{J}_{glob} = \begin{pmatrix} \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp,li} & D_{\tilde{\boldsymbol{\xi}}_{min}}\bar{\boldsymbol{\xi}}_{sorp,li} & -\boldsymbol{I} & \boldsymbol{0} & D_{\boldsymbol{\xi}_{kin}}\bar{\boldsymbol{\xi}}_{sorp,li} \\ D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{min,ld} & \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{min}}\bar{\boldsymbol{\xi}}_{min,ld} & \boldsymbol{0} & -\boldsymbol{I} & D_{\boldsymbol{\xi}_{kin}}\bar{\boldsymbol{\xi}}_{min,ld} \\ \theta\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \theta\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \theta\boldsymbol{I} \end{pmatrix}\,.$$

For $\Delta t = 0$ the determinant of $\boldsymbol{J}_{glob}$ is always nonzero. This can easily be seen by expanding the determinant along the last $(J_{sorp,li} + J_{min} + J^*_{1,kin})$ rows. So for Δt small enough the Jacobian matrix is always invertible.

We consider the case that there is one equilibrium sorption reaction and no other chemical reactions. Applying the reduction scheme gives the following transformed variables: one variable ξ_{sorp}, one variable $\bar{\xi}_{sorp}$, one variable $\tilde{\xi}_{sorp}$ and some variables η, $\bar{\eta}$ depending on the number of the species. The global problem consists of the equations

$$\begin{aligned}\tilde{\xi}_{sorp} - \xi_{sorp} + \bar{\xi}_{sorp}(\tilde{\xi}_{sorp}) &= 0\\ \partial_t(\theta\tilde{\xi}_{sorp}) + L\xi_{sorp} &= 0\,.\end{aligned}$$

The Jacobian matrix of the global problem is

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp} & -\boldsymbol{I} \\ \theta\boldsymbol{I} & \Delta t\boldsymbol{L}_h \end{pmatrix}. \tag{3.86}$$

In this case the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ in (3.80) are

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{S}_{1,sorp} \\ \boldsymbol{S}_{2,sorp} \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} -\boldsymbol{S}_{1,sorp} \\ \boldsymbol{0} \end{pmatrix}$$

where the matrices $\boldsymbol{S}_{1,sorp}$, $\boldsymbol{S}_{2,sorp}$ consist only of one column. So with the estimate (3.80) we get

$$-1 \leq \bar{\xi}'_{sorp}(\tilde{\xi}_{sorp}) \leq 0\,. \tag{3.87}$$

Now we will compute the spectral condition number of the global Jacobian for $\theta = 1$ and $\Delta t = 0$. In this case the Jacobian is

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp} & -\boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}.$$

To simplify the computations we calculate the condition number of

$$\tilde{\boldsymbol{J}} := \underbrace{\begin{pmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}}_{=:\boldsymbol{Q}} \boldsymbol{J} = \begin{pmatrix} -\boldsymbol{I} - D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}.$$

As the matrix $\boldsymbol{Q}$ is orthogonal the spectral condition numbers of $\boldsymbol{J}$ and $\tilde{\boldsymbol{J}}$ are the same. For the eigenvectors $\boldsymbol{v}_i$ we make the ansatz

$$\boldsymbol{v}_i = \begin{pmatrix} a_i\boldsymbol{e}_i \\ \boldsymbol{e}_i \end{pmatrix}$$

with $\boldsymbol{e}_i$ the i-th unit vector and $a_i \in \mathbb{R}$. So we get

$$\begin{aligned}\tilde{\boldsymbol{J}}\boldsymbol{v} &= \begin{pmatrix}(-a_i - a_i d_i + 1)\boldsymbol{e}_i \\ a_i\boldsymbol{e}_i\end{pmatrix} \overset{!}{=} \lambda_i \begin{pmatrix} a_i\boldsymbol{e}_i \\ \boldsymbol{e}_i\end{pmatrix} \\ &\Rightarrow a_i = \lambda_i\,, \quad -\lambda_i - \lambda_i d_i + 1 = \lambda_i^2\end{aligned}$$

with d_i the i-th entry of the diagonal matrix $D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$. It follows

$$\lambda_{i,1,2} = \frac{-(1+d_i) \pm \sqrt{(1+d_i)^2 + 4}}{2}\,.$$

We know that $d_i \in [-1;0]$ (see (3.87)). Hence we get the eigenvalue with the highest absolute value for "–" and $d_i = 0$, thus

$$|\lambda| \leq \left|\frac{-1-\sqrt{1^2+4}}{2}\right| = \frac{1+\sqrt{5}}{2}$$

and the eigenvalue with the smallest absolute value for "+" and $d_i = 0$, thus

$$|\lambda| \geq \left|\frac{-1+\sqrt{1^2+4}}{2}\right| = \frac{-1+\sqrt{5}}{2}\,.$$

Because $\tilde{\boldsymbol{J}}$ is symmetric we can calculate the spectral condition number of $\tilde{\boldsymbol{J}}$ by

$$\operatorname{cond}\tilde{\boldsymbol{J}} = \frac{\max|\lambda|}{\min|\lambda|} \leq \frac{1+\sqrt{5}}{-1+\sqrt{5}} = \frac{(1+\sqrt{5})^2}{4} \approx 2.6180\,. \tag{3.88}$$

So the condition number is bounded independent of the value of the reaction constant K and the values of the concentrations $(\boldsymbol{c}, \bar{\boldsymbol{c}})$.

3.5.3 Condition of the Problem for Large Δt

For the examination the same example as in the last section is used. As large time step sizes Δt are considered we assume that the term $\theta\boldsymbol{I}$ in the Jacobian matrix (3.86), which stems from the discretization of the time derivative, can be neglected. Note that the terms in $\boldsymbol{L}_h$ stemming from the discretization of the second derivative contain the factor $\frac{1}{h^2}$. So for example for $\Delta t = 0.5$ and $h = 0.01$, that are realistic discretization parameters, the terms of $\theta\boldsymbol{I}$ are small compared with that one of $\Delta t\boldsymbol{L}_h$ and the assumption is justified.

Neglecting $\theta\boldsymbol{I}$ the Jacobian matrix is

$$\boldsymbol{J} = \begin{pmatrix}\boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp} & -\boldsymbol{I} \\ \boldsymbol{0} & \Delta t\boldsymbol{L}_h\end{pmatrix}.$$

As a consequence the problem decomposes in two subproblems. The first subproblem has the matrix $\Delta t \boldsymbol{L}_h$. This subproblem is equivalent to solving an elliptic PDE. So it is possible to solve it.

The second subproblem has the matrix $\boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp}$. We know that the entries of the diagonal matrix $D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp}$ are between minus one and zero (see (3.87)). So it is possible that the diagonal matrix $\boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp}$ has one entry equal to ε and one equal to $1 - \varepsilon$. Hence the condition number of the matrix $\boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp}$ can be very large. But for the case of a diagonal matrix the condition analysis is to pessimistic (see [Kna02, Sec. 2.6]). As the matrix is diagonal every equation can be solved independent of the other ones. To solve one equation one only has to perform a division, which is a well conditioned operation. Hence it is possible to solve the second subproblem numerically without problems.

Altogether it is possible to solve the problem in the case that the term $\theta \boldsymbol{I}$ is neglected despite of the large condition number of the whole Jacobian matrix $\boldsymbol{J}$. Hence one can hope that the reduction scheme works well also for large time step sizes.

3.6 Variants

3.6.1 No Additional Variables

It is also possible to use the formulation (3.30)-(3.38) together with the retransformation (3.15) directly and not to introduce additional variables. This is done in [KK07] for the case no minerals. Without the additional variables we have to use another resolution function. In the case no equilibrium minerals it reads

$$(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{kin}) .$$

This approach leads to the remaining nonlinear system

$$\begin{aligned} \partial_t(\theta \boldsymbol{\xi}_{sorp}) + L\boldsymbol{\xi}_{sorp} &= \partial_t(\theta \bar{\boldsymbol{\xi}}_{sorp}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{kin})) + \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{kin}) \\ \partial_t(\theta \boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta \boldsymbol{A}_{1,kin} \boldsymbol{r}_{kin}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{kin}) \end{aligned}$$

which consists of less equations than that one which we get by use of the additional variables. By use of the additional variables we have to add the defining equations of the additional variables to the system and that way the coupled nonlinear system gets bigger. But it is not possible to use this smaller system for realistic problems with large reaction constants because it does not converge (see below).

In [Krä08] the case with equilibrium minerals is treated without use of the additional variables. The main difference is that the partitioning in local and global variables depends on the fact if the equilibrium minerals are present at this point or not. Not using the additional variables the global and local variables are

$$\boldsymbol{\xi}_{loc} = \begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}^{\mathcal{I}}_{min} \\ \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{kin} \end{pmatrix}, \qquad \boldsymbol{\xi}_{glob} = \begin{pmatrix} \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}^{\mathcal{A}}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix}$$

where $\boldsymbol{\xi}^{\mathcal{I}}_{min}$ denotes the vector which contains all entries of $\boldsymbol{\xi}_{min}$ for which the associated mineral is present and analogously $\boldsymbol{\xi}^{\mathcal{A}}_{min}$ contains all entries of $\boldsymbol{\xi}_{min}$ where the associated mineral is not present. The mineral concentrations $\bar{\boldsymbol{\xi}}_{min}$ are treated in a special way. They are computed a-posteriori by evaluating the left hand side of a PDE.

That the case differentiation is unavoidable when not using the additional variables should be illustrated with the following example. We consider the following simple example with one mobile species, one mineral and one mineral reaction in equilibrium:

$$\mathrm{A} \leftrightarrow \overline{\mathrm{B}}$$

In this example the transformed variables are

$$\xi_{min} = -c(\mathrm{A}), \qquad \bar{\xi}_{min} = \bar{c}(\overline{\mathrm{B}}), \qquad \tilde{\xi}_{min} = -c(\mathrm{A}) - \bar{c}(\overline{\mathrm{B}})$$

where the last one only appears by use of the additional variables.

Depending if we use the additional variables (right) or not (left) we have:

$$\begin{aligned} \partial_t \xi_{min} + L\xi_{min} &= \partial_t \bar{\xi}_{min} \\ \min\{\bar{\xi}_{min}, K + \xi_{min}\} &= 0 \end{aligned}$$

Resolving the min-function with respect to one variable is not possible. So it is necessary to make a case differentiation in the partitioning in local and global variables.

$$\begin{aligned} \tilde{\xi}_{min} + \bar{\xi}_{min} - \xi_{min} &= 0 \\ \partial_t \tilde{\xi}_{min} + L\xi_{min} &= 0 \\ \min\{\bar{\xi}_{min}, K + \tilde{\xi}_{min} + \bar{\xi}_{min}\} &= 0 \end{aligned}$$

Resolving the min-function with respect to $\bar{\xi}_{min}$ is possible:

$$\begin{aligned} \bar{\xi}_{min} &= 0 && \text{for } K + \tilde{\xi}_{min} \geq 0 \\ \bar{\xi}_{min} &= -K - \tilde{\xi}_{min} && \text{else} \end{aligned}$$

Without use of the additional variables the equations of the local problem are

$$\begin{aligned} \boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \boldsymbol{0} \\ \boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) &= \boldsymbol{0} \\ \boldsymbol{\psi}^{\mathcal{I}}(\boldsymbol{c}) &= \boldsymbol{0} \\ \frac{\theta\bar{\boldsymbol{\xi}}_{kin} - (\theta\bar{\boldsymbol{\xi}}_{kin})_{old}}{\Delta t} &= \theta \boldsymbol{A}_{2,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\,. \end{aligned}$$

For the case that there is no linear dependency of $\boldsymbol{S}_{1,sorp}$ and $\boldsymbol{S}_{1,min}$ (compare (3.2)) the global problem reads

$$\begin{aligned}
\partial_t(\theta\boldsymbol{\xi}_{sorp}) + L\boldsymbol{\xi}_{sorp} &= \partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}^{\mathcal{A}}_{min}, \boldsymbol{\xi}_{kin})) \\
&\quad + \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}^{\mathcal{A}}_{min}, \boldsymbol{\xi}_{kin}) \\
\partial_t(\theta\boldsymbol{\xi}^{\mathcal{A}}_{min}) + L\boldsymbol{\xi}^{\mathcal{A}}_{min} &= \partial_t(\theta\bar{\boldsymbol{\xi}}^{\mathcal{A}}_{min}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}^{\mathcal{A}}_{min}, \boldsymbol{\xi}_{kin})) + \theta\boldsymbol{A}^{\mathcal{A}}_{1,min}\boldsymbol{r}_{kin}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}^{\mathcal{A}}_{min}, \boldsymbol{\xi}_{kin}) \\
\partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}^{\mathcal{A}}_{min}, \boldsymbol{\xi}_{kin})\,.
\end{aligned}$$

Note that we have the second equation only on that parts of the domain on which the associated mineral is not present. The mineral concentrations $\bar{\boldsymbol{\xi}}_{min}$ are computed in the following way: When the mineral is not present we have the trivial equation $\bar{\boldsymbol{\xi}}^{\mathcal{A}}_{min} = \boldsymbol{0}$. When the mineral is present we compute the mineral concentrations $\bar{\boldsymbol{\xi}}^{\mathcal{I}}_{min}$ with help of the PDEs (see [Krä08] for details)

$$\partial_t(\theta\boldsymbol{\xi}^{\mathcal{I}}_{min}) + L\boldsymbol{\xi}^{\mathcal{I}}_{min} = \partial_t(\theta\bar{\boldsymbol{\xi}}^{\mathcal{I}}_{min}) + \theta\boldsymbol{A}^{\mathcal{I}}_{1,min}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\,.$$

The case differentiation in the variant with no additional variables causes some difficulties in the implementation that do not occur by the use of the additional variables. Firstly we have to solve the PDE for the variable $\boldsymbol{\xi}^{\mathcal{A}}_{min}$ only on a part of the domain. So the number of unknowns per node differs from node to node and the number of unknowns at one node changes if a mineral fully dissolves or precipitates. Secondly we have to evaluate the left hand side of a PDE during the computation of $\bar{\boldsymbol{\xi}}^{\mathcal{I}}_{min}$ with the PDE. That is an unusual way to use a PDE for numerical computations. So most platforms for solving PDEs do not provide this kind of using a PDE.

The main problem of the formulation without the additional variables is that the resulting method does not converge for realistic problems. For example it is not possible to compute the MoMaS–benchmark (see Chap. 4), not even for extremely small time step sizes.

Probably the high values of the derivatives $D_{\boldsymbol{\xi}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$ are the reason for the non-convergence. Because of the implicit elimination (see 3.3.2) this derivatives appear in the global Jacobian matrix.

We look at the same example with the chemistry of the "easy test case" of the MoMaS–benchmark, which is mentioned in Section 3.5.1. Using this variant the matrix of the linear system to calculate $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ is:

```
     1.00        -0.00         0.00         0.00        -0.00        -0.00        -0.00
    -0.00         1.00        -0.00         0.21         0.62        -0.00        -0.00
     0.16        -0.16         0.31         1.00        -0.52        -0.00        -0.00
     0.00         0.33         0.00         0.69         1.00        -0.00        -0.00
    -0.00         0.00        -0.00         0.00         1.00        -0.00        -0.00
    -0.00      6845.42        -0.00      6845.42     20536.27         1.00         0.00
     0.85        -0.85         1.07         4.05        -3.21         0.38         1.00
```

The associated right hand side is:

```
       0.00      -0.00
      -0.21       0.00
       0.48      -0.60
      -0.33      -0.00
      -0.00       0.00
   -6845.42       0.00
       2.56      -2.77
```

Therewith we get for $D_{\boldsymbol{\xi}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$:

```
  -3113.96         0.00
   1180.07        -0.71
```

The other entries in the global Jacobi matrix have order of magnitude one. So these entries destroy the convergence of the method.

By use of the additional variables one can prove that the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ are bounded and the bound only depends on stoichiometric coefficients that are integral and small (see Sec. 3.5.1).

Like in Section 3.5.2 we examine the condition number of the Jacobian matrix for $\Delta t = 0$. We consider the following easy example with one chemical reaction, one mobile and one immobile species. The chemical reaction is the equilibrium sorption reaction

$$\overline{\mathrm{B}} \leftrightarrow 2\mathrm{A}$$

with the equilibrium condition

$$\bar{c} = Kc^2\,.$$

Applying the reduction scheme gives the transformed variables

$$\xi_{sorp} = \frac{1}{2}c\,, \qquad\qquad \bar{\xi}_{sorp} = -\bar{c}\,.$$

Using this variant the global problem consists of the equation

$$\partial_t(\theta\xi_{sorp}) + L\xi_{sorp} = \partial_t(\theta\bar{\xi}_{sorp}(\xi_{sorp}))$$

with the resolution function

$$\bar{\xi}_{sorp}(\xi_{sorp}) = -4K\xi_{sorp}^2\,. \tag{3.89}$$

Here the Jacobian of the global problem is

$$\boldsymbol{J} = \theta\boldsymbol{I} + \Delta t\boldsymbol{L}_h - \theta D_{\boldsymbol{\xi}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$$

where

$$\bar{\xi}'_{sorp}(\xi_{sorp}) = -8K\xi_{sorp} = -4Kc\,. \tag{3.90}$$

As K and c are positive we know that

$$-\infty < \bar{\xi}'_{sorp}(\xi_{sorp}) \leq 0\,.$$

Like in Section 3.5.2 we consider the case $\theta = 1$ and $\Delta t = 0$. In this case the global Jacobian is

$$\boldsymbol{J} = \boldsymbol{I} - D_{\boldsymbol{\xi}_{sorp}} \bar{\boldsymbol{\xi}}_{sorp} = \boldsymbol{I} + 4K\,\mathrm{diag}(\boldsymbol{c})\,.$$

$\boldsymbol{J}$ is a diagonal matrix with only strictly positive entries. Hence the spectral condition number of $\boldsymbol{J}$ is the largest entry divided by the smallest entry. So we get

$$\mathrm{cond}\,\boldsymbol{J} = \frac{1 + 4Kc_{max}}{1 + 4Kc_{min}}\,.$$

Here the condition number depends on the equilibrium constant and the concentration values. In realistic scenarios a species is often only present in a part of the domain. Hence c_{min} is zero and the order of magnitude of c_{max} is roughly one. Furthermore the reaction constants K often assume very high values. In such a situation we get

$$\mathrm{cond}\,\boldsymbol{J} \approx 4K$$

and so we see that because of the large reaction constant the spectral condition number of the Jacobian matrix has a very high value. Hence this numerical method is not suitable for this problem when K has a high value because then convergence problems would occur. This is opposite to the case with the additional variables, where cond $\boldsymbol{J}$ is bounded by a fixed number (see (3.88)).

For this example the resolution function $\bar{\xi}_{sorp}(\tilde{\xi}_{sorp})$ can also be stated explicitly by use of additional variable $\tilde{\xi}_{sorp}$. Some simple calculations give

$$\bar{\xi}_{sorp}(\tilde{\xi}_{sorp}) = -\tilde{\xi}_{sorp} - \frac{1}{8K} + \sqrt{\frac{\tilde{\xi}_{sorp}}{4K} + \frac{1}{64K^2}}\,. \tag{3.91}$$

Hence the derivative is

$$\bar{\xi}'_{sorp}(\tilde{\xi}_{sorp}) = -1 + \frac{1}{\sqrt{16K\tilde{\xi}_{sorp} + 1}}\,. \tag{3.92}$$

The reason for the different condition numbers are the derivatives of the resolution functions (see (3.92), (3.90)). Let us consider the batch situation where $\tilde{\xi}_{sorp} = \xi_{sorp} - \bar{\xi}_{sorp} = \frac{1}{2}c + \bar{c}$ is a constant. The equilibrium condition is the limit for $k \to \infty$ of the ODE

$$\partial_t \bar{c} = k(r_f(2(\tilde{\xi}_{sorp} - \bar{c})) - r_b(\bar{c}))$$

with $r_f(c) = k_f c^2$ and $r_b(\bar{c}) = k_b \bar{c}$. Treating r_f and r_b implicitly gives

$$\frac{\bar{c} - \bar{c}_{old}}{\Delta t} = k(r_f(2(\tilde{\xi}_{sorp} - \bar{c})) - r_b(\bar{c}))\,.$$

Resolving for $\bar{c}$ and passing to the limit $k \to \infty$ gives

$$\bar{c} = \tilde{\xi}_{sorp} + \frac{1}{8K} - \sqrt{\frac{\tilde{\xi}_{sorp}}{4K} + \frac{1}{64K^2}}$$

with $K = k_f/k_b$. This corresponds to the resolution function by use of the additional variables (3.91). Treating r_f explicitly and r_b implicitly gives

$$\frac{\bar{c} - \bar{c}_{old}}{\Delta t} = k(r_f(2(\tilde{\xi}_{sorp} - \bar{c}_{old})) - r_b(\bar{c}))\,.$$

Resolving for $\bar{c}$ and passing to the limit $k \to \infty$ gives

$$\bar{c} = 4K\big(\tilde{\xi}_{sorp} - \bar{c}_{old}\big)^2\,.$$

This corresponds to the resolution function of the formulation without additional variables (3.89). So the different resolution functions are connected with the explicit or implicit treatment of the forward reaction rate r_f.

So we can conclude that it is necessary to use the additional variables $\tilde{\boldsymbol{\xi}}_{sorp}$ and $\tilde{\boldsymbol{\xi}}_{min}$. In the case no equilibrium mineral reactions two more variants are possible.

3.6.2 Eliminating $\boldsymbol{\xi}_{sorp}$

By using the additional variables $\tilde{\boldsymbol{\xi}}$ there are two possibilities to eliminate one block of variables. The first one is to resolve the defining equation of $\tilde{\boldsymbol{\xi}}_{sorp}$ (see (3.39)) for $\boldsymbol{\xi}_{sorp}$ and to plug the resulting equation in (3.31). The other possibility is described in the next section.

It is also possible to do the elimination of $\boldsymbol{\xi}_{sorp}$ not on the nonlinear level but on the linear level. First the equations are linearized and then $\boldsymbol{\xi}_{sorp}$ is eliminated. This technique is described in [Krä08, Sec. 4.4.3]. As both equations (3.39), (3.31) are linear in $\boldsymbol{\xi}_{sorp}$ this leads to the same equations after linearization of the equations resulting from the elimination on the nonlinear level.

Doing the elimination on the nonlinear level we get the new PDE

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}_{sorp}) + L\tilde{\boldsymbol{\xi}}_{sorp} = \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) - L\bar{\boldsymbol{\xi}}_{sorp}\,. \tag{3.93}$$

The retransformation is the same as by adding the defining equations of $\tilde{\boldsymbol{\xi}}_{sorp}$ as additional equation:

$$\begin{aligned}\boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{S}_{1,sorp}\tilde{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{1,kin}\boldsymbol{\xi}_{kin} + \boldsymbol{B}_1^\perp\boldsymbol{\eta} + \boldsymbol{S}_{1,sorp}\bar{\boldsymbol{\xi}}_{sorp}\\ \bar{\boldsymbol{c}} &= \boldsymbol{S}_{2,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \boldsymbol{B}_2^\perp\bar{\boldsymbol{\eta}}\end{aligned}$$

Also the resolution function is the same:

$$(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{kin})$$

So the existence of the resolution function is secured as we can apply the existence proof of Section 3.2.1.

Altogether we have to solve the same local problem as by adding the defining equations of the additional variables as additional equations. But now the global problem is

$$\begin{aligned}\partial_t(\theta\tilde{\boldsymbol{\xi}}_{sorp}) + L\tilde{\boldsymbol{\xi}}_{sorp} &= \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin}) - L\bar{\boldsymbol{\xi}}_{sorp}(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin})\\ \partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin})\,.\end{aligned}$$

The number of equations is the same as in the variant without the additional equations and so the number is less than by adding the defining equations of the additional variables as additional equations.

The advantage of this variant in comparison to the variant not using the additional variables $\tilde{\boldsymbol{\xi}}_{sorp}$ is that the derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ are bounded. Because of the use of the same resolution function as by adding the defining equations of $\tilde{\boldsymbol{\xi}}_{sorp}$ as additional equation all considerations about $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ of this case (see Sec. 3.5.1) are applicable.

The disadvantage of this variant is that the transport operator applied to immobile species ($L\bar{\boldsymbol{\xi}}_{sorp}$) appears. This term is physically not meaningful. The operator L describes the transport of a mobile species but $\bar{\boldsymbol{\xi}}_{sorp}$ is immobile, i.e., it is not transported. Also from a numerical point of view this term is problematic. Because of the use of a resolution function $\bar{\boldsymbol{\xi}}_{sorp}$ is a nonlinear function of $(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin})$ and it is known that nonlinearities under the transport operator cause numerical difficulties.

Numerical tests with the “easy test case” of the MoMaS–benchmark (see Chap. 4) have shown that this variant leads to a converging method but only for very small time step sizes. The restriction on the time step size is so heavy that it is preferable to use the standard method with the additional equations because the total CPU time is less than by use of this variant despite of the larger system of equations.

Applying this variant to the example of Sec. 3.5.3 leads to the equation

$$\partial_t(\theta\tilde{\xi}_{sorp}) + L\tilde{\xi}_{sorp} = -L\bar{\xi}_{sorp}(\tilde{\xi}_{sorp})\,.$$

Solving this nonlinear PDE with Newton's method one gets the Jacobian matrix

$$\boldsymbol{J} = \theta\boldsymbol{I} + \Delta t\boldsymbol{L}_h + \Delta t\boldsymbol{L}_h D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}\,.$$

Neglecting the term $\theta\boldsymbol{I}$ like it is done in Section 3.5.3 gives

$$\boldsymbol{J} = \Delta t\boldsymbol{L}_h(\boldsymbol{I} + D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp})\,.$$

This matrix can be ill conditions, e.g., when the diagonal matrix $D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp}$ has entries $-\varepsilon$ and $-1+\varepsilon$ (compare Sec. 3.5.3). But here the argumentation of Sec. 3.5.3, how the problem can be solved numerically despite of the large condition number, is not applicable. So using this variant numerical problems are expected for large time step sizes.

3.6.3 Eliminating $\bar{\boldsymbol{\xi}}_{sorp}$

The second possible variant is to eliminate the variables $\bar{\boldsymbol{\xi}}_{sorp}$. For this purpose the defining equation of $\tilde{\boldsymbol{\xi}}_{sorp}$ (see (3.39)) is solved for $\bar{\boldsymbol{\xi}}_{sorp}$ and the resulting equation is plugged in (3.31). So we get the new PDE

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}_{sorp}) + L\boldsymbol{\xi}_{sorp} = \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\,. \tag{3.94}$$

This time we have to change the partitioning in local and global variables. We take $\boldsymbol{\xi}_{sorp}$ as local variables instead of global ones. Also in the retransformation we have to eliminate $\bar{\boldsymbol{\xi}}_{sorp}$ with help of the defining equation of $\tilde{\boldsymbol{\xi}}_{sorp}$. This leads to the new retransformation

$$\begin{aligned}
\boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{S}_{1,sorp}\boldsymbol{\xi}_{sorp} + \boldsymbol{S}^*_{1,kin}\boldsymbol{\xi}_{kin} + \boldsymbol{B}_1^{\perp}\boldsymbol{\eta}\\
\bar{\boldsymbol{c}} &= \boldsymbol{S}_{2,sorp}\boldsymbol{\xi}_{sorp} - \boldsymbol{S}_{2,sorp}\tilde{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \boldsymbol{B}_2^{\perp}\bar{\boldsymbol{\eta}}\,.
\end{aligned}$$

With this new partitioning in local and global variables we have to show the existence of a resolution function

$$(\tilde{\boldsymbol{\xi}}_{sorp}, \boldsymbol{\xi}_{kin}) \mapsto (\boldsymbol{\xi}_{mob}, \boldsymbol{\xi}_{sorp}, \bar{\boldsymbol{\xi}}_{kin})\,.$$

This can be done analogously to the existence proof in Section 3.2.1. Mainly we have to replace $\bar{\boldsymbol{\xi}}_{sorp}$ by $\boldsymbol{\xi}_{sorp}$. Note that the variables the resolution function depends on are unchanged.

When we compute the solution of the local problem like it is described in Section 3.4.1 we have to solve the same equations. Only at the end we have to compute $\boldsymbol{\xi}_{sorp}$ by its defining equation instead of $\bar{\boldsymbol{\xi}}_{sorp}$. The global problem of this variant reads

$$\begin{aligned}\partial_t(\theta\tilde{\boldsymbol{\xi}}_{sorp}) + L\boldsymbol{\xi}_{sorp}(\tilde{\boldsymbol{\xi}}_{sorp},\boldsymbol{\xi}_{kin}) &= \theta(\boldsymbol{A}_{1,sorp} - \boldsymbol{A}_{2,sorp})\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp},\boldsymbol{\xi}_{kin})\\ \partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\tilde{\boldsymbol{\xi}}_{sorp},\boldsymbol{\xi}_{kin})\,.\end{aligned}$$

The disadvantage of this variant is that a nonlinearity under the transport operator ($L\boldsymbol{\xi}_{sorp}(\tilde{\boldsymbol{\xi}}_{sorp},\boldsymbol{\xi}_{kin})$) appears because of the new partitioning in local and global variables. It is known that a nonlinearity under the differential operator can cause numerical problems.

Numerical tests of this variant have shown that we get exactly the same Newton iterates as by using the variant where $\boldsymbol{\xi}_{sorp}$ is eliminated. This can be seen in the following way: In both cases the local problem contains the defining equations

$$\tilde{\boldsymbol{\xi}}_{sorp} = \left((\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c}\right)_{i=J_{mob}+1,\ldots,J_{eq}} - \left((\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}\right)_{i=1,\ldots,J_{sorp}}\,.$$

The local variables $\boldsymbol{\xi}_{sorp}$ in this variant and $\bar{\boldsymbol{\xi}}_{sorp}$ in the other variant are calculated after solving the local problem with their defining equations

$$\bar{\boldsymbol{\xi}}_{sorp} = \left((\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}\right)_{i=1,\ldots,J_{sorp}}, \qquad \boldsymbol{\xi}_{sorp} = \left((\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c}\right)_{i=J_{mob}+1,\ldots,J_{eq}},$$

respectively. If we solve the local problems of the two variants with the same values for the global variables we get that $\boldsymbol{\xi}_{sorp}$ of this variant minus $\bar{\boldsymbol{\xi}}_{sorp}$ of the variant where $\boldsymbol{\xi}_{sorp}$ is eliminated is equal to $\tilde{\boldsymbol{\xi}}_{sorp}$. Therewith we see immediately that calculating the defect of the PDEs (3.93) and (3.94) provides the same results.

The derivatives $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ are calculated with a linear system of equations of the form $\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B}\boldsymbol{X} = \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{C}$ (see (3.53)). For the variant where $\boldsymbol{\xi}_{sorp}$ is eliminated we have exactly the linear system (3.53) because we have the same local problem and the same retransformation as by adding the defining equations of the additional variables as additional equations. For this variant we can analogously derive a linear system of this form. Doing so we see that in the two variants only the matrix $\boldsymbol{C}$ differs. For this variant we get

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{S}_{1,kin}^* \\ \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}\,.$$

So we get by calculating the difference

$$\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B}(\boldsymbol{X}_1 - \boldsymbol{X}_2) = \boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}(\boldsymbol{C}_1 - \boldsymbol{C}_2)$$

where the index denotes the variants. We know that

$$\boldsymbol{C}_1 - \boldsymbol{C}_2 = \begin{pmatrix} -\boldsymbol{S}_{1,sorp} & \boldsymbol{0} \\ -\boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}.$$

As $\boldsymbol{B}(\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{B})^{-1}\boldsymbol{B}^T\tilde{\boldsymbol{\Lambda}}$ is a projection onto $\boldsymbol{B}$ (see (3.78)) and as the first column block of $\boldsymbol{C}_1 - \boldsymbol{C}_2$ is minus the second column block of $\boldsymbol{B}$ we get

$$\boldsymbol{X}_1 - \boldsymbol{X}_2 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \end{pmatrix}.$$

Particularly we have the identity

$$D_{\tilde{\boldsymbol{\xi}}_{sorp}}\bar{\boldsymbol{\xi}}_{sorp} - D_{\tilde{\boldsymbol{\xi}}_{sorp}}\boldsymbol{\xi}_{sorp} = -\boldsymbol{I}_{J_{sorp}}.$$

Therewith it is easy to see that the PDEs (3.93) and (3.94) provide the same contribution to the global Jacobi matrix.

In summary the first variant without the additional variables cannot be used for realistic problems because it does not converge. The next two variants, where the variables $\boldsymbol{\xi}_{sorp}$ and $\bar{\boldsymbol{\xi}}_{sorp}$, respectively, are eliminated, are inappropriate for realistic problems because they only converge for small time step sizes. Hence it is reasonable to use the global system (3.43)-(3.47) although its global system consists of more equations than the global systems of these three variants.

3.6.4 Elimination on Linear Level Instead Use of Resolution Function

Instead of using a resolution function to eliminate the local equations, like it is described in Section 3.2, it is also possible to eliminate the local equations on the linear level. For the case that no additional variables $\tilde{\boldsymbol{\xi}}$ are used a detailed description of this technique can be found in [Krä08, Sec. 4.4.3]. By use of the additional variables this elimination technique is applicable exactly in the same way.

The reason why here the elimination of the local equations is done on the nonlinear level with help of a resolution function is that using the resolution function the resulting method is more efficient. This can be seen in the following way: Let us assume that the equilibrium conditions are not fulfilled at one node and that twenty Newton-steps are necessary to equilibrate the concentrations. The equilibrium conditions of realistic scenarios are highly nonlinear. So the assumption that twenty Newton-steps are necessary is realistic.

Using the resolution function one has to perform twenty *local* Newton-steps. To do so one has to solve twenty linear systems with the number of equations equal to the number of concentrations. This does not take much time. But using the elimination on the linear level one has to perform twenty *global* Newton-steps. Here one has to solve twenty linear systems of the size number of nodes times number of global variables. This is very time consuming. So with the resolution function the method is much more efficient.

3.7 Implementation

The algorithm was implemented using a software kernel for parallel computations in the field of PDEs called M++ [Wie05]. M++ itself is an object oriented code based on C++ using MPI for the parallelization. So the code runs on clusters consisting of several machines. Running the code on a single computer all cores of a multi-core CPU can be used.

The code is implemented for 2D and 3D problems and uses finite elements on unstructured grids. For solving the nonlinear systems of equations Newton's method is used. Different iterative linear solvers (e.g., GMRES, BiCGStab, QM-RCGStab) and different standard preconditioners (e.g., SSOR) are implemented. The coarse grid must be specified in a text file with a certain format. It is possible to refine this grid regularly as often as it is wanted.

The user specifies the problem in a script file including ansatz spaces, the discretization parameters, type of linear solver/preconditioner, stopping criteria for nonlinear/linear solver, etc. As output M++ build text files for visualization with different programs (gnuplot, OpenDX, vtk-files for e.g. ParaView). How many output files of each type are built is also specified in the script file.

In the framework of this PhD thesis the reduction scheme was implemented using conformal finite elements like it is described in the previous sections including the special numerical treatment of the local problem described in Section 3.4.1, the starting value search out of Section 3.4.2, the cutting-off strategy of the global Newton step out of Section 3.4.3 (the simplified case for the normal formulation and the general case for the generalized formulation out of Section 3.9) and the FV-stabilization for the convection dominated case like in Section 3.4.4. So realistic problems with high reaction constants can be handled.

The implementation was carried out in such a way that also the number of species, the transport parameters and the chemical reactions are specified by entries in the script file. So changes of the simulated problem are easily possible. Furthermore a small C–program was written which generates grids for a given flow field according to the strategy described in Section 3.4.5.

The Darcy velocity $\boldsymbol{q}$, the transport operator L is based on, and the water content θ can either be provided by the user as an input or it is computed by solving Richards equation

$$\begin{aligned} \partial_t \theta(\psi) + \nabla \cdot \boldsymbol{q} &= 0 \\ \boldsymbol{q} &= -\boldsymbol{K}_s k_r(\psi) \nabla(\psi + z) \end{aligned}$$

where ψ denotes the pressure measured in meter water head, $\boldsymbol{K}_s$ the saturated hydraulic conductivity, k_r the relative hydraulic conductivity and z the height against the gravity direction.

In the fully saturated case Richards equation degenerates to the elliptic equation

$$-\nabla \cdot \boldsymbol{K}_s \nabla(\psi + z) = 0\,.$$

In this case the solver for the Richards equation is called only once at the beginning of the simulation.

For solving Richards equation Hybrid Mixed Finite Elements are used. The advantages of Mixed Finite Elements are that they are mass conserving and that they can handle discontinuous hydraulic conductivities $\boldsymbol{K}_s$ which occurs in realistic scenarios. Two types of hybrid mixed finite elements are available: Lowest order Raviart–Thomas (RT_0) and lowest order Brezzi–Douglas–Marini (BDM_1), which has a higher order of convergence in the flux variable. For details to the solver of the Richards equation see [Koh05].

When the Darcy velocity $\boldsymbol{q}$ is constant in time (e.g., in the fully saturated case) the discrete transport operator is only assembled once at the beginning of the computation and is stored in a matrix. Subsequently in every time step this precomputed matrix is used to set up that part of the global Jacobian matrix which is related to the transport operator. With this procedure CPU time is saved.

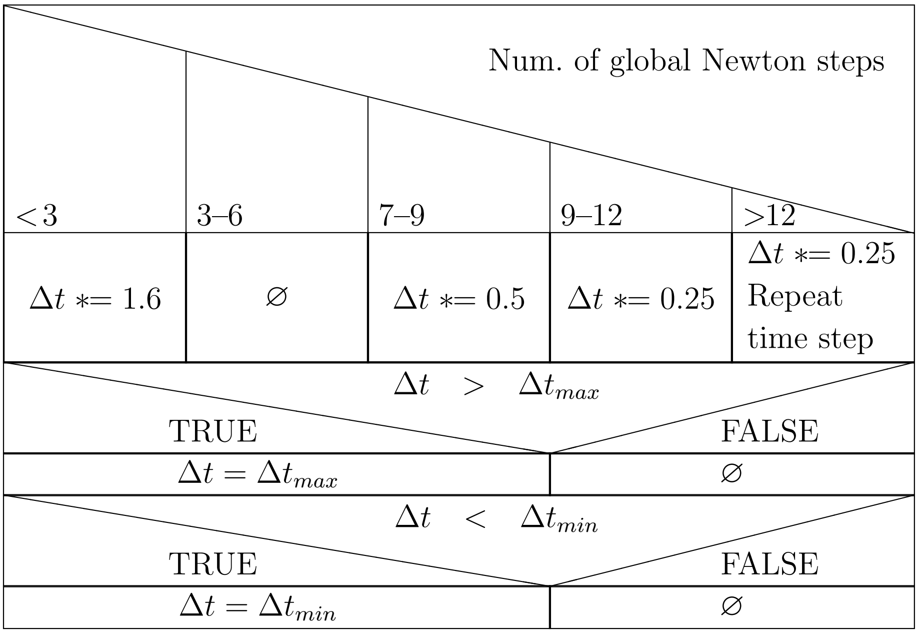
The stopping criterion for both the linear and the nonlinear solver is a combined criterion, i.e., the iteration is stopped if the residual $\boldsymbol{r}$ fulfills

$$|\boldsymbol{r}| < \max\{Eps, |\boldsymbol{r}_0| Red\} \tag{3.95}$$

where $\boldsymbol{r}_0$ is the initial residual and Eps, Red are user specified parameters. For each solver different parameters Eps, Red can be given.

An adaptive time stepping is implemented. The time step size is chosen between a given minimum and maximum value: $\Delta t_{min} \leq \Delta t \leq \Delta t_{max}$. The criterion of the time step choice depends on the number of Newton steps in the last time step. For a small number the time step size is enlarged and for a high number it is reduced. When it is necessary a time step is repeated with a smaller time step size. For the computations of the MoMaS benchmark (see Chap. 4) the following algorithm is used:

Adaptive time stepping


Num. of global Newton steps
<3
3–6
7–9
9–12
>12
$\Delta t *= 1.6$
$\varnothing$
$\Delta t *= 0.5$
$\Delta t *= 0.25$
$\Delta t *= 0.25$ Repeat time step
$\Delta t > \Delta t_{max}$
TRUE
FALSE
$\Delta t = \Delta t_{max}$
$\varnothing$
$\Delta t < \Delta t_{min}$
TRUE
FALSE
$\Delta t = \Delta t_{min}$
$\varnothing$

The starting values for both the η-problem and the nonlinear problem are calculated by extrapolation from two time levels. If the starting value of the nonlinear problem computed by extrapolation corresponds to negative concentrations the starting value search of Section 3.4.2 is used.

For some applications much CPU time can be saved by performing extrapolation in comparison to the usage of the value of the old time step as starting value. In Table 3.1 the normalized CPU time for a computation of the "advective easy test case" in 2D is shown. The used grid consists of 26660 triangles. In this computation 35% of the CPU time can be saved by using extrapolation. As proposed in [BBC$^+$] the CPU time is measured in normalized units (see Chap. 4).

	CPU time	time steps	Newton steps
with extrapolation	5838.7	13044	2.18
without extrapolation	8918.8	17100	2.88

Table 3.1: CPU time with and without extrapolation, "advective easy test case" in 2D

The speed-up of the parallelization is quite good. In Table 3.2 the CPU times for a computation using one processor and a computation using eight processors are shown. The considered problem is the "advective easy test case" in 2D on a grid consisting of 26660 triangles without extrapolation. In this performance test the speed-up factor is 6.05.

	CPU time	time steps	Newton steps
1 proc.	8918.8	17100	2.88
8 proc.s	1473.4	16987	2.89

Table 3.2: CPU time with different numbers of processors, “advective easy test case” in 2D without extrapolation

3.8 Link to Morel Formulation

The goal of this section is to show the connections between the reduction scheme presented in the Sections 3.1, 3.2 and the widely used Morel formulation (see e.g. [AK09], [dDEK09], [HKK09]).

To derive the Morel formulation it is necessary to transform the stoichiometric matrix $\boldsymbol{S}$ to standard form. It is always possible to transform the stoichiometric matrix $\boldsymbol{S}$ (compare (3.3)) such that it is of the form[4]

$$\boldsymbol{S} = \left(\begin{array}{cccc} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} & \boldsymbol{S}_{1,kin} \\ \hline \mathbf{0} & \boldsymbol{S}_{2,sorp} & \mathbf{0} & \tilde{\boldsymbol{S}}_{2,kin} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{J_{min}} & \mathbf{0} \end{array}\right)$$

$$\sim \left(\begin{array}{cccc} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} & \boldsymbol{F} \\ -\boldsymbol{I}_{J_{mob}} & \mathbf{0} & \mathbf{0} & \\ \hline \mathbf{0} & \hat{\boldsymbol{B}} & \mathbf{0} & \boldsymbol{G} \\ \mathbf{0} & -\boldsymbol{I}_{J_{sorp}} & \mathbf{0} & \\ \mathbf{0} & \mathbf{0} & -\boldsymbol{I}_{J_{min}} & \mathbf{0} \end{array}\right) = \left(\begin{array}{cccc} \boldsymbol{C}_1 & \boldsymbol{A}_1 & \boldsymbol{D}_1 & \boldsymbol{F}_1 \\ \boldsymbol{C}_2 & \boldsymbol{A}_2 & \boldsymbol{D}_2 & \boldsymbol{F}_2 \\ -\boldsymbol{I}_{J_{mob}} & \mathbf{0} & \mathbf{0} & \boldsymbol{F}_3 \\ \hline \mathbf{0} & \hat{\boldsymbol{B}}_1 & \mathbf{0} & \boldsymbol{G}_1 \\ \mathbf{0} & \hat{\boldsymbol{B}}_2 & \mathbf{0} & \boldsymbol{G}_2 \\ \mathbf{0} & -\boldsymbol{I}_{J_{sorp}} & \mathbf{0} & \boldsymbol{G}_3 \\ \mathbf{0} & \mathbf{0} & -\boldsymbol{I}_{J_{min}} & \mathbf{0} \end{array}\right)$$

where the blocks $\boldsymbol{A}_i$ have the substructure (with $\boldsymbol{A}_{ld}$ out of (3.2))

$$\begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{1,li} & \boldsymbol{D}_1\boldsymbol{A}_{ld} \\ \boldsymbol{A}_{2,li} & \boldsymbol{D}_2\boldsymbol{A}_{ld} \end{pmatrix} \tag{3.96}$$

[4] In order that the submatrices $\boldsymbol{B}_i$ of $\boldsymbol{S}$ can not be mixed up with the transformation matrices $\boldsymbol{B}_i$ out of Section 3.1 the submatrices of $\boldsymbol{S}$ are mark with a hat $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{B}}_i$.

and with $\begin{pmatrix} \boldsymbol{C}_2 & \boldsymbol{A}_{2,li} & \boldsymbol{D}_2 & \boldsymbol{F}_2^* \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_3^* \end{pmatrix}$, $\boldsymbol{G}_2^*$ and $\begin{pmatrix} \hat{\boldsymbol{B}}_2 & \boldsymbol{G}_2^* \\ -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{G}_3^* \end{pmatrix}$ invertible. The matrices $\boldsymbol{F}_i^*$ and $\boldsymbol{G}_i^*$ consist of some columns of $\boldsymbol{F}_i$ and $\boldsymbol{G}_i$, respectively, analogously to $\boldsymbol{S}_{i,kin}$ in (3.7) and (3.8), respectively. To get such a stoichiometric matrix one has to perform only the following operations: taking a multiple of one column, adding one column of the block of mobile equilibrium reactions to a column in the block of equilibrium reactions, adding one column of the block of sorption equilibrium reactions to another column in this block and interchanging rows within the block of mobile species or within the block of fixed nonmineral species. Adding a column to another one in the block of equilibrium reactions corresponds to adding the equilibrium conditions in the logarithmized form (or multiplying the equilibrium conditions in the nonlogarithmized form). Thus the equilibrium conditions for the transformed stoichiometric matrix are of the same form as the original ones.

In detail the transformation can be done in the following way:

(i) Replacing $\boldsymbol{S}_{1,kin}$ and $\tilde{\boldsymbol{S}}_{2,kin}$ by $\boldsymbol{S}_{1,kin}^*$ and $\boldsymbol{S}_{2,kin}^*$, respectively, such that the matrices $\boldsymbol{S}_i^*$ (see (3.7), (3.8)) are a maximal system of linear independent columns of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively.

(ii) Sorting the rows in the block of the nonminerals such that in the matrix $\begin{pmatrix} \boldsymbol{S}_{2,sorp,1} & \boldsymbol{S}_{2,kin,1}^* \\ \boldsymbol{S}_{2,sorp,2} & \boldsymbol{S}_{2,kin,2}^* \\ \boldsymbol{S}_{2,sorp,3} & \boldsymbol{S}_{2,kin,3}^* \end{pmatrix}$ the quadratic submatrices $\begin{pmatrix} \boldsymbol{S}_{2,sorp,2} & \boldsymbol{S}_{2,kin,2}^* \\ \boldsymbol{S}_{2,sorp,3} & \boldsymbol{S}_{2,kin,3}^* \end{pmatrix}$, $\boldsymbol{S}_{2,sorp,3}$ and $\boldsymbol{S}_{2,kin,2}^*$ are invertible.

It is always possible to sort the rows such that $\begin{pmatrix} \boldsymbol{S}_{2,sorp,2} & \boldsymbol{S}_{2,kin,2}^* \\ \boldsymbol{S}_{2,sorp,3} & \boldsymbol{S}_{2,kin,3}^* \end{pmatrix}$ is invertible because the columns of the original matrix $\begin{pmatrix} \boldsymbol{S}_{2,sorp} & \boldsymbol{S}_{2,kin}^* \end{pmatrix}$ are linear independent.

Moreover it is always possible to sort the rows of $\begin{pmatrix} \boldsymbol{S}_{2,sorp,2} & \boldsymbol{S}_{2,kin,2}^* \\ \boldsymbol{S}_{2,sorp,3} & \boldsymbol{S}_{2,kin,3}^* \end{pmatrix}$ such that $\boldsymbol{S}_{2,sorp,3}$ and $\boldsymbol{S}_{2,kin,2}^*$ are invertible because of the expansion theorem of Laplace (see e.g. [Zie97, Satz 5.3.21])

$$\det \boldsymbol{A} = \sum_{\gamma \in G_p} (\operatorname{sgn} \gamma)(\det \boldsymbol{A}_\gamma)(\det \boldsymbol{A}_\gamma^*)$$

with G_p the set of all permutations $\sigma \in S_n$ for which $\sigma(1) < \sigma(2) < \cdots < \sigma(p)$ and $\sigma(p+1) < \sigma(p+2) < \cdots < \sigma(n)$, $\boldsymbol{A}_\gamma$ the matrix which arises of $\boldsymbol{A}$ by removing the last $(n-p)$ columns and the rows with the indices

$\gamma(p+1), \dots, \gamma(n)$ and $\boldsymbol{A}^*_\gamma$ the matrix which arises of $\boldsymbol{A}$ by removing the first p columns and the rows with the indices $\gamma(1), \dots, \gamma(p)$. Applying this theorem with $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{S}_{2,sorp,2} & \boldsymbol{S}^*_{2,kin,2} \\ \boldsymbol{S}_{2,sorp,3} & \boldsymbol{S}^*_{2,kin,3} \end{pmatrix}$ gives: If there were no invertible matrices $\boldsymbol{S}_{2,sorp,3}$ and $\boldsymbol{S}^*_{2,kin,2}$ the product $(\det \boldsymbol{A}_\gamma)(\det \boldsymbol{A}^*_\gamma)$ would be zero for all γ and so the $\det \boldsymbol{A}$ would be zero. That is a contradiction to the invertibility of $\boldsymbol{A}$.

(iii) By applying the Gauss algorithm (with interchanging of rows if necessary) to $\begin{pmatrix} \boldsymbol{S}^T_{1,sorp} & \boldsymbol{S}^T_{2,sorp} \end{pmatrix}$ one can get the $-\boldsymbol{I}_{J_{sorp}}$ block in the submatrix with the sorption equilibrium reactions. Because of the invertibility of $\boldsymbol{S}_{2,sorp,3}$ the Gauss algorithm can always be performed. Now the immobile part is of the required form. By applying the Gauss algorithm the space spanned by the columns of $\boldsymbol{S}_{1,sorp}$ is unchanged. So as the columns of the original matrix $\boldsymbol{S}^*_1$ are linear independent the columns of the modified matrix $\boldsymbol{S}^*_1$ are still linear independent.

(iv) If the columns of $\begin{pmatrix} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}$ are linear dependent one has to divide the submatrix $\boldsymbol{S}_{1,sorp}$ in $\begin{pmatrix} \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min}\boldsymbol{A}_{ld} \end{pmatrix}$ such that the columns of $\begin{pmatrix} \boldsymbol{S}_{mob} & \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min} \end{pmatrix}$ are linear independent. When some columns are permuted the corresponding rows in the block of the nonminerals must also be permuted such that the $-\boldsymbol{I}_{J_{sorp}}$ block in submatrix with the sorption equilibrium reactions is preserved.

(v) Sorting the rows in the block of the mobile species such that in the matrix $\begin{pmatrix} \boldsymbol{S}_{1,mob,1} & \boldsymbol{S}_{1,sorp,li,1} & \boldsymbol{S}_{1,min,1} & \boldsymbol{S}^*_{1,kin,1} \\ \boldsymbol{S}_{1,mob,2} & \boldsymbol{S}_{1,sorp,li,2} & \boldsymbol{S}_{1,min,2} & \boldsymbol{S}^*_{1,kin,2} \\ \boldsymbol{S}_{1,mob,3} & \boldsymbol{S}_{1,sorp,li,3} & \boldsymbol{S}_{1,min,3} & \boldsymbol{S}^*_{1,kin,3} \end{pmatrix}$ the submatrices $\boldsymbol{S}_{1,mob,3}$ and $\begin{pmatrix} \boldsymbol{S}_{1,mob,2} & \boldsymbol{S}_{1,sorp,li,2} & \boldsymbol{S}_{1,min,2} & \boldsymbol{S}^*_{1,kin,2} \\ \boldsymbol{S}_{1,mob,3} & \boldsymbol{S}_{1,sorp,li,3} & \boldsymbol{S}_{1,min,3} & \boldsymbol{S}^*_{1,kin,3} \end{pmatrix}$ are invertible. This is always possible with the same argument as in (ii).

(vi) By applying the Gauss algorithm to $\boldsymbol{S}^T_{1,mob}$ one gets the $-\boldsymbol{I}_{J_{mob}}$ block in the submatrix with the mobile equilibrium reactions. Then by adding columns of the submatrix with the mobile equilibrium reactions to columns of the submatrix with sorption and mineral equilibrium reactions one can form the two zero blocks in the submatrix with the sorption and mineral equilibrium reactions. Now the whole matrix is of the required form.

In the Morel formulation the concentrations are split in primary and secondary concentrations

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_{prim} \\ \boldsymbol{c}_{sec} \end{pmatrix}, \qquad \bar{\boldsymbol{c}} = \begin{pmatrix} \bar{\boldsymbol{c}}_{nmin,prim} \\ \bar{\boldsymbol{c}}_{nmin,sec} \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix}$$

where the number of the mobile secondary concentrations $\boldsymbol{c}_{sec}$ is J_{mob} and the number of immobile secondary concentrations $\bar{\boldsymbol{c}}_{nmin,sec}$ is J_{sorp}. The variables used in the Morel formulation are the total concentrations $\boldsymbol{T}$ and the total fixed concentrations $\boldsymbol{W}$ that are defined by

$$\boldsymbol{T} := \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} + \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min} \tag{3.97}$$

$$\boldsymbol{W} := \bar{\boldsymbol{c}}_{nmin,prim} + \hat{\boldsymbol{B}}\bar{\boldsymbol{c}}_{nmin,sec}\,. \tag{3.98}$$

Furthermore we define the mobile part of the total concentrations $\boldsymbol{T}_M$ and the immobile part of the total concentrations $\boldsymbol{T}_F$

$$\boldsymbol{T}_M := \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} \tag{3.99}$$

$$\boldsymbol{T}_F := \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min}\,. \tag{3.100}$$

The equations of the Morel formulation can be split into the chemical problem

$$\begin{aligned} \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} + \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min} &= \boldsymbol{T} \\ \bar{\boldsymbol{c}}_{nmin,prim} + \hat{\boldsymbol{B}}\bar{\boldsymbol{c}}_{nmin,sec} &= \boldsymbol{W} \\ \boldsymbol{C}^T \ln(\boldsymbol{c}_{prim}) - \ln(\boldsymbol{c}_{sec}) &= \boldsymbol{k}_{mob} \\ \boldsymbol{A}^T \ln(\boldsymbol{c}_{prim}) + \hat{\boldsymbol{B}}^T \ln(\bar{\boldsymbol{c}}_{nmin,prim}) - \ln(\bar{\boldsymbol{c}}_{nmin,sec}) &= \boldsymbol{k}_{sorp} \\ \min\left\{\boldsymbol{D}^T \ln(\boldsymbol{c}_{prim}) - \boldsymbol{k}_{min}, \bar{\boldsymbol{c}}_{min}\right\} &= \boldsymbol{0} \end{aligned} \tag{3.101}$$

which consists of mass balance equations and equilibrium conditions and the transport problem

$$\partial_t(\theta \boldsymbol{T}) + L\boldsymbol{T}_M = \boldsymbol{0} \tag{3.102}$$

$$\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min}) \tag{3.103}$$

$$\partial_t(\theta \boldsymbol{W}) = \boldsymbol{0} \tag{3.104}$$

with $\boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min}) = \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min}$. The transport problem is formulated without kinetic reactions because in most papers dealing with the Morel formulation kinetic reactions are excluded.

Now we derive the reduction scheme for that case that the stoichiometric matrix is in standard form.[5] Using the transformed stoichiometric matrix the

[5] The link between the variables of the Morel formulation and the reduction scheme is much simpler in the case no kinetic reactions. This simpler case can be found in Appendix A.1

matrix $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$, consisting of the linear independent columns of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively, are of the form

$$\boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{E}_1 \\ \boldsymbol{C}_2 & \boldsymbol{E}_2 \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{E}_3 \end{pmatrix}, \qquad \boldsymbol{S}_2^* = \begin{pmatrix} \hat{\boldsymbol{B}}_1 & \boldsymbol{0} & \boldsymbol{G}_1^* \\ \hat{\boldsymbol{B}}_2 & \boldsymbol{0} & \boldsymbol{G}_2^* \\ -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} & \boldsymbol{G}_3^* \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix}$$

with the abbreviations $\boldsymbol{E}_1 := \begin{pmatrix} \boldsymbol{A}_{1,li} & \boldsymbol{D}_1 & \boldsymbol{F}_1^* \end{pmatrix}$, $\boldsymbol{E}_2 := \begin{pmatrix} \boldsymbol{A}_{2,li} & \boldsymbol{D}_2 & \boldsymbol{F}_2^* \end{pmatrix}$ and $\boldsymbol{E}_3 := \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_3^* \end{pmatrix}$. The entries of the concentrations vectors $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$ are partitioned analogously to the rows of $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$, respectively,

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_{prim,1} \\ \boldsymbol{c}_{prim,2} \\ \boldsymbol{c}_{sec} \end{pmatrix}, \qquad \bar{\boldsymbol{c}} = \begin{pmatrix} \bar{\boldsymbol{c}}_{nmin,prim,1} \\ \bar{\boldsymbol{c}}_{nmin,prim,2} \\ \bar{\boldsymbol{c}}_{nmin,sec} \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix}.$$

It is useful to choose as matrix $\boldsymbol{S}_1^\perp$, consisting of a maximal system of linear independent vectors that are orthogonal to all columns of $\boldsymbol{S}_1^*$, the following one

$$\boldsymbol{S}_1^\perp = \begin{pmatrix} \boldsymbol{I}_{I-J_{mob}-J_{sorp,li}-J_{min}-J_{1,kin}^*} \\ -(\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3)^{-T}(\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)^T \\ \boldsymbol{C}_1^T - \boldsymbol{C}_2^T(\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3)^{-T}(\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)^T \end{pmatrix}.$$

The existence of the inverse of $\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3$ is shown below. Calculating $(\boldsymbol{S}_1^*)^T\boldsymbol{S}_1^\perp$ one can see that the columns of $\boldsymbol{S}_1^\perp$ are orthogonal to those of $\boldsymbol{S}_1^*$. Furthermore it is useful to choose the following transformation matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_1^\perp$

$$\boldsymbol{B}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{sorp,li}+J_{min}+J_{1,kin}^*} \\ \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \end{pmatrix}, \quad \boldsymbol{B}_1^\perp = \begin{pmatrix} \boldsymbol{I}_{I-J_{mob}-J_{sorp,li}-J_{min}-J_{1,kin}^*} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}.$$

The condition that the columns of $\boldsymbol{B}_1$, $\boldsymbol{S}_1^\perp$ form a basis of the whole space is fulfilled. The standard form of the stoichiometric matrix is constructed in such a way that $\begin{pmatrix} \boldsymbol{C}_2 & \boldsymbol{E}_2 \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{E}_3 \end{pmatrix}$ is invertible. Hence also the inverse of $\boldsymbol{B}_1^T\boldsymbol{S}_1^* = \begin{pmatrix} -\boldsymbol{I}_{J_{mob}} & \boldsymbol{E}_3 \\ \boldsymbol{C}_2 & \boldsymbol{E}_2 \end{pmatrix}$ exists.

Let the inverse $(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}$ be partitioned in $\begin{pmatrix}\boldsymbol{N}_1 & \boldsymbol{N}_2\\ \boldsymbol{N}_3 & \boldsymbol{N}_4\end{pmatrix}$. We know that

$$\begin{aligned}\begin{pmatrix}\boldsymbol{I}_{J_{mob}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{J_{sorp,li}+J_{min}+J_{1,kin}^*}\end{pmatrix} &= \begin{pmatrix}-\boldsymbol{I}_{J_{mob}} & \boldsymbol{E}_3\\ \boldsymbol{C}_2 & \boldsymbol{E}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{N}_1 & \boldsymbol{N}_2\\ \boldsymbol{N}_3 & \boldsymbol{N}_4\end{pmatrix}\\ &= \begin{pmatrix}-\boldsymbol{N}_1+\boldsymbol{E}_3\boldsymbol{N}_3 & -\boldsymbol{N}_2+\boldsymbol{E}_3\boldsymbol{N}_4\\ \boldsymbol{C}_2\boldsymbol{N}_1+\boldsymbol{E}_2\boldsymbol{N}_3 & \boldsymbol{C}_2\boldsymbol{N}_2+\boldsymbol{E}_2\boldsymbol{N}_4\end{pmatrix}.\end{aligned}$$

Resolving the equation of the upper right block for $\boldsymbol{N}_2$ and plugging in the equation of the lower right block yields

$$(\boldsymbol{C}_2\boldsymbol{E}_3+\boldsymbol{E}_2)\boldsymbol{N}_4 = \boldsymbol{I}_{J_{sorp,li}+J_{min}+J_{1,kin}^*}\,.$$

Because the matrix on the right hand side is invertible both quadratic matrices on the left hand side must be invertible and we get $\boldsymbol{N}_4 = (\boldsymbol{C}_2\boldsymbol{E}_3+\boldsymbol{E}_2)^{-1}$. Again using the equation of the upper right block yields $\boldsymbol{N}_2 = \boldsymbol{E}_3(\boldsymbol{C}_2\boldsymbol{E}_3+\boldsymbol{E}_2)^{-1}$. With help of the identity

$$\begin{pmatrix}\boldsymbol{I}_{J_{mob}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{J_{sorp,li}+J_{min}+J_{1,kin}^*}\end{pmatrix} = \begin{pmatrix}\boldsymbol{N}_1 & \boldsymbol{N}_2\\ \boldsymbol{N}_3 & \boldsymbol{N}_4\end{pmatrix}\begin{pmatrix}-\boldsymbol{I}_{J_{mob}} & \boldsymbol{E}_3\\ \boldsymbol{C}_2 & \boldsymbol{E}_2\end{pmatrix}$$

one can derive analogously $\boldsymbol{N}_3 = (\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}\boldsymbol{C}_2$. If the matrix $\boldsymbol{E}_2$ is invertible one also gets the formula $\boldsymbol{N}_1 = -(\boldsymbol{I}_{J_{mob}}+\boldsymbol{E}_3\boldsymbol{E}_2^{-1}\boldsymbol{C}_2)^{-1}$.

Using this we get for the transformed variables $\boldsymbol{\xi}$ (compare (3.13), (3.14))

$$\begin{aligned}\begin{pmatrix}\boldsymbol{\xi}_{mob}\\ \begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\\ \boldsymbol{\xi}_{kin}\end{pmatrix}\end{pmatrix} &= (\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c}\\ &= \begin{pmatrix}\boldsymbol{E}_3(\boldsymbol{C}_2\boldsymbol{E}_3+\boldsymbol{E}_2)^{-1}\boldsymbol{c}_{prim,2}+\boldsymbol{N}_1\boldsymbol{c}_{sec}\\ (\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}\boldsymbol{c}_{prim,2}+(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}\boldsymbol{C}_2\boldsymbol{c}_{sec}\end{pmatrix}.\end{aligned} \tag{3.105}$$

And for the transformed variables $\boldsymbol{\eta}$ we have (compare (3.13))

$$\begin{aligned}\boldsymbol{\eta} &= \left((\boldsymbol{S}_1^\perp)^T\boldsymbol{B}_1^\perp\right)^{-1}(\boldsymbol{S}_1^\perp)^T\boldsymbol{c}\\ &= \boldsymbol{c}_{prim,1}-(\boldsymbol{E}_1+\boldsymbol{C}_1\boldsymbol{E}_3)(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}\boldsymbol{c}_{prim,2}\\ &\quad+\left(\boldsymbol{C}_1-(\boldsymbol{E}_1+\boldsymbol{C}_1\boldsymbol{E}_3)(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}\boldsymbol{C}_2\right)\boldsymbol{c}_{sec}\,.\end{aligned} \tag{3.106}$$

Now we consider the immobile species. As a basis of the orthogonal complement of $\boldsymbol{S}_2^*$ we choose

$$\boldsymbol{S}_2^\perp = \begin{pmatrix}\boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}-J_{2,kin}^*}\\ -(\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-T}(\boldsymbol{G}_1^*+\hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)^T\\ \hat{\boldsymbol{B}}_1^T-\hat{\boldsymbol{B}}_2^T(\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-T}(\boldsymbol{G}_1^*+\hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)^T\\ \boldsymbol{0}\end{pmatrix}.$$

The existence of the inverse of $\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*$ is shown below. Like in the case of mobile species it is easy to see that this matrix is orthogonal to $\boldsymbol{S}_2^*$. Here it is useful to choose as transformation matrices

$$\boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J^*_{2,kin}} \\ \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix}, \qquad \boldsymbol{B}_2^{\perp} = \begin{pmatrix} \boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}-J^*_{2,kin}} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}.$$

Like in the case of mobile species it is obvious that $\boldsymbol{B}_2$ and $\boldsymbol{S}_2^{\perp}$ form a basis of the whole space and due to the construction of the standard form of the stoichiometric matrix the inverse of $(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}$ exists.

With the same methods with that we have calculated $(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}$ one can compute that (Note that we know that $\boldsymbol{G}_2^*$ is invertible)

$$(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1} = \begin{pmatrix} -(\boldsymbol{I}_{J_{sorp}} + \boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1} & \boldsymbol{0} & \boldsymbol{G}_3^*(\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^* + \boldsymbol{G}_2^*)^{-1} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} & \boldsymbol{0} \\ (\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2 & \boldsymbol{0} & (\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^* + \boldsymbol{G}_2^*)^{-1} \end{pmatrix}.$$

Using this we get for the transformed variables $\bar{\boldsymbol{\xi}}$ (compare (3.13), (3.14))

$$\begin{aligned} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \\ \bar{\boldsymbol{\xi}}_{kin} \end{pmatrix} &= (\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}} \\ &= \begin{pmatrix} \boldsymbol{G}_3^*(\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\bar{\boldsymbol{c}}_{nmin,prim,2} - (\boldsymbol{I}_{J_{sorp}}+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1}\bar{\boldsymbol{c}}_{nmin,sec} \\ -\bar{\boldsymbol{c}}_{min} \\ (\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\bar{\boldsymbol{c}}_{nmin,prim,2} + (\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\bar{\boldsymbol{c}}_{nmin,sec} \end{pmatrix}. \end{aligned} \tag{3.107}$$

Moreover we get for the transformed variables $\bar{\boldsymbol{\eta}}$ (compare (3.13))

$$\begin{aligned} \bar{\boldsymbol{\eta}} &= \left((\boldsymbol{S}_2^{\perp})^T\boldsymbol{B}_2^{\perp}\right)^{-1}(\boldsymbol{S}_2^{\perp})^T\bar{\boldsymbol{c}} \\ &= \bar{\boldsymbol{c}}_{nmin,prim,1} - (\boldsymbol{G}_1^* + \hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)(\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\bar{\boldsymbol{c}}_{nmin,prim,2} \\ &\quad + \left(\hat{\boldsymbol{B}}_1 - (\boldsymbol{G}_1^* + \hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)(\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\right)\bar{\boldsymbol{c}}_{nmin,sec}\,. \end{aligned} \tag{3.108}$$

Now we define the additional variables $\tilde{\boldsymbol{\xi}}$ (compare (3.39))

$$\tilde{\boldsymbol{\xi}} = \begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\xi}_{sorp} - \bar{\boldsymbol{\xi}}_{sorp,li} + (\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{li} \\ \boldsymbol{\xi}_{min} - \bar{\boldsymbol{\xi}}_{min} - \boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} + \boldsymbol{A}_{ld}(\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{ld} \\ \boldsymbol{\xi}_{kin} \end{pmatrix}.$$

Differing from the proceeding in Section 3.2 these variables contain also a term with $\bar{\boldsymbol{\xi}}_{kin}$.

We want to replace the mass balance equations for $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}$, $\bar{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\xi}}$ in the local problem (see Sec. 3.4.1) by the following linear combinations of these equations $\boldsymbol{\eta}+(\boldsymbol{E}_1+\boldsymbol{C}_1\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}}$, $(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}}$, $\bar{\boldsymbol{\eta}}+(\boldsymbol{G}_1^*+\hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin}$ and $(\boldsymbol{G}_2^*+\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin}$. To calculate the right hand side of the first two equations we have to compute $(i=1,2)$:

$$
\begin{aligned}
&(\boldsymbol{E}_i+\boldsymbol{C}_i\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}}\\
&\quad=(\boldsymbol{E}_i{+}\boldsymbol{C}_i\boldsymbol{E}_3)\begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\\ \boldsymbol{\xi}_{kin}\end{pmatrix}-(\boldsymbol{E}_i{+}\boldsymbol{C}_i\boldsymbol{E}_3)\begin{pmatrix}\bar{\boldsymbol{\xi}}_{sorp,li}-(\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{li}\\ \bar{\boldsymbol{\xi}}_{min}+\boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld}-\boldsymbol{A}_{ld}(\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{ld}\\ \boldsymbol{0}\end{pmatrix}
\end{aligned}
$$

First we consider the second summand. Using the definitions of the matrices $\boldsymbol{E}_i=\begin{pmatrix}\boldsymbol{A}_{i,li} & \boldsymbol{D}_i & \boldsymbol{F}_i^*\end{pmatrix}$ and $\boldsymbol{E}_3=\begin{pmatrix}\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_3^*\end{pmatrix}$ we see that the second summand is

$$
\boldsymbol{A}_{i,li}\bar{\boldsymbol{\xi}}_{sorp,li}-\boldsymbol{A}_{i,li}(\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{li}+\boldsymbol{D}_i\bar{\boldsymbol{\xi}}_{min}+\boldsymbol{D}_i\boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld}-\boldsymbol{D}_i\boldsymbol{A}_{ld}(\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin})_{ld}\,.
$$

With help of the substructure of $\boldsymbol{A}_i=\begin{pmatrix}\boldsymbol{A}_{i,li} & \boldsymbol{D}_i\boldsymbol{A}_{ld}\end{pmatrix}$ we get

$$
=\boldsymbol{A}_i\bar{\boldsymbol{\xi}}_{sorp}+\boldsymbol{D}_i\bar{\boldsymbol{\xi}}_{min}-\boldsymbol{A}_i\boldsymbol{G}_3^*\bar{\boldsymbol{\xi}}_{kin}\,.
$$

Plugging in the definitions of the variables $\bar{\boldsymbol{\xi}}$ (see (3.107)) gives

$$
\begin{aligned}
&=\boldsymbol{A}_i\Big(\boldsymbol{G}_3^*(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\bar{\boldsymbol{c}}_{nmin,prim,2}-(\boldsymbol{I}_{J_{sorp}}{+}\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1}\bar{\boldsymbol{c}}_{nmin,sec}\Big)-\boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}\\
&\qquad-\boldsymbol{A}_i\boldsymbol{G}_3^*\left((\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\bar{\boldsymbol{c}}_{nmin,prim,2}+(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\bar{\boldsymbol{c}}_{nmin,sec}\right)\\
&=-\boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}-\boldsymbol{A}_i(\boldsymbol{I}_{J_{sorp}}+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1}\bar{\boldsymbol{c}}_{nmin,sec}\\
&\qquad-\boldsymbol{A}_i\boldsymbol{G}_3^*(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\bar{\boldsymbol{c}}_{nmin,sec}\\
&=-\boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}-\boldsymbol{A}_i(\boldsymbol{I}_{J_{sorp}}+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1}\\
&\qquad\left(\boldsymbol{I}_{J_{sorp}}+(\boldsymbol{G}_3^*+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\right)\bar{\boldsymbol{c}}_{nmin,sec}\\
&=-\boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}-\boldsymbol{A}_i(\boldsymbol{I}_{J_{sorp}}+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}\hat{\boldsymbol{B}}_2)^{-1}\\
&\qquad\left(\boldsymbol{I}_{J_{sorp}}+\boldsymbol{G}_3^*(\boldsymbol{G}_2^*)^{-1}(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)(\boldsymbol{G}_2^*{+}\hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)^{-1}\hat{\boldsymbol{B}}_2\right)\bar{\boldsymbol{c}}_{nmin,sec}\\
&=-\boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}-\boldsymbol{A}_i\bar{\boldsymbol{c}}_{nmin,sec}\,.
\end{aligned}
$$

Using this and the definition of $\boldsymbol{\xi}$ (see (3.105)) we get in summary

$$
\begin{aligned}
(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}}&=\boldsymbol{c}_{prim,2}+\boldsymbol{C}_2\boldsymbol{c}_{sec}+\boldsymbol{D}_2\bar{\boldsymbol{c}}_{min}+\boldsymbol{A}_2\bar{\boldsymbol{c}}_{nmin,sec}\\
(\boldsymbol{E}_1+\boldsymbol{C}_1\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}}&=(\boldsymbol{E}_1+\boldsymbol{C}_1\boldsymbol{E}_3)(\boldsymbol{E}_2+\boldsymbol{C}_2\boldsymbol{E}_3)^{-1}(\boldsymbol{c}_{prim,2}+\boldsymbol{C}_2\boldsymbol{c}_{sec})\\
&\qquad+\boldsymbol{D}_1\bar{\boldsymbol{c}}_{min}+\boldsymbol{A}_1\bar{\boldsymbol{c}}_{nmin,sec}\,.
\end{aligned}
$$

Adding the definition of $\boldsymbol{\eta}$ to the last one of the two equations leads to

$$\boldsymbol{\eta} + (\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}} = \boldsymbol{c}_{prim,1} + \boldsymbol{C}_1\boldsymbol{c}_{sec} + \boldsymbol{D}_1\bar{\boldsymbol{c}}_{min} + \boldsymbol{A}_1\bar{\boldsymbol{c}}_{nmin,sec} \,.$$

This equation and the first one of the two equations above will be used as mass balance equations in the local problem instead of the defining equations of $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}$.

Concerning the immobile species we compute with help of the definition of $\bar{\boldsymbol{\xi}}_{kin}$ (see (3.107)) and the definition of $\bar{\boldsymbol{\eta}}$ (see (3.108))

$$\begin{aligned}
\bar{\boldsymbol{\eta}} + (\boldsymbol{G}_1^* + \hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} &= \bar{\boldsymbol{c}}_{nmin,prim,1} + \hat{\boldsymbol{B}}_1\bar{\boldsymbol{c}}_{nmin,sec} \\
(\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} &= \bar{\boldsymbol{c}}_{nmin,prim,2} + \hat{\boldsymbol{B}}_2\bar{\boldsymbol{c}}_{nmin,sec} \,.
\end{aligned}$$

These equations will be used in the local problem instead of the definitions of $\bar{\boldsymbol{\eta}}$ and $\bar{\boldsymbol{\xi}}_{kin}$. For the moment we put the ODE for the variable $\bar{\boldsymbol{\xi}}_{kin}$ in the global problem instead of the local problem. Doing so and using the new mass balance equations, the local problem with the logarithmized mobile concentrations $\boldsymbol{l}$, the logarithmized nonmineral concentrations $\bar{\boldsymbol{l}}_{nmin}$ and the mineral concentrations $\bar{\boldsymbol{c}}_{min}$ as unknowns reads:

$$\begin{aligned}
\boldsymbol{C}_1^T\boldsymbol{l}_{prim,1} + \boldsymbol{C}_2^T\boldsymbol{l}_{prim,2} &= \boldsymbol{l}_{sec} + \boldsymbol{k}_{mob} \\
\boldsymbol{\eta} + (\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}} &= \exp(\boldsymbol{l}_{prim,1}) + \boldsymbol{C}_1\exp(\boldsymbol{l}_{sec}) \\
&\quad + \boldsymbol{A}_1\exp(\bar{\boldsymbol{l}}_{nmin,sec}) + \boldsymbol{D}_1\bar{\boldsymbol{c}}_{min} \\
(\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}} &= \exp(\boldsymbol{l}_{prim,2}) + \boldsymbol{C}_2\exp(\boldsymbol{l}_{sec}) \\
&\quad + \boldsymbol{A}_2\exp(\bar{\boldsymbol{l}}_{nmin,sec}) + \boldsymbol{D}_2\bar{\boldsymbol{c}}_{min} \\
\boldsymbol{A}_1^T\boldsymbol{l}_{prim,1} + \boldsymbol{A}_2^T\boldsymbol{l}_{prim,2} & \\
+ \hat{\boldsymbol{B}}_1^T\bar{\boldsymbol{l}}_{nmin,prim,1} + \hat{\boldsymbol{B}}_2^T\bar{\boldsymbol{l}}_{nmin,prim,2} &= \bar{\boldsymbol{l}}_{nmin,sec} + \boldsymbol{k}_{sorp} \\
\min\left\{\boldsymbol{D}_1^T\boldsymbol{l}_{prim,1} + \boldsymbol{D}_2^T\boldsymbol{l}_{prim,2} - \boldsymbol{k}_{min}, \bar{\boldsymbol{c}}_{min}\right\} &= \boldsymbol{0} \\
\bar{\boldsymbol{\eta}} + (\boldsymbol{G}_1^* + \hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} &= \exp(\bar{\boldsymbol{l}}_{nmin,prim,1}) + \hat{\boldsymbol{B}}_1\exp(\bar{\boldsymbol{l}}_{nmin,sec}) \\
(\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} &= \exp(\bar{\boldsymbol{l}}_{nmin,prim,2}) + \hat{\boldsymbol{B}}_2\exp(\bar{\boldsymbol{l}}_{nmin,sec})
\end{aligned}$$

Comparing the mass balance equations with the definition of the total concentrations $\boldsymbol{T}$ (3.97) and the total fixed concentrations $\boldsymbol{W}$ (3.98) one sees that

$$\begin{pmatrix} \boldsymbol{\eta} + (\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}} \\ (\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3)\tilde{\boldsymbol{\xi}} \end{pmatrix} = \boldsymbol{T} \tag{3.109}$$

$$\begin{pmatrix} \bar{\boldsymbol{\eta}} + (\boldsymbol{G}_1^* + \hat{\boldsymbol{B}}_1\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} \\ (\boldsymbol{G}_2^* + \hat{\boldsymbol{B}}_2\boldsymbol{G}_3^*)\bar{\boldsymbol{\xi}}_{kin} \end{pmatrix} = \boldsymbol{W}. \tag{3.110}$$

Using this and undoing the partitioning of the primary variables in two parts one gets

$$\begin{aligned}
\boldsymbol{C}^T\boldsymbol{l}_{prim} - \boldsymbol{l}_{sec} &= \boldsymbol{k}_{mob} \\
\exp(\boldsymbol{l}_{prim}) + \boldsymbol{C}\exp(\boldsymbol{l}_{sec}) + \boldsymbol{A}\exp(\bar{\boldsymbol{l}}_{nmin,sec}) + \boldsymbol{D}\bar{\boldsymbol{c}}_{min} &= \boldsymbol{T} \\
\boldsymbol{A}^T\boldsymbol{l}_{prim} + \hat{\boldsymbol{B}}^T\bar{\boldsymbol{l}}_{nmin,prim} - \bar{\boldsymbol{l}}_{nmin,sec} &= \boldsymbol{k}_{sorp} \\
\min\left\{\boldsymbol{D}^T\boldsymbol{l}_{prim} - \boldsymbol{k}_{min}, \bar{\boldsymbol{c}}_{min}\right\} &= \boldsymbol{0} \\
\exp(\bar{\boldsymbol{l}}_{nmin,prim}) + \hat{\boldsymbol{B}}\exp(\bar{\boldsymbol{l}}_{nmin,sec}) &= \boldsymbol{W}\,.
\end{aligned}$$

Thus the local problem is the same as the chemical problem of the Morel formulation (3.101). So a modular implementation of the reduction scheme is possible. That means any existing chemical solver (e.g. the solver of the chemical subproblem of a splitting code) can be reused to solve the local problem of the reduction scheme. This is an advantage of using the standard form of the stoichiometric matrix. Applying the reduction scheme to arbitrary stoichiometric matrices this would not be possible.

As the standard form of the stoichiometric matrix is used, it is possible to reduce the size of the local problem by resolving the equilibrium conditions of the mobile and sorption equilibrium reactions for the secondary concentrations $\boldsymbol{l}_{sec}$ and $\bar{\boldsymbol{l}}_{nmin,sec}$, respectively, and then plugging these equations in the other ones. This is also an advantage of using the standard form of the stoichiometric matrix. Applying the reduction scheme to arbitrary stoichiometric matrices the reduction of the local problem size is not possible. Doing so the smaller local problem reads

$$\begin{aligned}
\exp\left(\boldsymbol{l}_{prim}\right) + \boldsymbol{C}\exp\left(\boldsymbol{C}^T\boldsymbol{l}_{prim} - \boldsymbol{k}_{mob}\right)& \\
+\boldsymbol{A}\exp\left(\boldsymbol{A}^T\boldsymbol{l}_{prim} + \hat{\boldsymbol{B}}^T\bar{\boldsymbol{l}}_{nmin,prim} - \boldsymbol{k}_{sorp}\right) + \boldsymbol{D}\bar{\boldsymbol{c}}_{min} - \boldsymbol{T} &= \boldsymbol{0} \\
\min\left\{\boldsymbol{D}^T\boldsymbol{l}_{prim} - \boldsymbol{k}_{min}, \bar{\boldsymbol{c}}_{min}\right\} &= \boldsymbol{0} \\
\hat{\boldsymbol{B}}\exp\left(\boldsymbol{A}^T\boldsymbol{l}_{prim} + \hat{\boldsymbol{B}}^T\bar{\boldsymbol{l}}_{nmin,prim} - \boldsymbol{k}_{sorp}\right) + \exp\left(\bar{\boldsymbol{l}}_{nmin,prim}\right) - \boldsymbol{W} &= \boldsymbol{0}\,.
\end{aligned}$$

If variables $\bar{\boldsymbol{\xi}}_{kin}$ appear in the problem there are two possibilities to treat them. The first one is to move the ODEs for $\bar{\boldsymbol{\xi}}_{kin}$ to the global problem. Then an existing chemical solver can be used without changes but one has to solve a larger global problem. The second possibility is to solve the ODEs for $\bar{\boldsymbol{\xi}}_{kin}$ in the local problem. Then the global problem is not enlarged but modifications at the chemical solver are necessary.

Now we consider the global problem. The global problem of the reduction scheme without kinetic reactions reads

$$\begin{aligned}
\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} &= \mathbf{0} && (3.111)\\
\partial_t(\theta\tilde{\boldsymbol{\xi}}_{sorp}) + L\boldsymbol{\xi}_{sorp} &= \mathbf{0} && (3.112)\\
\partial_t(\theta\tilde{\boldsymbol{\xi}}_{min}) + L\boldsymbol{\xi}_{min} &= \mathbf{0} && (3.113)\\
\tilde{\boldsymbol{\xi}}_{sorp} &= \boldsymbol{\xi}_{sorp} + \bar{\boldsymbol{\xi}}_{sorp}(\boldsymbol{\eta}, \tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}) && (3.114)\\
\tilde{\boldsymbol{\xi}}_{min} &= \boldsymbol{\xi}_{min} + \bar{\boldsymbol{\xi}}_{min}(\boldsymbol{\eta}, \tilde{\boldsymbol{\xi}}_{sorp}, \tilde{\boldsymbol{\xi}}_{min}) && (3.115)\\
\partial_t(\theta\bar{\boldsymbol{\eta}}) &= \mathbf{0}. && (3.116)
\end{aligned}$$

As $\begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\end{pmatrix}$ is the mobile part of $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\eta}$ is a linear combination of only mobile species, it follows from (3.109) for the mobile part of the total concentrations $\boldsymbol{T}_M$

$$\begin{pmatrix} \boldsymbol{\eta} + (\boldsymbol{E}_1 + \boldsymbol{C}_1\boldsymbol{E}_3)\begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\end{pmatrix} \\ (\boldsymbol{E}_2 + \boldsymbol{C}_2\boldsymbol{E}_3)\begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\end{pmatrix} \end{pmatrix} = \boldsymbol{T}_M\,.$$

Using this and (3.109) it can be seen that the equations of the first block of the Morel formulation $\partial_t(\theta\boldsymbol{T}) + L\boldsymbol{T}_M = \mathbf{0}$ (see (3.102)) are linear combinations of the equations (3.111)-(3.113). Furthermore the equations of the second block of the Morel formulation $\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min})$ (see (3.103)) are linear combinations of the equations (3.114), (3.115) and the trivial equations $\boldsymbol{\eta} = \boldsymbol{\eta}$. Note that the number of equations in the block $\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min})$ is equal to the number of equations in the blocks (3.114) and (3.115) plus the number of the variables $\boldsymbol{\eta}$. Because of (3.110) we have $\bar{\boldsymbol{\eta}} = \boldsymbol{W}$ (for this comparison we have assumed that there are no kinetic reactions). So the equations (3.104) are the same as (3.116).

Here one can see clearly the three advantages of the reduction scheme. The first one is that because of the use of the resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ no splitting techniques are needed. By plugging in the resolution function one gets the global system (3.111)-(3.116) which depends only on the variables $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{min}$, $\boldsymbol{\xi}_{sorp}$, $\boldsymbol{\xi}_{min}$, $\bar{\boldsymbol{\eta}}$ while in the equations of the Morel formulation (3.102)-(3.104) the concentration values $\bar{\boldsymbol{c}}_{nmin,sec}$, $\bar{\boldsymbol{c}}_{min}$ appear.

The second advantage of the reduction scheme is that the number of equations is smaller by the number of the variables $\boldsymbol{\eta}$ because the blocks (3.114), (3.115) consist of less equations than the block (3.103). The third advantage of the reduction scheme that the equation of block (3.111) decouple from the rest of the system and can be solved independent of the rest of the system. So one gets a smaller coupled nonlinear system.

3.9 Generalization of the Reduction Scheme

The goal of this section is to derive a more general formulation of the reduction scheme such that the normal formulation of the reduction scheme and the Morel formulation are special cases of the new formulation.

To derive a more general formulation the block of the PDEs (3.4) is multiplied with the matrices ${\tilde{\boldsymbol{S}}_1^{\perp}}^T$ and $\boldsymbol{C}_1^T$ instead of $\left({\boldsymbol{S}_1^{\perp}}^T \boldsymbol{B}_1^{\perp}\right)^{-1} {\boldsymbol{S}_1^{\perp}}^T$ and $(\boldsymbol{B}_1^T \boldsymbol{S}_1^*)^{-1} \boldsymbol{B}_1^T$ (compare Sec. 3.1). The matrices $\tilde{\boldsymbol{S}}_1^{\perp}$ and $\boldsymbol{C}_1$ must fulfil the following conditions:

(i) The number of rows of both matrices is I

(ii) All columns of $\tilde{\boldsymbol{S}}_1^{\perp}$ and all columns of $\boldsymbol{C}_1$ are linear independent

(iii) All columns of $\tilde{\boldsymbol{S}}_1^{\perp}$ are orthogonal to all columns of $\boldsymbol{S}_1^*$ (but it is not necessary that $\tilde{\boldsymbol{S}}_1^{\perp}$ is a maximal system of linear independent vectors that are orthogonal to all columns of $\boldsymbol{S}_1^*$)

(iv) $\operatorname{span}\left\{\boldsymbol{C}_1, \tilde{\boldsymbol{S}}_1^{\perp}\right\} = \mathbb{R}^I$

(v) There is a matrix $\boldsymbol{D}_1$ such that

$$\boldsymbol{C}_1^T \boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_1 \end{pmatrix} \tag{3.117}$$

(the block $\boldsymbol{D}_1$ can be rectangular)

The size of the matrix $\boldsymbol{C}_1$ is $I \times (J_{mob} + N_*)$ where $J_{sorp,li} + J_{min} + J_{1,kin}^* \leq N_* \leq I - J_{mob}$ (otherwise the conditions (i)-(iv) can not be true), the size of the block $\boldsymbol{D}_1$ is $N_* \times (J_{sorp,li} + J_{min} + J_{1,kin}^*)$. The size of the matrix $\tilde{\boldsymbol{S}}_1^{\perp}$ must be $I \times (I - J_{mob} - N_*)$ because otherwise it is not possible that the conditions (ii) and (iv) are fulfilled.

For the choice ${\tilde{\boldsymbol{S}}_1^{\perp}}^T = \left({\boldsymbol{S}_1^{\perp}}^T \boldsymbol{B}_1^{\perp}\right)^{-1} {\boldsymbol{S}_1^{\perp}}^T$ and $\boldsymbol{C}_1^T = (\boldsymbol{B}_1^T \boldsymbol{S}_1^*)^{-1} \boldsymbol{B}_1^T$ all these conditions are fulfilled. In this case the fourth condition is fulfilled because of the assumption that $\boldsymbol{B}_1$, $\boldsymbol{S}_1^{\perp}$ form a basis of the whole space (see Sec. 3.1) and the block $\boldsymbol{D}_1$ in the fifth condition is the identity matrix. So the transformation done in the normal formulation of the reduction scheme is a special case of the transformation done here.

Multiplication of the block of the PDEs (3.4) with $\tilde{\boldsymbol{S}}_1^{\perp^T}$ and $\boldsymbol{C}_1^T$ gives with use of condition (iii)

$$\begin{aligned}
&\partial_t\left(\theta\tilde{\boldsymbol{S}}_1^{\perp^T}\boldsymbol{c}\right)+L\tilde{\boldsymbol{S}}_1^{\perp^T}\boldsymbol{c}=\boldsymbol{0}\\
&\partial_t\left(\theta\boldsymbol{C}_1^T\boldsymbol{c}\right)+L\boldsymbol{C}_1^T\boldsymbol{c}=\theta\boldsymbol{C}_1^T\boldsymbol{S}_1^*\boldsymbol{A}_1\begin{pmatrix}\boldsymbol{r}_{eq}\\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}.
\end{aligned}$$

Using the fifth condition and the structure of $\boldsymbol{A}_1$ (see (3.10)) it follows

$$\begin{aligned}
&\partial_t\left(\theta\tilde{\boldsymbol{S}}_1^{\perp^T}\boldsymbol{c}\right)+L\tilde{\boldsymbol{S}}_1^{\perp^T}\boldsymbol{c}=\boldsymbol{0}\\
&\partial_t\left(\theta\boldsymbol{C}_1^T\boldsymbol{c}\right)+L\boldsymbol{C}_1^T\boldsymbol{c}=\theta\begin{pmatrix}\boldsymbol{r}_{mob}+\boldsymbol{A}_{1,mob}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\\ \boldsymbol{R}\begin{pmatrix}\boldsymbol{r}_{sorp}\\ \boldsymbol{r}_{min}\end{pmatrix}+\boldsymbol{A}_{1,*}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\end{pmatrix}
\end{aligned}$$

where

$$\boldsymbol{R}:=\boldsymbol{D}_1\begin{pmatrix}\boldsymbol{I}_{J_{sorp,li}} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{ld} & \boldsymbol{I}_{J_{min}}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\end{pmatrix},\qquad \boldsymbol{A}_{1,*}:=\boldsymbol{D}_1\begin{pmatrix}\boldsymbol{A}_{1,sorp}\\ \boldsymbol{A}_{1,min}\\ \boldsymbol{A}_{1,kin}\end{pmatrix}. \tag{3.118}$$

The size of $\boldsymbol{R}$ is $N_*\times(J_{sorp}+J_{min})$ and the size of $\boldsymbol{A}_{1,*}$ is $N_*\times J_{kin}$.

This motivates the following definition of the transformed variables

$$\boldsymbol{\eta}:=\tilde{\boldsymbol{S}}_1^{\perp^T}\boldsymbol{c},\qquad \boldsymbol{\xi}=\begin{pmatrix}\boldsymbol{\xi}_{mob}\\ \boldsymbol{\xi}_*\end{pmatrix}:=\boldsymbol{C}_1^T\boldsymbol{c}. \tag{3.119}$$

The number of the variables $\boldsymbol{\eta}$ is $I-J_{mob}-N_*$, the number of the variables $\boldsymbol{\xi}_{mob}$ is J_{mob} and the number of the variables $\boldsymbol{\xi}_*$ is N_*. Using these new variables it holds

$$\begin{aligned}
&\partial_t(\theta\boldsymbol{\eta})+L\boldsymbol{\eta}=\boldsymbol{0}\\
&\partial_t(\theta\boldsymbol{\xi}_*)+L\boldsymbol{\xi}_*=\theta\boldsymbol{R}\begin{pmatrix}\boldsymbol{r}_{sorp}\\ \boldsymbol{r}_{min}\end{pmatrix}+\theta\boldsymbol{A}_{1,*}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}).
\end{aligned} \tag{3.120}$$

Because of the conditions (ii) and (iv) the matrix $\begin{pmatrix}\boldsymbol{C}_1^T\\ \tilde{\boldsymbol{S}}_1^{\perp^T}\end{pmatrix}$ is invertible. Hence the retransformation can be written as

$$\boldsymbol{c}=\begin{pmatrix}\boldsymbol{C}_1^T\\ \tilde{\boldsymbol{S}}_1^{\perp^T}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\xi}_{mob}\\ \boldsymbol{\xi}_*\\ \boldsymbol{\eta}\end{pmatrix}. \tag{3.121}$$

Partitioning the columns of $\begin{pmatrix} \boldsymbol{C}_1^T \\ \tilde{\boldsymbol{S}}_1^{\perp T} \end{pmatrix}^{-1}$ analogously to the entries of the vector $\begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}_* \\ \boldsymbol{\eta} \end{pmatrix}$ gives

$$\begin{pmatrix} \boldsymbol{C}_1^T \\ \tilde{\boldsymbol{S}}_1^{\perp T} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y} & \boldsymbol{Z} \end{pmatrix}. \tag{3.122}$$

The size of the block $\boldsymbol{X}$ is $I \times J_{mob}$, the size of the block $\boldsymbol{Y}$ is $I \times N_*$ and the size of the block $\boldsymbol{Z}$ is $I \times (I - J_{mob} - N_*)$. With help of the conditions (iii) and (v) one gets

$$\begin{pmatrix} \boldsymbol{C}_1^T \boldsymbol{S}_1^* \\ \tilde{\boldsymbol{S}}_1^{\perp T} \boldsymbol{S}_1^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_1 \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

Multiplying with $\begin{pmatrix} \boldsymbol{C}_1^T \\ \tilde{\boldsymbol{S}}_1^{\perp T} \end{pmatrix}^{-1}$ from left, using the partitioning of this matrix and multiplying with $\boldsymbol{A}_1$ from right yields

$$\boldsymbol{S}_1^* \boldsymbol{A}_1 = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}\boldsymbol{D}_1 \end{pmatrix} \boldsymbol{A}_1.$$

With (3.9), the block structure of $\boldsymbol{A}_1$ (see (3.10)) and the definitions of the matrices $\boldsymbol{R}$ and $\boldsymbol{A}_{1,*}$ (see (3.118)) it follows

$$\boldsymbol{S}_1 = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}\boldsymbol{R} & \boldsymbol{Y}\boldsymbol{A}_{1,*} \end{pmatrix}.$$

Especially it holds (see (2.8), (3.1) for the block structure of $\boldsymbol{S}_1$)

$$\boldsymbol{X} = \boldsymbol{S}_{1,mob}\,, \qquad \boldsymbol{Y}\boldsymbol{R} = \begin{pmatrix} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}. \tag{3.123}$$

With the first relation and the partitioning of $\begin{pmatrix} \boldsymbol{C}_1^T \\ \tilde{\boldsymbol{S}}_1^{\perp T} \end{pmatrix}^{-1}$ the retransformation can be rewritten as

$$\boldsymbol{c} = \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\boldsymbol{\xi}_* + \boldsymbol{Z}\boldsymbol{\eta}. \tag{3.124}$$

The transformation of the equations related to the immobile species remain unchanged (see Sec. 3.1 for this transformation). In this general setting the additional variables are defined as

$$\tilde{\boldsymbol{\xi}} := \boldsymbol{\xi}_* - \boldsymbol{R} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix}. \tag{3.125}$$

With these variables the new retransformation is (see Sec. 3.2 how the retransformation is modified in connection with the additional variables)

$$\begin{aligned}\boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\tilde{\boldsymbol{\xi}} + \boldsymbol{Y}\boldsymbol{R}\begin{pmatrix}\bar{\boldsymbol{\xi}}_{sorp}\\ \bar{\boldsymbol{\xi}}_{min}\end{pmatrix} + \boldsymbol{Z}\boldsymbol{\eta}\\ &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\tilde{\boldsymbol{\xi}} + \boldsymbol{S}_{1,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,min}\bar{\boldsymbol{\xi}}_{min} + \boldsymbol{Z}\boldsymbol{\eta}.\end{aligned} \tag{3.126}$$

Solving the ODEs (3.23)-(3.25) (remember the partitioning of $\bar{\boldsymbol{\xi}}_{sorp}$ (3.16)) for $\boldsymbol{r}_{sorp}$ and $\boldsymbol{r}_{min}$ and plugging this in the PDEs (3.120) gives

$$\partial_t(\theta\boldsymbol{\xi}_*) + L\boldsymbol{\xi}_* = \boldsymbol{R}\begin{pmatrix}\partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp}) - \theta\boldsymbol{A}_{2,sorp}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}})\\ \partial_t(\theta\bar{\boldsymbol{\xi}}_{min})\end{pmatrix} + \theta\boldsymbol{A}_{1,*}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}).$$

This can be rewritten as

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}) + L\boldsymbol{\xi}_* = \theta\boldsymbol{A}_*\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \tag{3.127}$$

with

$$\boldsymbol{A}_* := \boldsymbol{A}_{1,*} - \boldsymbol{R}\begin{pmatrix}\boldsymbol{A}_{2,sorp}\\ \boldsymbol{0}\end{pmatrix}. \tag{3.128}$$

Altogether one has the equations

$$\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} = \boldsymbol{0} \tag{3.129}$$

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}_* - \boldsymbol{R}\begin{pmatrix}\bar{\boldsymbol{\xi}}_{sorp}\\ \bar{\boldsymbol{\xi}}_{min}\end{pmatrix} \tag{3.130}$$

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}) + L\boldsymbol{\xi}_* = \theta\boldsymbol{A}_*\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \tag{3.131}$$

$$\partial_t(\theta\bar{\boldsymbol{\eta}}) = \boldsymbol{0} \tag{3.132}$$

$$\partial_t(\theta\bar{\boldsymbol{\xi}}_{kin}) = \theta\boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \tag{3.133}$$

$$\boldsymbol{\phi}_{mob}(\boldsymbol{c}) = \boldsymbol{0} \tag{3.134}$$

$$\boldsymbol{\phi}_{sorp}(\boldsymbol{c},\bar{\boldsymbol{c}}_{nmin}) = \boldsymbol{0} \tag{3.135}$$

$$\boldsymbol{\phi}_{min}(\boldsymbol{c},\bar{\boldsymbol{c}}_{min}) = \boldsymbol{0}\,. \tag{3.136}$$

with

$$\begin{aligned}\boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\tilde{\boldsymbol{\xi}} + \boldsymbol{S}_{1,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}_{1,min}\bar{\boldsymbol{\xi}}_{min} + \boldsymbol{Z}\boldsymbol{\eta}\\ \bar{\boldsymbol{c}} &= \begin{pmatrix}\boldsymbol{S}_{2,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \tilde{\boldsymbol{B}}_2^{\perp}\bar{\boldsymbol{\eta}}\\ \bar{\boldsymbol{\xi}}_{min}\end{pmatrix}.\end{aligned}$$

The proof for the existence of a resolution function

$$(\tilde{\boldsymbol{\xi}},\bar{\boldsymbol{\xi}}_{kin}) \mapsto (\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min})$$

can be taken over from the normal formulation of the reduction scheme (see Sec. 3.2.1) because the local variables $\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}$ appear in the retransformation with the same coefficients as in the normal formulation and so the derivatives $\dfrac{\partial \boldsymbol{c}}{\partial \boldsymbol{\xi}_{loc}}$ are unchanged. Also the proof for the existence of a resolution function

$$\tilde{\boldsymbol{\xi}} \mapsto (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})$$

is the same as in the normal formulation.

The derivatives $D_{\tilde{\boldsymbol{\xi}}} \boldsymbol{\xi}_{loc}$ needed to assemble the global Jacobian (compare Sec. 3.3.2) can be computed by solving a linear system of equation. This system has the same structure (3.53) as in the normal formulation of the reduction scheme because the resolution function is unchanged. The only difference is that the matrix $\boldsymbol{C}$ now must be chosen as

$$\boldsymbol{C} = \begin{pmatrix} -\boldsymbol{Y} \\ \boldsymbol{0} \end{pmatrix}$$

because in this generalization the variables $\tilde{\boldsymbol{\xi}}$, the variables the resolution function depends on, appear in the retransformation with the coefficients $\boldsymbol{Y}$ instead of $\begin{pmatrix} \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min} & \boldsymbol{S}^*_{1,kin} \end{pmatrix}$.

Morel formulation as special case of generalized reduction scheme

If there are no kinetic reactions and the stoichiometric matrix has the standard form (compare Sec. 3.8)

$$\boldsymbol{S} = \left(\begin{array}{ccc} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} \\ -\boldsymbol{I}_{J_{mob}} & 0 & 0 \\ \hline 0 & \hat{\boldsymbol{B}} & 0 \\ 0 & -\boldsymbol{I}_{J_{sorp}} & 0 \\ 0 & 0 & -\boldsymbol{I}_{J_{min}} \end{array}\right)$$

it is possible to achieve that

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{T}, \quad \boldsymbol{\xi}_* = \boldsymbol{T}_M, \quad \boldsymbol{R} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix} = -\boldsymbol{T}_F, \quad \bar{\boldsymbol{\eta}} = \boldsymbol{W}.$$

For this purpose one has to choose the transformation matrices (note the choice of the transformation matrices for the immobile species is the same as in Sec. 3.8)

$$\boldsymbol{C}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{N_*} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{C}^T \end{pmatrix} \tag{3.137}$$

$$\boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{pmatrix}, \qquad \boldsymbol{B}_2^\perp = \begin{pmatrix} \boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \tag{3.138}$$

with $N_* = I - J_{mob}$. Note that in this case the matrix $\tilde{\boldsymbol{S}}_1^{\perp}$ consists of zero columns and so disappears. Then for the variables $\boldsymbol{\xi}$ defined in (3.119) it holds

$$\begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}_* \end{pmatrix} = \boldsymbol{C}_1^T \boldsymbol{c} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_{J_{mob}} \\ \boldsymbol{I}_{N_*} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \boldsymbol{c}_{prim} \\ \boldsymbol{c}_{sec} \end{pmatrix}$$
$$= \begin{pmatrix} -\boldsymbol{c}_{sec} \\ \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} \end{pmatrix} .$$

Especially one has $\boldsymbol{\xi}_* = \boldsymbol{T}_M$ (see (3.99) for the definition of $\boldsymbol{T}_M$). The choice of $\boldsymbol{B}_2$, $\boldsymbol{B}_2^{\perp}$ leads to (see Sec. 3.8, note that in this section kinetic reactions are excluded)

$$\bar{\boldsymbol{\xi}}_{sorp} = -\bar{\boldsymbol{c}}_{nmin,sec} \, , \quad \bar{\boldsymbol{\xi}}_{min} = -\bar{\boldsymbol{c}}_{min} \, , \quad \bar{\boldsymbol{\eta}} = \boldsymbol{W} .$$

As all columns of $\boldsymbol{S}_1$ are linear independent it holds $\boldsymbol{S}_1^* = \boldsymbol{S}_1$ and $\boldsymbol{A}_1 = \boldsymbol{I}_{J_{eq}}$. So by computing the matrix product

$$\boldsymbol{C}_1^T \boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_{J_{mob}} \\ \boldsymbol{I}_{N_*} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{D} \end{pmatrix}$$

ones sees with the definitions of the matrices $\boldsymbol{D}_1$ (3.117) and $\boldsymbol{R}$ (3.118) that

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{D} \end{pmatrix} .$$

Using this it follows for $\tilde{\boldsymbol{\xi}}$ defined in (3.125)

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{\xi}_* - \boldsymbol{R} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix} = \boldsymbol{\xi}_* - \begin{pmatrix} \boldsymbol{A} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix}$$
$$= \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} + \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min} .$$

Therefore it holds $\tilde{\boldsymbol{\xi}} = \boldsymbol{T}$ and $\boldsymbol{R} \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix} = -\boldsymbol{T}_F$ (see (3.97) and (3.100) for the definitions of $\boldsymbol{T}$ and $\boldsymbol{T}_F$, respectively).

With help of the identities for the variables it is easy to see that also the equations coincide

$$\begin{aligned} \partial_t(\theta \boldsymbol{T}) + L\boldsymbol{T}_M &= \boldsymbol{0} \\ \boldsymbol{T} &= \boldsymbol{T}_M + \boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min}) \\ \partial_t(\theta \boldsymbol{W}) &= \boldsymbol{0} \\ \boldsymbol{\phi}_{mob}(\boldsymbol{c}) &= \boldsymbol{0} \\ \boldsymbol{\phi}_{sorp}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{nmin}) &= \boldsymbol{0} \\ \boldsymbol{\phi}_{min}(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) &= \boldsymbol{0} . \end{aligned}$$

Using the reduction scheme one has a resolution function $(\bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})(\tilde{\boldsymbol{\xi}})$. Using the identities between the total concentrations and the transformed variables one gets a resolution function

$$\boldsymbol{T}_F(\boldsymbol{T}).$$

Plugging in this resolution function gives

$$\begin{aligned}\partial_t(\theta\boldsymbol{T}) + L\boldsymbol{T}_M &= \mathbf{0}\\ \boldsymbol{T} &= \boldsymbol{T}_M + \boldsymbol{T}_F(\boldsymbol{T})\end{aligned}$$

Solving this system is called *global-ODE approach* (see [dDEK09]). The formulation used here is very similar to the formulation in [AK09]. The only difference is that the equation $\boldsymbol{T}_F = \boldsymbol{\psi}(\boldsymbol{T})$, appearing in the formulation of [AK09], is plugged in the other equations of the formulation.

Normal formulation of the reduction scheme as special case of generalized reduction scheme

To get the normal reduction scheme one has to choose

$$\tilde{\boldsymbol{S}}_1^{\perp^T} = \left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^{\perp}\right)^{-1}\boldsymbol{S}_1^{\perp^T}, \qquad \boldsymbol{C}_1^T = \left(\boldsymbol{B}_1^T\boldsymbol{S}_1^*\right)^{-1}\boldsymbol{B}_1^T. \tag{3.139}$$

In this case one gets $\boldsymbol{D}_1 = \boldsymbol{I}_{N_*}$ (compare (3.117)) and so it follows using (3.118)

$$\boldsymbol{R} = \begin{pmatrix}\boldsymbol{I}_{J_{sorp,li}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{A}_{ld} & \boldsymbol{I}_{J_{min}}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{pmatrix}. \tag{3.140}$$

Comparing the definitions of the transformed variables for the generalized formulation (3.119) and the normal formulation (3.14) one gets that

$$\boldsymbol{\xi}_* = \begin{pmatrix}\boldsymbol{\xi}_{sorp}\\ \boldsymbol{\xi}_{min}\\ \boldsymbol{\xi}_{kin}\end{pmatrix}. \tag{3.141}$$

Using all this and the definition of $\boldsymbol{A}_{1,*}$ (see (3.118)) one sees that the PDEs (3.120)

$$\partial_t(\theta\boldsymbol{\xi}_*) + L\boldsymbol{\xi}_* = \theta\boldsymbol{R}\begin{pmatrix}\boldsymbol{r}_{sorp}\\ \boldsymbol{r}_{min}\end{pmatrix} + \theta\boldsymbol{A}_{1,*}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})$$

are the same as

$$\begin{aligned}\partial_t(\theta\boldsymbol{\xi}_{sorp}) + L\boldsymbol{\xi}_{sorp} &= \theta(\boldsymbol{r}_{sorp,li} + \boldsymbol{A}_{1,sorp}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}))\\ \partial_t(\theta\boldsymbol{\xi}_{min}) + L\boldsymbol{\xi}_{min} &= \theta(\boldsymbol{r}_{min} + \boldsymbol{A}_{ld}\boldsymbol{r}_{sorp,ld} + \boldsymbol{A}_{1,min}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}))\\ \partial_t(\theta\boldsymbol{\xi}_{kin}) + L\boldsymbol{\xi}_{kin} &= \theta\boldsymbol{A}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}).\end{aligned}$$

These are exactly the equations (3.19)-(3.21).

In this case it holds (see (3.122) for the definitions of $\boldsymbol{Y}$ and $\boldsymbol{Z}$)

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{S}_{1,sorp,li} & \boldsymbol{S}_{1,min} & \boldsymbol{S}^*_{1,kin} \end{pmatrix} , \qquad \boldsymbol{Z} = \boldsymbol{B}_1^\perp$$

because of the relation

$$\begin{pmatrix} \boldsymbol{C}_1^T \\ \tilde{\boldsymbol{S}}_1^{\perp T} \end{pmatrix}^{-1} = \begin{pmatrix} \left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \\ \left(\boldsymbol{S}_1^{\perp T} \boldsymbol{B}_1^\perp\right)^{-1} \boldsymbol{S}_1^{\perp T} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{S}_1^* & \boldsymbol{B}_1^\perp \end{pmatrix} .$$

That the second relation is true can be seen by multiplying from left with $\begin{pmatrix} \left(\boldsymbol{B}_1^T \boldsymbol{S}_1^*\right)^{-1} \boldsymbol{B}_1^T \\ \left(\boldsymbol{S}_1^{\perp T} \boldsymbol{B}_1^\perp\right)^{-1} \boldsymbol{S}_1^{\perp T} \end{pmatrix}$.

Plugging $\boldsymbol{R}$ (3.140) and the partitioning of $\boldsymbol{\xi}_*$ (3.141) in the definition of the additional variables (3.125) of the generalized reduction scheme gives

$$\begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \\ \tilde{\boldsymbol{\xi}}_{kin} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix} - \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp,li} \\ \boldsymbol{A}_{ld} \bar{\boldsymbol{\xi}}_{sorp,ld} + \bar{\boldsymbol{\xi}}_{min} \\ \boldsymbol{0} \end{pmatrix} .$$

This is not exactly the same as in the normal formulation of the reduction scheme (compare (3.39)) where the (trivial) additional variables $\tilde{\boldsymbol{\xi}}_{kin}$ do not appear. But in the generalized formulation it is not possible to assign the variables $\boldsymbol{\xi}_*$ to the chemical reactions. So one has to define additional variables $\tilde{\boldsymbol{\xi}}$ for all variables $\boldsymbol{\xi}_*$ and it is not possible to do this only for that variables that are related to equilibrium reactions like it is done in the normal formulation.

With this choice of the transformation matrices one gets the normal formulation of the reduction scheme with the only difference that there are the variables $\tilde{\boldsymbol{\xi}}_{kin}$ that are defined by the trivial equation $\tilde{\boldsymbol{\xi}}_{kin} = \boldsymbol{\xi}_{kin}$. These variables do not appear in the normal formulation of the reduction scheme.

Chapter 4

MoMaS–Benchmark

The research group GdR MoMaS[1] set up a numerically extremely challenging benchmark for reactive-transport problems (see [BBC$^+$], [CKK09]). The posed problems are in the style of real hydro-geochemical problems. Both the transport and the chemical reactions are of high complexity. The numerical difficulty arise from the fact that very high equilibrium constants (the largest is 10^{35}, the smallest one 10^{-12}) and large stoichiometric coefficients (the largest one is 10) appear and the order of magnitude of the concentration values differ much. The equilibrium conditions are chosen in such a way that also the small concentration values are crucial for the question which of the chemical reactions are effectively running.

This benchmark was solved by several research groups using their own codes. The codes use different methods (non-iterative operator splitting, iterative operator splitting, one-step method) and different numerics (Finite Differences, Finite Volumes, conform Finite Elements, Mixed Finite Elements). For a comparison of the results see [CHK$^+$10].

4.1 Problem Formulation

There is a 1D and a 2D scenario. In the 1D scenario a thin layer with low permeability and high reactivity is situated in the middle of the domain (see Fig. 4.1). In the 2D case the layer with the low permeability is ranged over 90% of the height of the domain such that a small passage is formed. The water flow of the 2D scenario is shown in Fig. 4.2.

The parameters of the two media in the convective case are shown in Table 4.1. In the diffusive case the dispersion is larger by a factor of 1000.

[1] GdR MoMaS: Groupement de Recherche Modélisations Mathématiques et Simulations Numériques liées aux problèmes de gestion des déchets nucléaires

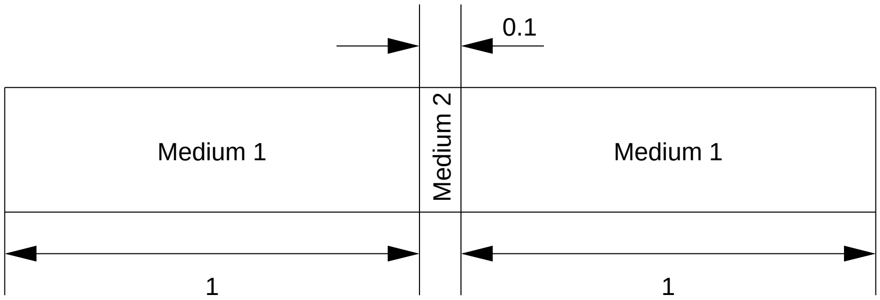


Figure 4.1: Partitioning of the domain in the two media in the 1D scenario

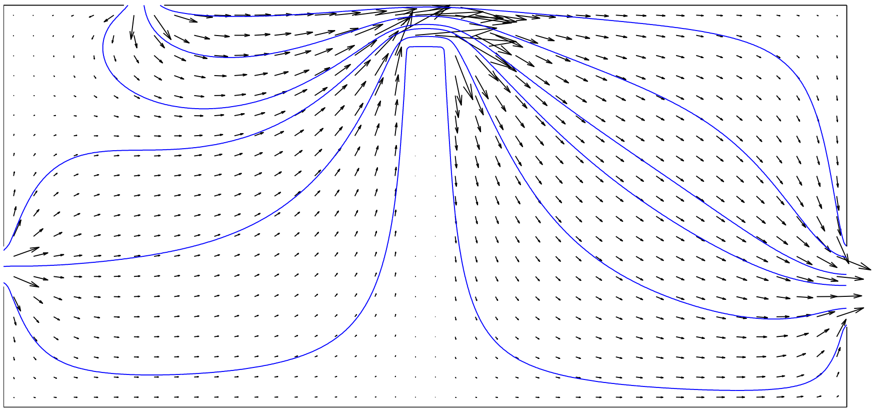


Figure 4.2: Water flow of the 2D scenario

	Medium A	Medium B
Porosity ω	0.25	0.5
Dispersion β	10^{-2}	$6 \cdot 10^{-2}$

Table 4.1: Parameters of the two media

In the 1D case the water flow is constant with the value $\boldsymbol{q} = 5.5 \cdot 10^{-3}$. On the inflow boundary there are Dirichlet boundary conditions and at the outflow boundary there are homogeneous Neumann boundary conditions. At $t = 5000$ the values for the Dirichlet boundary conditions change. The end time of the computation is $T = 6000$.

There are three different chemical reaction networks, that are named “easy test case”, “medium test case” and “hard test case”. The “easy test case” contains only equilibrium reactions in the mobile phase and equilibrium sorption reactions, the “medium test case” additionally contains a kinetic mineral reaction and the “hard test case” additionally has mineral reactions in equilibrium and a decay reaction.

In the “easy test case” there are nine mobile and three immobile species. The five equilibrium reactions in the mobile phase and the two equilibrium sorption reactions are

$$\begin{aligned}
\mathrm{C_1} + \mathrm{X_2} &\longleftrightarrow && K = 10^{-12}\\
\mathrm{C_2} &\longleftrightarrow \mathrm{X_2} + \mathrm{X_3} && K = 1\\
\mathrm{C_3} + \mathrm{X_2} &\longleftrightarrow \mathrm{X_4} && K = 1\\
\mathrm{C_4} + 4\mathrm{X_2} &\longleftrightarrow \mathrm{X_3} + 3\mathrm{X_4} && K = 0.1\\
\mathrm{C_5} &\longleftrightarrow 4\mathrm{X_2} + 3\mathrm{X_3} + \mathrm{X_4} && K = 10^{35}\\[1ex]
\mathrm{CS_1} &\longleftrightarrow 3\mathrm{X_2} + \mathrm{X_3} + \mathrm{S} && K = 10^{6}\\
\mathrm{CS_2} + 3\mathrm{X_2} &\longleftrightarrow \mathrm{X_4} + 2\mathrm{S} && K = 0.1
\end{aligned}$$

where $\mathrm{C_i}$ denotes the secondary mobile species, $\mathrm{X_i}$ the primary mobile species, S the free sorption sites and $\mathrm{CS_i}$ the sorbed species.

Applying the reduction scheme to this problem leads to two linear decoupled partial differential equations (η-problem), one decoupled ODE ($\bar{\eta}$-problem), seven algebraic equilibrium conditions (local problem) and two coupled nonlinear partial differential equations (global problem).

In the “medium test case” there are eleven mobile and four immobile species. Furthermore there are two more equilibrium reactions in the mobile phase and one kinetic mineral reaction. The additional reactions are

$$\begin{aligned}
\mathrm{C_6} &\longleftrightarrow 10\mathrm{X_2} + 3\mathrm{X_3} && K = 10^{32}\\
\mathrm{C_7} + 8\mathrm{X_2} &\longleftrightarrow 2\mathrm{X_4} && K = 10^{-4}\\[1ex]
3\mathrm{C_3} &\longleftrightarrow \mathrm{Cc} + 2\mathrm{X_4}
\end{aligned}$$

where Cc denotes the kinetic mineral. The reaction rate for the mineral reaction is

$$r(\mathrm{C_3}, \mathrm{X_4}) = \left(0.2\frac{\mathrm{C_3}^3}{\mathrm{X_4}^2} - 1\right) k$$

with

$$k = \begin{cases} 10^{-2} & \text{for } 0.2\frac{\mathrm{C_3}^3}{\mathrm{X_4}^2} \geq 1 \\ 10 & \text{else} \end{cases} .$$

Applying the reduction scheme we get three more local equations compared to the “easy test case”: two algebraic equilibrium conditions and one ODE describing the kinetics of the mineral reaction. The number of the other equations does not change.

In the “hard test case” there are twelve mobile and six immobile species. In addition to the reactions of the “medium test case” there are two mineral reactions at equilibrium and one decay reaction

$$\begin{aligned}
\mathrm{CP}_1 &\longleftrightarrow 3\mathrm{X}_2 + \mathrm{X}_4 && K = 10^8 \\
\mathrm{CP}_2 &\longleftrightarrow \mathrm{X}_2 + \mathrm{X}_5 && K = 20 \\
\mathrm{X}_5 &\longrightarrow 2\mathrm{X}_2 + \mathrm{X}_3
\end{aligned}$$

with the decay rate

$$r(\mathrm{X}_5, \mathrm{CP}_2) = 0.05\mathrm{X}_5 + 5\mathrm{CP}_2$$

where CP_i denotes the equilibrium minerals.

Using the reduction method leads to a system of three coupled nonlinear partial differential equations, two linear decoupled partial differential equations, one decoupled ODE and twelve local equations (nine algebraic equations, two complementarity conditions and one ODE).

4.2 Transformation

In the “easy test case” of this benchmark the stoichiometric matrices and one possible choice for the orthogonal complement corresponding to mobile species are

$$\boldsymbol{S}_{1,mob} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 4 & -4 \\ 0 & -1 & 0 & -1 & -3 \\ 0 & 0 & -1 & -3 & -1 \end{pmatrix}, \quad \boldsymbol{S}_{1,sorp} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -3 & 3 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{S}_1^{\perp} = \begin{pmatrix} \frac{1}{36} & 0 \\ \frac{2}{36} & 0 \\ -\frac{2}{36} & 0 \\ -\frac{2}{36} & 0 \\ \frac{2}{36} & 0 \\ 0 & 1 \\ -\frac{1}{36} & 0 \\ \frac{3}{36} & 0 \\ -\frac{3}{36} & 0 \end{pmatrix}$$

and for the immobile ones

$$\boldsymbol{S}_{2,sorp} = \begin{pmatrix} -1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{S}_2^{\perp} = \frac{1}{6}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

The transformation matrices are chosen in the following way

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \qquad \boldsymbol{B}_1^{\perp} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\boldsymbol{B}_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{B}_2^{\perp} = \frac{1}{6}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The first five unit vectors in $\boldsymbol{B}_1$ cause that the transformed variables $\boldsymbol{\xi}_{mob}$ correspond to the secondary species C_i (see (3.13) and (3.14) for the definition of the transformed variables). The choice of the last two columns of $\boldsymbol{B}_1$ bring about that $\tilde{\boldsymbol{\xi}}_{sorp}$ corresponds except for the sign to the total concentrations T_3 and T_4. Furthermore the unit vectors in $\boldsymbol{B}_2$ cause that $\bar{\boldsymbol{\xi}}_{sorp}$ corresponds to the secondary immobile species $\mathrm{CS_i}$. Altogether we get for the transformed variables in the "easy test case"

$$\begin{aligned} \xi_{mob,i} &= \mathrm{C_i} \qquad (i = 1, \ldots, 5) \\ \eta_1 &= -\mathrm{C}_1 - 2\mathrm{C}_2 + 2\mathrm{C}_3 + 2\mathrm{C}_4 - 2\mathrm{C}_5 + \mathrm{X}_2 - 3\mathrm{X}_3 + 3\mathrm{X}_4 \\ &= \mathrm{T}_2 - 3\mathrm{T}_3 + 3\mathrm{T}_4 \\ \eta_2 &= \mathrm{X}_1 = T_1 \\ \tilde{\xi}_{sorp,1} &= -\mathrm{C}_2 - \mathrm{C}_4 - 3\mathrm{C}_5 - \mathrm{X}_3 - \mathrm{CS}_1 = -\mathrm{T}_3 \\ \tilde{\xi}_{sorp,2} &= -\mathrm{C}_3 - 3\mathrm{C}_4 - \mathrm{C}_5 - \mathrm{X}_4 - \mathrm{CS}_2 = -\mathrm{T}_4 \\ \bar{\eta} &= 6\mathrm{S} + 6\mathrm{CS}_1 + 12\mathrm{CS}_2 = 6\mathrm{TS} \\ \bar{\xi}_{sorp,i} &= \mathrm{CS_i} \qquad (i = 1, 2). \end{aligned} \tag{4.1}$$

Most of the transformed variables correspond to one of the original variables the benchmark is formulated with. So comparisons with the results of other groups are easily possible.

Also in the "medium test case" and in the "hard test case" the matrices $\boldsymbol{B}_i$ and $\boldsymbol{B}_i^{\perp}$ are chosen in such a way that every column has only one nonzero entry. Again this has the consequence that most transformed variables coincide with original variables.

4.3 Numerical Results

The reduction scheme is implemented in 2D. So the 2D code is used to emulate the 1D problems by replacing the 1D computational domain by a narrow 2D computational domain. The width of the domain is chosen such that the width matches the size of two cells at the coarsest part of the grid (compare Fig. 4.3). All presented results of the 1D problems are cuts at the middle of the domain (red line in Fig. 4.3).

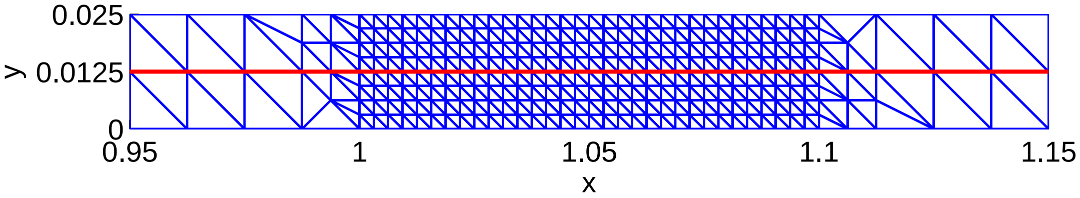


Figure 4.3: Detail of a grid for the 1D problem

For all 1D advective cases we use a preadapted mesh with different step sizes on the two media: step size h_1 on medium A and step size h_2 on medium B with $h_1 = 4h_2$ (For an example see Fig. 4.3).

The reason for this is that by doing so oscillations in the elution curve of C5 can be avoided. The oscillations depend only on the step size h_2 (see Fig. 4.4). So it is not necessary to choose a smaller value for h_1.

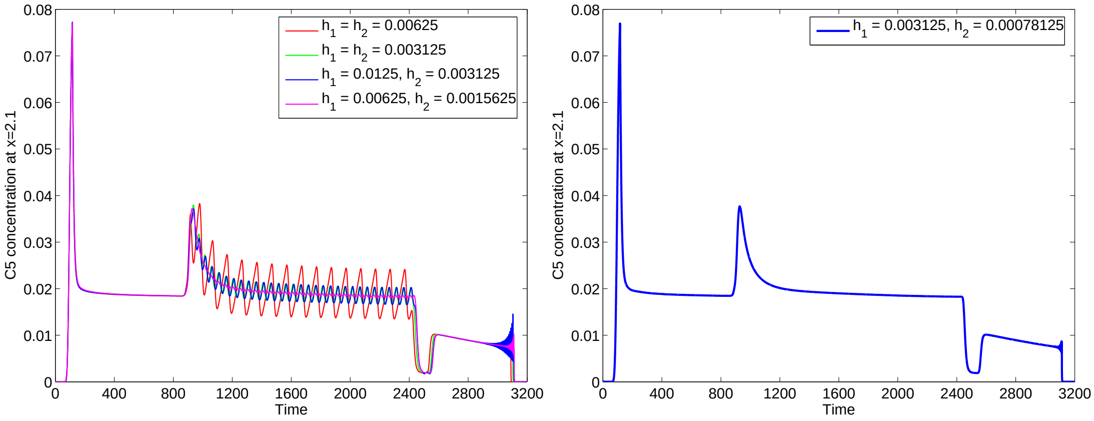


Figure 4.4: Elution curve of C5 for different step sizes

It can be observed that the number of oscillations is equal to the number of cells in the interval $(1, 1.1)$ (see Fig. 4.5).

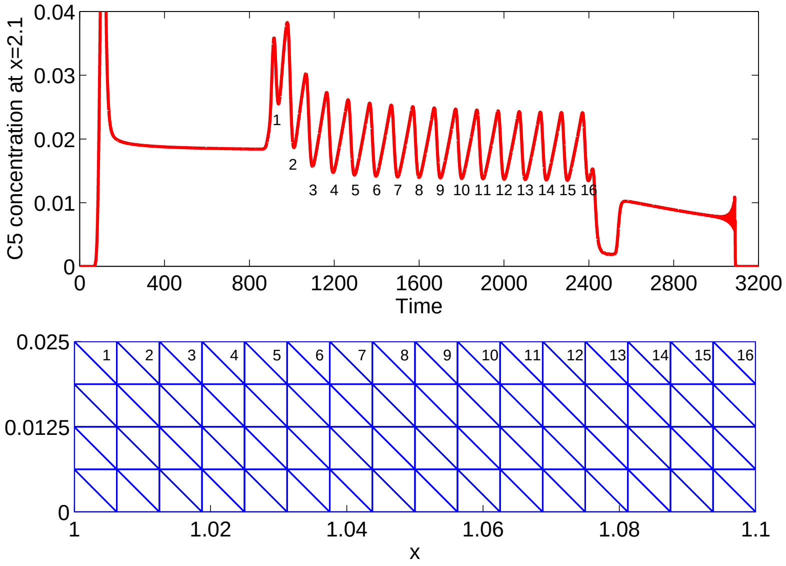


Figure 4.5: Elution curve of C5 (top) and a detail of the grid used for this computation (bottom)

For the 2D "advective test case" we also use a preadapted mesh. Here the mesh is refined in the high velocity zone (the small passage above the layer with the low permeability) and near the outflow. It turned out that this is useful to reduce oscillations in the concentration profile of C5. For all other cases the computations are carried out with regular grids.

The 2D problem is convection dominated. So it is necessary to use a stabilization. Here the FV stabilization described in Section 3.4.4 with full upwinding is used. The 1D problem is not convection dominated. So the FV stabilization is switched off for the 1D cases.

The η-equations reach a steady-state much earlier than the nonlinear system. For example in the 1D "advective easy test case" this happens at $t \approx 200$ and $t \approx 5200$ respectively, whereas the nonlinear system reaches a steady-state not until $t = 3200$. So one advantage of the used reduction scheme is that for a large part of the simulation (here for $200 < t < 3200$ and $5200 < t$) less equations (those for η) have to be solved.

We have carried out computations for the 1D and the 2D "advective easy test case", the 1D "diffusive easy test case", all four "medium test cases" and the 1D and the 2D "advective hard test case". The normalized CPU time (one unit is the CPU time for the multiplication of two 1000 × 1000 matrices), the number of cells of the used grid, the number of time steps and the average number of global Newton steps for all these computations is presented in the Tables 4.2 and 4.3

(see also [HKK09]). In the 1D cases also the number of nodes in x-direction is given so that the results can be compared to pure 1D codes.

	nodes in x-direction	cells 2D-grid	CPU time	time steps	Newton steps
advective easy	777	4638	1484.9	10683	2.80
advective easy	1165	6942	3405.0	12494	3.23
diffusive easy	547	2184	1894.9	6873	3.32
diffusive easy	673	2688	3398.6	8235	3.57
advective medium	389	2334	179.9	2361	2.39
advective medium	874	5214	479.5	2353	2.72
diffusive medium	337	1344	187.5	1674	2.13
diffusive medium	505	2016	452.4	1738	2.26
advective hard	777	4638	4102.2	31091	1.60[2]
advective hard	874	5214	4860.9	31758	1.61[2]

Table 4.2: CPU time, time steps, Newton steps of the 1D problems

	cells	CPU time	time steps	Newton steps
advective easy	38016	10645.6	13338	2.73
advective easy	107520	45092.6	18990	3.14
advective medium	26880	6991.9	12810	2.07
diffusive medium	26880	7436.0	7880	2.98
advective hard	26880	19212.7	27199	2.17

Table 4.3: CPU time, time steps, Newton steps of the 2D problems

The computations were carried out on a Linux cluster with 18 dual-processor nodes. Every node has two Intel Xeon processors (NetBurst architecture) with 2.4 GHz and 1GB RAM. On this computer one unit of the normalized CPU time corresponds to 15.0s. For each simulation only one processor is used to make the comparison of the CPU time with other groups easier. The implementation of the reduction scheme used for these simulations is based on the old version of M++.

[2] For $2566.6 \leq t \leq 5000$ (12167 time steps) only one Newton step is required

4.4 Comparison of the CPU Time with Other Groups

A detailed comparison of the easy test case results of all participants of the benchmark will be published in a synthesis article [CHK+10]. So here only a short comparison will be given.

In the 1D "advective easy test case" the reduction scheme was the fastest code of six participants despite of the disadvantage that we used a 2D code to solve a 1D problem. Fig. 4.6 shows the normalized CPU times of the different participants in dependency of the number of cells. This figure is taken from a preliminary version of [CHK+10].

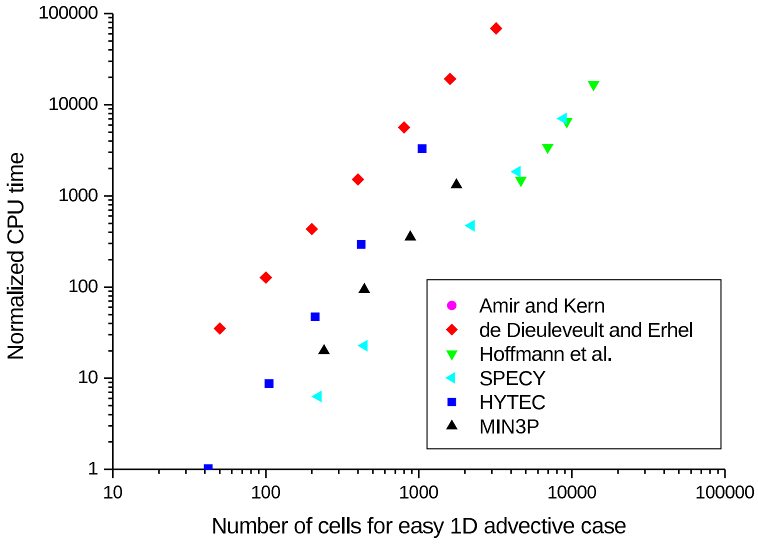


Figure 4.6: CPU time in the 1D "advective easy test case"

In the 2D "advective easy test case" two other groups presented results. Vincent Lagneau and Jan van der Lee presented results computed with the code HYTEC, which uses iterative operator splitting (SIA). K. Ulrich Mayer and Kerry T. B. MacQuarrie presented results computed with the code MIN3P, which uses the direct substitution approach (DSA). For a short description of the method used by the benchmark participants see Section 4.6. In Fig. 4.7 the normalized CPU times of the three participants in dependency of the number of cells can be seen. Again this figure is taken from a preliminary version of [CHK+10].

The other groups give only results for coarser grids. By extrapolating the lines one can see that the reduction scheme is faster by a factor greater than five compared with the second fastest code.

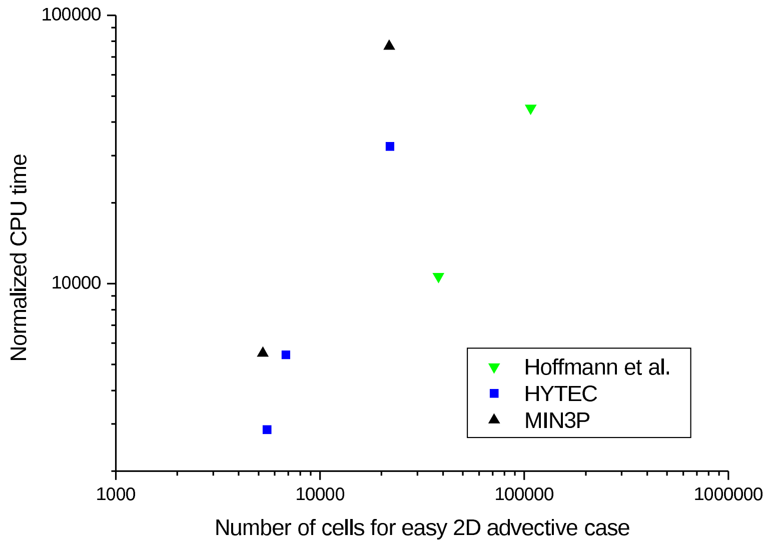


Figure 4.7: CPU time in the 2D “advective easy test case”

4.5 Comparison of the CPU Time with Other Methods

In this section the implementation of the generalization of the reduction scheme (see Sec. 3.9) is used to compare the CPU times of the reduction scheme with those of the global–ODE approach and those of iterative operator splitting.

In a first test the variables and equations of the reduction scheme are used but the η-equations are added to the global problem and so are solved simultaneously with the global problem. Here one can see the saving of CPU time by the decoupling of the η-equations.

To achieve that the η-equations and the global problem is solved in one system the transformation matrices $\tilde{\boldsymbol{S}}_1^{\perp}$, $\boldsymbol{C}_1$ of the generalization of the reduction scheme must be chosen in the following way. The matrix $\boldsymbol{C}_1$ must be chosen as

$$\boldsymbol{C}_1 = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -1.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & -2.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 2.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & -1.0 & -3.0 & 0.0 & 2.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & -3.0 & -1.0 & 0.0 & -2.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 & -3.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 3.0 \end{pmatrix}$$

and the matrix $\tilde{\boldsymbol{S}}_1^\perp$ must be the empty matrix. The transformed variables are linear combinations of the concentrations. The columns of the matrix $\boldsymbol{C}_1$ contains the coefficients of the linear combinations of the variables $\boldsymbol{\xi}_{mob}$, $\boldsymbol{\xi}_{sorp}$, $\boldsymbol{\eta}$ (see (4.1) for the transformed variables of the easy test case). In Table 4.4 the results are named "no decoupling of η-equations".

To get the global–ODE approach one has to choose $\boldsymbol{C}_1$ analogously to (3.137) (note that here $\boldsymbol{S}_1$ has a different form as in Sec. 3.9, here the identity block is on the top and has no minus sign). For the easy test case one gets

$$\boldsymbol{C}_1 = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & -1.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & -4.0 & 1.0 & 3.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 4.0 & 3.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}.$$

To do SIA with the implementation of the generalized reduction scheme one has to use same matrix $\boldsymbol{C}_1$ as in global–ODE approach but has to do some modifications in the implementation. The main thing is to comment out the terms with $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ in the Jacobi matrix and to plug equation (3.130) in (3.131).

Moreover the stopping criterion (compare (3.95)) of the global Newton, which is the same as the stopping criterion for the coupling of the transport and the chemical problem in a pure SIA code, is changed. The reduction *Red* is enlarged by a factor of ten and two different values for the absolute criterion $Eps = 2 \cdot 10^{-8}, 10^{-8}$ are taken. The second value is by a factor of hundred larger than the value taken in reduction scheme and the global–ODE approach.

Furthermore the terms with $D_{\boldsymbol{\xi}_{glob}}\boldsymbol{\xi}_{loc}$ in the update of the concentration values must be commented out and line search in the global Newton must be switched off. Additionally it is necessary to adjust the adaption of the time step size (compare Sec. 3.7) because the number of iteration steps doing SIA is much larger than the number of Newton steps using the reduction scheme or the global–ODE approach. Here the time step size is enlarged up to the maximal time step size $\Delta t_{max} = 0.5$ when the number of iteration steps in the previous time step was less than twenty and it is reduced when more than thirty iteration steps were needed.

All computations presented in this section are done with an implementation based on the new version of M++ and exponential upwinding is used in the FV stabilization. Because of these two points the results of the normal formulation of the reduction scheme are slightly different to those given in Table 4.3.

In Table 4.4 the CPU time, the number of time steps and the average number of global Newton steps (iteration steps in the case of SIA) are given.

	cells	CPU time	time steps	Newton steps
reduction scheme	26880	6260.5	12921	2.14
reduction scheme	38016	11269.3	14602	2.51
no decoupling of η-equations	26880	10266.0	13123	2.15
no decoupling of η-equations	38016	18498.8	14779	2.52
global–ODE approach	26880	10844.5	13077	2.25
global–ODE approach	38016	19647.9	14986	2.63
SIA ($Eps = 2 \cdot 10^{-8}$)	9504	12031.0	14979	15.3
SIA ($Eps = 10^{-8}$)	9504	14404.3	16726	17.1
SIA ($Eps = 2 \cdot 10^{-8}$)	26880	35437.4	15680	16.6
SIA ($Eps = 10^{-8}$)	26880	42782.3	17949	18.4

Table 4.4: CPU time, time steps, Newton steps for different methods

Comparing the results of the reduction scheme and the case “no decoupling of η-equations” one can see that the gain of CPU time by the decoupling of the η-equations is bit more than one third. Furthermore it can be seen, if one does not want to use the reduction scheme, then the global–ODE approach is the best thing one can do.

The SIA approach has two drawbacks. The first one is that the CPU time is much higher than all the other methods. Compared with the reduction scheme the CPU time is higher by a factor greater than five. The second drawback of the SIA approach is that it is not possible to get solutions that are as precise as the solutions of the other methods because it is not possible to choose the same stopping criterion.

In these computations it is necessary to enlarge the stopping parameter *Eps* by a factor of hundred because otherwise the convergence gets so slow that the method is practically unusable. How much more CPU time is needed when the stopping parameter *Eps* is diminished only by a factor of two can be seen in Table 4.4.

Analyzing the SIA method one sees that it has a linear convergence behavior like it is expected. After some iteration steps (about five) in each iteration the residual is reduced by a approximately constant factor. The reason why a strong stopping criteria is practically unusable is that sometimes this factor is only 1.3 depending on the problem and the time step size. So it would take 18 iteration steps to reduce the residual by a factor of hundred. The much larger numbers of iteration steps lead to much smaller time step sizes because of the adaptive time stepping. For example in a test computation with the stopping parameter *Eps* reduced by a factor of hundred the time step sizes are one third of the original ones. This much smaller time step sizes together with the higher number of iteration steps leads to the very high CPU times of the SIA method in case of strong stopping criteria.

4.6 The Different Methods Used by the Participants

We consider the simple situation that there are no minerals and that all chemical reactions are equilibrium reactions like it is the case in the “easy test case”. Then the stoichiometric matrix $\boldsymbol{S}$ has the form

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_1 \\ \hline \boldsymbol{S}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} \\ \hline \boldsymbol{0} & \boldsymbol{S}_{2,sorp} \end{pmatrix} = \left(\begin{array}{cc} \boldsymbol{C} & \boldsymbol{A} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \boldsymbol{B} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{sorp}} \end{array}\right).$$

Written in terms of logarithmized concentrations of primary/secondary species and total concentrations the chemical subsystem consisting of equilibrium conditions and mass balance equations is

$$\begin{aligned} \boldsymbol{C}^T \boldsymbol{l}_{prim} - \boldsymbol{l}_{sec} - \boldsymbol{k}_{mob} &= \boldsymbol{0} \\ \exp(\boldsymbol{l}_{prim}) + \boldsymbol{C} \exp(\boldsymbol{l}_{sec}) + \boldsymbol{A} \exp(\bar{\boldsymbol{l}}_{sec}) - \boldsymbol{T} &= \boldsymbol{0} \\ \boldsymbol{A}^T \boldsymbol{l}_{prim} + \boldsymbol{B}^T \bar{\boldsymbol{l}}_{prim} - \bar{\boldsymbol{l}}_{sec} - \boldsymbol{k}_{sorp} &= \boldsymbol{0} \\ \exp(\bar{\boldsymbol{l}}_{prim}) + \boldsymbol{B} \exp(\bar{\boldsymbol{l}}_{sec}) - \boldsymbol{W} &= \boldsymbol{0} \end{aligned} \tag{4.2}$$

where $\boldsymbol{T}$ denotes the total concentrations and $\boldsymbol{W}$ the total fixed concentrations.

The secondary concentrations $\boldsymbol{l}_{sec}$ and $\bar{\boldsymbol{l}}_{sec}$ can be eliminated from the system by resolving the equilibrium conditions for $\boldsymbol{l}_{sec}$ and $\bar{\boldsymbol{l}}_{sec}$, respectively, and plugging in the other equations. The total fixed concentrations $\boldsymbol{W}$ are always equal to their initial values. Hence from now on we will handle $\boldsymbol{W}$ as a constant. So we can write system (4.2) shortly as

$$\boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{T}) = \boldsymbol{0}$$

with $\boldsymbol{X} := \begin{pmatrix} \boldsymbol{l}_{prim} \\ \bar{\boldsymbol{l}}_{prim} \end{pmatrix}$.

The mobile part of the total concentrations is named with $\boldsymbol{T}_M$ and the immobile part of the total concentrations with $\boldsymbol{T}_F$. It holds

$$\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{T}_F\,.$$

$\boldsymbol{T}_M$ and $\boldsymbol{T}_F$ can be computed with help of the functions

$$\begin{aligned}\boldsymbol{T}_M(\boldsymbol{X}) &:= \exp(\boldsymbol{l}_{prim}) + \boldsymbol{C}\exp(\boldsymbol{C}^T\boldsymbol{l}_{prim} - \boldsymbol{k}_{mob})\\ \boldsymbol{T}_F(\boldsymbol{X}) &:= \boldsymbol{A}\exp(\boldsymbol{A}^T\boldsymbol{l}_{prim} + \boldsymbol{B}^T\bar{\boldsymbol{l}}_{prim} - \boldsymbol{k}_{sorp})\,.\end{aligned}$$

Let us define the resolution function of the chemical problem $\boldsymbol{\Psi}(\boldsymbol{T})$. It is defined by

$$\boldsymbol{\Psi}(\boldsymbol{T}) = \boldsymbol{T}_F(\boldsymbol{X}^*)$$

with $\boldsymbol{X}^*$ being the solution of $\boldsymbol{\Phi}(\boldsymbol{X}^*, \boldsymbol{T}) = \boldsymbol{0}$.

The PDEs describing the transport are

$$\partial_t\boldsymbol{T} + L\boldsymbol{T}_M = \boldsymbol{0}$$

or alternatively they can be written as

$$\partial_t(\boldsymbol{T}_M + \boldsymbol{T}_F) + L\boldsymbol{T}_M = \boldsymbol{0}\,.$$

First we compare the reduction scheme used in this work with the formulation in terms of primary/secondary species and total concentrations. The global problem of the reduction scheme is

$$\partial_t\boldsymbol{\eta} + L\boldsymbol{\eta} = \boldsymbol{0} \tag{4.3}$$

$$\partial_t\tilde{\boldsymbol{\xi}}_{sorp} + L\boldsymbol{\xi}_{sorp} = \boldsymbol{0} \tag{4.4}$$

$$\tilde{\boldsymbol{\xi}}_{sorp} = \boldsymbol{\xi}_{sorp} - \bar{\boldsymbol{\xi}}_{sorp}(\boldsymbol{\eta}, \tilde{\boldsymbol{\xi}}_{sorp})\,. \tag{4.5}$$

For a certain choice of the transformation matrices (see Sec. 3.8) one gets the connection between the variables $\boldsymbol{\eta}$, $\boldsymbol{\xi}_{sorp}$, $\tilde{\boldsymbol{\xi}}_{sorp}$, $\bar{\boldsymbol{\xi}}_{sorp}$ and $\boldsymbol{T}_M$, $\boldsymbol{T}$, $\boldsymbol{T}_F$

$$\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{0}\end{pmatrix} + \boldsymbol{A}\boldsymbol{\xi}_{sorp} = \boldsymbol{T}_M\,, \qquad \begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{0}\end{pmatrix} + \boldsymbol{A}\tilde{\boldsymbol{\xi}}_{sorp} = \boldsymbol{T}\,, \qquad -\boldsymbol{A}\bar{\boldsymbol{\xi}}_{sorp} = \boldsymbol{T}_F\,.$$

So in this case the equations $\partial_t\boldsymbol{T} + L\boldsymbol{T}_M = \boldsymbol{0}$ are linear combinations of the equations (4.3) and (4.4). Furthermore the equations $\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{\Psi}(\boldsymbol{T})$ are linear combinations of the equations (4.5) and the trivial equations $\boldsymbol{\eta} = \boldsymbol{\eta}$. Remember that the number of equations in $\boldsymbol{T} = \boldsymbol{T}_M + \boldsymbol{\Psi}(\boldsymbol{T})$ is equal to the number of equations in (4.5) plus the number of the variables $\boldsymbol{\eta}$. For the spatial discretization conform finite elements are used with a finite volume stabilization for convection-dominated problems. When the stabilization is used the discretization scheme is equivalent to a cell-centered finite volume scheme. The time discretization is done with the implicit Euler method.

Now let us have a short look on all the different methods used by the participants of the MoMaS–benchmark. The software SPECY (see [Car01], [Car09]) uses the non iterative splitting scheme (SNIA)

$$\begin{aligned}\frac{\boldsymbol{T}_M^* - \boldsymbol{T}_M^n}{\Delta t} + \frac{\boldsymbol{T}_F^n - \boldsymbol{T}_F^n}{\Delta t} + L_{adv}\boldsymbol{T}_M^n + L_{disp}\boldsymbol{T}_M^* &= \mathbf{0}\\ \boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{T}_M^* + \boldsymbol{T}_F^n) &= \mathbf{0}\\ \boldsymbol{T}_M^{n+1} &= \boldsymbol{T}_M(\boldsymbol{X})\\ \boldsymbol{T}_F^{n+1} &= \boldsymbol{T}_F(\boldsymbol{X})\,.\end{aligned}$$

In every time step the transport equations are solved first. So one gets new values for the mobile part of the total concentrations $\boldsymbol{T}_M^*$. Then the immobile part of the total concentrations from the last time step $\boldsymbol{T}_F^n$ are added and with these values for the total concentrations the chemical subsystem is solved. Subsequently new values for the mobile and immobile parts of the total concentrations $\boldsymbol{T}_M$, $\boldsymbol{T}_F$ are computed. For the spatial discretization discontinuous finite elements time explicit for advection and mixed hybrid finite elements, time implicit for dispersion are used.

The software HYTEC (see [LvdL09]) uses the iterative splitting scheme (SIA)

$$\begin{aligned}\frac{\boldsymbol{T}_M^{n+1,2m+1} - \boldsymbol{T}_M^n}{\Delta t} + \alpha L\boldsymbol{T}_M^{n+1,2m+1} + (1-\alpha)L\boldsymbol{T}_M^n &= -\frac{\boldsymbol{T}_F^{n+1,2m} - \boldsymbol{T}_F^n}{\Delta t}\\ \boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{T}_M^{n+1,2m+1} + \boldsymbol{T}_F^{n+1,2m}) &= \mathbf{0}\\ \boldsymbol{T}_M^{n+1,2m+2} &= \boldsymbol{T}_M(\boldsymbol{X})\\ \boldsymbol{T}_F^{n+1,2m+2} &= \boldsymbol{T}_F(\boldsymbol{X})\end{aligned}$$

where m denotes the number of the iteration step. In every iteration step first the transport equations are solved and then the chemical system. The code uses a finite volume scheme based on a Voronoi (nearest-neighbour) spatial discretization. The discretization scheme is centered in space. For the time-discretization the semi-implicit Crank–Nicholson method is used.

In [dD08], [dDEK09] and [dDE09] a method following the DAE approach is described. The whole system

$$\begin{aligned}\partial_t\boldsymbol{T} + L\boldsymbol{T}_M &= \mathbf{0}\\ \boldsymbol{\Phi}(\boldsymbol{X}, \boldsymbol{T}) &= \mathbf{0}\\ \boldsymbol{T}_M &= \boldsymbol{T}_M(\boldsymbol{X})\end{aligned}$$

is solved, after it is discretized in space, using a DAE solver. For the space discretization a cell-centered finite volume scheme is used. For the advective term a first-order upwind scheme and for the diffusion term a second-order centered scheme is applied. The DAE system is solved with a BDF–method with variable order (up to 5).

The software MIN3P (see [MFB02], [May99] and [MM09]) follows the DSA approach. The resulting system of equations is

$$\partial_t \boldsymbol{T}_M(\boldsymbol{X}) + \partial_t \boldsymbol{T}_F(\boldsymbol{X}) + L\boldsymbol{T}_M(\boldsymbol{X}) = \boldsymbol{0}\,.$$

Note that the unknowns are the *logarithms* of the primary concentrations. If one formally sets $L = 0$ and replaces the time derivatives by a difference quotient one gets the same equations as in the chemical subproblem of a splitting method. Spatial discretization is performed using a control volume method with half-cells on the boundary, in which for the advective transport upstream weighting is used. The code uses implicit time weighting.

In [AK09] a two level global algorithm following the global implicit approach (GIA) is introduced. The system of equations is

$$\begin{aligned}
\frac{\boldsymbol{T}_M^{n+1} - \boldsymbol{T}_M^{n}}{\Delta t} + L_{adv}\boldsymbol{T}_M^{n} + L_{diff}\boldsymbol{T}_M^{n+1} &= \frac{\boldsymbol{T}_F^{n+1} - \boldsymbol{T}_F^{n}}{\Delta t} \\
\boldsymbol{T}^{n+1} &= \boldsymbol{T}_M^{n+1} + \boldsymbol{T}_F^{n+1} \\
\boldsymbol{T}_F^{n+1} &= \boldsymbol{\Psi}(\boldsymbol{T}^{n+1})
\end{aligned}$$

Like in the reduction scheme a resolution function is used to handle the chemical problem. Between the resolution functions there is the connection

$$-\boldsymbol{A}\bar{\boldsymbol{\xi}}_{sorp}(\boldsymbol{\eta}, \tilde{\boldsymbol{\xi}}_{sorp}) = \boldsymbol{\Psi}(\boldsymbol{T})\,.$$

Here the system of equations consists of more equations than in the case of the reduction scheme because of three reasons. Firstly in the reduction scheme the equations $\boldsymbol{T}_F = \boldsymbol{\Psi}(\boldsymbol{T})$ are plugged into the other equations. Secondly in the reduction scheme linear combinations of the equations are taken in such a way that some linear partial differential equations decouple from the system. Thirdly by taking linear combinations of the equations in the second block some equations get trivial by use of the reduction scheme ($\boldsymbol{\eta} = \boldsymbol{\eta}$) and can be left out.

A cell-centered finite volume scheme is used for the space discretization. The diffusive flux is discretized with a centered approximation with harmonic averages for the diffusion coefficient and for the advective flux an upwind approximation is used. In the time discretization the diffusive terms are treated implicitly and the advective terms are handled explicitly. For the advective term a splitting scheme with sub-time steps is used.

4.7 Suggestion for a Benchmark 2.0

Up to now in both inflows the same species come in. But it is an interesting case when in both inflow zones different species come in, because then the mixing of the species due to transversal dispersion is crucial for the question which reactions proceed. And the effective transversal dispersion depends strongly on the numerical method used to solve the PDEs (see [BK04]). So large differences in the results of the different softwares can be expected.

To do so the species X_3 is replaced by the two species X_{3a}, X_{3b}. In the chemical reaction forming C_2 the species X_3 is replaces by X_{3a}, in the reactions forming C_4 and C_5, respectively, X_3 is replaced by X_{3b} and in the reaction forming CS_1 the species X_3 is replaced by $X_{3a} + X_{3b}$. So the species CS_1 can only be formed in that regions where both species X_{3a} and X_{3b} are present. Altogether there are the chemical reactions

$$\begin{aligned}
C_1 + X_2 &\longleftrightarrow && K = 10^{-12}\\
C_2 &\longleftrightarrow X_2 + X_{3a} && K = 1\\
C_3 + X_2 &\longleftrightarrow X_4 && K = 1\\
C_4 + 4X_2 &\longleftrightarrow X_{3b} + 3X_4 && K = 0.1\\
C_5 &\longleftrightarrow 4X_2 + 3X_{3b} + X_4 && K = 10^{35}\\
CS_1 &\longleftrightarrow 3X_2 + X_{3a} + X_{3b} + S && K = 10^{6}\\
CS_2 + 3X_2 &\longleftrightarrow X_4 + 2S && K = 0.1\,.
\end{aligned}$$

In the modified scenario the transversal dispersion factor is doubled compared to the advective test case

$$\beta_{t,A} = 0.002\,, \qquad \beta_{t,B} = 0.012$$

so that there is a higher mixing of the species. The layer with the low permeability is ranged only over 60% of the height of the domain. Otherwise in the high velocity zone everything gets mixed and the desired effect can not be seen. The water flow of the modified scenario is plotted in Fig. 4.8.

The main difference to the original formulation is that in the two inflow zones different species come in. At the inflow 1 (at the top of the domain) the species X_{3b} comes in but not the species X_{3a} and at the inflow 2 (at the left side) it is vice versa, the species X_{3a} comes in but not the species X_{3b}. Altogether the boundary

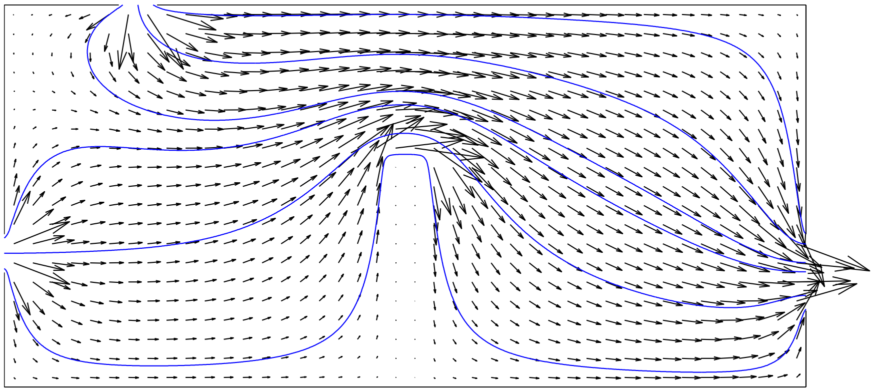


Figure 4.8: Water flow of the modified scenario

conditions for the primary mobile species are at inflow 1

$$
\begin{aligned}
\mathrm{X}_1 &= 0.3\\
\mathrm{X}_2 &= 0.2416198487\\
\mathrm{X}_{3\mathrm{a}} &= 0\\
\mathrm{X}_{3\mathrm{b}} &= 0.2416198487\\
\mathrm{X}_4 &= 0
\end{aligned}
$$

and at inflow 2

$$
\begin{aligned}
\mathrm{X}_1 &= 0.3\\
\mathrm{X}_2 &= 0.2416198487\\
\mathrm{X}_{3\mathrm{a}} &= 0.2416198487\\
\mathrm{X}_{3\mathrm{b}} &= 0\\
\mathrm{X}_4 &= 0\,.
\end{aligned}
$$

The end time of this modified scenario is $T = 2500$ and there is no change in the boundary conditions. All other parameters and the initial conditions are the same as in the 2D “advective easy test case”. Plots of all concentrations are given at the times $t = 50$, $t = 750$, $t = 1300$, $t = 2500$. The computations were done with the implementation of the reduction scheme based on the new version of M++ and exponential upwinding is used in the FV stabilization.

In the left column of Fig. 4.9 the concentration profiles of the species CS_1, which can only be formed in that regions where the species $\mathrm{X}_{3\mathrm{a}}$ and $\mathrm{X}_{3\mathrm{b}}$ are mixed due to transversal dispersion, are given at the times $t = 750$, $t = 1300$, $t = 2500$. The plots are generated out of the results of the computation with 107520 cells.

The concentration profiles of all species and some transformed variables can be found in Appendix B.4. The CPU time, the number of time steps and the number of global Newton steps for the modified scenario can be found in Table 4.5.

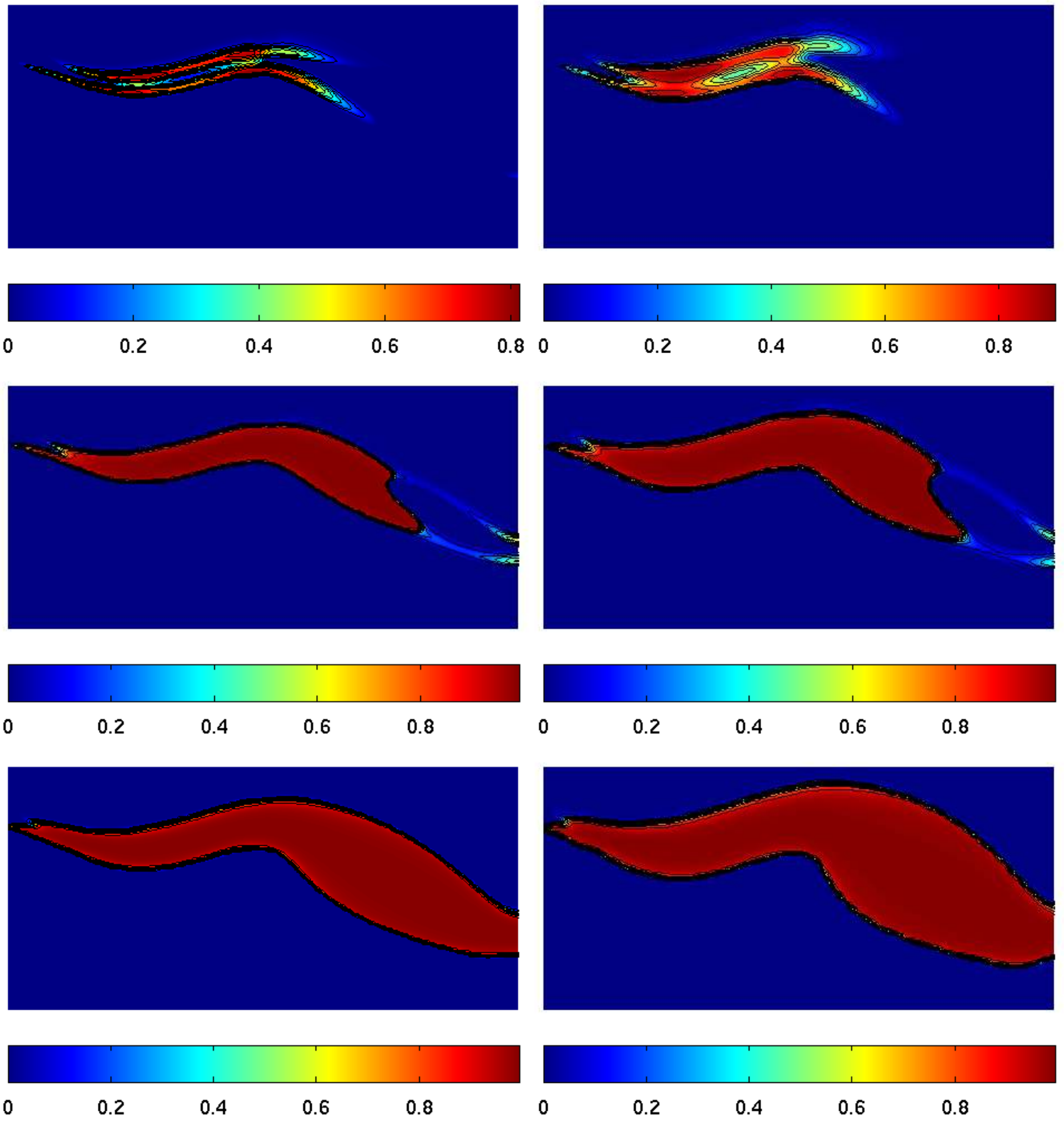


Figure 4.9: The species CS_1 at the times $t = 750$, $t = 1300$, $t = 2500$ (left column: exponential upwinding, grid with 107520 cells; right column: full upwinding, grid with 26880 cells)

To show that there are differences in the results depending on the used numerical method and the number of cells, a simulation with full upwinding (instead

cells	CPU time	time steps	Newton steps
26880	3115.9	5108	2.17
107520	23539.8	6790	3.32

Table 4.5: CPU time, time steps, Newton steps for the modified scenario

of exponential upwinding) and 26880 cells (instead of 107520 cells) was carried out. It is known that full upwinding induces more numerical diffusion than exponential upwinding does and that a coarser grid leads also to more numerical diffusion. The concentration profiles of the species CS_1 at the times $t = 750$, $t = 1300$, $t = 2500$ can be found in the right column of Fig. 4.9. Comparing the right and the left column of Fig. 4.9 some differences can clearly be seen. In the plot at the time $t = 750$ using exponential upwinding there are two regions where the species is presented, that are nearly disconnected, while using full upwinding the two regions are connected. In the plot at $t = 2500$ the plume is much wider when full upwinding instead of exponential upwinding is used.

Chapter 5

Kinetic Mineral Reactions

In this section we consider a model problem with two mobile species A, B, one immobile species $\overline{\mathrm{C}}$ and one kinetic mineral reaction

$$n\mathrm{A} + m\mathrm{B} \leftrightarrow \overline{\mathrm{C}}\,.$$

For the mobile species we get by mass balance the partial differential equations

$$\begin{aligned}\partial_t(\theta c_1) + Lc_1 &= -n\rho\partial_t\bar{c}\\ \partial_t(\theta c_2) + Lc_2 &= -m\rho\partial_t\bar{c}\,.\end{aligned}$$

We assume that the precipitation rate r_p is given by *law of mass action* with ideal activity coefficients

$$r_p(c_1, c_2) = k_p c_1^n c_2^m$$

and that the dissolution rate r_d is a constant k_d if the mineral is present

$$r_d = k_d, \text{ if } \bar{c} > 0\,.$$

For the equation describing the concentration of the mineral and hence the reaction kinetics different formulations are used.

Formulation with set-valued rate function

In [KvDH95], [vDK97], [vDKS98], [vDP04] a formulation with a set-valued rate function is used

$$\begin{aligned}\rho\partial_t\bar{c} &= \theta(r_p(c_1, c_2) - k_d w)\\ w &\in H(\bar{c})\end{aligned} \tag{5.1}$$

where H is the set-valued Heaviside "function"

$$H(u) = \begin{cases}\{1\} & \text{for } u > 0\\ [0,1] & \text{for } u = 0\\ \{0\} & \text{for } u < 0\,.\end{cases}$$

Formulation with complementarity condition

Like in the equilibrium case it is also possible to formulate this problem as a complementarity problem (see [Krä08, chapter 4] for the equilibrium case)

$$\begin{aligned} \bar{c}\,(\rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d)) &= 0 \\ \bar{c} \geq 0,\ \rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d) &\geq 0\,. \end{aligned} \tag{5.2}$$

Formulation with discontinuous rate function

Formulations with discontinuous rate functions are used as well. In [FR92] the following formulation with a case differentiation can be found:

$$\rho\partial_t\bar{c} = \begin{cases} \theta(r_p(c_1,c_2) - k_d) & \text{for } (\bar{c} > 0) \vee (r_p(c_1,c_2) - k_d > 0) \\ 0 & \text{for } (\bar{c} = 0) \wedge (r_p(c_1,c_2) - k_d \leq 0) \end{cases} \tag{5.3}$$

Whereas the formulation in [BEHM07] uses the sign function and the positive and negative part of the rate $F(c_1,c_2) := r_p(c_1,c_2) - k_d$, which is valid when mineral is present. Therefore we define the following notation:

$$x^+ := \max\{0,x\}, \qquad x^- := (-x)^+, \qquad \text{sign}(x) := \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases}$$

It holds $x = x^+ - x^-$. With this notation the equation describing the concentration of the kinetic mineral is

$$\rho\partial_t\bar{c} = \theta(F^+(c_1,c_2) - \text{sign}^+(\bar{c})F^-(c_1,c_2))\,. \tag{5.4}$$

The two formulations with the discontinuous rates (5.3) and (5.4) are identical for $\bar{c} \geq 0$. This can be seen in the following way: For $t \in (0,T)$ such that $\bar{c}(t) > 0$ we have $\text{sign}^+(\bar{c}(t)) = 1$ and so the right hand side of (5.4) becomes $\theta F(c_1(t),c_2(t))$. This coincides with (5.3). For $t \in (0,T)$ such that $\bar{c}(t) = 0$ we have $\text{sign}^+(\bar{c}(t)) = 0$ and so the right hand side of (5.4) becomes $\theta F^+(c_1(t),c_2(t))$, i.e., the right hand side is $\theta F(c_1(t),c_2(t))$ for $F(c_1(t),c_2(t)) > 0$ and 0 for $F(c_1(t),c_2(t)) \leq 0$. This also coincides with (5.3). For $\bar{c} < 0$ (5.3) is not defined.

5.1 Equivalence of the Different Formulations

States of equilibrium

In the following it is always assumed that $r_p(c_1,c_2)$ is nonnegative.

The formulation with the set-valued rate function (5.1) is constructed in such a way that the states of equilibrium are

$$((r_p(c_1,c_2) = k_d) \wedge (\bar{c} > 0)) \vee ((r_p(c_1,c_2) \leq k_d) \wedge (\bar{c} = 0)) \tag{5.5}$$

(see [KvDH95]). Formally the formulation with the set-valued rate function has the additional state of equilibrium $((r_p(c_1,c_2) = 0) \wedge (\bar{c} < 0))$.

The formulation with the complementarity condition (5.2) leads to the states of equilibrium

$$\begin{gathered} \bar{c}\,(r_p(c_1,c_2) - k_d) = 0 \\ \bar{c} \geq 0,\ -(r_p(c_1,c_2) - k_d) \geq 0\,. \end{gathered}$$

It is obvious that these states of equilibrium are the same as (5.5).

And the formulation with the discontinuous rate function (5.4) leads to the states of equilibrium

$$F^+(c_1,c_2) - \text{sign}^+(\bar{c})F^-(c_1,c_2) = 0\,.$$

For $\bar{c} > 0$ it holds $\text{sign}^+(\bar{c}) = 1$ and so we get $F(c_1,c_2) = 0$. With the definition of F it follows $r_p(c_1,c_2) = k_d$. For $\bar{c} = 0$ we have $\text{sign}^+(\bar{c}) = 0$. This yields $F^+(c_1,c_2) = 0$. That is equivalent to $F(c_1,c_2) \leq 0$. Plugging in the definition of F leads to $r_p(c_1,c_2) \leq k_d$. So for nonnegative mineral concentration $\bar{c}$ the states of equilibrium are exactly (5.5). For $\bar{c} < 0$ there are the states of equilibrium $r_p(c_1,c_2) \leq k_d$.

Pointwise considerations

The different formulations for the kinetic mineral problem are not pointwisely equivalent. As a consequence solutions of the formulation with the set-valued rate function are not always solutions of the formulation with the discontinuous rate function. For example, if travelling wave solutions of the formulation with the set-valued rate function with continuous from the right derivatives are considered, like it is done in [vDK97], then these solutions are not solutions of the formulation with the discontinuous rate function.

This can be seen in the following way: A travelling wave solution is a function of the variable $\eta := x - at$ with the wave speed $a > 0$. Let η_d be a point of discontinuity of $\partial_t\bar{c}$ with $\bar{c}(\eta) = 0$ for $\eta \leq \eta_d$ and $r_p(\eta_d) < k_d$. Such points η_d exist in travelling wave solutions (see [KvDH95, Sec. 3] for travelling wave solutions). Because of $\bar{c}(\eta) = 0$ for $\eta \leq \eta_d$ it holds $\partial_t\bar{c}(\eta) = 0$ for $\eta < \eta_d$. As $\partial_t\bar{c}$ is continuous from the right $\partial_t\bar{c}(\eta_d) = \lim_{\eta\searrow\eta_d} \partial_t\bar{c}$. This limit is not zero because we have assumed that $\partial_t \bar{c}$ is discontinuous at η_d. But (5.3) yields $\partial_t \bar{c}(\eta_d) = 0$ because of $\bar{c}(\eta_d) = 0$. So the travelling wave solution of the formulations with the set-valued rate function is not a solution of the formulation with the discontinuous rate function.

If we consider weak solutions we will see that the three formulations are equivalent.

Weak solutions

We assume that the concentrations of the mobile species c_1 and c_2 are given such that $r_p(c_1, c_2) \in L^\infty(0,T)$. In this section we study weak solutions of the three different formulations. A weak solution of the set-valued formulation is a pair of functions $(\bar{c}, w) \in H^1(0,T) \times L^\infty(0,T)$ which fulfills

$$\int_0^T (\rho \partial_t \bar{c} - \theta(r_p(c_1,c_2) - k_d w))\phi \, dt = 0 \qquad \forall \phi \in C_0^\infty(0,T) \tag{5.6}$$

$$w \in H(\bar{c}) \qquad \text{a.e. in } (0,T) \tag{5.7}$$

$$\bar{c}(0) = \bar{c}_0 \, . \tag{5.8}$$

In case of the complementarity formulation $\bar{c} \in H^1(0,T)$ is a weak solution if

$$\bar{c}\,(\rho \partial_t \bar{c} - \theta(r_p(c_1,c_2) - k_d)) = 0 \qquad \text{a.e. in } (0,T) \tag{5.9}$$

$$\bar{c} \geq 0 \qquad \text{in } (0,T) \tag{5.10}$$

$$\rho \partial_t \bar{c} - \theta(r_p(c_1,c_2) - k_d) \geq 0 \qquad \text{a.e. in } (0,T) \tag{5.11}$$

$$\bar{c}(0) = \bar{c}_0 \, . \tag{5.12}$$

And using the formulation with a discontinuous rate function $\bar{c} \in H^1(0,T)$ is a weak solution if

$$\int_0^T (\rho \partial_t \bar{c} - \theta(F^+(c_1,c_2) - \operatorname{sign}^+(\bar{c}) F^-(c_1,c_2)))\phi \, dt = 0 \quad \forall \phi \in C_0^\infty(0,T) \tag{5.13}$$

$$\bar{c}(0) = \bar{c}_0 \, . \tag{5.14}$$

Lemma 5.1. *A weak solution $\bar{c}$ of the formulation with the set-valued rate function (5.6)-(5.8) is nonnegative if the initial value $\bar{c}_0$ is nonnegative.*

Proof. Using

$$\phi(s) = \begin{cases} -\bar{c}^-(s) & \text{for } s \leq t \\ 0 & \text{for } s > t \end{cases}$$

as test function in (5.6) yields

$$\int_0^t \rho\partial_t\bar{c}(-\bar{c}^-) - \theta(\underbrace{r_p(c_1,c_2)}_{\geq 0} - k_d w)(\underbrace{-\bar{c}^-}_{\leq 0})\; ds = 0\,.$$

Because of

$$\int_0^t \partial_t\bar{c}(-\bar{c}^-)\; ds = \frac{1}{2}\int_0^t \partial_t(\bar{c}^-)^2\; ds = \frac{1}{2}(\bar{c}^-(t))^2 - \frac{1}{2}(\bar{c}^-(0))^2$$

we get the estimate

$$\frac{1}{2}\rho(\bar{c}^-(t))^2 \leq \int_0^t \theta k_d w\bar{c}^-\; ds + \frac{1}{2}\rho(\bar{c}^-(0))^2 = 0\,.$$

The first term on the right hand side is zero because one of the factors w, $\bar{c}^-$ is zero a.e. due to (5.7) and the second term is zero due to the assumption that $\bar{c}_0$ is nonnegative. That concludes the proof. □

The proofs of the next two theorems are adapted from [BEHM07, Proposition 3.4] where the equivalence of the formulation with a discontinuous rate function to a formulation similar to a complementarity condition is shown.

Theorem 5.2. *The formulation with the set-valued rate function* (5.6)-(5.8) *and the formulation with the complementarity condition* (5.9)-(5.12) *are equivalent.*

Proof. "⇐": Let $\bar{c} \in H^1(0,T)$ be a weak solution of the complementarity formulation (5.9)-(5.12). First we define the set $A := \{t \in (0,T) |\, \bar{c}(t) = 0\}$ and its complement $\bar{A} := \{t \in (0,T) |\, \bar{c}(t) > 0\}$. We set

$$w = \begin{cases} \frac{1}{k_d} r_p(c_1,c_2) & \text{on } A \\ 1 & \text{on } \bar{A}\,. \end{cases}$$

Because of (5.10) we can split the integral in (5.6) in an integral over A and an integral over $\bar{A}$:

$$\begin{aligned}&\int_0^T (\rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d w))\phi\; dt \\ &= \int_A (\rho\partial_t\bar{c} - \theta(\underbrace{r_p(c_1,c_2) - r_p(c_1,c_2)}_{=0}))\phi\; dt + \int_{\bar{A}} (\rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d))\phi\; dt\end{aligned}$$

By Stampacchia's theorem, on A we have $\partial_t\bar{c} = 0$ and so the first integral vanishes. Because of the complementarity condition (5.9) $\rho\partial_t\bar{c} - \theta(k_p r(c_1,c_2) - k_d)$ is zero a.e. on $\bar{A}$ and so the second integral vanishes, too. That proofs (5.6).

Using again that $\partial_t \bar{c} = 0$ on A by Stampacchia's theorem it follows from (5.11) that

$$\begin{aligned} 0 - \theta(r_p(c_1,c_2) - k_d) &\geq 0 \qquad \text{a.e. on } A \\ \Leftrightarrow \frac{1}{k_d} r_p(c_1,c_2) &\leq 1 \qquad \text{a.e. on } A\,. \end{aligned}$$

That proofs (5.7).

"⇒": Let $(\bar{c}, w) \in H^1(0,T) \times L^\infty(0,T)$ be a weak solution of the formulation with the set-valued rate function (5.6)-(5.8). According to Lemma 5.1 $\bar{c}$ is nonnegative. So the inequality (5.10) is valid. From (5.6) it follows

$$\begin{aligned} \rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d w) &= 0 && \text{a.e. in } (0,T) \\ \Leftrightarrow \rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d) &= \theta k_d(\underbrace{1-w}_{\geq 0}) && \text{a.e. in } (0,T)\,. \end{aligned}$$

$1 - w$ is nonnegative a.e. due to (5.7). That proofs (5.11). Furthermore we get with the relation above

$$\bar{c}\,(\rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d)) = \bar{c}\,\theta k_d(1-w) = 0 \qquad \text{a.e. in } (0,T)\,.$$

The product on the right hand side is zero a.e. because $\bar{c}$ is nonnegative and so one of the factors $\bar{c}$, $1 - w$ is zero a.e. due to (5.7). This proofs (5.9). □

Theorem 5.3. *The formulation with the discontinuous rate function* (5.13)-(5.14) *and the formulation with the complementarity condition* (5.9)-(5.12) *are equivalent.*

Proof. "⇐": Let $\bar{c} \in H^1(0,T)$ be a weak solution of the complementarity formulation (5.9)-(5.12). First we define the set $A := \{t \in (0,T) |\, \bar{c}(t) = 0\}$ and its complement $\bar{A} := \{t \in (0,T) |\, \bar{c}(t) > 0\}$. It holds

$$F^+(c_1,c_2) - \operatorname{sign}^+(\bar{c})F^-(c_1,c_2) = \begin{cases} F^+(c_1,c_2) & \text{on } A \\ F(c_1,c_2) & \text{on } \bar{A}\,. \end{cases}$$

Because of (5.10) we can split the integral in (5.13) in an integral over A and an integral over $\bar{A}$:

$$\begin{aligned} &\int_0^T (\rho\partial_t\bar{c} - \theta(F^+(c_1,c_2) - \operatorname{sign}^+(\bar{c})F^-(c_1,c_2)))\phi\, dt \\ &= \int_A (\rho\partial_t\bar{c} - \theta F^+(c_1,c_2))\phi\, dt + \int_{\bar{A}} (\rho\partial_t\bar{c} - \theta(r_p(c_1,c_2) - k_d))\phi\, dt \qquad (5.15) \end{aligned}$$

Because of the complementarity condition (5.9) $\rho\partial_t\bar{c} - \theta(r_p(c_1, c_2) - k_d)$ is zero a.e. in $\bar{A}$ and so the second integral vanishes.

By Stampacchia's theorem, on A we have $\partial_t\bar{c} = 0$. Using this it follows from (5.11) that

$$\begin{aligned} 0 - \theta F(c_1, c_2) &\geq 0 && \text{a.e. on } A \\ \Leftrightarrow F(c_1, c_2) &\leq 0 && \text{a.e. on } A \\ \Rightarrow F^+(c_1, c_2) &= 0 && \text{a.e. on } A\,. \end{aligned}$$

So the first integral in (5.15) vanishes, too. That proofs (5.13).

"⇒": Let $\bar{c}$ be a weak solution of the formulation with the discontinuous rate function (5.13)-(5.14). If for $t \in (0, T)$ it holds $\bar{c}(t) \leq 0$ then it follows that a.e.

$$\begin{aligned} \rho\partial_t\bar{c}(t) &= \theta(F^+(c_1(t), c_2(t)) - \underbrace{\text{sign}^+(\bar{c}(t))}_{=0} F^-(c_1(t), c_2(t))) \\ &= \theta F^+(c_1(t), c_2(t)) \geq 0\,. \end{aligned}$$

If $\bar{c}_0 \geq 0$ it follows that $\bar{c}$ is nonnegative. That proofs (5.10).

Due to (5.13) and the definition of F we have a.e. in $(0, T)$

$$\begin{aligned} &\rho\partial_t\bar{c} - \theta(r_p(c_1, c_2) - k_d) \\ &\quad = \theta(F^+(c_1, c_2) - \text{sign}^+(\bar{c})F^-(c_1, c_2)) - \theta(F^+(c_1, c_2) - F^-(c_1, c_2)) \\ &\quad = \theta(\underbrace{1 - \text{sign}^+(\bar{c})}_{\geq 0})\underbrace{F^-(c_1, c_2)}_{\geq 0} \geq 0\,. \end{aligned}$$

That proofs (5.11). Furthermore using this identity we get

$$\int_0^T \bar{c}\,(\rho\partial_t\bar{c} - \theta(r_p(c_1, c_2) - k_d))\phi\;dt = \int_0^T \bar{c}\,\theta(1 - \text{sign}^+(\bar{c}))F^-(c_1, c_2)\phi\;dt\,.$$

One of the factors $\bar{c}$, $1 - \text{sign}^+(\bar{c})$ is always zero because for $t \in (0, T)$ such that $\bar{c}(t) > 0$ it holds $\text{sign}^+(\bar{c}(t)) = 1$. So the integral on the right hand side vanishes. That proofs (5.9). □

5.2 Algorithmic Examination of the Formulations

In this section we discretize the equations (5.2), (5.3), (5.4) with the implicit Euler method and solve the resulting nonlinear equation with Newton's method. Again we assume that the mobile concentrations are given.

Complementarity formulation

It is well known that a complementarity condition can be replace by a equivalent equation (see e.g. [Kan04]). So the complementarity formulation (5.2) is equivalent to

$$\min\{\bar{c},\, \rho\partial_t\bar{c} - \theta(r_p(c_1,c_2)-k_d)\} = 0\,.$$

Discretization with the implicit Euler method with constant time step size Δt leads to

$$\min\left\{\bar{c}(t^{n+1}),\, \rho\frac{\bar{c}(t^{n+1})-\bar{c}(t^n)}{\Delta t} - \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d)\right\} = 0\,.$$

The upper index denotes the number of the time step. This equation is equivalent to

$$\min\left\{\rho\frac{\bar{c}(t^{n+1})}{\Delta t},\, \rho\frac{\bar{c}(t^{n+1})-\bar{c}(t^n)}{\Delta t} - \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d)\right\} = 0\,. \tag{5.16}$$

If $\rho\frac{\bar{c}(t^n)}{\Delta t} + \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d) < 0$ the minimum is attained in the first argument and vice versa. Graphically this means it is not possible that more mineral is dissolved than it is present. Note the decision if the minimum is attained in the first or in the second argument is independent of $\bar{c}(t^{n+1})$. So we can treat the two cases separately.

Equation (5.16) is solved with Newton's method. First we treat the case $\rho\frac{\bar{c}(t^n)}{\Delta t} + \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d) < 0$. In this case we have to solve the linear equation

$$\rho\frac{\bar{c}(t^{n+1})}{\Delta t} = 0\,.$$

So after one Newton step we get the solution $\bar{c}(t^{n+1}) = 0$. In the case $\rho\frac{\bar{c}(t^n)}{\Delta t} + \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d) > 0$ we also have to solve a equation which is linear in $\bar{c}(t^{n+1})$. This time the equation reads

$$\rho\frac{\bar{c}(t^{n+1})-\bar{c}(t^n)}{\Delta t} - \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d) = 0\,.$$

So again after one Newton step we get the solution

$$\bar{c}(t^{n+1}) = \bar{c}(t^n) + \frac{\Delta t}{\rho}\theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d)\,.$$

As we are in the case $\rho\frac{\bar{c}(t^n)}{\Delta t} + \theta(r_p(c_1(t^{n+1}),c_2(t^{n+1}))-k_d) > 0$ it is ensured that $\bar{c}(t^{n+1})$ is nonnegative.

In the very unlikely case that $\rho\frac{\bar{c}(t^n)}{\Delta t} + \theta(r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d)$ is exactly zero it does not matter which one of the two cases is used. Both cases lead to the solution $\bar{c}(t^{n+1}) = 0$.

In summary only one Newton step is always needed and it is ensured that the discrete solution is nonnegative. So the complementarity formulation is a proper formulation to solve a kinetic mineral problem numerically.

Formulation with discontinuous rate function

A discretization of (5.3) is

$$\rho\frac{\bar{c}(t^{n+1}) - \bar{c}(t^n)}{\Delta t}$$
$$= \begin{cases} \theta(r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d) & \text{for } (\bar{c}(t^{n+1}) > 0) \vee (r_p(c_1(t^{n+1}), c_2(t^{n+1})) > k_d) \\ 0 & \text{for } (\bar{c}(t^{n+1}) \leq 0) \wedge (r_p(c_1(t^{n+1}), c_2(t^{n+1})) \leq k_d) \end{cases}$$

which is again obtained by the implicit Euler method. This equation should be solved with Newton's method.

Let $\bar{c}_k^{n+1}$ denote the k-th Newton iterate. If we are in the case $(\bar{c}_k^{n+1} \leq 0) \wedge (r_p(c_1(t^{n+1}), c_2(t^{n+1})) \leq k_d)$ we get the linear system

$$\frac{\rho}{\Delta t}(\bar{c}_{k+1}^{n+1} - \bar{c}_k^{n+1}) = -\rho\frac{\bar{c}_k^{n+1} - \bar{c}^n}{\Delta t}\,.$$

This leads to

$$\bar{c}_{k+1}^{n+1} = \bar{c}^n \tag{5.17}$$

In the case $(\bar{c}_k^{n+1} > 0) \vee (r_p(c_1(t^{n+1}), c_2(t^{n+1})) > k_d)$ applying Newton's method yields

$$\frac{\rho}{\Delta t}(\bar{c}_{k+1}^{n+1} - \bar{c}_k^{n+1}) = -\rho\frac{\bar{c}_k^{n+1} - \bar{c}^n}{\Delta t} + \theta(r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d)\,.$$

So we get

$$\bar{c}_{k+1}^{n+1} = \bar{c}^n + \frac{\Delta t}{\rho}\theta(r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d)\,. \tag{5.18}$$

It can happen that $\bar{c}_{k+1}^{n+1}$ is negative because we are always in this case when $\bar{c}_k^{n+1}$ is positive independent of the sign of the right hand side in (5.18).

If it happens that the Newton iterate $\bar{c}_1^{n+1}$ resulting from (5.18) is negative then the next Newton iterate $\bar{c}_2^{n+1}$ is computed according to (5.17) and so is equal to the starting value. This has the consequence that $\bar{c}_3^{n+1}$ is equal to $\bar{c}_1^{n+1}$. So the Newton's method is in an infinite loop and will not converge. Hence, this formulation can not be used for numerical computations.

One possibility to circumvent the problem with the infinite loop is to replace $\bar{c}(t^{n+1})$ in the case distinction by $\bar{c}(t^n)$. Then the decision which case is used is independent of the Newton iterate. But in the case $(\bar{c}(t^n) > 0) \vee (r_p(c_1(t^{n+1}), c_2(t^{n+1})) > k_d)$ it is still possible that the mineral concentration gets negative.

In [BEHM07] the formulation (5.4) with a discontinuous rate function is used on the continuous level. But in the discretized form no discretization of the discontinuous rate function appears. Instead of that the formulation

$$\bar{c}(t^{n+1}) = \left(\bar{c}(t^n) + \frac{\Delta t}{\rho}\theta(r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d)\right)^+$$

is used. This formulation is equivalent to the discretization (5.16) of the complementarity formulation. So all considerations of the complementarity formulation are applicable.

Formulation with set-valued rate function

In [DPvDC08] for numerical computations on the pore scale a formulation with a regularized Heaviside function is used. As argument of the regularized Heaviside function the mineral concentration at the old time level is used. The discrete formulation reads

$$\rho\frac{\bar{c}(t^{n+1}) - \bar{c}(t^n)}{\Delta t} = r_p(c_1(t^{n+1}), c_2(t^{n+1})) - k_d H_\delta(\bar{c}(t^n))$$

with

$$H_\delta(v) = \begin{cases} 0 & \text{for } v \leq 0 \\ v/\delta & \text{for } v \in (0, \delta) \\ 1 & \text{for } v \geq \delta\,. \end{cases}$$

In the pore scale model the equation for the kinetic mineral is valid on the moving boundary. So computations on the pore scale are not comparable to computations with the model of this work.

Applied to our problem this formulation has the disadvantage that in the undersaturated case $(r_p(c_1, c_2) < k_d)$ $\bar{c} \equiv 0$ is no solution of the discrete formulation. Furthermore is not secured that the mineral concentration $\bar{c}$ is always nonnegative. For our problem the regularization is more useful for theoretical considerations and will be used in Section 5.4.

5.3 Implementation with the Reduction Scheme

In the reaction rate we introduce the additional parameter k

$$k(r_p(c_1, c_2) - k_d)\,.$$

The objective is to get an implementation with which it is possible to solve the kinetic mineral problem also for high values of k. For high values of k we expect that we will get results very similar to the equilibrium case. So for $k \to \infty$ the implementation of the kinetic mineral problem should be as similar as possible to the implementation of the equilibrium case (see Chap. 3 for the implementation of the equilibrium case).

Starting point is the formulation with the complementarity condition. For the implementation we assume that the unit of the mineral concentrations is the same as the unit of the mobile concentrations (like in Chap. 3). So the system of equations reads

$$\begin{aligned}\partial_t(\theta c_1) + Lc_1 &= -n\partial_t(\theta\bar{c})\\ \partial_t(\theta c_2) + Lc_2 &= -m\partial_t(\theta\bar{c})\\ \bar{c}\,(\partial_t(\theta\bar{c}) - \theta k(r_p(c_1, c_2) - k_d)) &= 0\\ \bar{c} \geq 0,\ \partial_t(\theta\bar{c}) - \theta k(r_p(c_1, c_2) - k_d) &\geq 0\,.\end{aligned} \tag{5.19}$$

First the complementarity condition is replaced by an equivalent equation with the minimum function

$$\min\{\partial_t(\theta\bar{c}) - \theta k(k_p c_1^n c_2^m - k_d), \bar{c}\} = 0\,.$$

To apply the reduction scheme we need the matrices

$$\boldsymbol{S}_1 = \begin{pmatrix}-n\\-m\end{pmatrix}, \qquad \boldsymbol{S}_1^\perp = \frac{1}{n+m}\begin{pmatrix}m\\-n\end{pmatrix}, \qquad \boldsymbol{B}_1 = \boldsymbol{S}_1, \qquad \boldsymbol{B}_1^\perp = \boldsymbol{S}_1^\perp\,.$$

With the definitions of the transformed variables (3.13) and (3.39) we get

$$\begin{aligned}\eta &= \frac{n+m}{n^2+m^2}(mc_1 - nc_2)\\ \xi_{min} &= \frac{1}{n^2+m^2}(-nc_1 - mc_2)\\ \bar{\xi}_{min} &= \bar{c}\\ \tilde{\xi}_{min} &= \frac{1}{n^2+m^2}(-nc_1 - mc_2) - \bar{c}\,.\end{aligned} \tag{5.20}$$

The associated retransformation is

$$c_1 = -n(\tilde{\xi}_{min} + \bar{\xi}_{min}) + \frac{m}{n+m}\eta \tag{5.21}$$

$$c_2 = -m(\tilde{\xi}_{min} + \bar{\xi}_{min}) - \frac{n}{n+m}\eta \tag{5.22}$$

$$\bar{c} = \bar{\xi}_{min} \,. \tag{5.23}$$

Hence, applying the reduction scheme we get the system of equations

$$\begin{aligned} \partial_t(\theta\eta) + L\eta &= 0 \\ \tilde{\xi}_{min} &= \xi_{min} - \bar{\xi}_{min} \\ \partial_t(\theta\tilde{\xi}_{min}) + L\xi_{min} &= 0 \\ \min\left\{\partial_t(\theta\bar{c}) - \theta k(k_p c_1^n c_2^m - k_d), \bar{c}\right\} &= 0 \,. \end{aligned}$$

Then we define

$$k_{inv} := 1/k \,.$$

Now the equilibrium case corresponds to $k_{inv} = 0$. With this definition the time discrete problem reads

$$\frac{\theta\eta - (\theta\eta)_{old}}{\Delta t} + L\eta = 0 \tag{5.24}$$

$$\tilde{\xi}_{min} = \xi_{min} - \bar{\xi}_{min} \tag{5.25}$$

$$\frac{\theta\tilde{\xi}_{min} - (\theta\tilde{\xi}_{min})_{old}}{\Delta t} + L\xi_{min} = 0 \tag{5.26}$$

$$\min\left\{k_{inv}\frac{\theta\bar{c} - (\theta\bar{c})_{old}}{\Delta t} - \theta(k_p c_1^n c_2^m - k_d), \frac{\bar{c}}{\Delta t}\right\} = 0 \,. \tag{5.27}$$

In the next step a resolution function $\bar{\xi}_{min}(\tilde{\xi}_{min})$ defined by (5.27) and the retransformation (5.21)-(5.23) is needed. To show the existence of such a resolution function we have to differentiate (5.27) with respect to $\tilde{\xi}_{min}$ after plugging in (5.21)-(5.23). If the minimum in (5.27) is attained in the first argument we get

$$\begin{aligned} &\frac{\theta k_{inv}}{\Delta t}\bar{\xi}'_{min}(\tilde{\xi}_{min}) - \theta\left(nk_p c_1^{n-1} c_2^m \quad mk_p c_1^n c_2^{m-1}\right) \boldsymbol{S}_{1,min}(1 + \bar{\xi}'_{min}(\tilde{\xi}_{min})) = 0 \\ &\Leftrightarrow \left(\frac{\theta k_{inv}}{\Delta t} + \theta n^2 k_p c_1^{n-1} c_2^m + \theta m^2 k_p c_1^n c_2^{m-1}\right) \bar{\xi}'_{min}(\tilde{\xi}_{min}) \\ &\qquad + \theta n^2 k_p c_1^{n-1} c_2^m + \theta m^2 k_p c_1^n c_2^{m-1} = 0 \,. \end{aligned}$$

The factor in front of $\bar{\xi}'_{min}(\tilde{\xi}_{min})$ is always positive. So with the implicit function theorem the resolution function exists. If the minimum in (5.27) is attained in the second argument we get

$$\frac{1}{\Delta t}\bar{\xi}'_{min}(\tilde{\xi}_{min}) = 0\,.$$

Again the factor in front of $\bar{\xi}'_{min}(\tilde{\xi}_{min})$ is always positive and so also in this case the resolution function exists.

Using the resolution function $\bar{\xi}_{min}(\tilde{\xi}_{min})$ we have to solve the global problem

$$\begin{aligned}\tilde{\xi}_{min} - \xi_{min} + \bar{\xi}_{min}(\tilde{\xi}_{min}) &= 0\\ \frac{\theta\tilde{\xi}_{min} - (\theta\tilde{\xi}_{min})_{old}}{\Delta t} + L\xi_{min} &= 0\,.\end{aligned}$$

To solve the local problem the concentrations are used as variables instead of $\bar{\xi}_{min}$ and the defining equations of $\tilde{\xi}_{min}$ and η are added as additional equations. This is the same approach as in the equilibrium case (see Sec. 3.4.1). Then the local problem reads

$$\begin{aligned}\eta - \frac{n+m}{n^2+m^2}(mc_1 - nc_2) &= 0\\ \tilde{\xi}_{min} - \frac{1}{n^2+m^2}(-nc_1 - mc_2) + \bar{c} &= 0\\ \min\left\{k_{inv}\frac{\theta\bar{c} - (\theta\bar{c})_{old}}{\Delta t} - \theta(k_p c_1^n c_2^m - k_d), \frac{\bar{c}}{\Delta t}\right\} &= 0\,.\end{aligned}$$

Contrary to the equilibrium case it is not possible to calculate the mineral concentration $\bar{c}$ a posteriori because the mineral concentration appears not only in the defining equation of $\tilde{\xi}_{min}$ and the second argument of the minimum function but also in the first argument of the minimum function. Furthermore it is not possible to take the logarithm of the first argument of the minimum function because of the additional summand with the different quotient of $\bar{c}$.

But like in the equilibrium case one can use the logarithms of the mobile concentrations l_1, l_2 as unknowns instead of the concentrations c_1, c_2:

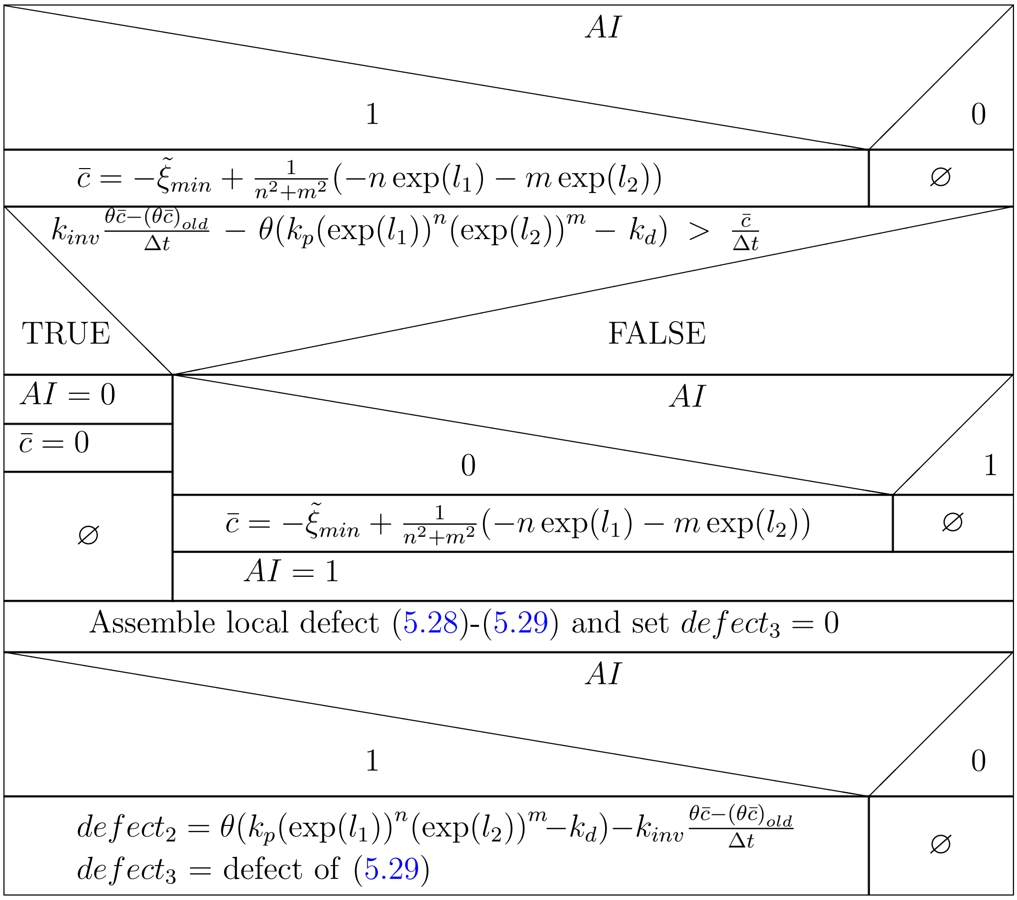
$$\eta - \frac{n+m}{n^2+m^2}(m\exp(l_1) - n\exp(l_2)) = 0 \tag{5.28}$$

$$\tilde{\xi}_{min} - \frac{1}{n^2+m^2}(-n\exp(l_1) - m\exp(l_2)) + \bar{c} = 0 \tag{5.29}$$

$$\min\left\{k_{inv}\frac{\theta\bar{c} - (\theta\bar{c})_{old}}{\Delta t} - \theta(k_p(\exp(l_1))^n(\exp(l_2))^m - k_d), \frac{\bar{c}}{\Delta t}\right\} = 0 \tag{5.30}$$

To keep the method as similar as possible to the equilibrium case the following algorithm is used to compute the local defect (compare Sec. 3.4.1 for solving the local problem in the equilibrium case and the notation used in the following structured chart):

Exact algorithm for calculating the local defect



```
AI
  1: $\bar{c} = -\tilde{\xi}_{min} + \frac{1}{n^2+m^2}(-n\exp(l_1) - m\exp(l_2))$
  0: $\varnothing$
$k_{inv}\frac{\theta\bar{c}-(\theta\bar{c})_{old}}{\Delta t} - \theta(k_p(\exp(l_1))^n(\exp(l_2))^m - k_d) > \frac{\bar{c}}{\Delta t}$
  TRUE:
    $AI = 0$
    $\bar{c} = 0$
    $\varnothing$
  FALSE:
    AI
      0: $\bar{c} = -\tilde{\xi}_{min} + \frac{1}{n^2+m^2}(-n\exp(l_1) - m\exp(l_2))$
      1: $\varnothing$
    $AI = 1$
Assemble local defect (5.28)-(5.29) and set $defect_3 = 0$
AI
  1: $defect_2 = \theta(k_p(\exp(l_1))^n(\exp(l_2))^m - k_d) - k_{inv}\frac{\theta\bar{c}-(\theta\bar{c})_{old}}{\Delta t}$
     $defect_3$ = defect of (5.29)
  0: $\varnothing$
```

To assemble the Jacobian matrix of the global problem we need the derivative $\bar{\xi}'_{min}(\tilde{\xi}_{min})$. If the minimum of (5.27) is attained in the first argument we can compute this derivative by plugging (5.21)-(5.23) in (5.27) and differentiating with respect to $\tilde{\xi}_{min}$. This results in the linear system

$$\frac{\theta k_{inv}}{\Delta t}\bar{\xi}'_{min}(\tilde{\xi}_{min}) - \theta\left(nk_p c_1^{n-1} c_2^m \quad mk_d c_1^n c_2^{m-1}\right) \boldsymbol{S}_{1,min}(1 + \bar{\xi}'_{min}(\tilde{\xi}_{min})) = 0\,.$$

Alternatively we can write (5.27) in the form $k_{inv}\frac{\theta\bar{c}-(\theta\bar{c})_{old}}{\Delta t} + \theta k_d = \theta k_p c_1^n c_2^m$, take the logarithm on both sides and then differentiate. This leads to the linear system

$$\frac{1}{k_{inv}\frac{\theta\bar{c}-(\theta\bar{c})_{old}}{\Delta t} + \theta k_d}\frac{\theta k_{inv}}{\Delta t}\bar{\xi}'_{min}(\tilde{\xi}_{min}) + \boldsymbol{S}^T_{1,min}\begin{pmatrix} 1/c_1 & 0 \\ 0 & 1/c_2 \end{pmatrix}\boldsymbol{S}_{1,min}(1+\bar{\xi}'_{min}(\tilde{\xi}_{min})) = 0.$$

For $k_{inv} = 0$ this linear system is the same as in the equilibrium case while the first linear system results in a different one.

A problem which only occurs in the kinetic case is that the local problem do not have a nonnegative solution for all $\tilde{\xi}_{min} \le 0$. This can be seen in the following way. With the equation for the kinetic mineral (5.27) we get

$$\begin{aligned} k_{inv}\frac{\theta\bar{c}-(\theta\bar{c})_{old}}{\Delta t} &= \theta(k_p c_1^n c_2^m - k_d) \geq -\theta k_d \\ \Leftrightarrow \bar{c} &\geq -\frac{\Delta t}{k_{inv}}k_d + \frac{(\theta\bar{c})_{old}}{\theta}\,. \end{aligned} \tag{5.31}$$

Using the retransformation (5.21) and the fact that c_1 is nonnegative we get

$$\frac{m}{n(n+m)}\eta \geq (\tilde{\xi}_{min} + \xi_{min})\,.$$

Plugging this in the retransformation (5.22) leads to

$$\begin{aligned} c_2 &\geq -\frac{m^2}{n(n+m)}\eta - \frac{n}{n+m}\eta \\ &\geq -\frac{m^2+n^2}{n(n+m)}\eta \\ \Rightarrow c_2 &\geq \frac{m^2+n^2}{n(n+m)}\eta^-\,. \end{aligned} \tag{5.32}$$

The last conclusion is true because c_2 is nonnegative.

Analogously we can calculate that

$$c_1 \geq \frac{n^2+m^2}{m(n+m)}\eta^+\,. \tag{5.33}$$

With the definition of $\tilde{\xi}_{min}$ (5.20) and the previous estimates (5.31), (5.32), (5.33) and $\bar{c} \geq 0$ we get

$$\begin{aligned} \tilde{\xi}_{min} &= \frac{1}{n^2+m^2}(-nc_1 - mc_2) - \bar{c} \\ &\leq -\frac{n}{m(n+m)}\eta^+ - \frac{m}{n(n+m)}\eta^- + \min\left\{\frac{\Delta t}{k_{inv}}k_d - \frac{(\theta\bar{c})_{old}}{\theta}, 0\right\}\,. \end{aligned}$$

So if $\tilde{\xi}_{min}$ exceeds this bound we have to cut $\tilde{\xi}_{min}$ because otherwise there is no solution with nonnegative mobile concentrations of the local problem. This bound is strict. It is easy to see that for $\tilde{\xi}_{min}$ equal to this bound $c_1 = \frac{n^2+m^2}{m(n+m)}\eta^+$, $c_2 = \frac{m^2+n^2}{n(n+m)}\eta^-$ and $\bar{c} = \max\big\{-\frac{\Delta t}{k_{inv}}k_d + \frac{(\theta\bar{c})_{old}}{\theta}, 0\big\}$ is a nonnegative solution of the local problem.

To check the implementation computations with an easy example are carried out. The example is taken from [BEHM07, Sec. 7.1]. It is a 1D problem on the domain $\Omega = (0,1)$. The needed parameters are $k_p = 100$, $k_d = 1$, $n = m = 1$, $D = 1$, $q = 0$, $\theta = 1$ and the initial values are $c_{1,0} = c_{2,0} = 0.1$, $\bar{c}_0 = 5$. The boundary conditions are Dirichlet boundary conditions at $x = 0$ with the boundary values $c_{1,D} = c_{2,D} = 0$ and homogeneous Neumann boundary conditions at $x = 1$.

For the equilibrium case ($k_{inv} = 0$) the exact solution is known. It is

$$c_1(x,t) = c_2(x,t) = \begin{cases} \sqrt{\frac{k_d}{k_p}}\,\mathrm{erf}(x/(2\sqrt{t}))(\mathrm{erf}(\zeta_0/2))^{-1} & \text{for } 0 < x < \zeta_0\sqrt{t} \\ \sqrt{\frac{k_d}{k_p}} & \text{for } \zeta_0\sqrt{t} < x \end{cases}$$

$$\bar{c}(x,t) = \begin{cases} 0 & \text{for } 0 < x < \zeta_0\sqrt{t} \\ \bar{c}_0 & \text{for } \zeta_0\sqrt{t} < x \end{cases}$$

with ζ_0 the positive solution of

$$\sqrt{\frac{k_d}{k_p}} = \zeta_0\bar{c}_0 \exp(\zeta_0^2/4)\frac{\sqrt{\pi}}{2}\,\mathrm{erf}(\zeta_0/2)$$

($\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,dt$). One can compute numerically that $\zeta_0 \approx 0.1993$. In Fig. 5.1 the mineral concentration $\bar{c}$ is plotted as a function of time and space for different values of k_{inv}. For the smallest value $k_{inv} = 10^{-4}$ the solution looks like the exact solution of the equilibrium case. That is the expected behavior.

Travelling Waves

As a second test of the implementation computations of travelling waves in one space dimension are conducted. A travelling wave solution can be written as a function of one variable $\eta := x - at$ with the wave speed a. So the partial differential equations in (5.19) become ordinary differential equations

$$\begin{aligned} -a(\theta c_1)' - Dc_1'' + qc_1' &= na(\theta\bar{c})' \\ -a(\theta c_2)' - Dc_2'' + qc_2' &= ma(\theta\bar{c})'\,. \end{aligned}$$

The travelling wave solutions for an unbounded spatial domain represent the long time behavior of an initial-boundary-value problem. So for a sufficiently large computational domain and a sufficiently long simulation time the travelling waves can be taken as a reference solution for the initial-boundary-value problem. This is done in this section. On the one hand the solution is computed with help of the ordinary differential equations and on the other hand the solution of the initial-boundary-value problem is computed with the implementation of the kinetic mineral problem. Then the two results are compared.

As test problem the so called "reference case" out of [KvDH95, Sec. 3] is used. For the initial-boundary-value problem the domain $\Omega = (0, 40)$ is used. The parameters of the problem are $k_p = 1$, $k_d = 3.86884 \cdot 10^{-7}$, $k_{inv} = 10$,

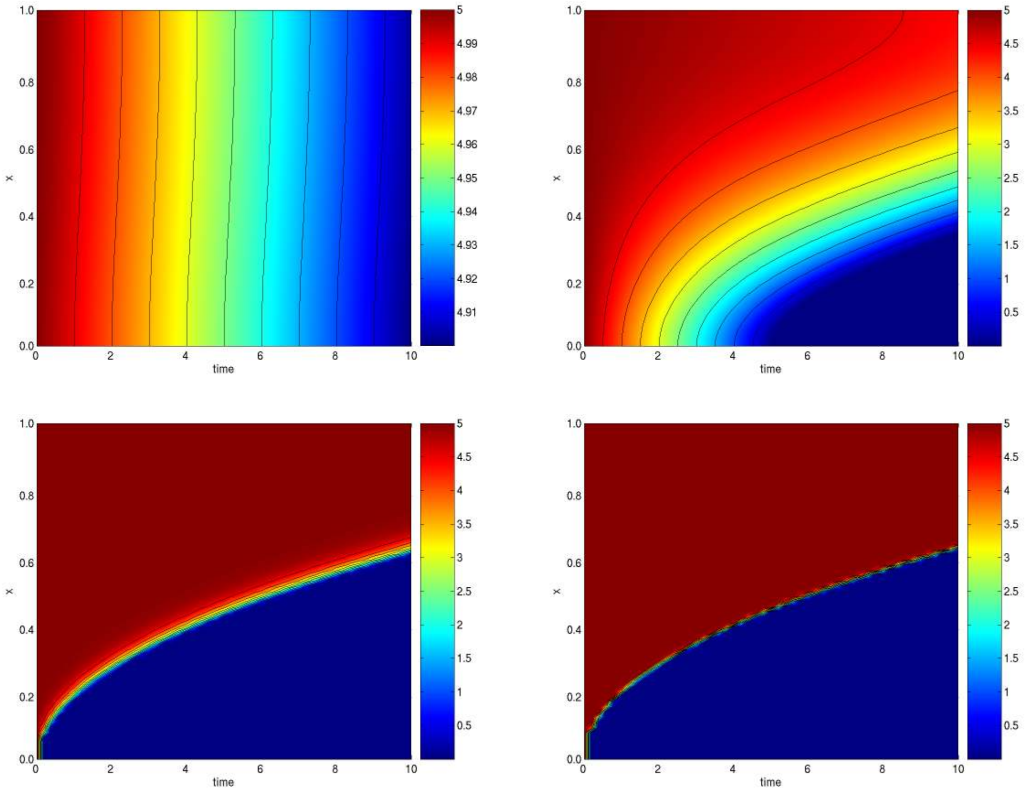


Figure 5.1: Mineral concentration for $k_{inv} = 100, 1, 10^{-2}, 10^{-4}$

$n = m = 1$, $D = 6.25 \cdot 10^{-4}$, $q = 0.3 \cdot 10^{-3}$, $\theta = 0.32$. For the initial-boundary-value problem the initial values are $c_{1,0} = c_{2,0} = 6.22 \cdot 10^{-4}$, $\bar{c}_0 = 2.7562 \cdot 10^{-4}$, the boundary conditions are Dirichlet boundary conditions at $x = 0$ with the boundary values $c_{1,D} = c_{2,D} = 2 \cdot 10^{-5}$ and homogeneous Neumann boundary conditions at $x = 40$. For the system of ordinary differential equations the values $c_{i,0}$, $\bar{c}_0$ are the boundary conditions at ∞ and the values $c_{i,D}$ together with $\bar{c} = 0$ are the boundary conditions at $-\infty$.

To write down the solution strategy, which uses the ordinary differential equations, we define the new variables

$$\begin{aligned} u &:= c_1 \\ v &:= \frac{n}{\theta}\bar{c} \\ c &:= mc_1 - nc_2 \, . \end{aligned}$$

For c being constant and having the value $mc_{1,D} - nc_{1,D}$ the wave speed a is constant and has the value $a = \frac{u_0 - u_D}{u_0 - u_D + v_0}$ (see [KvDH95]). According to [KvDH95] the travelling wave can be calculated by finding a number u_L such that the solution of

$$\begin{aligned} u' &= \frac{\frac{q}{\theta} - a}{D}(u - u_D) - \frac{a}{D}v && \text{for } \eta > L \\ v' &= \frac{nk_p}{k_{inv}a}\left(\frac{k_d}{k_p} - u^n\left(\frac{mu - c}{n}\right)^m\right) && \text{for } \eta > L \\ u(L) &= u_L \\ v(L) &= 0 \end{aligned} \tag{5.34}$$

fulfills

$$u(\infty) = u_0, \qquad v(\infty) = v_0, \qquad v(\eta) > 0 \quad \text{for } \eta > L\,.$$

This solution can be extended to the desired solution by solving

$$\begin{aligned} u' &= \frac{\frac{q}{\theta} - a}{D}(u - u_D) && \text{for } \eta < L \\ u(L) &= u_L\,. \end{aligned}$$

The explicit solution is

$$u(\eta) = (u_L - u_D)\exp\left(\frac{\frac{q}{\theta} - a}{D}\eta\right) + u_D \qquad \text{for } \eta \leq L\,.$$

We define the following three cases which can happen during the computation of u, v according to (5.34):

- Case A: There is a $\bar{\eta}$ such that $v(\bar{\eta}) = \frac{u(\bar{\eta}) - u_D}{u_0 - u_D}v_0$, $u(\bar{\eta}) < u_0$
- Case B: There is a $\bar{\eta}$ such that $u(\bar{\eta}) = u_0$, $v(\bar{\eta}) < \frac{u(\bar{\eta}) - u_D}{u_0 - u_D}v_0$
- Case C: Neither case A nor case B happen

Therewith the solution of the travelling wave problem can be computed with the following shooting algorithm

Shooting algorithm

```
Choose $u_D < \underline{u}_L, \overline{u}_L < u_0$ such that for $u_L = \underline{u}_L$ case A occurs
and for $u_L = \overline{u}_L$ case B occurs
$|\overline{u}_L - \underline{u}_L|$ not small enough
    Compute $u, v$ according to (5.34) with $u_L = \frac{1}{2}(\underline{u}_L + \overline{u}_L)$
    Check which case occurs
        For case A
            $\underline{u}_L := \frac{1}{2}(\underline{u}_L + \overline{u}_L)$
        For case B
            $\overline{u}_L := \frac{1}{2}(\underline{u}_L + \overline{u}_L)$
        For case C
            STOP
```

The end time for the numerical computations is chosen in such a way that at the end time the mineral is completely dissolved in the interval $(0, 25)$ and the mineral is present in the interval $(25, 40)$. For the "reference case" this is the case at $T = 42400$. The results of the shooting algorithm and of the initial-boundary-value problem can be seen in Fig. 5.2. The two results of the two different methods are in good agreement.

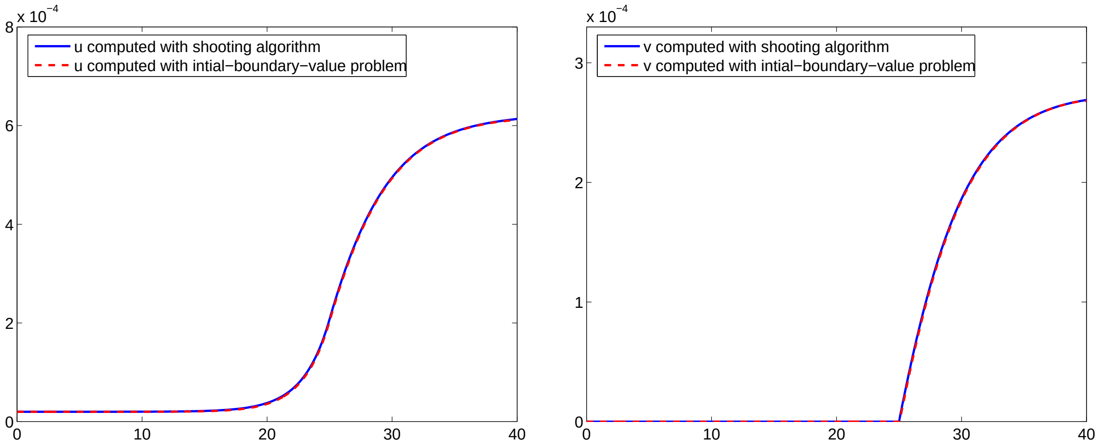


Figure 5.2: Results of the travelling wave problem

Then computations with smaller values of k_{inv} are carried out. These computations are closer to the equilibrium case than the problem above. Again the computations are done with the shooting algorithm and with the initial-boundary-value problem and then the results are compared. The results for the values $k_{inv} = 1, 0.1, 0.01$ are presented in Fig. 5.3. For all these values of k_{inv} the results of the two methods are in good agreement.

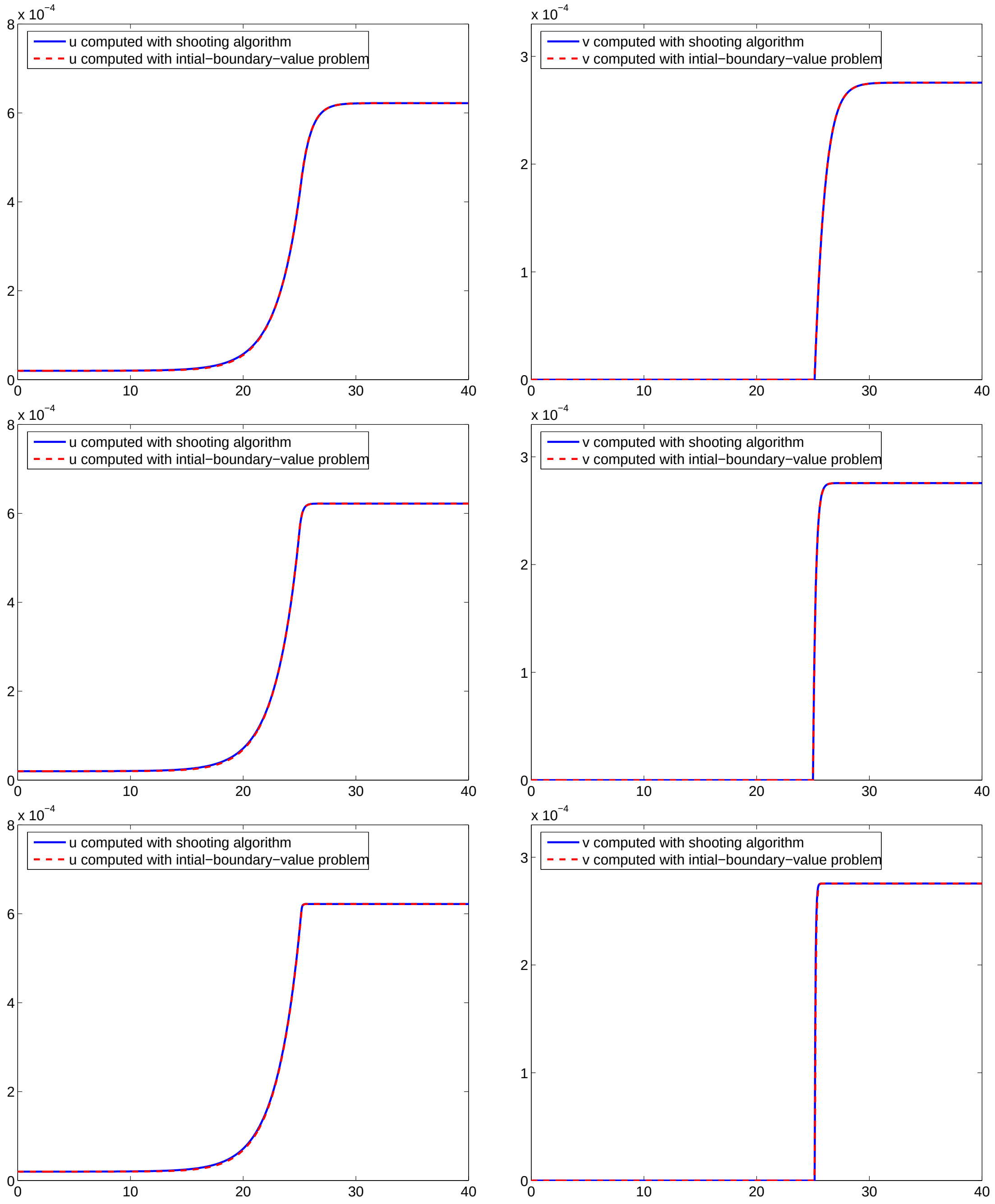


Figure 5.3: Results of the travelling wave problem for $k_{inv} = 1, 0.1, 0.01$

5.4 Existence

The goal is to prove the existence of a global solution for the more general problem

$$\begin{aligned}
\partial_t \boldsymbol{c} + L\boldsymbol{c} &= \boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}) + \boldsymbol{S}_{1,min}\partial_t\bar{\boldsymbol{c}}_{min} && \text{on } Q_T\\
\partial_t\bar{\boldsymbol{c}}_{min} &\in \boldsymbol{r}_{min}(\boldsymbol{c},\bar{\boldsymbol{c}}_{min}) && \text{on } Q_T\\
\boldsymbol{c}(\cdot,0) &= \boldsymbol{c}_0 && \text{on } \overline{\Omega}\\
\bar{\boldsymbol{c}}_{min}(\cdot,0) &= \bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega}\\
d\partial_{\boldsymbol{\nu}}\boldsymbol{c} &= \beta(\boldsymbol{c}-\boldsymbol{c}^*) && \text{on } S_T
\end{aligned}$$

under certain assumptions. In the problem formulation we have the linear transport operator $Lu := -\nabla\cdot(d\nabla u) + \boldsymbol{q}\cdot\nabla u$, Ω domain in $\mathbb{R}^n$, $Q_T := \Omega\times(0,T)$, $S_T := \partial\Omega\times(0,T)$ and the reaction rates $\boldsymbol{r}_{kin}$ according to law of mass action (compare (2.4)). It is assumed that only mobile species take part in the kinetic reactions. Furthermore there are the mineral reaction rates $\boldsymbol{r}_{min}$(It is assumed that the stoichiometric coefficient of the mineral is positive) that are of the form

$$r_{min,j}(\boldsymbol{c},\bar{\boldsymbol{c}}_{min}) = k_{p,j}\prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} c_i^{-S_{1,min,ij}} - k_{d,j}\prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} c_i^{S_{1,min,ij}} H(\bar{c}_{min,j})$$

with the set-valued Heaviside "function"

$$H(s) := \begin{cases} \{1\} & \text{for } s>0\\ [0,1] & \text{for } s=0\\ \{0\} & \text{for } s<0 \end{cases}.$$

This problem is denoted by $\boldsymbol{(P)}$. The boundary conditions include the cases (i) flux boundary conditions ($\beta = \boldsymbol{q}\cdot\boldsymbol{\nu}$) and (ii) homogeneous Neumann boundary conditions ($\beta = 0$) (compare [Kna86, (2.10)]).

To prove an a priori estimate with help of the maximum principle the following assumption is needed:

Assumptions 5.4. *There is a vector $\boldsymbol{s}^\perp$ with only strictly positive entries which is perpendicular to all columns of $\boldsymbol{S}_1 = \begin{pmatrix}\boldsymbol{S}_{1,kin} & \boldsymbol{S}_{1,min}\end{pmatrix}$ except of those columns of $\boldsymbol{S}_1$ that have only nonpositive entries.*

When no species are neglected in the kinetic reactions with only mobile species there is always a vector $\boldsymbol{s}^\perp$ which is perpendicular to all columns of $\boldsymbol{S}_{1,kin}$. In this case the number of atoms in one molecule of the i-th species is a possible choice for the components of $\boldsymbol{s}^\perp$ because the number of atoms is a conserved quantity regarding the chemical reactions. When additionally all entries of $\boldsymbol{S}_{1,min}$ are nonpositive, i.e., in the reaction equation of each mineral reaction no mobile species are on the same side as the mineral, the assumption is always fulfilled. Note that it is not necessary that no species are neglected in the reactions with only mobile species and that is also not necessary that all entries of $\boldsymbol{S}_{1,min}$ are nonpositive. For example the reaction $OH^- + H_3O^+ \leftrightarrow (2H_2O)$ neglecting the concentration of H_2O is also allowed. And mineral reactions with mobile species in the reaction equation on the same side as the mineral are allowed when there is a vector $\boldsymbol{s}^\perp$ which is also perpendicular to the associated column of $\boldsymbol{S}_{1,min}$.

In [Krä08, Sec. 3.2] the existence of a solution for the batch problem is shown with help of a vector $\boldsymbol{s}^\perp$ which must be perpendicular to all columns of the stoichiometric matrix. Here we handle a much more complex problem (with transport and with mineral reactions) and we need only a weaker assumption, it is not necessary that the vector $\boldsymbol{s}^\perp$ is perpendicular to all columns of the stoichiometric matrix.

Furthermore the following assumptions on the data of the problem are needed (compare [Kna86, Assumption 2.2]):

Assumptions 5.5.

(i) $d > \delta = \text{const} > 0$

(ii) $d, \partial_{x_k} d \in C^{\alpha,\alpha/2}(\overline{Q}_T)$ $(k = 1, \ldots, n)$, $\boldsymbol{q} \in C^{\alpha,\alpha/2}(\overline{Q}_T)^n$ *for some* $\alpha \in (0, 1)$

(iii) $c_{0,i} \in W_p^{2-2/p}(\Omega)$ *for some* $p > (n + 2)/2$, $p \geq 2$, $q \neq n + 2$; $c_{0,i}$ *is continuously differentiable in a neighborhood of* $\partial\Omega$ $(i = 1, \ldots, I)$

(iv) $\bar{c}_{min,0,j} \in C^\alpha(\overline{\Omega})$ $(j = 1, \ldots, J_{min})$

(v) $c_i^* \in W^{1-1/p,(1-1/p)/2}(S_T) \cap C(\overline{S}_T)$ $(i = 1, \ldots, I)$

(vi) $\beta \in C^{1-1/p+\epsilon,(1-1/p+\epsilon)/2}(\overline{S}_T)$ *for some* $\epsilon > 0$

(vii) *If* $p > 3$*:* $\partial_{\boldsymbol{\nu}} c_{0,i} = \beta(\cdot, 0)(c_{0,i} - c_i^*(\cdot, 0))$ *on* $\partial\Omega$ $(i = 1, \ldots, I)$

(viii) $\partial\Omega \in C^{2+\alpha}$

(ix) $\boldsymbol{c}_0, \bar{\boldsymbol{c}}_{min,0}, \boldsymbol{c}^* \geq 0$

Let us consider the modified problem $(\boldsymbol{P}^+)$ where the rate functions $\boldsymbol{r}_{kin}(\boldsymbol{c})$ are replaced by $\boldsymbol{r}_{kin}(\boldsymbol{c}^+)$ and the rates $\boldsymbol{r}_{min}(\boldsymbol{c}, \bar{\boldsymbol{c}})$ are replaced by $\boldsymbol{r}_{min}(\boldsymbol{c}^+, \bar{\boldsymbol{c}})$.

Additionally we define the regularized problem

$$\begin{aligned}
\partial_t \boldsymbol{c} + L\boldsymbol{c} &= \boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}^+) + \boldsymbol{S}_{1,min}\boldsymbol{r}_{\varepsilon,min}(\boldsymbol{c}^+,\bar{\boldsymbol{c}}_{min}) && \text{on } Q_T && (5.35)\\
\partial_t \bar{\boldsymbol{c}}_{min} &= \boldsymbol{r}_{\varepsilon,min}(\boldsymbol{c}^+,\bar{\boldsymbol{c}}_{min}) && \text{on } Q_T && (5.36)\\
\boldsymbol{c}(\cdot,0) &= \boldsymbol{c}_0 && \text{on } \overline{\Omega} && (5.37)\\
\bar{\boldsymbol{c}}_{min}(\cdot,0) &= \bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega} && (5.38)\\
d\partial_{\boldsymbol{\nu}}\boldsymbol{c} &= \beta(\boldsymbol{c}-\boldsymbol{c}^*) && \text{on } S_T && (5.39)
\end{aligned}$$

with the regularized rate functions

$$r_{\varepsilon,min,j}(\boldsymbol{c}^+,\bar{\boldsymbol{c}}_{min}) = k_{p,j}\prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I}(c_i^+)^{-S_{1,min,ij}} - k_{d,j}\prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I}(c_i^+)^{S_{1,min,ij}} H_\varepsilon(\bar{c}_{min,j}) \tag{5.40}$$

where H_ε is the regularized Heaviside function

$$H_\varepsilon(s) := \begin{cases} 1 & \text{for } s \geq \varepsilon \\ s/\varepsilon & \text{for } 0 < s < \varepsilon \\ 0 & \text{for } s \leq 0 \end{cases}.$$

with $\varepsilon > 0$. This problem is denoted by $(\boldsymbol{P_\varepsilon^+})$.

Nonnegativity

Lemma 5.6. *Let $(\boldsymbol{c},\bar{\boldsymbol{c}}_{min})$ be a solution of problem $(\boldsymbol{P_\varepsilon^+})$. Then $\boldsymbol{c}$ is nonnegative.*

Proof. The proof of this lemma is adapted from [Krä08, Lemma 3.2]. Let $\Omega_i^- = \Omega_i^-(t)$ be the support of $c_i^-(\cdot,t)$. Testing the i-th PDE with $-c_i^-$ yields

$$\begin{aligned}
&\frac{1}{2}\partial_t \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} + \int_{\Omega_i^-} (d|\nabla c_i^-|^2 + \boldsymbol{q}\cdot\nabla c_i^- c_i^-)\, d\boldsymbol{x}\\
&= -\sum_{j=1}^{J_{kin}} S_{1,kin,ij}\int_{\Omega_i^-}\Bigg(k_{f,j}\prod_{\substack{k=1\\ S_{1,kin,kj}<0}}^{I}(c_k^+)^{-S_{1,kin,kj}} - k_{b,j}\prod_{\substack{k=1\\ S_{1,kin,kj}>0}}^{I}(c_k^+)^{S_{1,kin,kj}}\Bigg)c_i^-\, d\boldsymbol{x}\\
&-\sum_{j=1}^{J_{min}} S_{1,min,ij}\int_{\Omega_i^-}\Bigg(k_{p,j}\prod_{\substack{k=1\\ S_{1,min,kj}<0}}^{I}(c_k^+)^{-S_{1,min,kj}} - k_{d,j}\prod_{\substack{k=1\\ S_{1,min,kj}>0}}^{I}(c_k^+)^{S_{1,min,kj}} H_\varepsilon(\bar{c}_{min,j})\Bigg)c_i^-\, d\boldsymbol{x}.
\end{aligned}$$

We know that $c_i^+ \equiv 0$ on the domain of integration Ω_i^-. Using this we get in the case $S_{1,kin,ij} > 0$ (Note that in this case c_i^+ is a factor of the second product)

$$-S_{1,kin,ij} \int_{\Omega_i^-} \Bigg(k_{f,j} \prod_{\substack{k=1 \\ S_{1,kin,kj}<0}}^{I} (c_k^+)^{-S_{1,kin,kj}} - k_{b,j} \prod_{\substack{k=1 \\ S_{1,kin,kj}>0}}^{I} (c_k^+)^{S_{1,kin,kj}} \Bigg) c_i^- \, d\boldsymbol{x}$$
$$= -S_{1,kin,ij} \int_{\Omega_i^-} \Bigg(k_{f,j} \prod_{\substack{k=1 \\ S_{1,kin,kj}<0}}^{I} (c_k^+)^{-S_{1,kin,kj}} \Bigg) c_i^- \, d\boldsymbol{x} \le 0 \,.$$

Analogously one gets that the term is also nonpositive in the case $S_{1,kin,ij} < 0$. In the same way one can show that the term with the mineral reaction rate is nonpositive, too. So we get

$$\frac{1}{2}\partial_t \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} + \int_{\Omega_i^-} (d|\nabla c_i^-|^2 + \boldsymbol{q} \cdot \nabla c_i^- c_i^-) \, d\boldsymbol{x} \le 0 \,.$$

Using Young's inequality it follows

$$\frac{1}{2}\partial_t \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} + \int_{\Omega_i^-} d|\nabla c_i^-|^2 \, d\boldsymbol{x} \le \frac{Q^2}{2\delta} \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} + \frac{\delta}{2} \int_{\Omega_i^-} |\nabla c_i^-|^2 \, d\boldsymbol{x}$$

with $Q := \|\boldsymbol{q}\|_{L^\infty(Q_T)^n}$. Absorbing the term with $\delta/2$ on the left hand side gives (Note that $d > \delta$, see Assumptions 5.5 (i))

$$\partial_t \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} + \int_{\Omega_i^-} d|\nabla c_i^-|^2 \, d\boldsymbol{x} \le \frac{Q^2}{\delta} \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} \,.$$

Especially it holds

$$\partial_t \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} \le \frac{Q^2}{\delta} \int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} \,.$$

Because of the assumption that the initial values are nonnegative (Assumptions 5.5 (ix)) it follows that $\int_{\Omega_i^-} |c_i^-|^2 \, d\boldsymbol{x} \equiv 0$ for all $t \ge 0$ and hence it holds $c_i \ge 0$ a.e. in Q_T. □

Lemma 5.7. *Let $(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min})$ be a solution of problem $(\boldsymbol{P}_{\boldsymbol{\varepsilon}}^{+})$. Then $\bar{\boldsymbol{c}}_{min}$ is nonnegative.*

Proof. Using

$$\phi_j(s) = \begin{cases} -\bar{c}_{min,j}^{-}(\cdot, s) & \text{for } s \le t \\ 0 & \text{for } s > t \end{cases}$$

as test function in the j-th ODE yields

$$\int_0^t \partial_t \bar{c}_{min,j}(-\bar{c}_{min,j}^-)\, ds =$$
$$\int_0^t \Bigg(\underbrace{k_{p,j} \prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} (c_i^+)^{-S_{1,min,ij}}}_{\geq 0} - k_{d,j} \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} H_\varepsilon(\bar{c}_{min,j}) \Bigg) \underbrace{(-\bar{c}_{min,j}^-)}_{\leq 0}\, ds\,.$$

Because of

$$\int_0^t \partial_t \bar{c}_{min,j}(-\bar{c}_{min,j}^-)\, ds = \frac{1}{2}\int_0^t \partial_t |\bar{c}_{min,j}^-|^2\, ds = \frac{1}{2}|\bar{c}_{min,j}^-(\cdot,t)|^2 - \frac{1}{2}|\bar{c}_{min,j}^-(\cdot,0)|^2$$

we get the estimate

$$\frac{1}{2}|\bar{c}_{min,j}^-(\cdot,t)|^2 \leq \int_0^t k_{d,j} \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} H_\varepsilon(\bar{c}_{min,j})\bar{c}_{min,j}^-\, ds + \frac{1}{2}|\bar{c}_{min,j}^-(\cdot,0)|^2 = 0.$$

The first summand on the right hand side is zero because one of the factors $H_\varepsilon(\bar{c}_{min,j})$, $\bar{c}_{min,j}^-$ is zero a.e. due to the definition of H_ε and the second one is zero due to the assumption that the initial value $\bar{c}_{min,0,j}$ is nonnegative (Assumptions 5.5 (*ix*)). □

Remark 5.8. *The assertions of the last two lemmas are also true for the problem* $(\boldsymbol{P^+})$.

We want to prove the existence of a global solution of the modified and regularized problem with help of Schaefer's fixed point theorem (e.g. [Eva98]).

The Fixed Point Operator

We define the fixed point operator $\mathcal{Z}$:

$$\begin{aligned} \mathcal{Z} : W_p^{2,1}(Q_T)^I &\longrightarrow W_p^{2,1}(Q_T)^I \\ \hat{\boldsymbol{c}} &\longmapsto \boldsymbol{c} = \mathcal{Z}(\hat{\boldsymbol{c}}) \end{aligned}$$

with $p > (n+2)/2$ and $\boldsymbol{c}$ being the solution of the problem:

$$\begin{aligned} \partial_t \boldsymbol{c} + L\boldsymbol{c} &= \boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\hat{\boldsymbol{c}}^+) + \boldsymbol{S}_{1,min}\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^+, \bar{\boldsymbol{c}}_{min}) && \text{on } Q_T \\ \partial_t \bar{\boldsymbol{c}}_{min} &= \boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^+, \bar{\boldsymbol{c}}_{min}) && \text{on } Q_T \\ \boldsymbol{c}(\cdot,0) &= \boldsymbol{c}_0 && \text{on } \overline{\Omega} \\ \bar{\boldsymbol{c}}_{min}(\cdot,0) &= \bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega} \\ d\partial_{\boldsymbol{\nu}}\boldsymbol{c} &= \beta(\boldsymbol{c} - \boldsymbol{c}^*) && \text{on } S_T \end{aligned}$$

Lemma 5.9. *The fixed point operator $\mathcal{Z}$ is well-posed.*

Proof. Some ideas of this proof are taken from the argumentation in [Krä08, page 104]. First we will show that a solution $\bar{\boldsymbol{c}}_{min}$ of the ODE subsystem exists and that this solution is in $C(\overline{Q}_T)^{J_{min}}$. From the embedding theorem $W_p^{2,1}(Q_T) \hookrightarrow C^{\alpha,\alpha/2}(\overline{Q}_T)$ with $0 < \alpha \leq 2-(n+2)/p$ for $p > (n+2)/2$ (e.g. [WYW06, Thm. 1.4.1]) the continuity of $\hat{\boldsymbol{c}}$ follows. So for fixed $\boldsymbol{x} \in \overline{\Omega}$ the right hand side $\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},t), \bar{\boldsymbol{c}}_{min})$ (see (5.40) for the definition of $\boldsymbol{r}_{\varepsilon,min}$) as a function of t and $\bar{\boldsymbol{c}}_{min}$ is continuous in t. The Lipschitz continuity of H_ε yields that $\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},t), \bar{\boldsymbol{c}}_{min})$ is Lipschitz continuous in $\bar{\boldsymbol{c}}_{min}$ with a Lipschitz constant independent of $\boldsymbol{x}$ and t:

$$\begin{aligned}
&|r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},t),y) - r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},t),\tilde{y})| \\
&\qquad = \left| k_{d,j} \prod_{\substack{i=1 \\ S_{1,min,ij}>0}}^{I} (\hat{c}_i^+(\boldsymbol{x},t))^{S_{1,min,ij}} \right| |H_\varepsilon(y) - H_\varepsilon(\tilde{y})| \\
&\qquad \leq C(M) L_\varepsilon |y-\tilde{y}|
\end{aligned}$$

with M a bound for the $C(\overline{Q}_T)^I$-norm of $\hat{\boldsymbol{c}}$ and L_ε the Lipschitz constant of H_ε. Such a bound $M < \infty$ exists because of $\hat{\boldsymbol{c}} \in C^{\alpha,\alpha/2}(\overline{Q}_T)^I$. So the theorem of Picard-Lindelöf proves that $\bar{\boldsymbol{c}}_{min}$ exists on the whole interval $[0,T]$, i.e., $\bar{\boldsymbol{c}}_{min}(\boldsymbol{x},\cdot) \in C([0,T])^{J_{min}}$ for fixed $\boldsymbol{x} \in \overline{\Omega}$.

To complete the proof that $\bar{\boldsymbol{c}}_{min} \in C(\overline{Q}_T)^{J_{min}}$ we will show that $\bar{\boldsymbol{c}}_{min}$ is Hölder continuous in $\boldsymbol{x}$, uniformly in $\overline{Q}_T$. Let $y \in C^1([0,T])$ be the solution of

$$\begin{aligned}
y' &= r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},\cdot),y) \\
y(0) &= \bar{c}_{min,0,j}(\boldsymbol{x})
\end{aligned}$$

and $\tilde{y} \in C^1([0,T])$ be the solution of

$$\begin{aligned}
\tilde{y}' &= r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\tilde{\boldsymbol{x}},\cdot),\tilde{y}) \\
\tilde{y}(0) &= \bar{c}_{min,0,j}(\tilde{\boldsymbol{x}}) \,.
\end{aligned}$$

As $\bar{c}_{min,0,j}$ is Hölder continuous (see Assumptions 5.5 (iv)) we know that

$$|\bar{c}_{min,0,j}(\boldsymbol{x}) - \bar{c}_{min,0,j}(\tilde{\boldsymbol{x}})| \leq K_1 |\boldsymbol{x}-\tilde{\boldsymbol{x}}|^\alpha \,.$$

Because of the Hölder continuity of $\hat{\boldsymbol{c}}$ we know that (Note that the product of two functions that are Hölder continuous with exponent α is also Hölder continuous with exponent α.)

$$|r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\boldsymbol{x},t),y) - r_{\varepsilon,min,j}(\hat{\boldsymbol{c}}^+(\tilde{\boldsymbol{x}},t),y)| \leq K_2 |\boldsymbol{x}-\tilde{\boldsymbol{x}}|^\alpha \quad \forall t \in [0,T] \,.$$

With [Heu89, Chap. 3, Satz 13.1] we get

$$|y(t) - \tilde{y}(t)| \leq \max\{K_1, K_2\}\Big(e^{LT} + \frac{1}{L}(e^{LT} - 1)\Big)|\boldsymbol{x} - \tilde{\boldsymbol{x}}|^{\alpha} \quad \forall t \in [0, T]$$

with $L = C(M)L_{\varepsilon}$ the Lipschitz constant of $\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^{+}(\boldsymbol{x}, t), \cdot)$ and so $\bar{\boldsymbol{c}}_{min}$ is Hölder continuous in $\boldsymbol{x}$, uniformly in $\overline{Q}_T$:

$$|\bar{c}_{min,j}(\boldsymbol{x}, t) - \bar{c}_{min,j}(\tilde{\boldsymbol{x}}, t)| \leq \max\{K_1, K_2\}\Big(e^{LT} + \frac{1}{L}(e^{LT} - 1)\Big)|\boldsymbol{x} - \tilde{\boldsymbol{x}}|^{\alpha} \quad \forall t \in [0, T]$$

The right hand side of the ODE subsystem $\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^{+}, \bar{\boldsymbol{c}}_{min})$ is an element of $C(\overline{Q}_T)^{J_{min}}$. So we get

$$\bar{\boldsymbol{c}}_{min} \in \mathcal{C}(\overline{Q}_T)^{J_{min}}$$

with $\mathcal{C}(\overline{Q}_T) := \big\{v \in C(\overline{Q}_T) | \partial_t v \in C(\overline{Q}_T)\big\}$.

Now we consider the PDEs. The right hand side of the PDE subsystem $\boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\hat{\boldsymbol{c}}^{+}) + \boldsymbol{S}_{1,min}\boldsymbol{r}_{\varepsilon,min}(\hat{\boldsymbol{c}}^{+}, \bar{\boldsymbol{c}}_{min})$ is an element of $C(\overline{Q}_T)^{I}$. It follows that the right hand side is an element of $L^{q}(Q_T)^{I}$ for all $1 \leq q \leq \infty$. Using the linear parabolic theory (compare [LSU68, IV, 9]), we get a solution of the PDE subsystem $\boldsymbol{c} \in W_p^{2,1}(Q_T)^{I}$. □

A Priori Estimates

We have to construct a bound holding for arbitrary solutions $\boldsymbol{c} \in W_p^{2,1}(Q_T)^{I}$ of the equation

$$\boldsymbol{c} = \lambda \mathcal{Z}(\boldsymbol{c})$$

with $\lambda \in [0, 1]$. So we have to find a bound for the solutions of

$$\begin{aligned} \partial_t \boldsymbol{c} + L\boldsymbol{c} &= \lambda\left(\boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}^{+}) + \boldsymbol{S}_{1,min}\boldsymbol{r}_{\varepsilon,min}(\boldsymbol{c}^{+}, \bar{\boldsymbol{c}}_{min})\right) && \text{on } Q_T \\ \partial_t \bar{\boldsymbol{c}}_{min} &= \boldsymbol{r}_{\varepsilon,min}(\boldsymbol{c}^{+}, \bar{\boldsymbol{c}}_{min}) && \text{on } Q_T \\ \boldsymbol{c}(\cdot, 0) &= \lambda \boldsymbol{c}_0 && \text{on } \overline{\Omega} \\ \bar{\boldsymbol{c}}_{min}(\cdot, 0) &= \bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega} \\ d\partial_{\boldsymbol{\nu}} \boldsymbol{c} &= \beta(\boldsymbol{c} - \lambda \boldsymbol{c}^{*}) && \text{on } S_T \end{aligned} \tag{5.41}$$

To derive the needed a priori estimate we want to use the maximum principle (e.g. [LSU68, I, Thm. 2.2/2.3]), which is also used in [Kna86, Sec. 3]. Here we will construct an upper bound $\tilde{\eta}$ for the mobile concentrations c_i with help of the maximum principle. $\tilde{\eta}$ will be the solution of a PDE and again using the maximum principle one can show that there is a bound for the $C(\overline{Q}_T)$-norm of $\tilde{\eta}$ which is independent of the solution. As $\tilde{\eta}$ is an upper bound for every c_i we have found a bound for the $C(\overline{Q}_T)^{I}$-norm of $\boldsymbol{c}$. Then it follows from the linear parabolic theory that there is also a bound for the $W_p^{2,1}(Q_T)^I$-norm for arbitrary solutions. For applying the maximum principle it is needed that the solution of the PDE is a classical solution. To show this the existence theorem [Fri64, Chap. 5, p. 147, Cor. 2] will be used.

Let $\tilde{\eta}$ be the solution of

$$\begin{aligned} \partial_t\tilde{\eta} + L\tilde{\eta} &= \lambda\left(\boldsymbol{s}^\perp \cdot (\boldsymbol{S}_{1,kin}^-(-\boldsymbol{k}_b^-)) + \boldsymbol{s}^\perp \cdot (\boldsymbol{S}_{1,min}^-(-\boldsymbol{k}_d^-))\right) && \text{on } Q_T \\ \tilde{\eta}(\cdot,0) &= \lambda\boldsymbol{s}^\perp \cdot \boldsymbol{c}_0 && \text{on } \overline{\Omega} \\ d\partial_{\boldsymbol{\nu}}\tilde{\eta} &= \beta(\tilde{\eta} - \lambda\boldsymbol{s}^\perp \cdot \boldsymbol{c}^*) && \text{on } S_T \end{aligned} \tag{5.42}$$

where $\boldsymbol{S}_{1,kin}^-$ and $\boldsymbol{S}_{1,min}^-$ are the submatrices of $\boldsymbol{S}_{1,kin}$ and $\boldsymbol{S}_{1,min}$, respectively, that contain all columns with only nonpositive entries. The vectors $\boldsymbol{k}_b^-$ and $\boldsymbol{k}_d^-$ contain all reaction constants $k_{b,j}$ and $k_{d,j}$, respectively, that correspond to a column of $\boldsymbol{S}_1$ with only nonpositive entries. $\tilde{\eta}$ is the solution of a parabolic PDE with constant right hand side. According to [Fri64, Chap. 5, p. 147, Cor. 2] a classical solution exists.

To apply the existence theorem to the PDE subsystem of (5.41) we have to show that the right hand side is Hölder continuous in $\boldsymbol{x}$, uniformly in $\overline{Q}_T$. From the embedding $W_p^{2,1}(Q_T) \hookrightarrow C^{\alpha,\alpha/2}(\overline{Q}_T)$ with $0 < \alpha \leq 2 - (n+2)/p$ for $p > (n+2)/2$ (e.g. [WYW06, Thm. 1.4.1]) we know that $\boldsymbol{c}$ is Hölder continuous. In the proof of Lemma 5.9 we have already shown that $\bar{\boldsymbol{c}}_{min}$ is Hölder continuous in $\boldsymbol{x}$, uniformly in $\overline{Q}_T$. Hence the requirements of the existence theorem [Fri64, Chap. 5, p. 147, Cor. 2] are fulfilled. So we get that $\boldsymbol{c}$ is a classical solution of the PDE subsystem.

As next step we examine the function

$$u := \boldsymbol{s}^\perp \cdot \boldsymbol{c} - \tilde{\eta}\,.$$

By taking linear combinations of the PDEs in (5.41) and using the PDE in (5.42) one can see that u is a solution of (Remember that $\boldsymbol{s}^\perp$ is perpendicular to all columns of $\boldsymbol{S}_1$ except of those with only nonpositive entries)

$$\begin{aligned} \partial_t u + Lu = \lambda\boldsymbol{s}^\perp\cdot\big(&\boldsymbol{S}_{1,kin}^-(\boldsymbol{r}_{kin}^-(\boldsymbol{c}^+) + \boldsymbol{k}_b^-) \\ &+ \boldsymbol{S}_{1,min}^-(\boldsymbol{r}_{\varepsilon,min}^-(\boldsymbol{c}^+, \bar{\boldsymbol{c}}_{min}) + \boldsymbol{k}_d^-)\big) && \text{on } Q_T \\ u(\cdot,0) = 0 & && \text{on } \overline{\Omega} \\ d\partial_{\boldsymbol{\nu}}u = \beta u & && \text{on } S_T \end{aligned} \tag{5.43}$$

where $\boldsymbol{r}_{kin}^-$ and $\boldsymbol{r}_{\varepsilon,min}^-$ contain all reaction rates $r_{kin,j}$ and $r_{\varepsilon,min,j}$, respectively, that correspond to a column of $\boldsymbol{S}_1$ with only nonpositive entries. Because of this all components of $\boldsymbol{r}_{kin}^{-}$ have the form

$$r_{kin,j}^{-}(\boldsymbol{c}^{+}) = k_{f,j} \prod_{\substack{i=1 \\ S_{1,kin,ij}^{-}<0}}^{I} (c_i^{+})^{-S_{1,kin,ij}^{-}} - k_{b,j}$$

and one sees immediately that all components of the vector $(\boldsymbol{r}_{kin}^{-}(\boldsymbol{c}^{+}) + \boldsymbol{k}_b^{-})$ are nonnegative. Regarding the components of $\boldsymbol{r}_{\varepsilon,min}^{-}$ one knows that the second product in (5.40) is empty. So one component of the vector $(\boldsymbol{r}_{\varepsilon,min}^{-}(\boldsymbol{c}^{+}, \bar{\boldsymbol{c}}_{min}) + \boldsymbol{k}_d^{-})$ is of the form

$$k_{p,j} \prod_{\substack{i=1 \\ S_{1,min,ij}^{-}<0}}^{I} (c_i^{+})^{-S_{1,min,ij}^{-}} + k_{d,j}(1 - H_\varepsilon(\bar{c}_{min,j}))\,.$$

Because of $H_\varepsilon \leq 1$ it follows that all components of $(\boldsymbol{r}_{\varepsilon,min}^{-}(\boldsymbol{c}^{+}, \bar{\boldsymbol{c}}_{min}) + \boldsymbol{k}_d^{-})$ are nonnegative. Furthermore by definition all components of $\boldsymbol{s}^{\perp}$ are positive and all entries of $\boldsymbol{S}_{1,kin}^{-}$ and $\boldsymbol{S}_{1,min}^{-}$ are nonpositive. Altogether we get that the right hand side of the PDE for u is nonpositive. So applying the maximum principle[1] (compare [LSU68, I, Thm. 2.2/2.3]) yields

$$\sup_{Q_T} u \leq 0\,.$$

As all components of $\boldsymbol{s}^{\perp}$ are positive and all mobile concentrations are nonnegative (see Lemma 5.6) it follows

$$c_i \leq \frac{1}{s_i^{\perp}}\tilde{\eta} \qquad \forall i = 1, \dots, I$$

So $\max_i \frac{1}{s_i^{\perp}}\tilde{\eta}$ is an upper bound for the mobile concentrations.

$\tilde{\eta}$ is the solution of (5.42). Applying the maximum principle [LSU68, I, Thm. 2.3] to it gives

$$\sup_{Q_T} |\tilde{\eta}| \leq K_1$$

with a constant K_1 only depending on β, $\boldsymbol{c}^*$, $\boldsymbol{c}_0$, $\boldsymbol{s}^{\perp}$, $\boldsymbol{S}_{1,kin}^{-}\boldsymbol{k}_b^{-}$, $\boldsymbol{S}_{1,min}^{-}\boldsymbol{k}_d^{-}$, $\|d\|_{C(\overline{Q}_T)}$, $\|\boldsymbol{q}\|_{C(\overline{Q}_T)^n}$ and the boundary of Ω. So we have found an bound for the $C(\overline{Q}_T)^I$-norm of $\boldsymbol{c}$ for an arbitrary solution of (5.41).

It follows that the $C(\overline{Q}_T)^I$-norm of the right hand side of the PDE in (5.41) is bounded by a constant K_2 only depending on K_1, $\boldsymbol{S}_1$ and the reaction constants $k_{f,j}$, $k_{b,j}$, $k_{p,j}$, $k_{d,j}$. In particular, every $L^q(Q_T)^I$-norm $(1 \leq q \leq \infty)$ of the right

[1] see Appendix A.2 for a detailed description of the application of the maximum principle

hand side is bounded independent of the solution $(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min})$. Then with the linear parabolic theory (compare [Sol64, Thm. 17], [LSU68, IV, 9]) it follows

$$\|\boldsymbol{c}\|_{W_p^{2,1}(Q_T)^I} \leq K_3 \,. \tag{5.44}$$

with a constant K_3 independent of the solution $(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min})$.

Compactness

Theorem 5.10. *The operator $\mathcal{Z}$ is continuous and compact.*

Proof. The proof of this theorem is adapted from the proof of [Krä08, Theorem 3.17]. Let $(\hat{\boldsymbol{c}}^n)$ be a sequence bounded in $W_p^{2,1}(Q_T)^I$. Due to the compact embedding $W_p^{2,1}(Q_T) \hookrightarrow\hookrightarrow C(\overline{Q}_T)$ for $p > (n+2)/2$ there is a subsequence, again denoted by $(\hat{\boldsymbol{c}}^n)$, which is convergent in $C(\overline{Q}_T)^I$. First we consider the ODE subproblem

$$\begin{aligned} \partial_t \bar{\boldsymbol{c}}^n_{min} &= \boldsymbol{r}_{\varepsilon,min}((\hat{\boldsymbol{c}}^n)^+, \bar{\boldsymbol{c}}^n_{min}) && \text{on } Q_T \\ \bar{\boldsymbol{c}}^n_{min}(\cdot,0) &= \bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega}\,. \end{aligned}$$

Let $y \in C^1([0,T])$ be the solution of

$$\begin{aligned} y' &= r_{\varepsilon,min,j}((\hat{\boldsymbol{c}}^l)^+(\boldsymbol{x},\cdot), y) \\ y(0) &= \bar{\boldsymbol{c}}_{min,0,j} \end{aligned}$$

and $\tilde{y}$ be the solution of

$$\begin{aligned} \tilde{y}' &= r_{\varepsilon,min,j}((\hat{\boldsymbol{c}}^m)^+(\boldsymbol{x},\cdot), \tilde{y}) \\ \tilde{y}(0) &= \bar{\boldsymbol{c}}_{min,0,j}\,. \end{aligned}$$

Let us define

$$r_{1,j}(\boldsymbol{c}^+) = k_{p,j} \prod_{\substack{i=1 \\ S_{1,min,ij}<0}}^{I} (c_i^+)^{-S_{1,min,ij}}\,, \quad r_{2,j}(\boldsymbol{c}^+) = k_{d,j} \prod_{\substack{i=1 \\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}}\,.$$

As both functions $r_{1,j}$ and $r_{2,j}$ are uniformly continuous it holds

$$\begin{aligned} &|r_{\varepsilon,min,j}((\hat{\boldsymbol{c}}^l)^+(\boldsymbol{x},t), y) - r_{\varepsilon,min,j}((\hat{\boldsymbol{c}}^m)^+(\boldsymbol{x},t), y)| \\ &\quad \leq |r_{1,j}((\hat{\boldsymbol{c}}^l)^+(\boldsymbol{x},t)) - r_{1,j}((\hat{\boldsymbol{c}}^m)^+(\boldsymbol{x},t))| \\ &\qquad + |r_{2,j}((\hat{\boldsymbol{c}}^l)^+(\boldsymbol{x},t)) - r_{2,j}((\hat{\boldsymbol{c}}^m)^+(\boldsymbol{x},t))| \underbrace{|H_\varepsilon(y)|}_{\leq 1} \\ &\quad \leq \omega(r_{1,j}, |\hat{\boldsymbol{c}}^l(\boldsymbol{x},t) - \hat{\boldsymbol{c}}^m(\boldsymbol{x},t)|) + \omega(r_{2,j}, |\hat{\boldsymbol{c}}^l(\boldsymbol{x},t) - \hat{\boldsymbol{c}}^m(\boldsymbol{x},t)|) \end{aligned}$$

with ω the modulus of continuity.

The Lipschitz constant L of $r_{\varepsilon,min,j}((\hat{\boldsymbol{c}}^n)^+(\boldsymbol{x},t),\cdot)$ is $C(M)L_\varepsilon$ with $M<\infty$ a bound for the $C(\overline{Q}_T)^I$-norm of $\boldsymbol{c}$ and L_ε the Lipschitz constant of H_ε (see proof of Lemma 5.9). Such a bound M exists because the sequence $(\boldsymbol{c}^n)$ is convergent in $C(\overline{Q}_T)^I$. So we get with [Heu89, Chap. 3, Satz 13.1] for all $t\in[0,T]$

$$|y(t)-\tilde{y}(t)|\le\frac{\omega(r_{1,j},|\hat{\boldsymbol{c}}^l(\boldsymbol{x},t)-\hat{\boldsymbol{c}}^m(\boldsymbol{x},t)|)+\omega(r_{2,j},|\hat{\boldsymbol{c}}^l(\boldsymbol{x},t)-\hat{\boldsymbol{c}}^m(\boldsymbol{x},t)|)}{L}(e^{LT}-1)\,.$$

It follows that

$$|\bar{c}^l_{min,j}(\boldsymbol{x},t)-\bar{c}^m_{min,j}(\boldsymbol{x},t)|\le\frac{\omega(r_{1,j},h)+\omega(r_{2,j},h)}{L}(e^{LT}-1)\quad\forall t\in[0,T]\,,\ \forall\boldsymbol{x}\in\Omega$$

with $h:=\|\hat{\boldsymbol{c}}^l-\hat{\boldsymbol{c}}^m\|_{C(\overline{Q}_T)^I}$. So $(\bar{\boldsymbol{c}}^n_{min})$ converges in $C(\overline{Q}_T)^{J_{min}}$.

Now the right hand sides of the PDE system

$$\partial_t\boldsymbol{c}^n+L\boldsymbol{c}^n=\boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}((\hat{\boldsymbol{c}}^n)^+)+\boldsymbol{S}_{1,min}\boldsymbol{r}_{\varepsilon,min}((\hat{\boldsymbol{c}}^n)^+,\bar{\boldsymbol{c}}^n_{min})$$

converges in $C(\overline{Q}_T)^I$. In particular, the right hand side converges in $L^q(Q_T)^I$ for all $1\le q\le\infty$. From the linear parabolic theory we know that the sequence of solutions $(\boldsymbol{c}^n)$ then converges in $W^{2,1}_p(Q_T)^I$, $(n+2)/2<p<\infty$. □

Theorem 5.11. *The problem* $(\boldsymbol{P}^+_{\boldsymbol{\varepsilon}})$ *has a solution.*

Proof. We apply Schaefer's fixed point theorem to the operator $\mathcal{Z}$, using the a priori estimate (5.44) and Theorem 5.10, to obtain a solution of $(\boldsymbol{P}^+_{\boldsymbol{\varepsilon}})$. □

Passing to the Limit

It remains to show that a solution of $(\boldsymbol{P}^+_{\boldsymbol{\varepsilon}})$ converge to a solution of $(\boldsymbol{P}^+)$ for $\varepsilon\searrow 0$. Then it follows with remark 5.8 that it is also a solution of problem $(\boldsymbol{P})$. Thereto we show that the tuple $(\boldsymbol{c}_\varepsilon,\bar{\boldsymbol{c}}_{\varepsilon,min},\boldsymbol{w}_\varepsilon)$ with $\boldsymbol{w}_\varepsilon:=H_\varepsilon(\bar{\boldsymbol{c}}_{\varepsilon,min})$ converges for $\varepsilon\searrow 0$ to the tuple $(\boldsymbol{c},\bar{\boldsymbol{c}}_{min},\boldsymbol{w})\in W^{2,1}_p(Q_T)^I\times\mathcal{L}(Q_T)^{J_{min}}\times L^\infty(Q_T)^{J_{min}}$ with $\mathcal{L}(Q_T):=\{v\in L^\infty(Q_T)|\partial_t v\in L^\infty(Q_T)\}$ which fulfills

$$\begin{aligned}\partial_t\boldsymbol{c}+L\boldsymbol{c}&=\boldsymbol{S}_{1,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}^+)+\boldsymbol{S}_{1,min}\tilde{\boldsymbol{r}}_{min}(\boldsymbol{c}^+,\boldsymbol{w}) && \text{on } Q_T && (5.45)\\ \partial_t\bar{\boldsymbol{c}}_{min}&=\tilde{\boldsymbol{r}}_{min}(\boldsymbol{c}^+,\boldsymbol{w}) && \text{on } Q_T && (5.46)\\ \boldsymbol{w}&\in H(\bar{\boldsymbol{c}}_{min}) && \text{on } Q_T && (5.47)\\ \boldsymbol{c}(\cdot,0)&=\boldsymbol{c}_0 && \text{on } \overline{\Omega} && (5.48)\\ \bar{\boldsymbol{c}}_{min}(\cdot,0)&=\bar{\boldsymbol{c}}_{min,0} && \text{on } \overline{\Omega} && (5.49)\\ d\partial_{\boldsymbol{\nu}}\boldsymbol{c}&=\beta(\boldsymbol{c}-\boldsymbol{c}^*) && \text{on } S_T && (5.50)\end{aligned}$$

with

$$\tilde{r}_{min,j}(\boldsymbol{c}^+,\boldsymbol{w}) = k_{p,j} \prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} (c_i^+)^{-S_{1,min,ij}} - k_{d,j} \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_j$$

From the a priori estimate (5.44) we know that the $W_p^{2,1}(Q_T)^I$-norm of $\boldsymbol{c}_\varepsilon$ is bounded with a bound independent of ε. From the embedding $W_p^{2,1}(Q_T) \hookrightarrow C(\overline{Q}_T)$ we get that $\boldsymbol{c}_\varepsilon$ is also bounded in the $C(\overline{Q}_T)^I$-norm.

With

$$\partial_t \bar{c}_{\varepsilon,min,j} = r_{\varepsilon,min,j}(\boldsymbol{c}^+,\bar{\boldsymbol{c}}_{\varepsilon,min}) \begin{cases} \le k_{p,j} \prod\limits_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} (c_i^+)^{-S_{1,min,ij}} \\ \ge -k_{d,j} \prod\limits_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} \end{cases}$$

we get that $\partial_t \bar{c}_{\varepsilon,min,j}$ is bounded in the $L^\infty(Q_T)$-norm. As $\bar{\boldsymbol{c}}_{min,0} \in C^\alpha(\overline{\Omega})^{J_{min}}$ also $\bar{c}_{\varepsilon,min,j}$ is bounded in the $L^\infty(Q_T)$-norm. Because of the definition of $\boldsymbol{w}_\varepsilon$ we have $0 \le w_{\epsilon,j} \le 1$.

By passing to a subsequence if necessary we see that:

$$\begin{aligned}
\boldsymbol{c}_\varepsilon &\longrightarrow \boldsymbol{c} && \text{weakly in } W_p^{2,1}(Q_T)^I \\
\boldsymbol{c}_\varepsilon &\longrightarrow \boldsymbol{c} && \text{strongly in } C(\overline{Q}_T)^I \\
\bar{c}_{\varepsilon,min,j} &\longrightarrow \bar{c}_{min,j} && \text{weakly-star in } L^\infty(Q_T) \quad (j=1,\dots,J_{min}) \\
\partial_t\bar{c}_{\varepsilon,min,j} &\longrightarrow \partial_t\bar{c}_{min,j} && \text{weakly-star in } L^\infty(Q_T) \quad (j=1,\dots,J_{min}) \\
w_{\varepsilon,j} &\longrightarrow w_j && \text{weakly-star in } L^\infty(Q_T) \quad (j=1,\dots,J_{min})
\end{aligned}$$

For the equations (5.45)-(5.50) except of (5.47) it is obvious that the limits fulfill these equations. To show (5.47) we adapt some ideas of the proof of [vDP04, Thm. 2.21]. First we introduce

$$\underline{\bar{c}}_{min,j}(\boldsymbol{x},t) := \liminf_{\varepsilon \searrow 0} \bar{c}_{\varepsilon,min,j}(\boldsymbol{x},t) \ge 0 \qquad \text{a.e. in } Q_T$$

and decompose $Q_T = S_{j,1} \cup S_{j,2}$, where (in the almost everywhere sense)

$$S_{j,1} = \{\underline{\bar{c}}_{min,j} > 0\} \text{ and } S_{j,2} = \{\underline{\bar{c}}_{min,j} = 0\}\,.$$

We will show that $\bar{c}_{min,j} > 0$ and $w_j = 1$ in $S_{j,1}$, while $\bar{c}_{min,j} = 0$ and $w_j \in [0,1]$ in $S_{j,2}$.

Because of $\bar{c}_{min,j} \geq \underline{\bar{c}}_{min,j}$ it follows that $\bar{c}_{min,j} > 0$ in $S_{j,1}$. Then we choose $(\boldsymbol{x}, t) \in S_{j,1}$ such that $\underline{\bar{c}}_{min,j}(\boldsymbol{x}, t) > 2\mu > 0$ for μ sufficiently small. So we have $\bar{c}_{\varepsilon,min,j}(\boldsymbol{x}, t) > \mu$ and $w_\varepsilon(\boldsymbol{x}, t) = 1$ for all ε small enough. Hence it holds $w(\boldsymbol{x}, t) = 1$.

Now we exclude that $\bar{c}_{min,j} > 0$ in $S_{j,2}$. As $\boldsymbol{c}$ is bounded in $L^\infty(Q_T)^I$ we have

$$\int_0^t \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_{\varepsilon,j}\, ds \to \int_0^t \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_j\, ds$$

weakly-star in $L^\infty(Q_T)$, It follows

$$\liminf_{\varepsilon \searrow 0} \int_0^t \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_{\varepsilon,j}\, ds \leq \int_0^t \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_j\, ds \quad \text{a.e. in } Q_T\,.$$

Using (5.36) and (5.46), that are valid a.e. in Q_T, and integrating in time gives a.e. in Q_T

$$\begin{aligned}
\bar{c}_{\varepsilon,min,j} &= \bar{c}_{min,0,j} + \int_0^t \tilde{r}_{min,j}(\boldsymbol{c}_\varepsilon^+, \boldsymbol{w}_\varepsilon)\, ds \\
&= \bar{c}_{min,j} + \int_0^t \tilde{r}_{min,j}(\boldsymbol{c}_\varepsilon^+, \boldsymbol{w}_\varepsilon)\, ds - \int_0^t \tilde{r}_{min,j}(\boldsymbol{c}^+, \boldsymbol{w})\, ds \\
&= \bar{c}_{min,j} + \int_0^t k_{p,j} \Bigg(\prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} (c_{\varepsilon,i}^+)^{-S_{1,min,ij}} - \prod_{\substack{i=1\\ S_{1,min,ij}<0}}^{I} (c_i^+)^{-S_{1,min,ij}} \Bigg)\, ds \\
&\qquad - \int_0^t k_{d,j} \Bigg(\prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_{\varepsilon,i}^+)^{S_{1,min,ij}} w_{\varepsilon,j} - \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_{\varepsilon,j} \Bigg)\, ds \\
&\qquad - \int_0^t k_{d,j} \Bigg(\prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_{\varepsilon,j} - \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_j \Bigg)\, ds\,.
\end{aligned}$$

We consider this identity on $S_{j,2}$. First we take the $\liminf_{\varepsilon \searrow 0}$ of this relation. As $\boldsymbol{c}_\varepsilon$ converges pointwisely and $w_{\varepsilon.j}$ is bounded in $L^\infty(Q_T)$ it holds

$$\begin{aligned}
&0 = \bar{c}_{min,j} - k_{d,j} \liminf_{\varepsilon \searrow 0} \int_0^t \Bigg(\prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_{\varepsilon,j} - \prod_{\substack{i=1\\ S_{1,min,ij}>0}}^{I} (c_i^+)^{S_{1,min,ij}} w_j \Bigg)\, ds \\
&\quad \geq \bar{c}_{min,j} \qquad \text{a.e. in } S_2\,.
\end{aligned}$$

Therefore $\bar{c}_{min,j} = 0$ in $S_{j,2}$. $0 \leq w \leq 1$ is valid because $0 \leq w_\varepsilon \leq 1$ for all ε.

Appendix A

A.1 Link between the variables of the Morel formulation and the reduction scheme in the case no kinetic reactions

Without kinetic reactions the standard form of the stoichiometric matrix $\boldsymbol{S}$ is

$$\boldsymbol{S} = \left(\begin{array}{ccc} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \\ \hline \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{array}\right)$$

$$\sim \left(\begin{array}{ccc} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \hat{\boldsymbol{B}} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} \end{array}\right) = \left(\begin{array}{ccc} \boldsymbol{C}_1 & \boldsymbol{A}_1 & \boldsymbol{D}_1 \\ \boldsymbol{C}_2 & \boldsymbol{A}_2 & \boldsymbol{D}_2 \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \hat{\boldsymbol{B}} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} \end{array}\right)$$

where the blocks $\boldsymbol{A}_i$ have the substructure (with $\boldsymbol{A}_{ld}$ out of (3.2))

$$\begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{1,li} & \boldsymbol{D}_1\boldsymbol{A}_{ld} \\ \boldsymbol{A}_{2,li} & \boldsymbol{D}_2\boldsymbol{A}_{ld} \end{pmatrix}$$

and with $\begin{pmatrix} \boldsymbol{A}_{2,li} & \boldsymbol{D}_2 \end{pmatrix}$ invertible.

Using the transformed stoichiometric matrix the matrix $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$, consisting of the linear independent columns of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively, are of the form

$$\boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{E}_1 \\ \boldsymbol{C}_2 & \boldsymbol{E}_2 \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \end{pmatrix}, \qquad \boldsymbol{S}_2^* = \begin{pmatrix} \hat{\boldsymbol{B}} & \boldsymbol{0} \\ -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} \end{pmatrix}$$

with the abbreviations $\boldsymbol{E}_1 := \begin{pmatrix} \boldsymbol{A}_{1,li} & \boldsymbol{D}_1 \end{pmatrix}$ and $\boldsymbol{E}_2 := \begin{pmatrix} \boldsymbol{A}_{2,li} & \boldsymbol{D}_2 \end{pmatrix}$. The entries of the concentrations vectors $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$ are partitioned analogously to the rows of $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$, respectively,

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_{prim,1} \\ \boldsymbol{c}_{prim,2} \\ \boldsymbol{c}_{sec} \end{pmatrix}, \qquad \bar{\boldsymbol{c}} = \begin{pmatrix} \bar{\boldsymbol{c}}_{nmin,prim} \\ \bar{\boldsymbol{c}}_{nmin,sec} \\ \bar{\boldsymbol{c}}_{min} \end{pmatrix}.$$

It is useful to choose as matrix $\boldsymbol{S}_1^\perp$, consisting of a maximal system of linear independent vectors that are orthogonal to all columns of $\boldsymbol{S}_1^*$, the following one

$$\boldsymbol{S}_1^\perp = \begin{pmatrix} \boldsymbol{I}_{I-J_{mob}-J_{sorp,li}-J_{min}} \\ -{\boldsymbol{E}_2}^{-T}{\boldsymbol{E}_1}^T \\ \boldsymbol{C}_1^T - \boldsymbol{C}_2^T{\boldsymbol{E}_2}^{-T}{\boldsymbol{E}_1}^T \end{pmatrix}.$$

Calculating $(\boldsymbol{S}_1^*)^T\boldsymbol{S}_1^\perp$ one can see easily that the columns of $\boldsymbol{S}_1^\perp$ are orthogonal to those of $\boldsymbol{S}_1^*$. Furthermore it is useful to choose the following transformation matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_1^\perp$

$$\boldsymbol{B}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{sorp,li}+J_{min}} \\ \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \end{pmatrix}, \qquad \boldsymbol{B}_1^\perp = \begin{pmatrix} \boldsymbol{I}_{I-J_{mob}-J_{sorp,li}-J_{min}} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}.$$

The condition that the columns of $\boldsymbol{B}_1$, $\boldsymbol{S}_1^\perp$ form a basis of the whole space is fulfilled. The standard form of the stoichiometric matrix is constructed in such a way that $\boldsymbol{E}_2$ is invertible. Hence also the inverse of $\boldsymbol{B}_1^T\boldsymbol{S}_1^* = \begin{pmatrix} -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \\ \boldsymbol{C}_2 & \boldsymbol{E}_2 \end{pmatrix}$ exists.

For the inverse we get

$$(\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1} = \begin{pmatrix} -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{E}_2^{-1}\boldsymbol{C}_2 & \boldsymbol{E}_2^{-1} \end{pmatrix}.$$

Using this we get for the transformed variables $\boldsymbol{\xi}$ (compare (3.13), (3.14))

$$\begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \begin{pmatrix} \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \end{pmatrix} \end{pmatrix} = (\boldsymbol{B}_1^T\boldsymbol{S}_1^*)^{-1}\boldsymbol{B}_1^T\boldsymbol{c} = \begin{pmatrix} -\boldsymbol{c}_{sec} \\ \boldsymbol{E}_2^{-1}\boldsymbol{c}_{prim,2} + \boldsymbol{E}_2^{-1}\boldsymbol{C}_2\boldsymbol{c}_{sec} \end{pmatrix}. \tag{A.1}$$

And for the transformed variables $\boldsymbol{\eta}$ we have (compare (3.13))

$$\begin{aligned} \boldsymbol{\eta} &= \left((\boldsymbol{S}_1^\perp)^T\boldsymbol{B}_1^\perp\right)^{-1}(\boldsymbol{S}_1^\perp)^T\boldsymbol{c} \\ &= \boldsymbol{c}_{prim,1} - \boldsymbol{E}_1\boldsymbol{E}_2^{-1}\boldsymbol{c}_{prim,2} + \left(\boldsymbol{C}_1 - \boldsymbol{E}_1\boldsymbol{E}_2^{-1}\boldsymbol{C}_2\right)\boldsymbol{c}_{sec}. \end{aligned}$$

Now we consider the immobile species. As a basis of the orthogonal complement of $\boldsymbol{S}_2^*$ we choose

$$\boldsymbol{S}_2^{\perp} = \begin{pmatrix} \boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}} \\ \hat{\boldsymbol{B}}^T \\ \boldsymbol{0} \end{pmatrix} .$$

Like in the case of mobile species it is easy to see that this matrix is orthogonal to $\boldsymbol{S}_2^*$. Here it is useful to choose as transformation matrices

$$\boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{pmatrix} , \qquad \boldsymbol{B}_2^{\perp} = \begin{pmatrix} \boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} .$$

Like in the case of mobile species it is obvious that $\boldsymbol{B}_2$ and $\boldsymbol{S}_2^{\perp}$ form a basis of the whole space and due to the construction of the standard form of the stoichiometric matrix the inverse of $(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}$ exists.

One can compute that

$$(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1} = \begin{pmatrix} -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} \end{pmatrix} .$$

Using this we get for the transformed variables $\bar{\boldsymbol{\xi}}$ (compare (3.13), (3.14))

$$\begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix} = (\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}} = \begin{pmatrix} -\bar{\boldsymbol{c}}_{nmin,sec} \\ -\bar{\boldsymbol{c}}_{min} \end{pmatrix} . \tag{A.2}$$

Moreover we get for the transformed variables $\bar{\boldsymbol{\eta}}$ (compare (3.13))

$$\bar{\boldsymbol{\eta}} = \left((\boldsymbol{S}_2^{\perp})^T\boldsymbol{B}_2^{\perp}\right)^{-1}(\boldsymbol{S}_2^{\perp})^T\bar{\boldsymbol{c}} = \bar{\boldsymbol{c}}_{nmin,prim} + \hat{\boldsymbol{B}}\bar{\boldsymbol{c}}_{nmin,sec} .$$

Now we define the additional variables $\tilde{\boldsymbol{\xi}}$ (compare (3.39))

$$\tilde{\boldsymbol{\xi}} = \begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\xi}_{sorp} - \bar{\boldsymbol{\xi}}_{sorp,li} \\ \boldsymbol{\xi}_{min} - \bar{\boldsymbol{\xi}}_{min} - \boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} \end{pmatrix} .$$

Then we take the linear combinations $\boldsymbol{\eta} + \boldsymbol{E}_1\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{E}_2\tilde{\boldsymbol{\xi}}$. Thereto we have to compute $(i = 1, 2)$:

$$\boldsymbol{E}_i\tilde{\boldsymbol{\xi}} = \boldsymbol{E}_i\begin{pmatrix} \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \end{pmatrix} - \boldsymbol{E}_i\begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp,li} \\ \bar{\boldsymbol{\xi}}_{min} + \boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} \end{pmatrix}$$

First we consider the second summand. Using the definitions of the matrices $\boldsymbol{E}_i = \begin{pmatrix} \boldsymbol{A}_{i,li} & \boldsymbol{D}_i \end{pmatrix}$ we see that the second summand is

$$\boldsymbol{A}_{i,li}\bar{\boldsymbol{\xi}}_{sorp,li} + \boldsymbol{D}_i\bar{\boldsymbol{\xi}}_{min} + \boldsymbol{D}_i\boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} .$$

With help of the substructure of $\boldsymbol{A}_i = \begin{pmatrix}\boldsymbol{A}_{i,li} & \boldsymbol{D}_i\boldsymbol{A}_{ld}\end{pmatrix}$ we get

$$= \boldsymbol{A}_i\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{D}_i\bar{\boldsymbol{\xi}}_{min}.$$

Plugging in the definitions of the variables $\bar{\boldsymbol{\xi}}$ gives

$$= -\boldsymbol{A}_i\bar{\boldsymbol{c}}_{nmin,sec} - \boldsymbol{D}_i\bar{\boldsymbol{c}}_{min}\,.$$

Using this and the definition of $\boldsymbol{\xi}$ we get in summary

$$\begin{aligned}\boldsymbol{E}_2\tilde{\boldsymbol{\xi}} &= \boldsymbol{c}_{prim,2} + \boldsymbol{C}_2\boldsymbol{c}_{sec} + \boldsymbol{D}_2\bar{\boldsymbol{c}}_{min} + \boldsymbol{A}_2\bar{\boldsymbol{c}}_{nmin,sec}\\ \boldsymbol{E}_1\tilde{\boldsymbol{\xi}} &= \boldsymbol{E}_1\boldsymbol{E}_2^{-1}(\boldsymbol{c}_{prim,2} + \boldsymbol{C}_2\boldsymbol{c}_{sec}) + \boldsymbol{D}_1\bar{\boldsymbol{c}}_{min} + \boldsymbol{A}_1\bar{\boldsymbol{c}}_{nmin,sec}\,.\end{aligned}$$

Adding the definition of $\boldsymbol{\eta}$ to the last one of the two equations leads to

$$\boldsymbol{\eta} + \boldsymbol{E}_1\tilde{\boldsymbol{\xi}} = \boldsymbol{c}_{prim,1} + \boldsymbol{C}_1\boldsymbol{c}_{sec} + \boldsymbol{D}_1\bar{\boldsymbol{c}}_{min} + \boldsymbol{A}_1\bar{\boldsymbol{c}}_{nmin,sec}\,.$$

Comparing with the definition of the total concentrations $\boldsymbol{T}$ (3.97) and the total fixed concentrations $\boldsymbol{W}$ (3.98) one sees that

$$\begin{pmatrix}\boldsymbol{\eta} + \boldsymbol{E}_1\tilde{\boldsymbol{\xi}}\\ \boldsymbol{E}_2\tilde{\boldsymbol{\xi}}\end{pmatrix} = \boldsymbol{T} \tag{A.3}$$

$$\bar{\boldsymbol{\eta}} = \boldsymbol{W}. \tag{A.4}$$

A.2 Applying the Maximum Principle

In [LSU68, I, Thm. 2.2/2.3] the boundary conditions are of the form

$$\left(\sum_{i=1}^{n} b_i(\boldsymbol{x},t)\partial_i u + b(\boldsymbol{x},t)u\right)\Big|_{S_T} = \psi(\boldsymbol{s},t)\,.$$

Using the boundary conditions of Section 5.4 we have

$$b_i = d\nu_i\,, \qquad\qquad b = -\beta\,.$$

One assumption of [LSU68, I, Thm. 2.2] is $b|_{S_T} > 0$, which is not fulfilled for $\beta = 0$. In [LSU68, I, Thm. 2.3] there is the assumption $b|_{S_T} \geq -b_0$ with $b_0 = \text{const} \geq 0$, which is fulfilled for $\beta = 0$. But in [LSU68, I, Thm. 2.3] the assertion is simplified and depends on $|f|$ while in [LSU68, I, Thm. 2.2] the assertion depends only on $\max f$. The problem is that we only know that $f \leq 0$ but we do not know a lower bound for f. So we have to look inside the proof of [LSU68, I, Thm. 2.3]. The proof of [LSU68, I, Thm. 2.3] is done by applying [LSU68, I, Thm. 2.2] to the function $w(\boldsymbol{x},t) := u(\boldsymbol{x},t)\varphi(\boldsymbol{x})$ with $\varphi \in O^2(\overline{\Omega})$ ($O^2(\overline{\Omega})$ is the set of all continuous functions in $\overline{\Omega}$ having continuous derivatives in $\overline{\Omega}$ up to order 1, with the derivatives of order 1 having a first differential at each point of $\overline{\Omega}$ and the derivatives of order 2 being bounded in $\overline{\Omega}$) a function that satisfies

$$\min_{\Omega} \varphi(\boldsymbol{x}) \geq \frac{1}{2}\,, \qquad \varphi|_{\partial\Omega} = 1\,, \qquad -\partial_{\boldsymbol{\nu}}\varphi|_{\partial\Omega} = m$$

where $m = \text{const} > b_0/\delta$ with δ out of Assumptions 5.5 (i). So we get for any $t_1 \in [0,T]$

$$w(\boldsymbol{x},t_1) \leq \inf_{\lambda > a_0} \max\left\{0; \max_{S_{t_1}} \frac{\psi\varphi e^{\lambda(t_1-t)}}{b - b_i\frac{\partial_i\varphi}{\varphi}}; e^{\lambda t_1}\max_{\Omega} w(\boldsymbol{x},0); \frac{1}{\lambda - a_0}\max_{Q_{t_1}} f e^{\lambda(t_1-t)}\right\}$$

with $a_0 = \max_{Q_T}(-a(\boldsymbol{x},t))$ where a is the 0th order coefficient of the PDE.

Now let u be the solution of (5.43). Because of the boundary condition $d\partial_{\boldsymbol{\nu}}u = \beta u$ on S_T it holds $\psi \equiv 0$, because of the initial condition $u(\cdot,0) = 0$ on $\overline{\Omega}$ it holds $w(\boldsymbol{x},0) = 0$ for all $\boldsymbol{x} \in \Omega$ and as there is no 0th order term in the PDE we have $a_0 = 0$. This yields

$$w(\boldsymbol{x},t_1) \leq \inf_{\lambda > 0} \max\left\{0; \frac{1}{\lambda}\max_{Q_{t_1}} f e^{\lambda(t_1-t)}\right\}\,.$$

As the right hand side f of the PDE for u is nonpositive it follows $w(\boldsymbol{x},t_1) \leq 0$. Because of $\varphi \geq 1/2$ this yields

$$u(\boldsymbol{x},t) \leq 0\,.$$

This is the assertion needed to derive the a priori estimate.

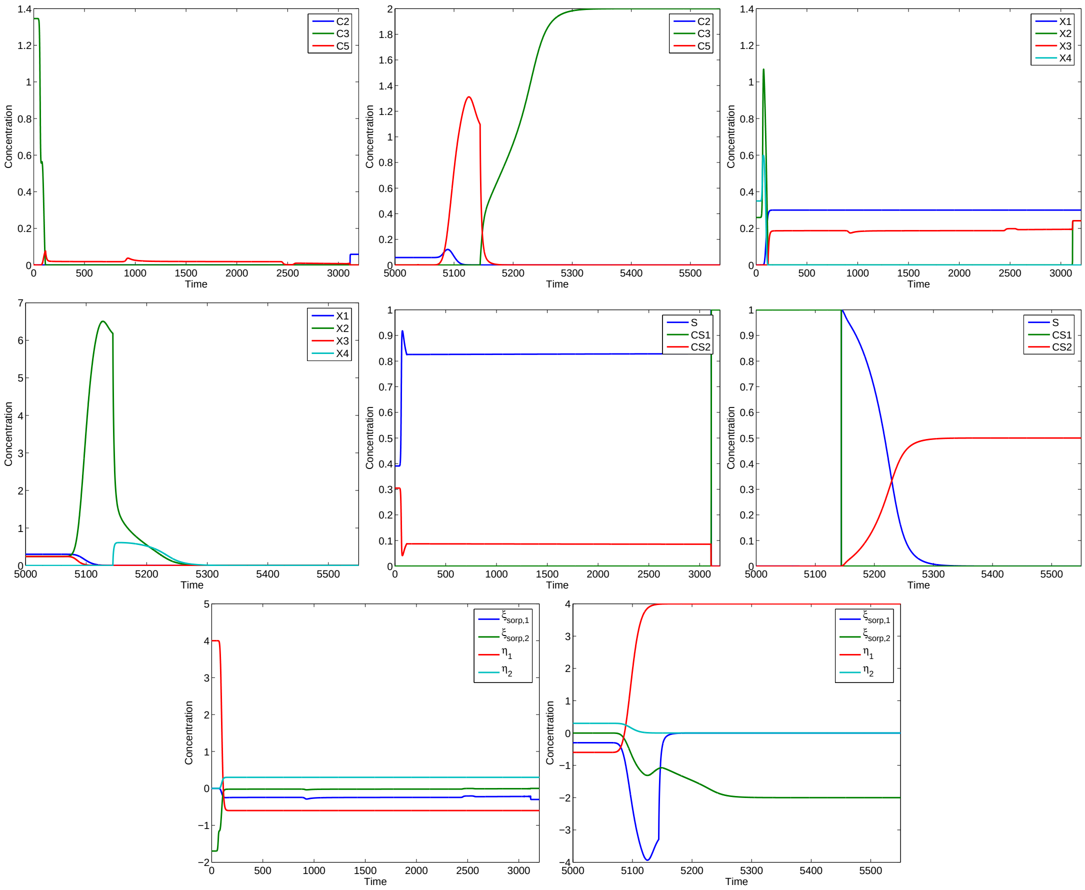
Appendix B

Results MoMaS–Benchmark

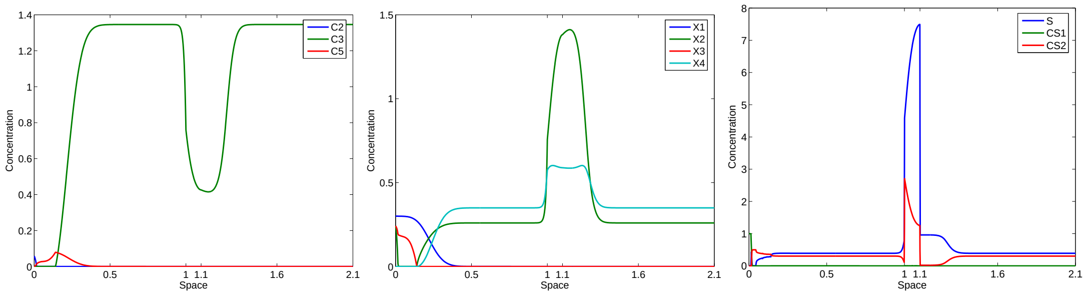
B.1 Easy Test Cases

B.1.1 Advective Easy 1D

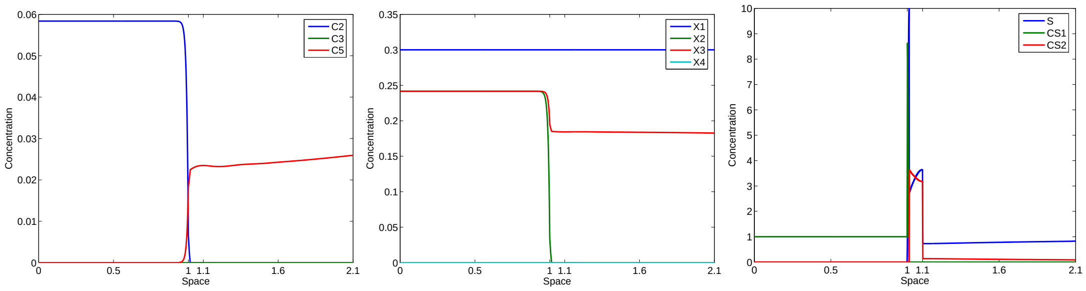
Elution curves



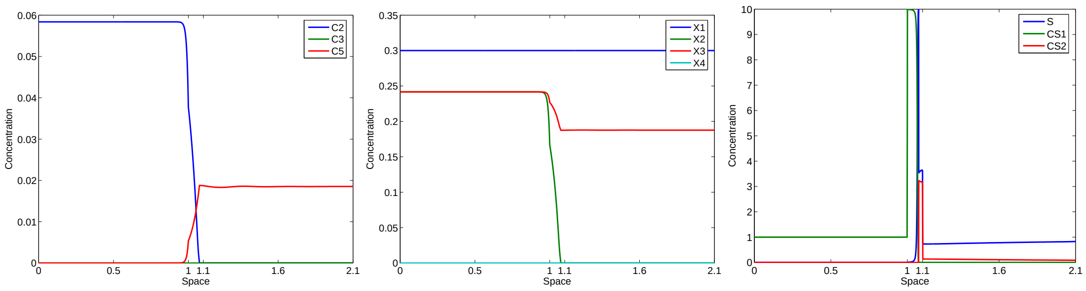
Concentration profile at $t = 10$



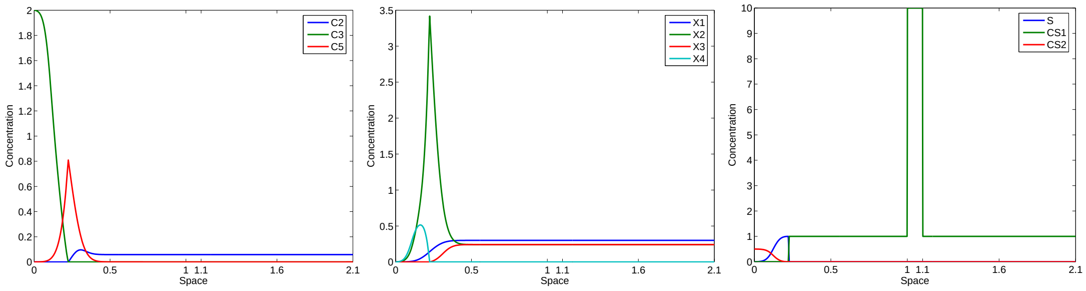
Concentration profile at $t = 1000$



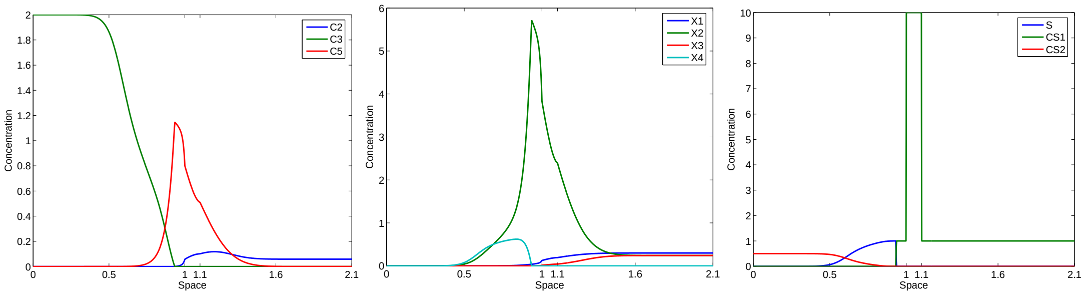
Concentration profile at $t = 2000$



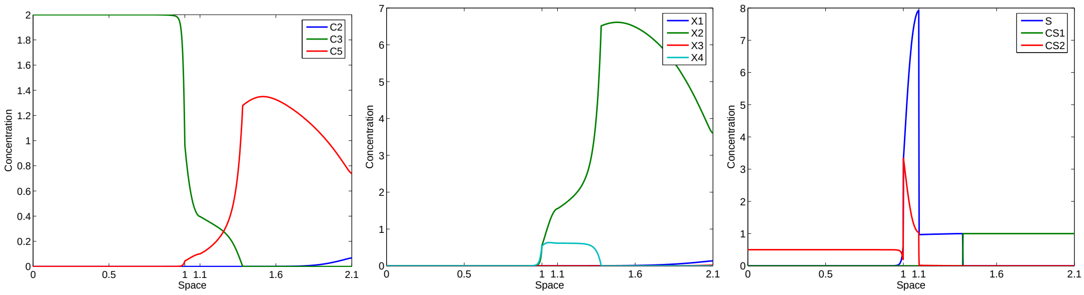
Concentration profile at $t = 5010$



Concentration profile at $t = 5050$

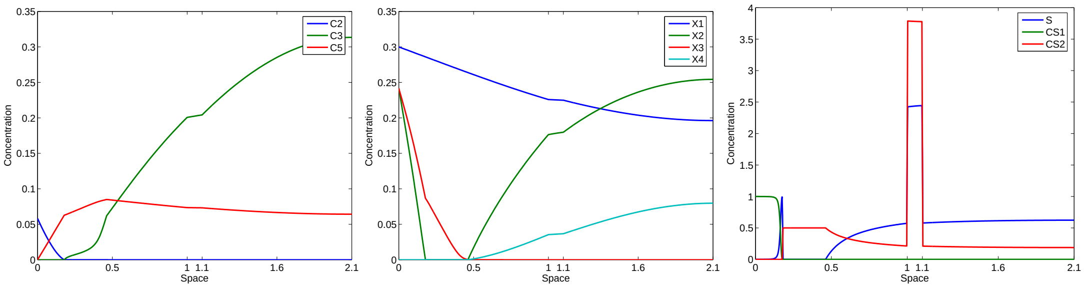


Concentration profile at $t = 5100$

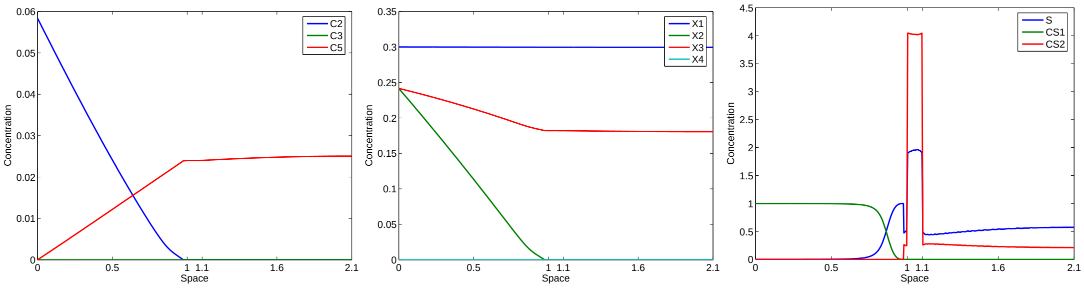


B.1.2 Diffusive Easy 1D

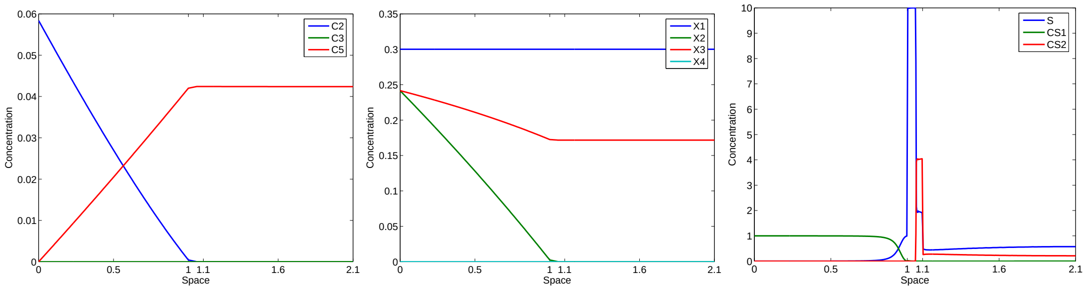
Concentration profile at $t = 10$



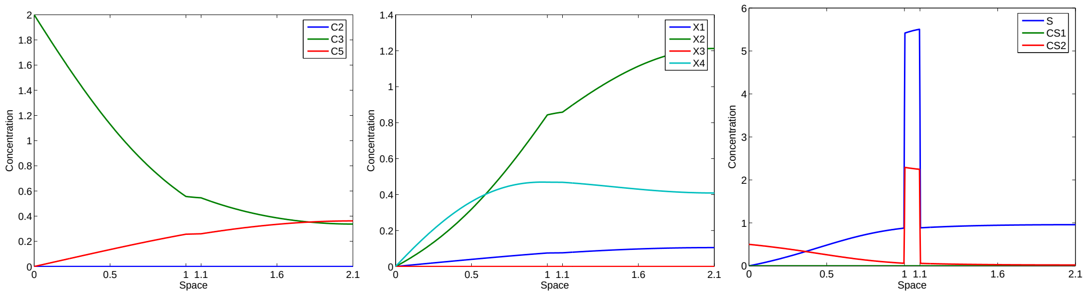
Concentration profile at $t = 50$



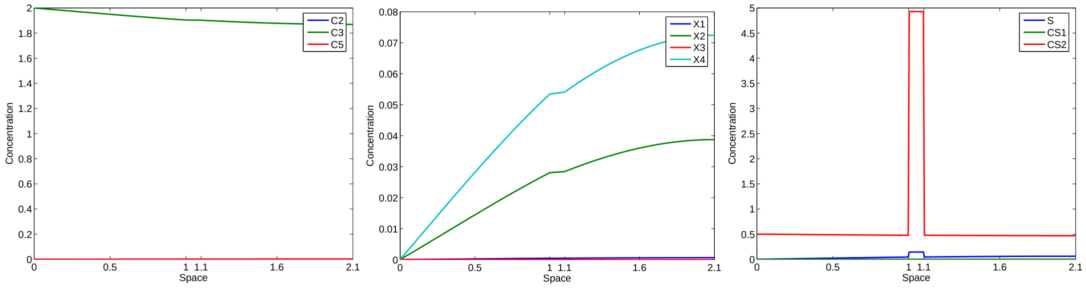
Concentration profile at $t = 100$



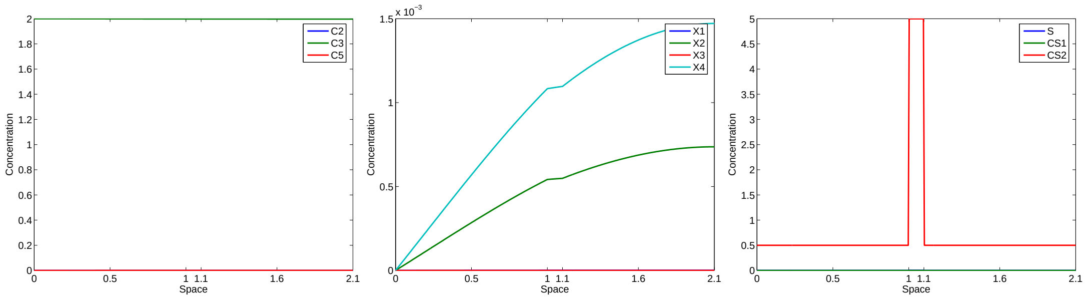
Concentration profile at $t = 5010$



Concentration profile at $t = 5050$

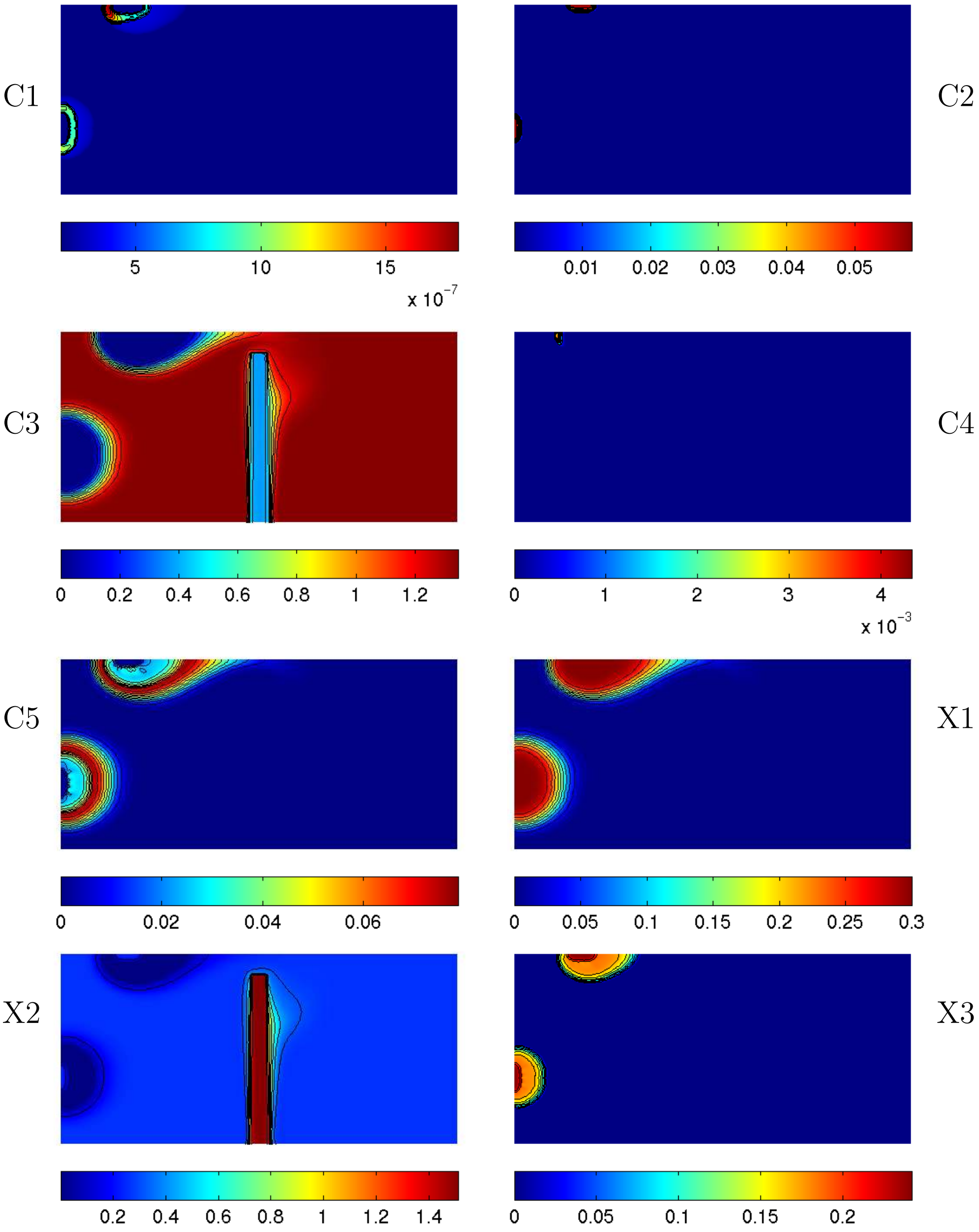


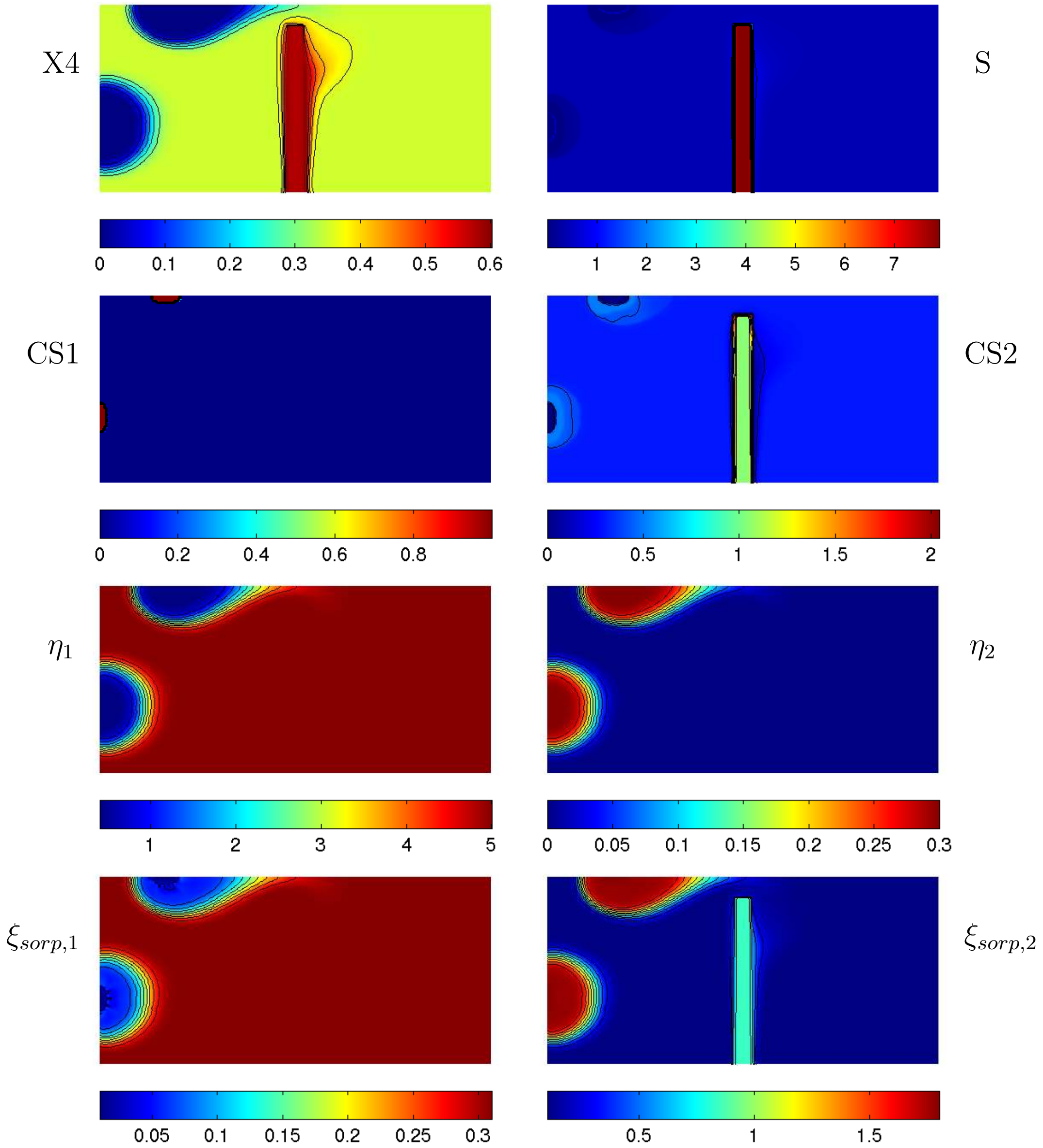
Concentration profile at $t = 5100$



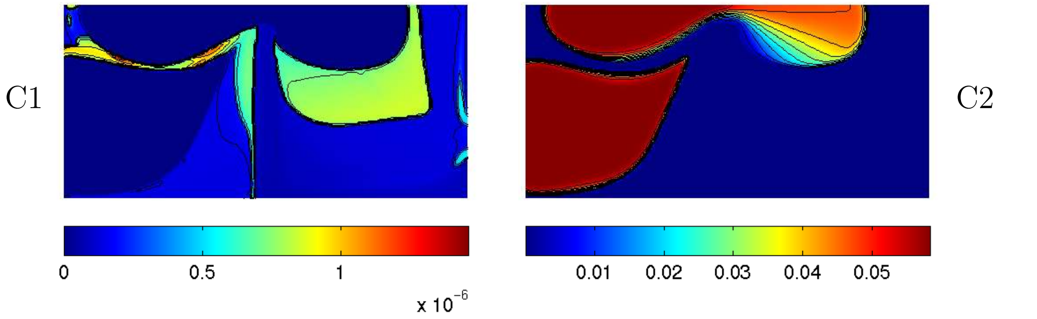
B.1.3 Advective Easy 2D

Concentration profile at $t = 10$

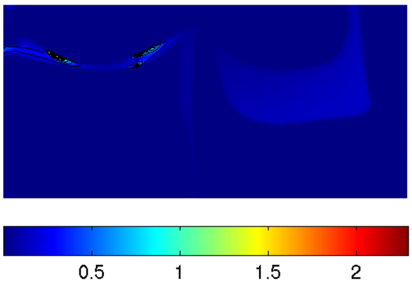




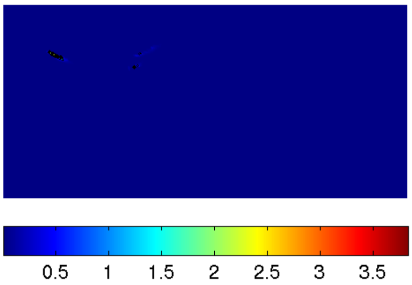
Concentration profile at $t = 1000$



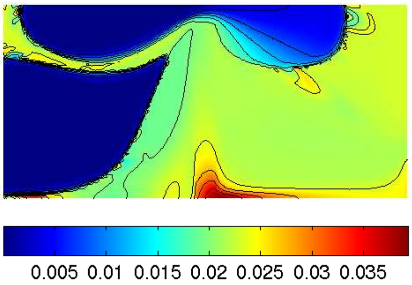
C3



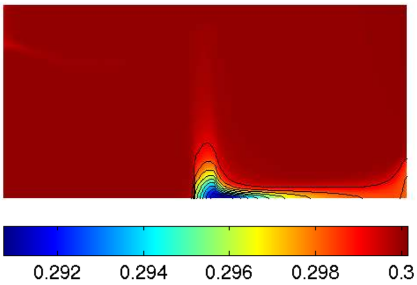
C4



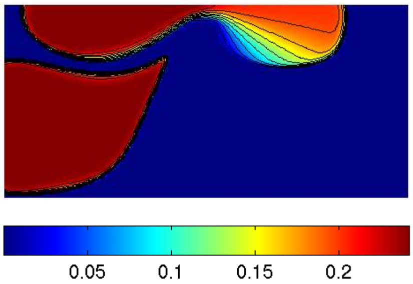
C5



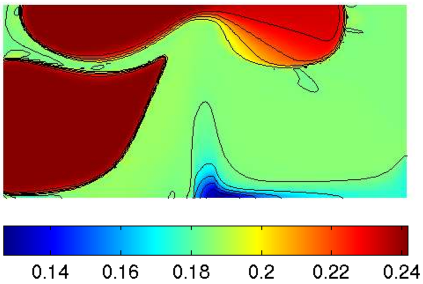
X1



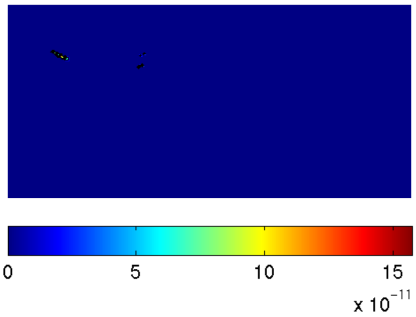
X2



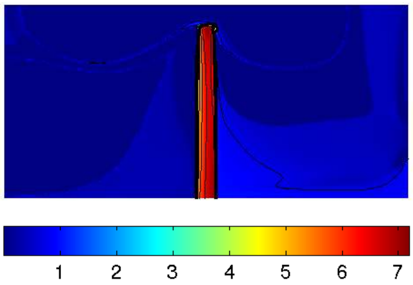
X3



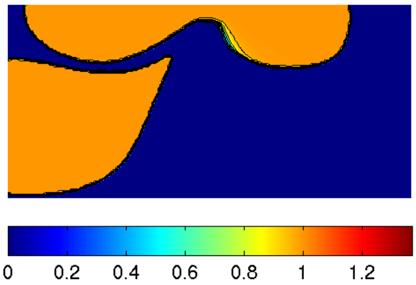
X4



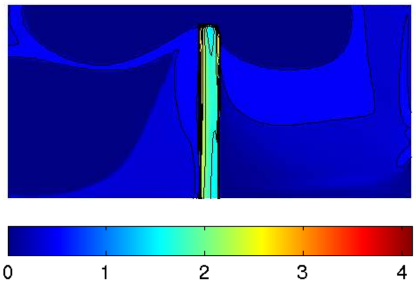
S

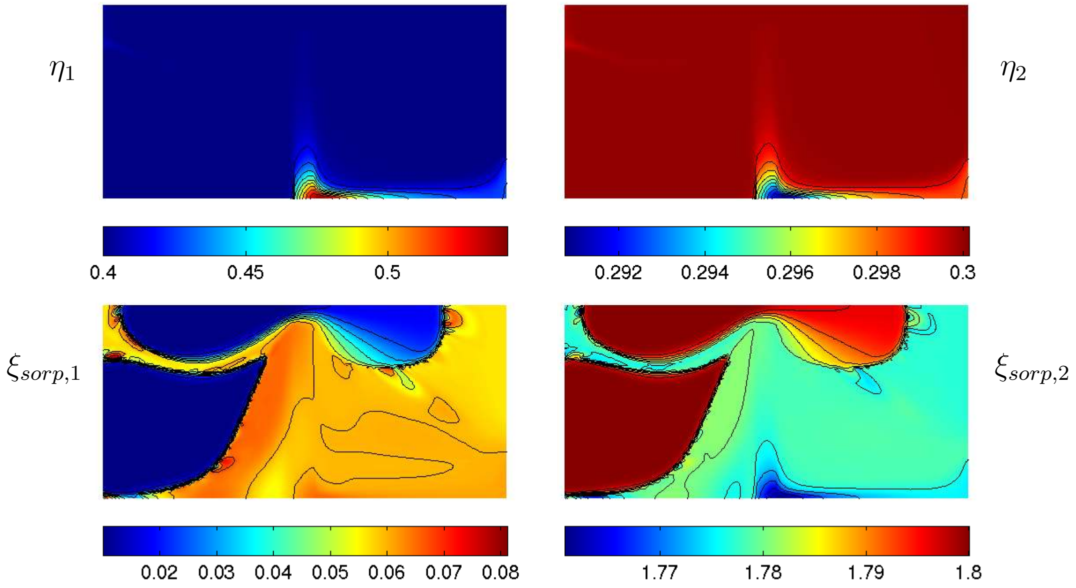


CS1

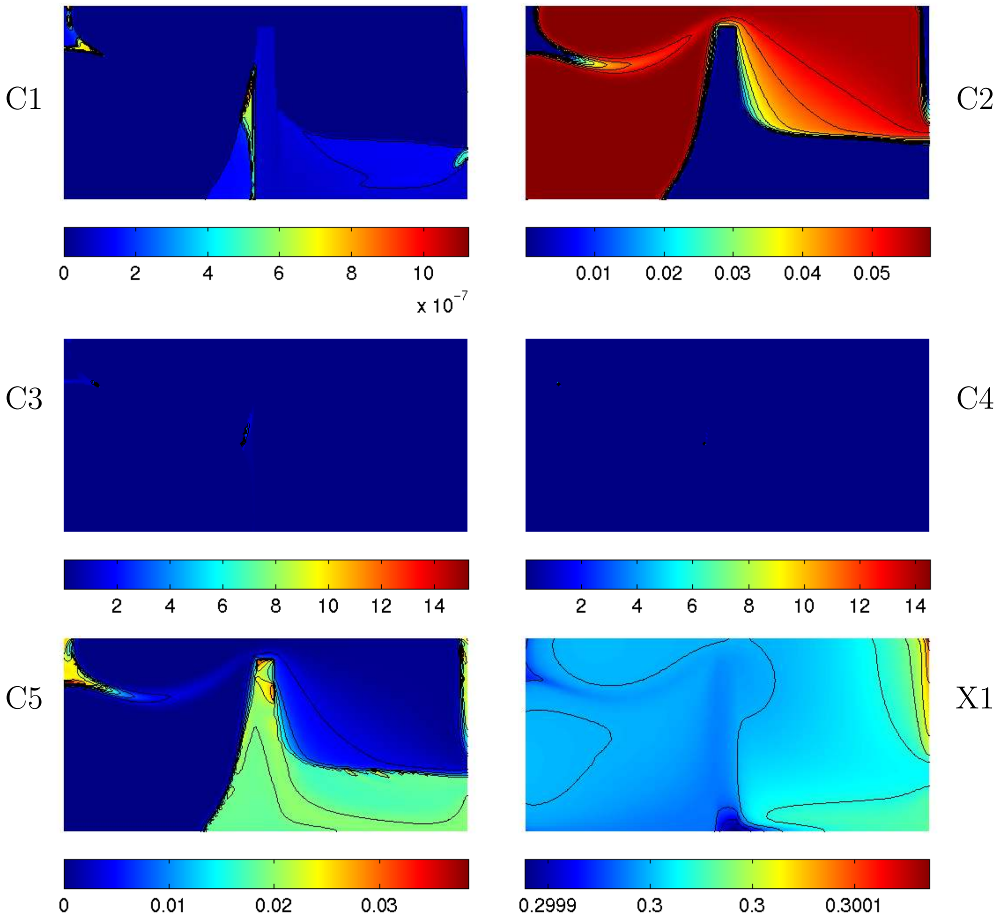


CS2

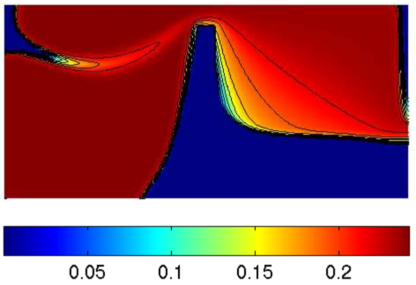


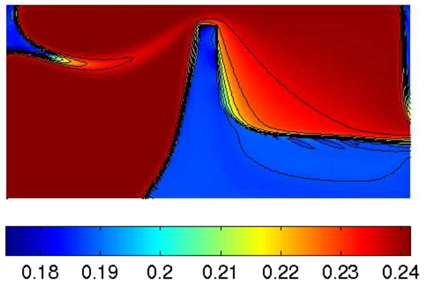


Concentration profile at $t = 2000$



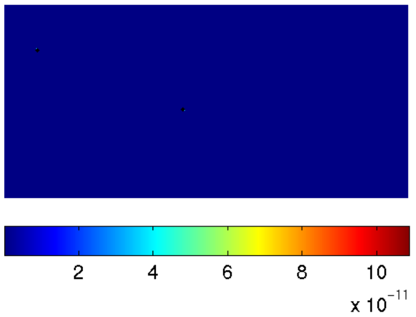
X2

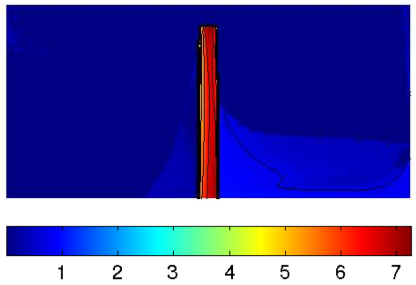




X3

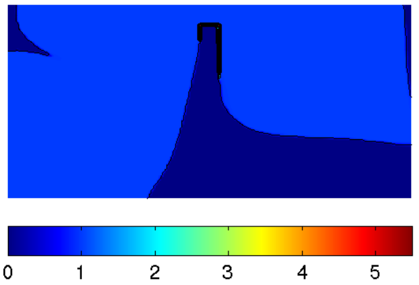
X4

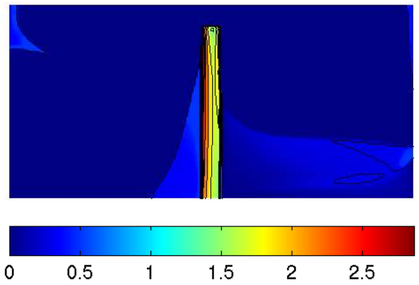




S

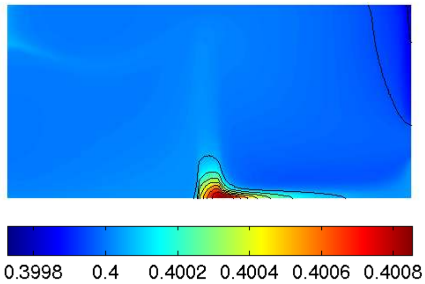
CS1

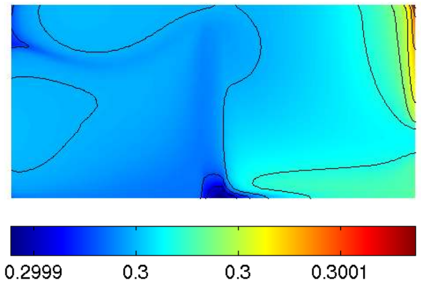




CS2

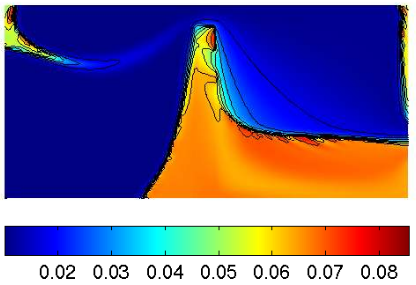
η_1

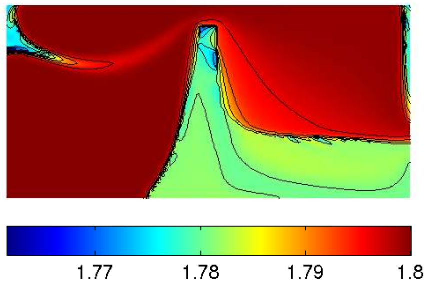




η_2

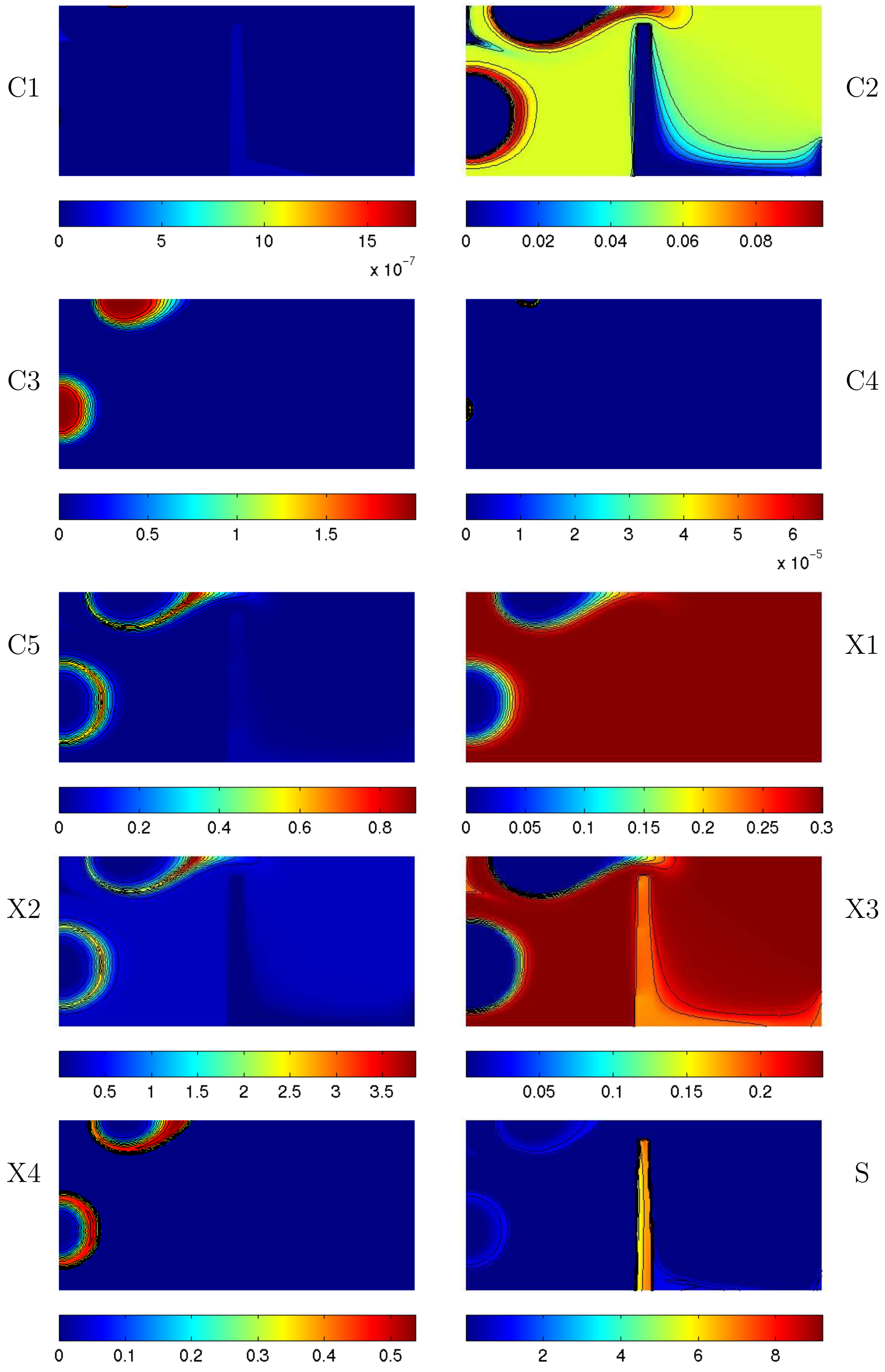
$\xi_{sorp,1}$

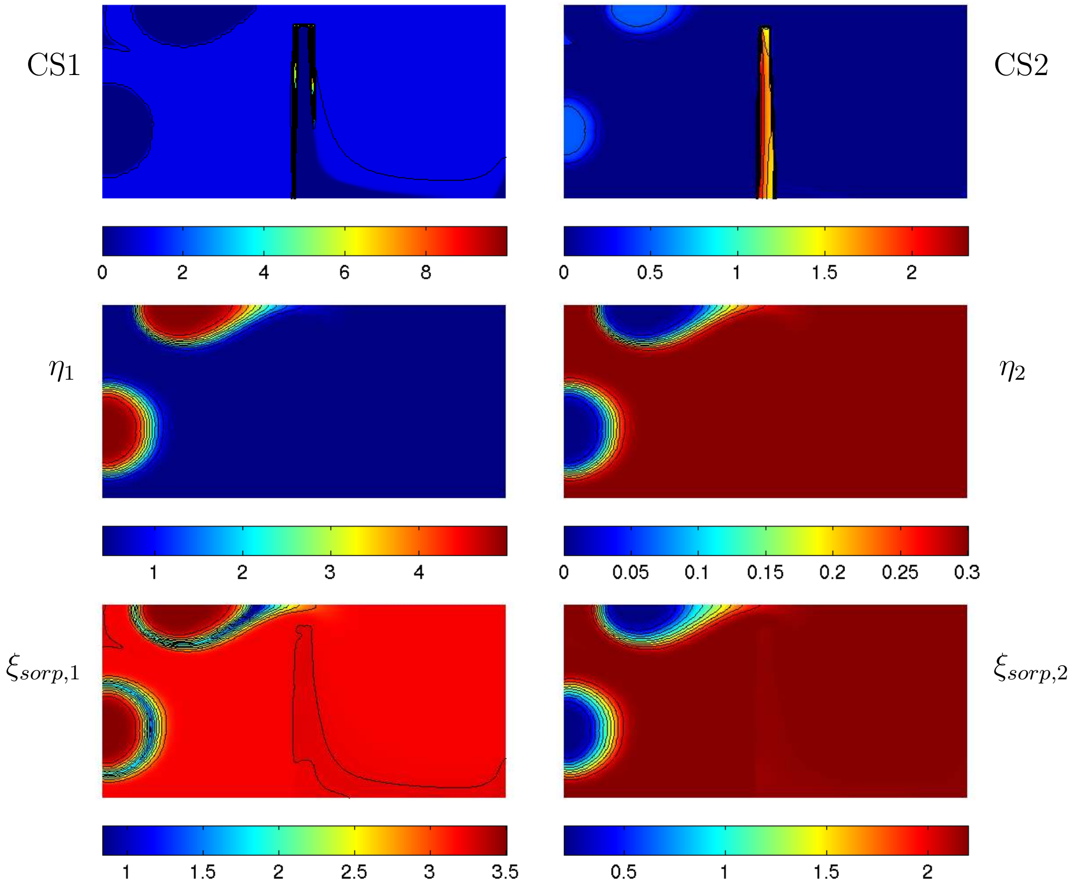




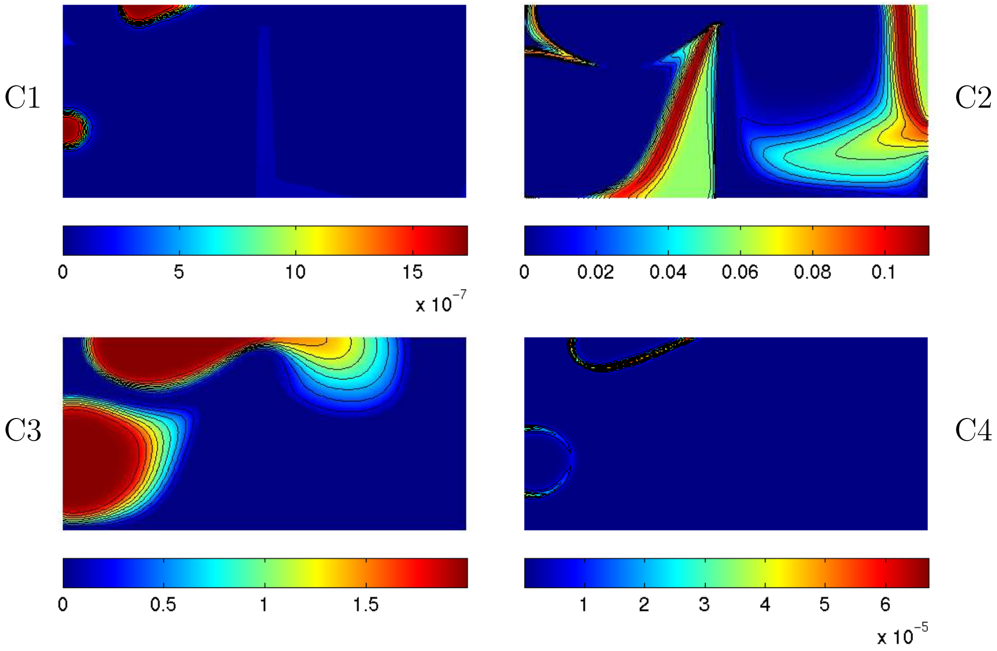
$\xi_{sorp,2}$

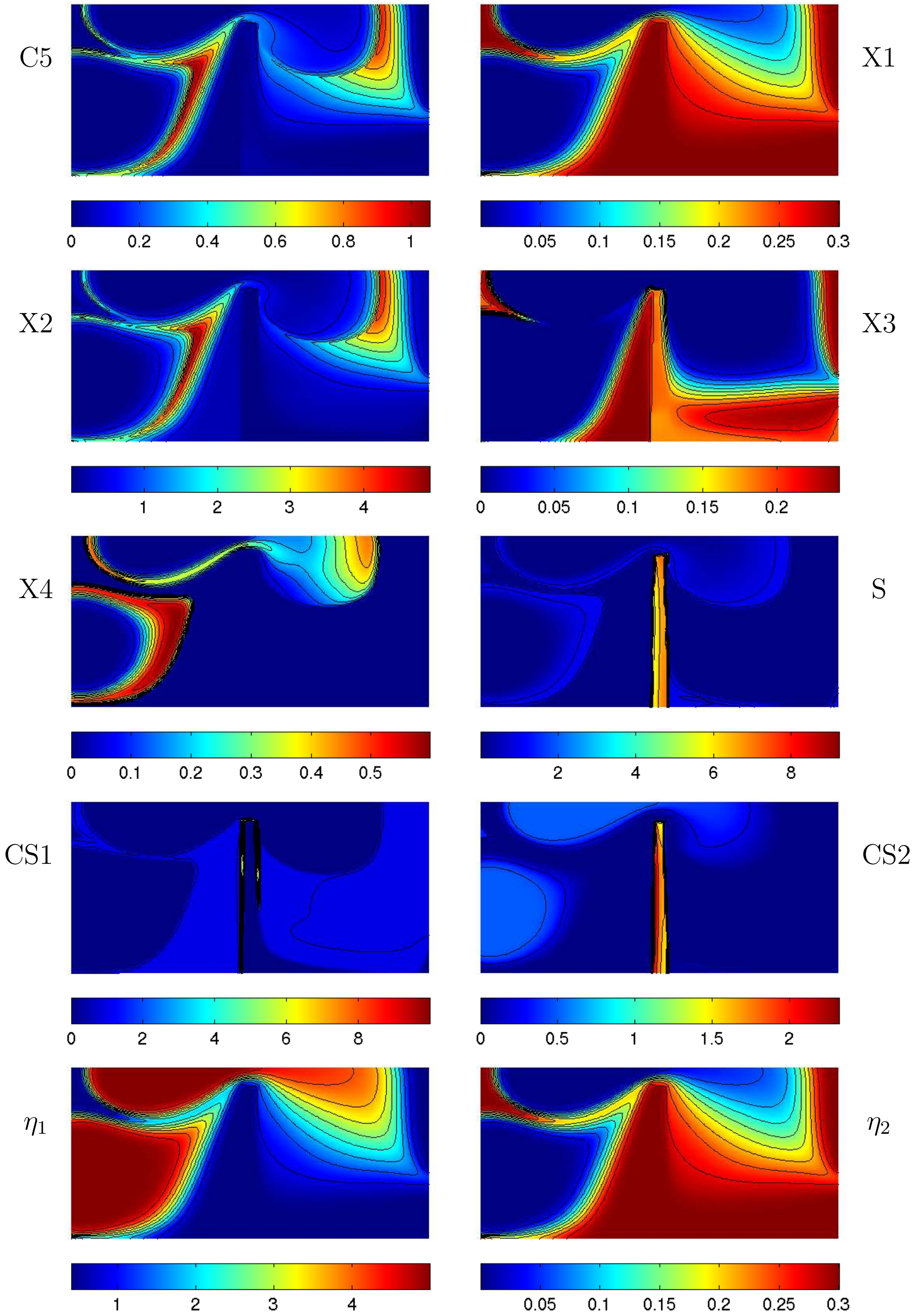
Concentration profile at $t = 5010$

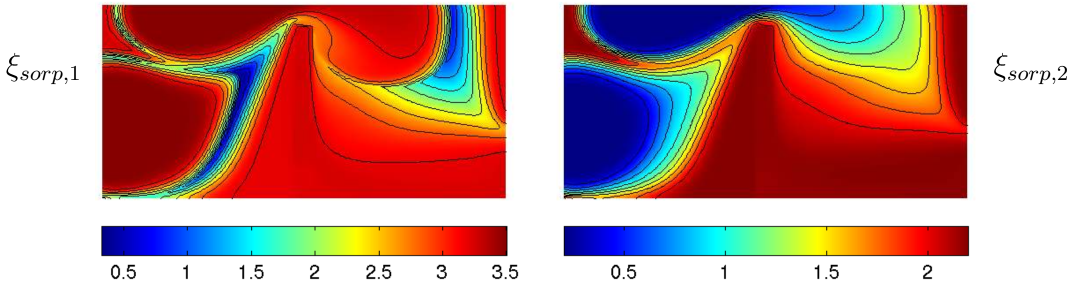




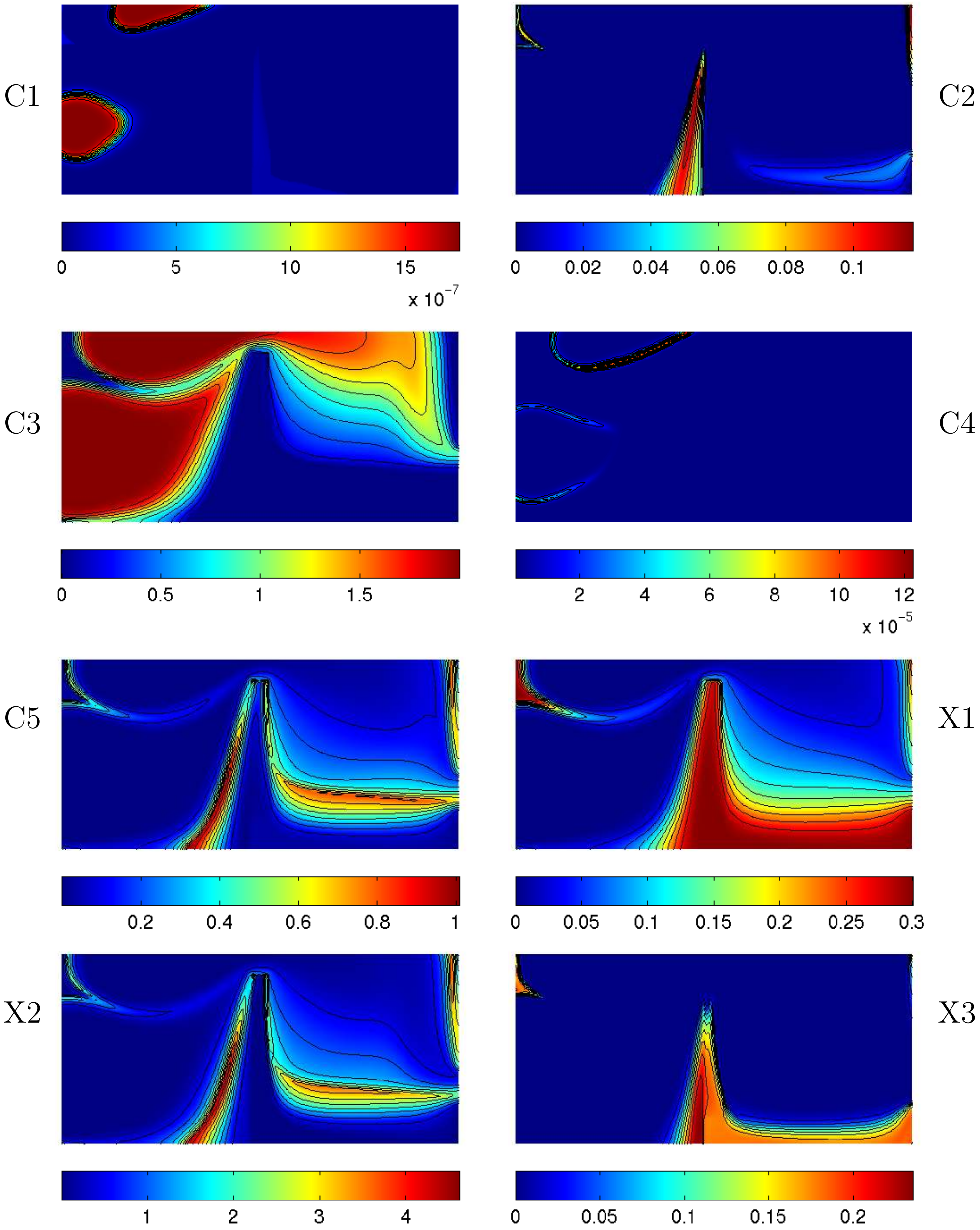
Concentration profile at $t = 5050$

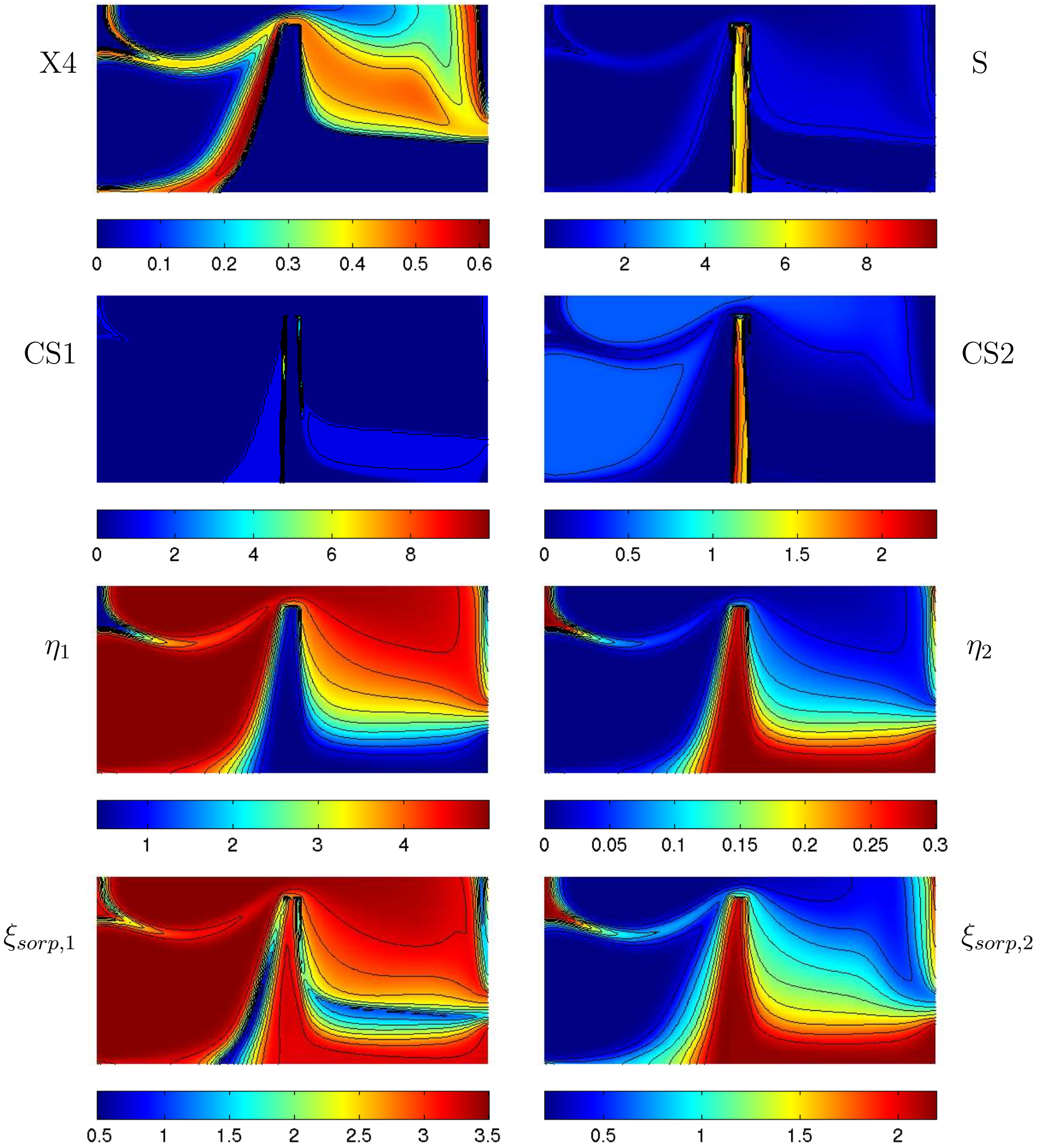



C5
0 0.2 0.4 0.6 0.8 1
X1
0.05 0.1 0.15 0.2 0.25 0.3
X2
1 2 3 4
X3
0 0.05 0.1 0.15 0.2
X4
0 0.1 0.2 0.3 0.4 0.5
S
2 4 6 8
CS1
0 2 4 6 8
CS2
0 0.5 1 1.5 2
η_1
1 2 3 4
η_2
0.05 0.1 0.15 0.2 0.25 0.3



Concentration profile at $t = 5100$

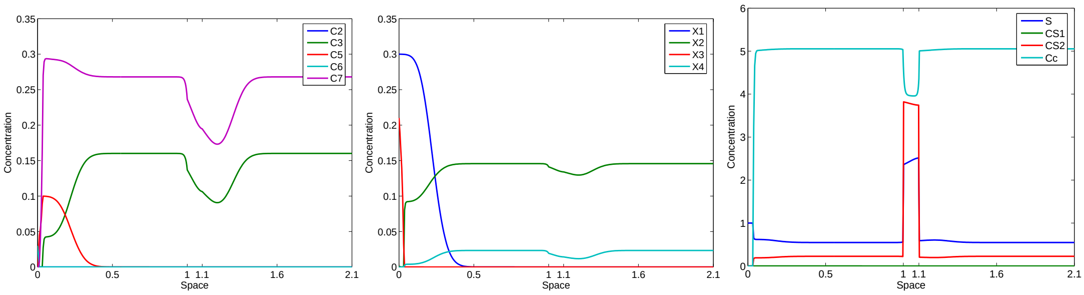



X4
0
0.1
0.2
0.3
0.4
0.5
0.6
S
2
4
6
8
CS1
0
2
4
6
8
CS2
0
0.5
1
1.5
2
η_1
1
2
3
4
η_2
0
0.05
0.1
0.15
0.2
0.25
0.3
$\xi_{sorp,1}$
0.5
1
1.5
2
2.5
3
3.5
$\xi_{sorp,2}$
0.5
1
1.5
2

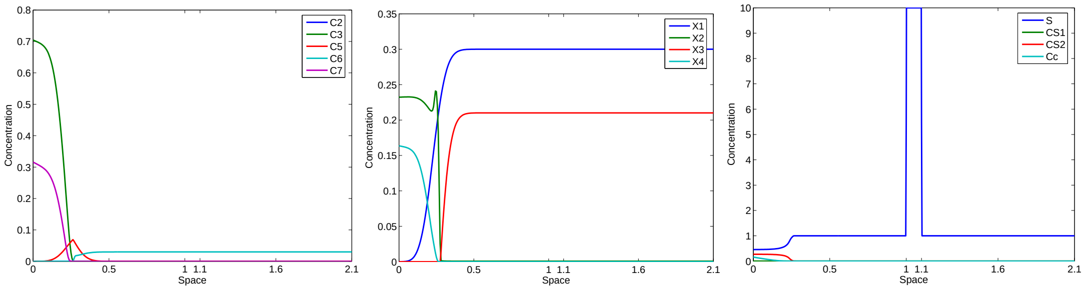
B.2 Medium Test Cases

B.2.1 Advective Medium 1D

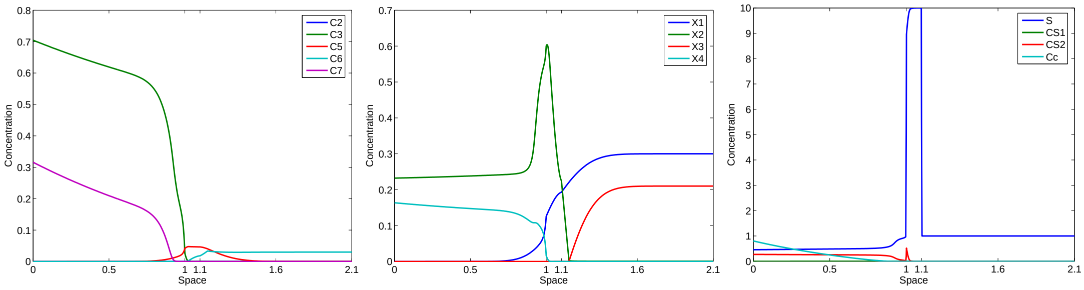
Concentration profile at $t = 10$



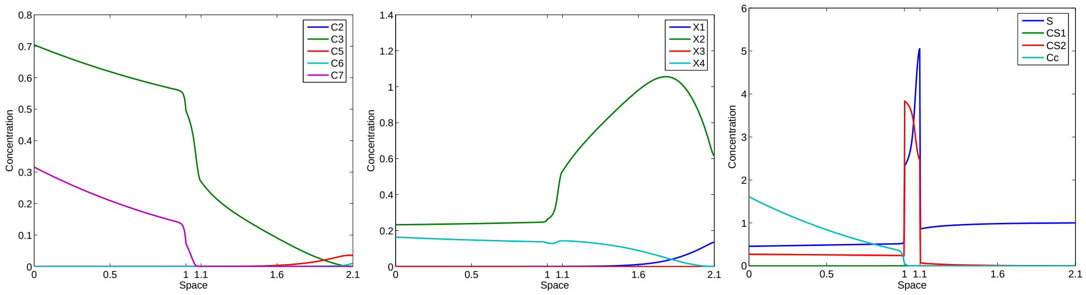
Concentration profile at $t = 5010$



Concentration profile at $t = 5050$

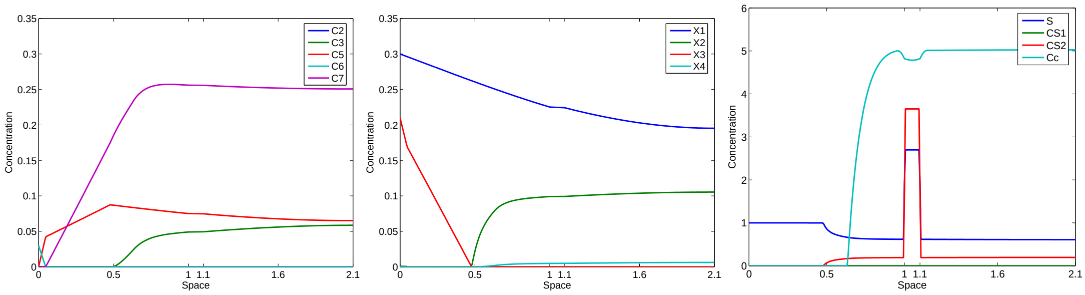


Concentration profile at $t = 5100$

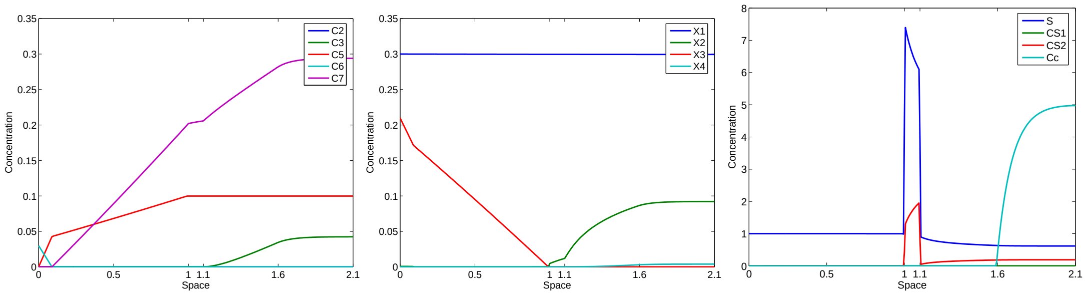


B.2.2 Diffusive Medium 1D

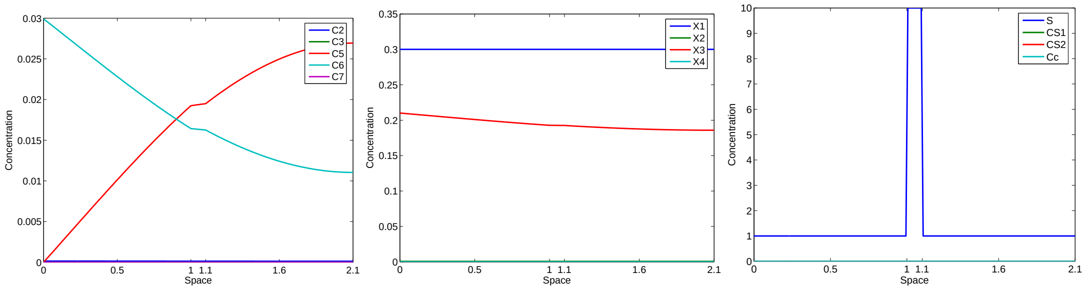
Concentration profile at $t = 10$



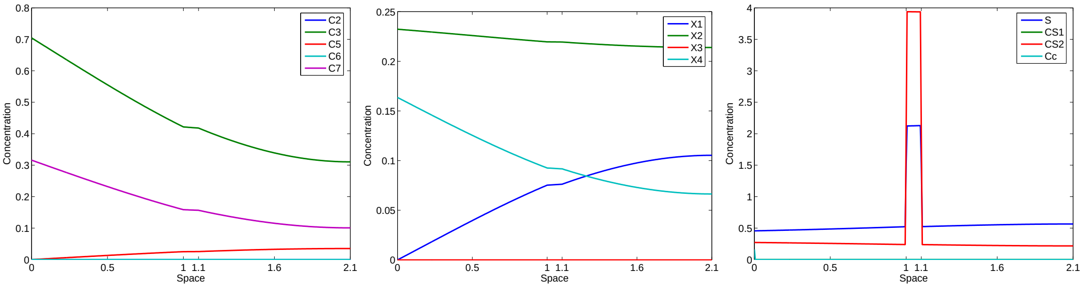
Concentration profile at $t = 50$



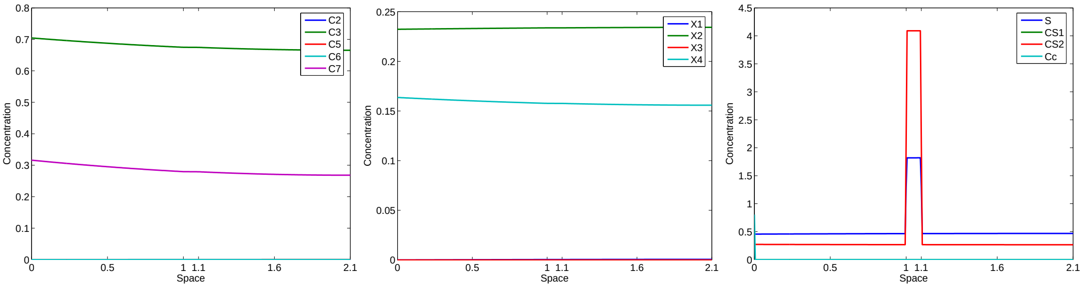
Concentration profile at $t = 100$



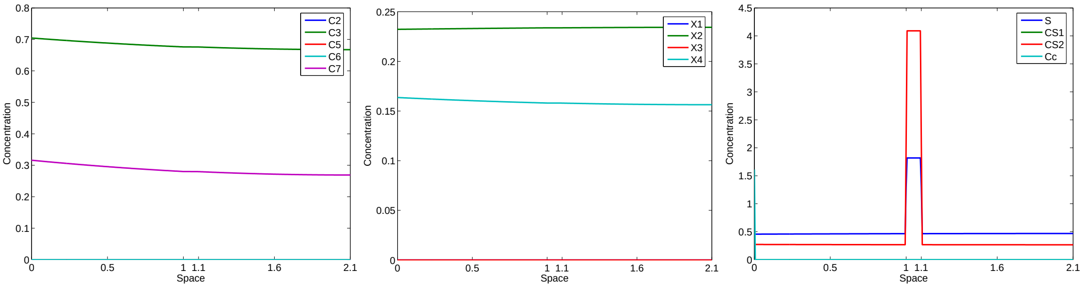
Concentration profile at $t = 5010$



Concentration profile at $t = 5050$

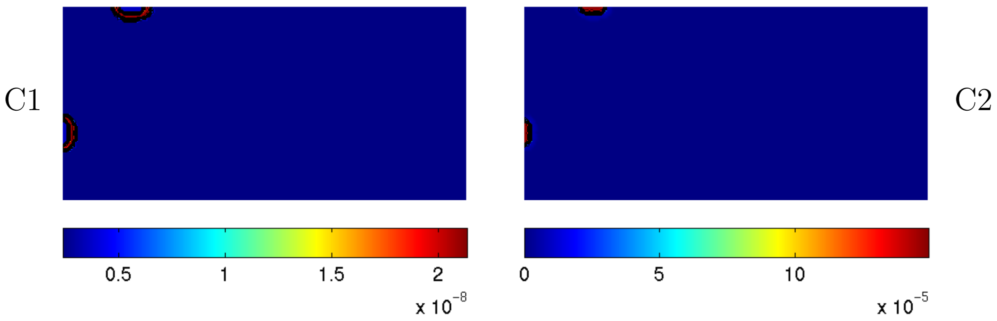


Concentration profile at $t = 5100$

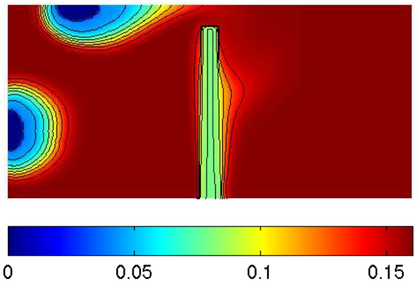


B.2.3 Advective Medium 2D

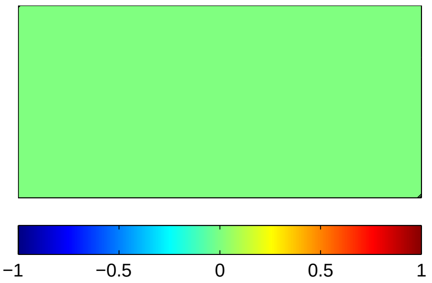
Concentration profile at $t = 10$



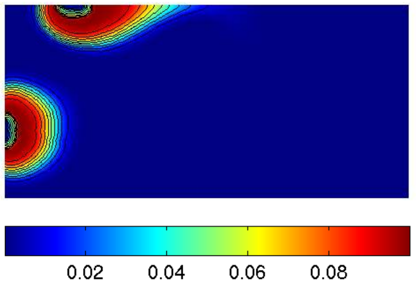
C3



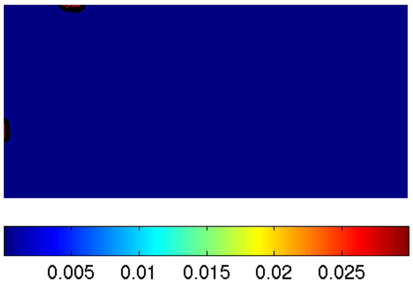
C4



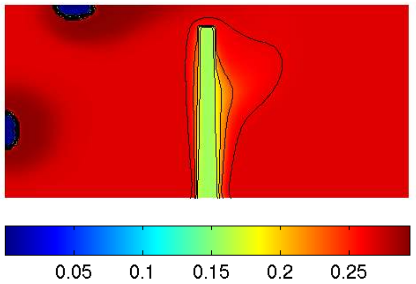
C5



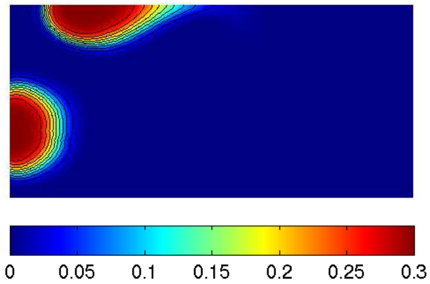
C6



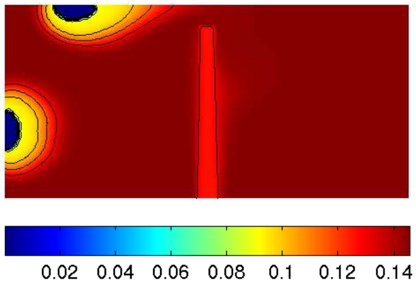
C7



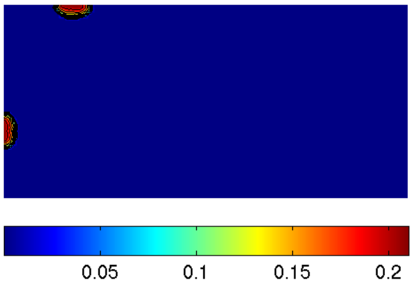
X1



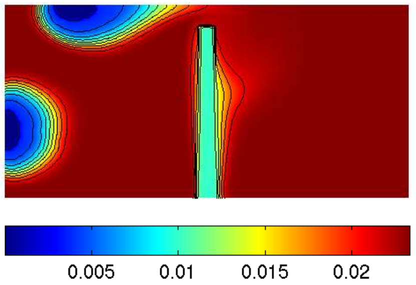
X2



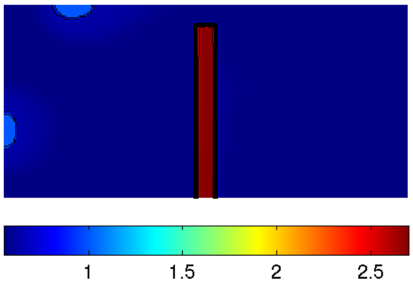
X3

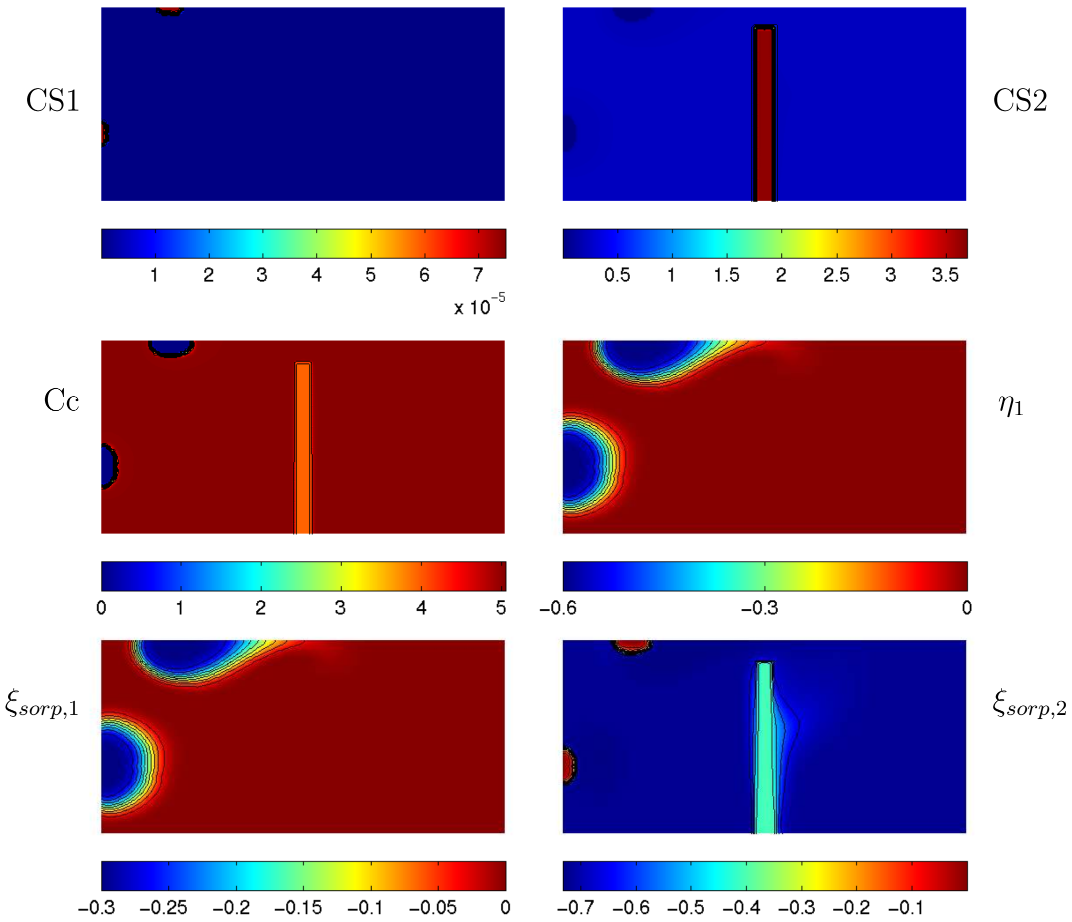


X4

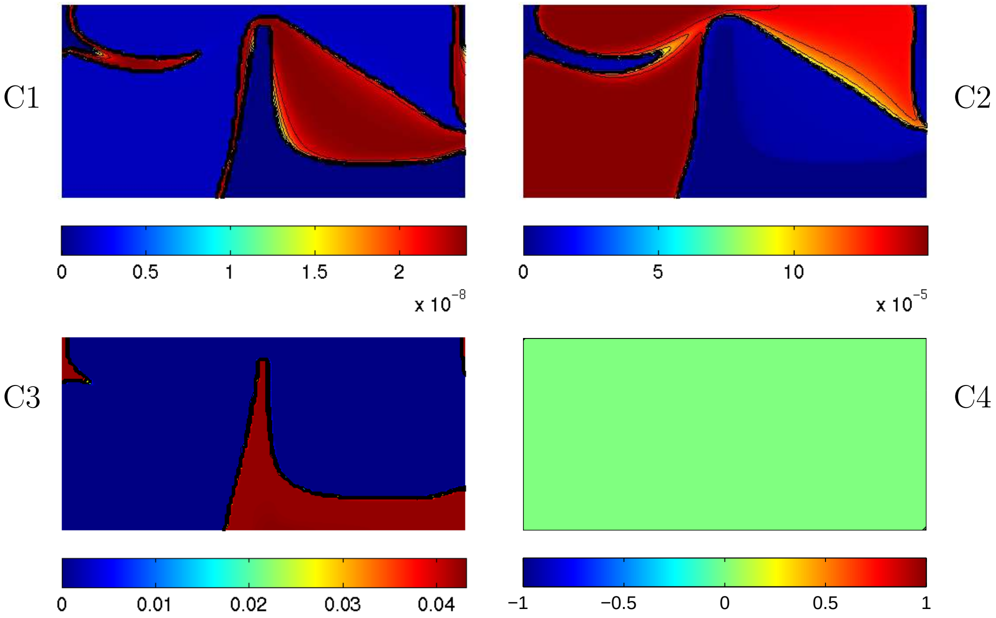


S

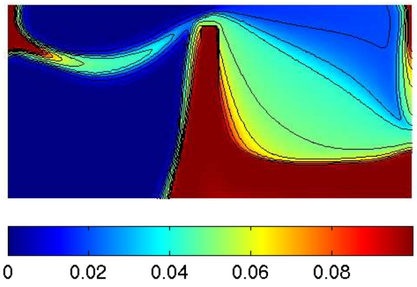


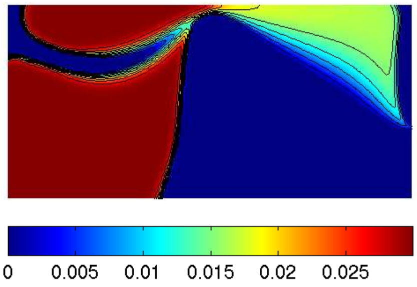


Concentration profile at $t = 1000$



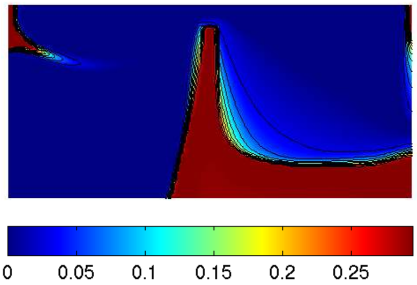
C5

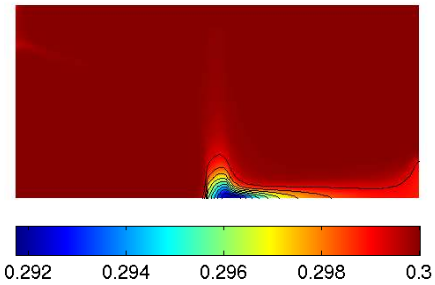




C6

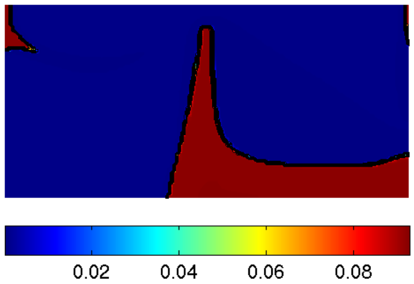
C7

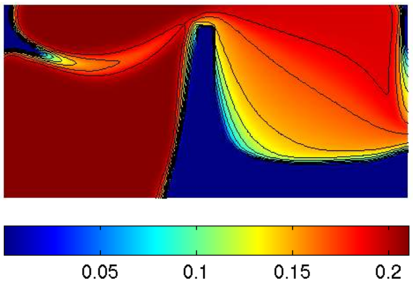




X1

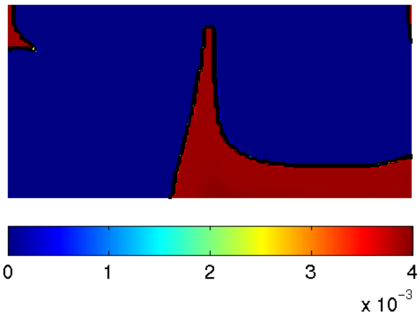
X2

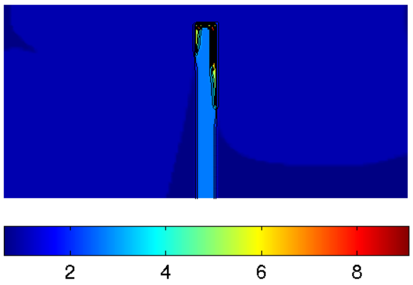




X3

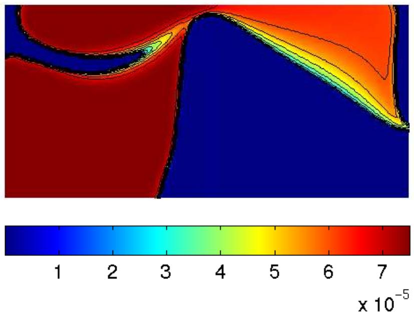
X4

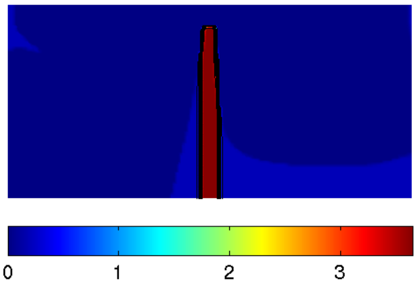




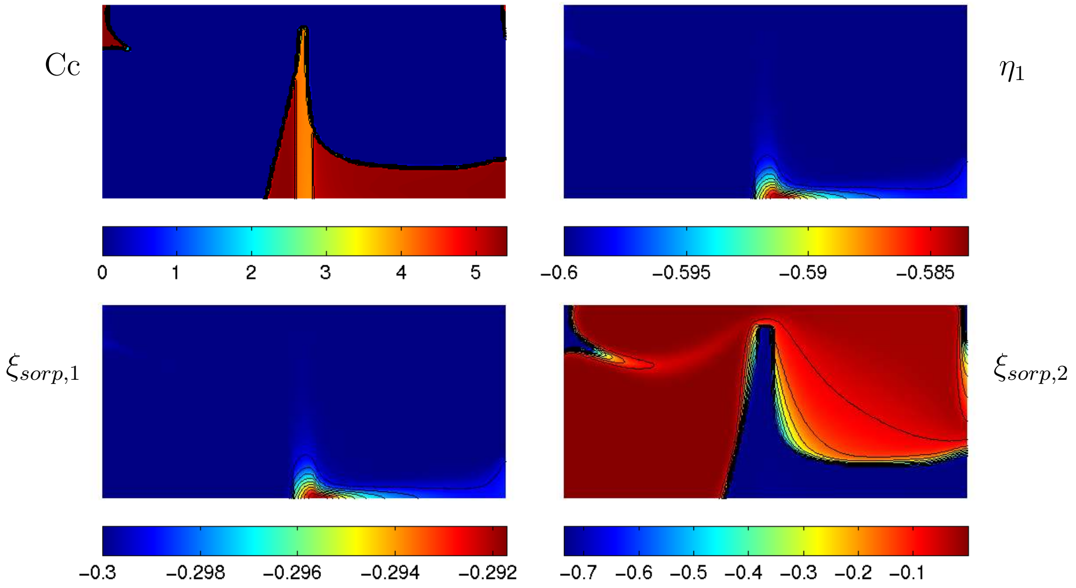
S

CS1

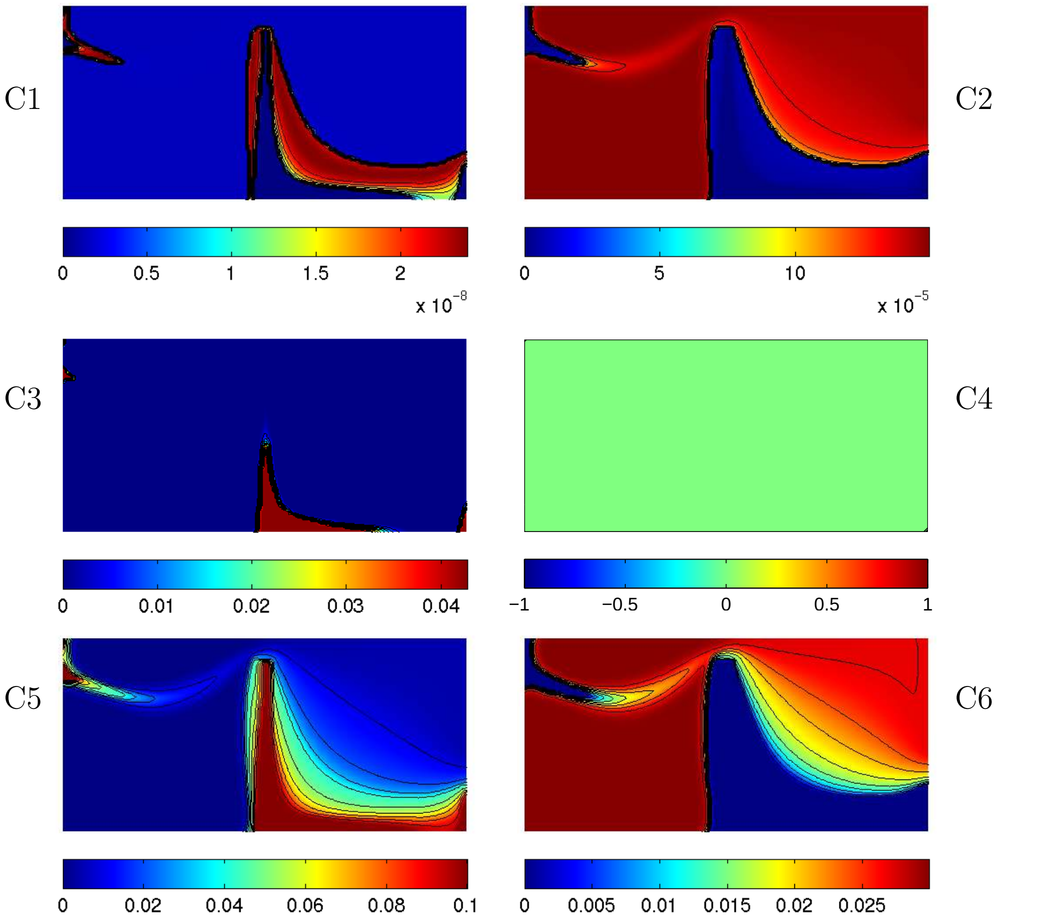




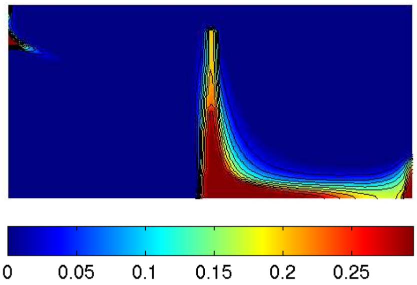
CS2



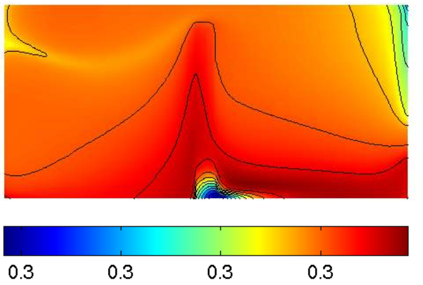
Concentration profile at $t = 2000$



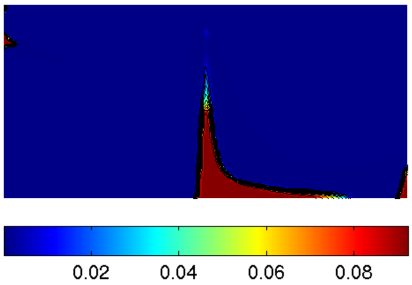
C7



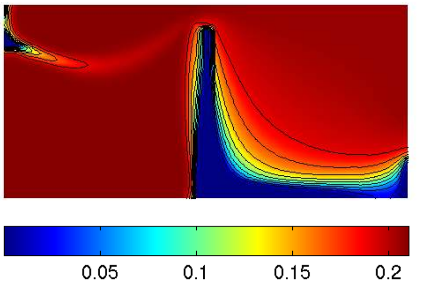
X1



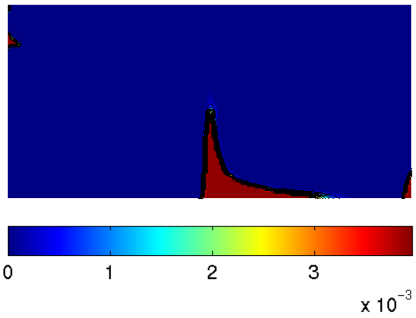
X2



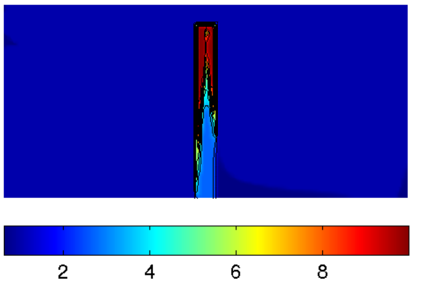
X3



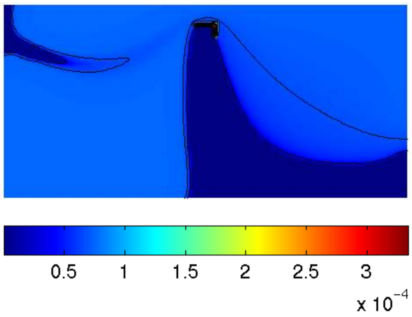
X4



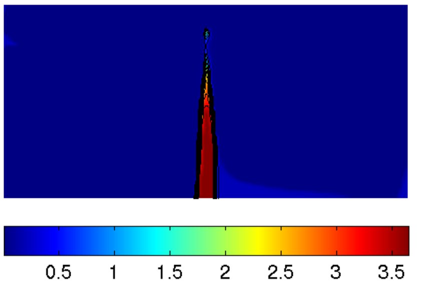
S



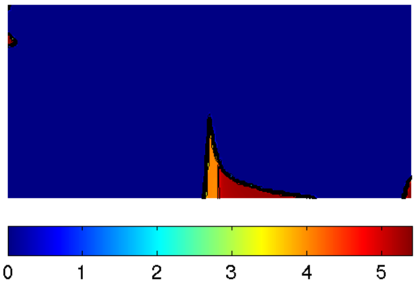
CS1



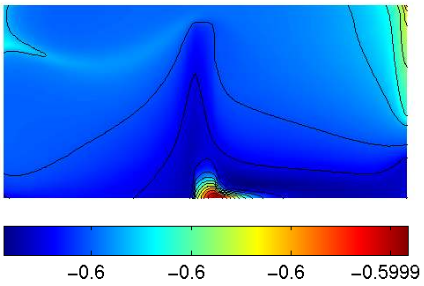
CS2



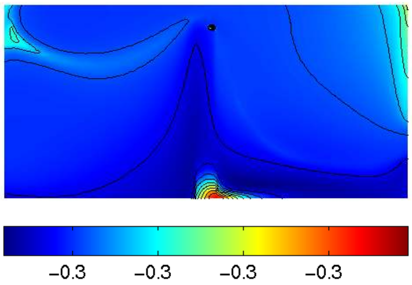
Cc



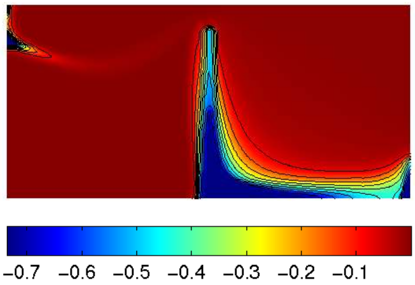
η_1



$\xi_{sorp,1}$

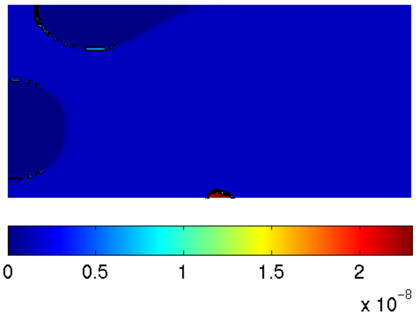


$\xi_{sorp,2}$

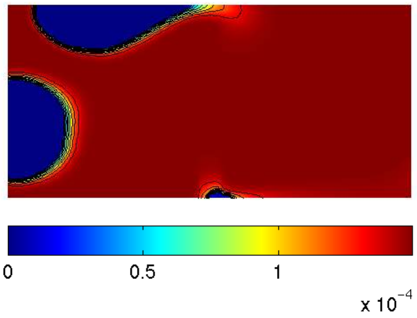


Concentration profile at $t = 5010$

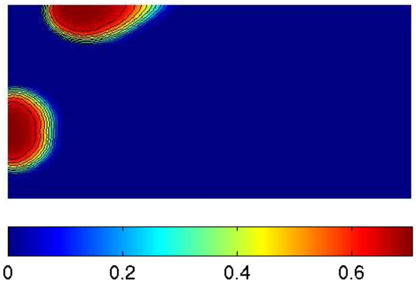
C1



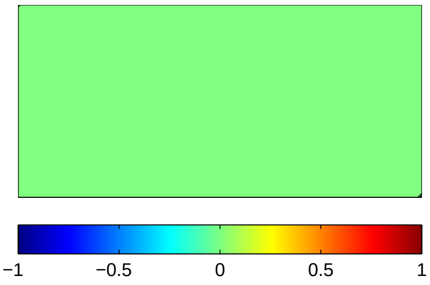
C2



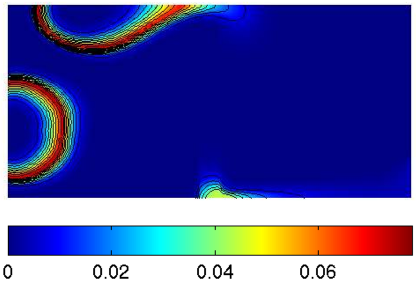
C3



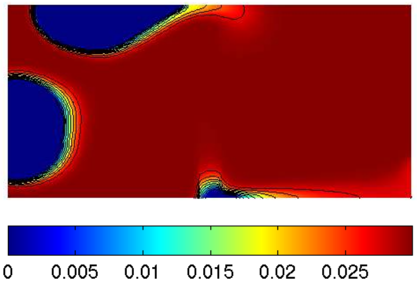
C4



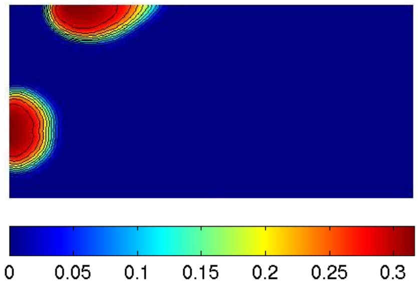
C5



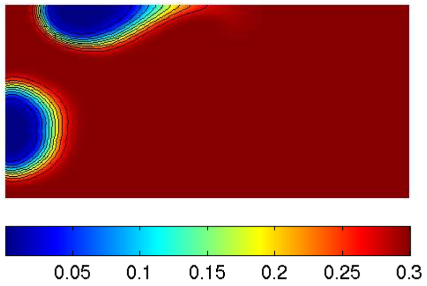
C6

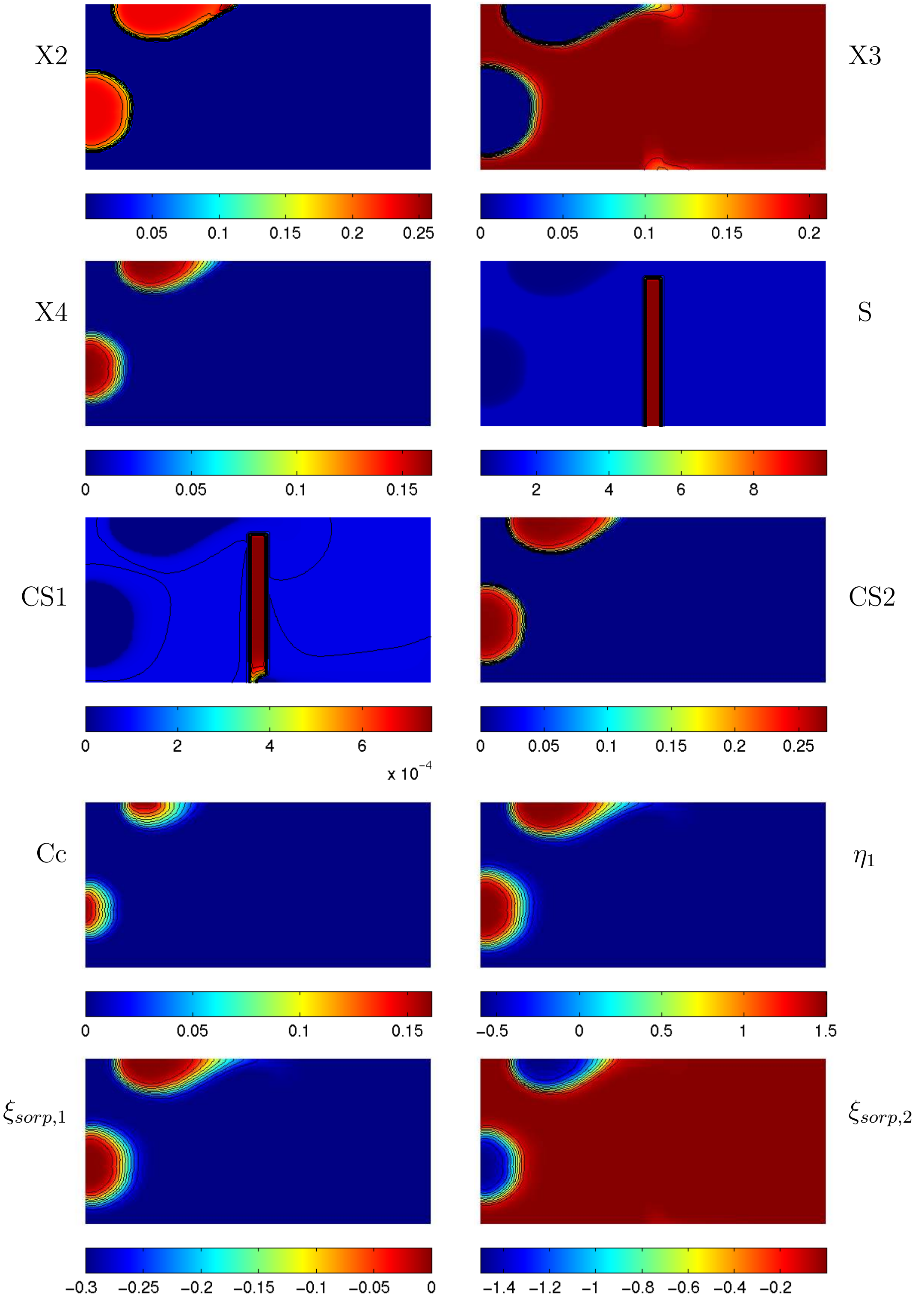


C7

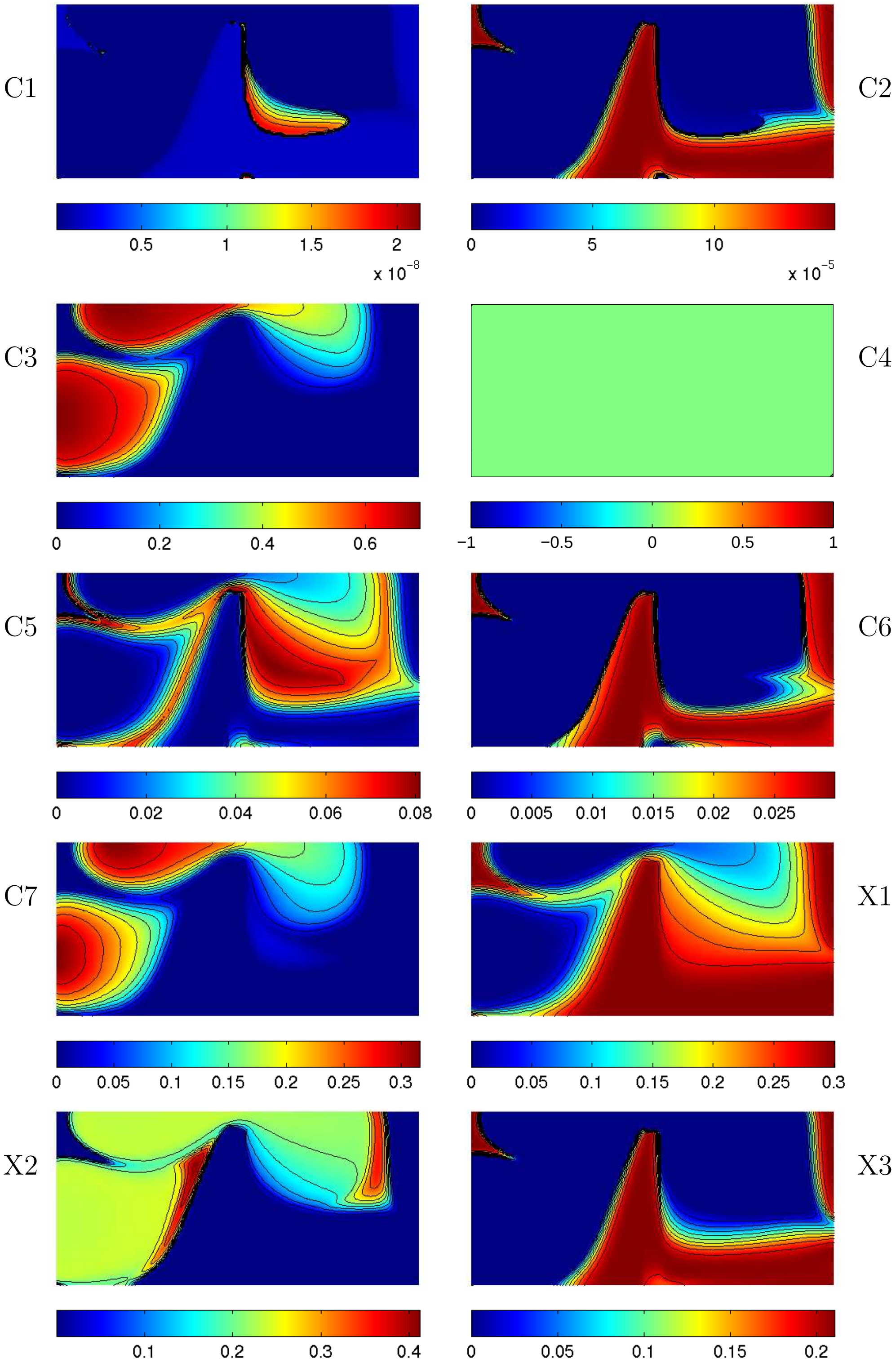


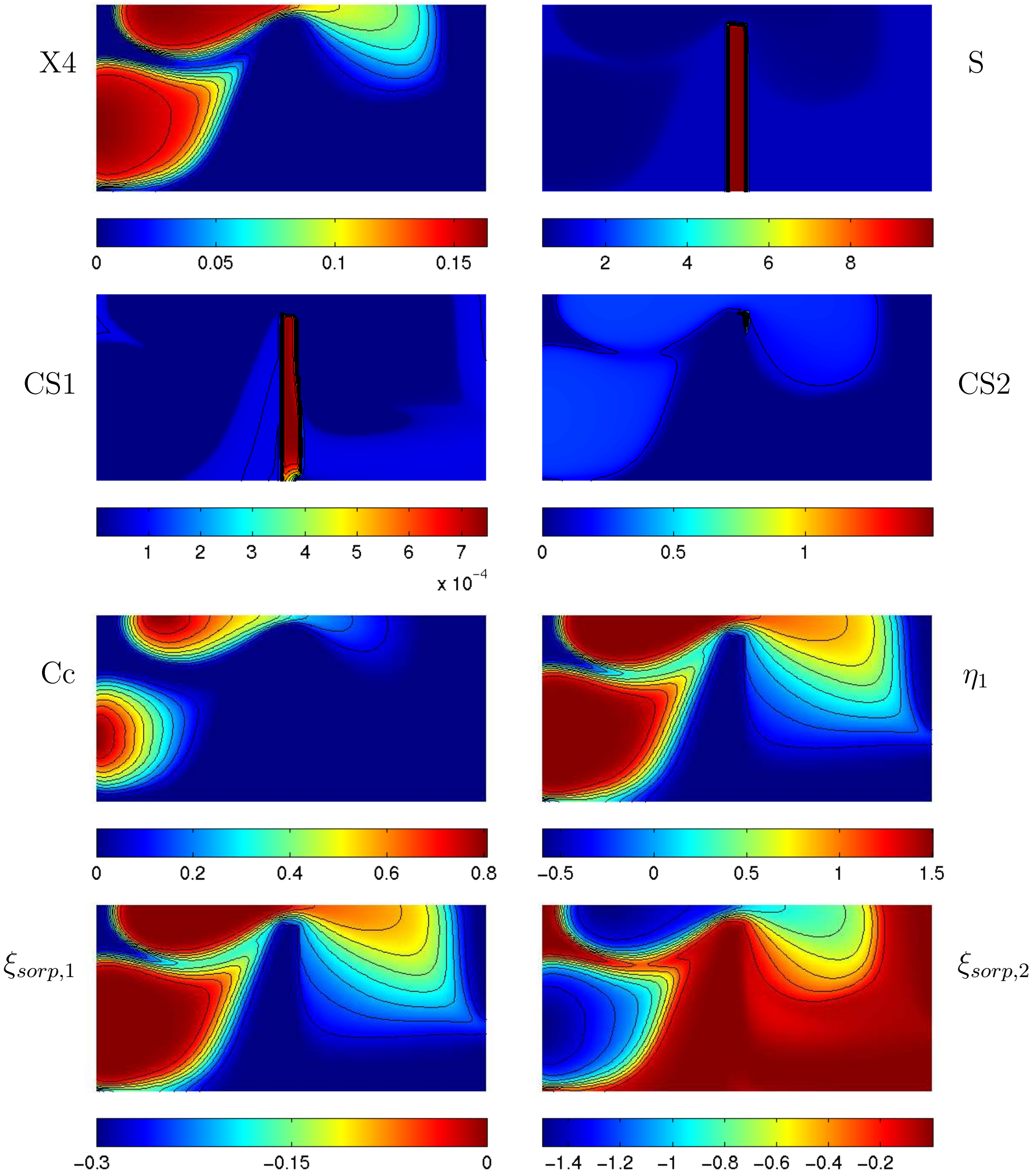
X1



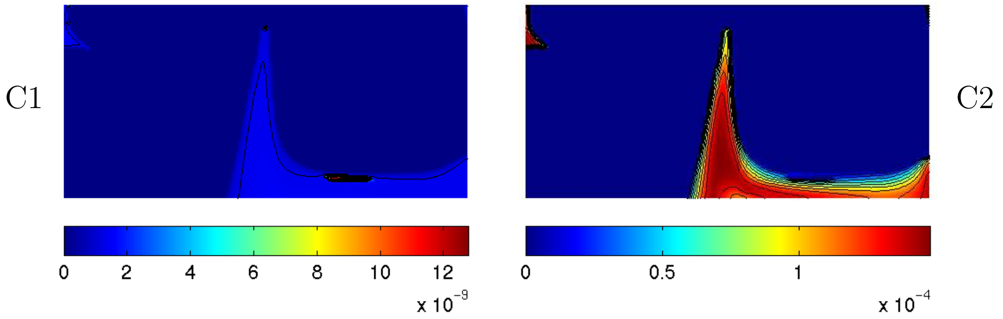

X2
0.05 0.1 0.15 0.2 0.25
X3
0 0.05 0.1 0.15 0.2
X4
0 0.05 0.1 0.15
S
2 4 6 8
CS1
0 2 4 6
x 10^{-4}
CS2
0 0.05 0.1 0.15 0.2 0.25
Cc
0 0.05 0.1 0.15
η_1
−0.5 0 0.5 1 1.5
$\xi_{sorp,1}$
−0.3 −0.25 −0.2 −0.15 −0.1 −0.05 0
$\xi_{sorp,2}$
−1.4 −1.2 −1 −0.8 −0.6 −0.4 −0.2

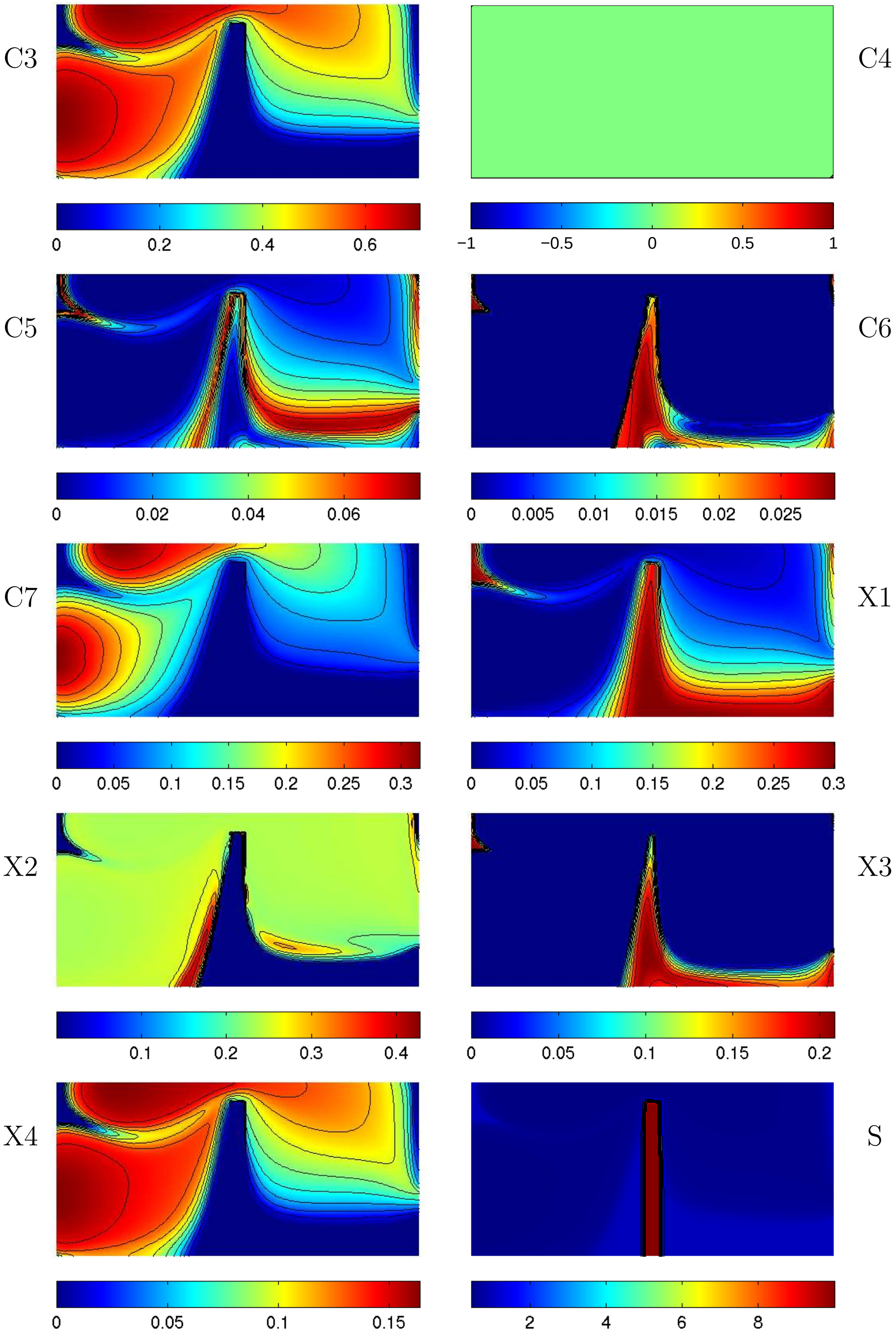
Concentration profile at $t = 5050$

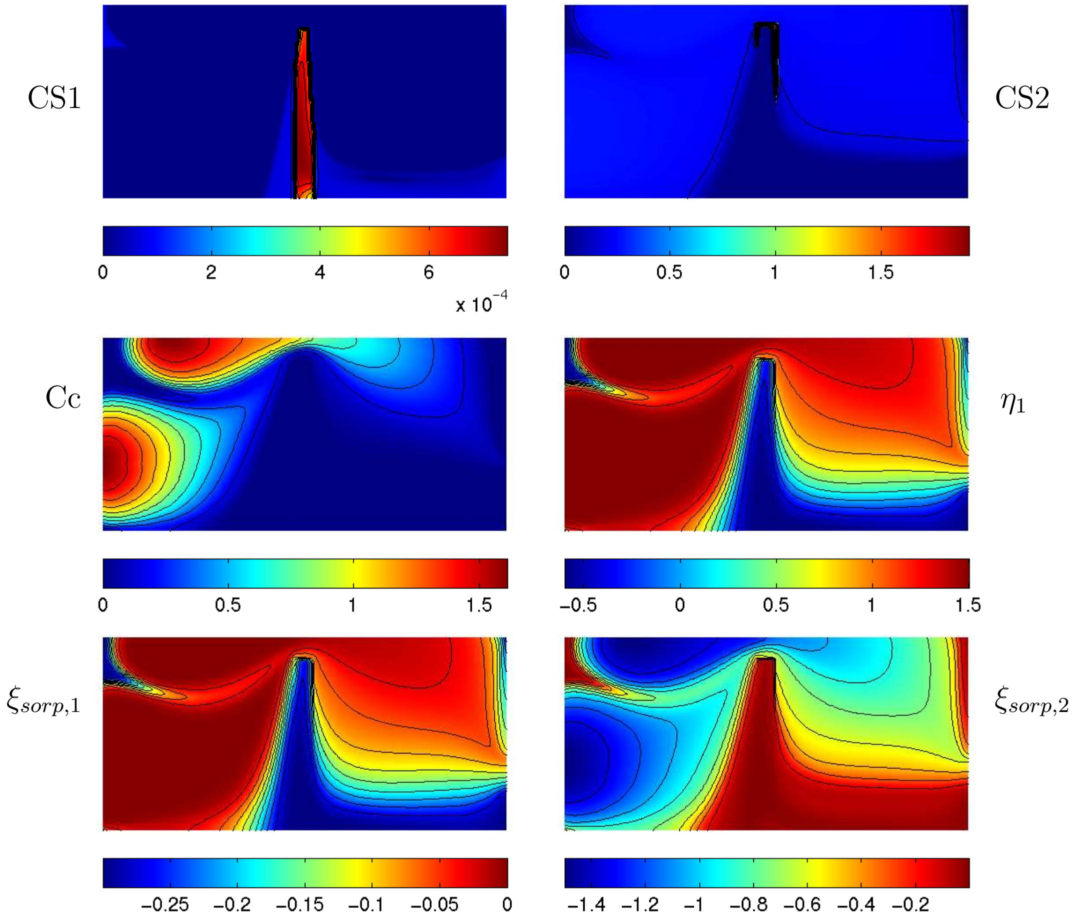




Concentration profile at $t = 5100$

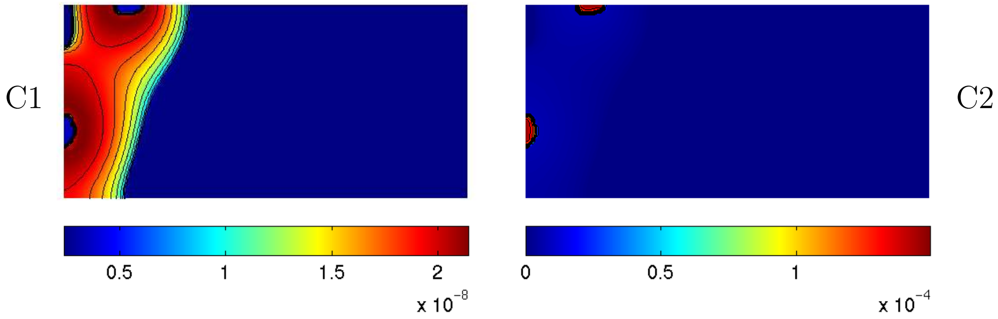


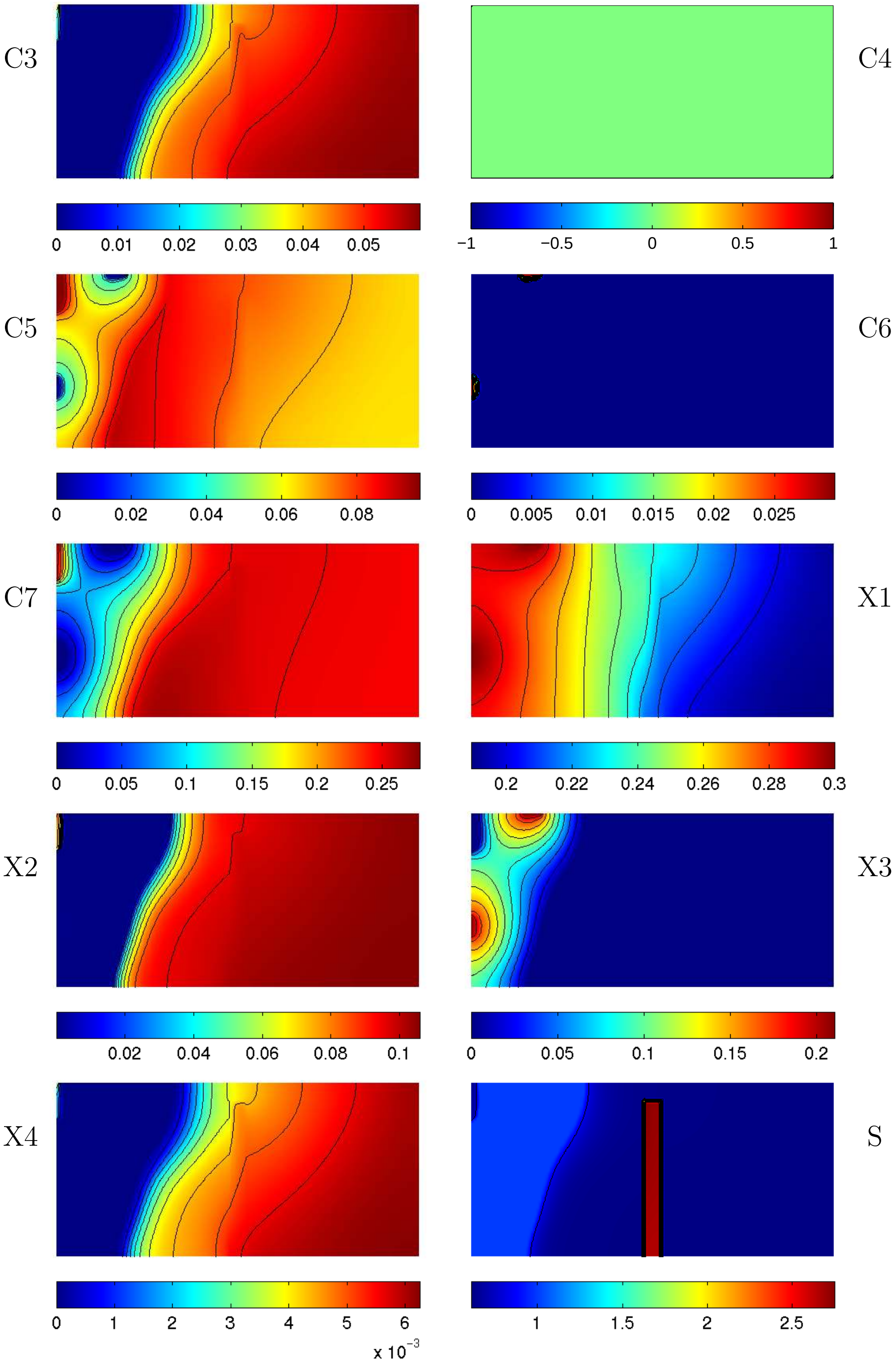

C3
C4
0 0.2 0.4 0.6
−1 −0.5 0 0.5 1
C5
C6
0 0.02 0.04 0.06
0 0.005 0.01 0.015 0.02 0.025
C7
X1
0 0.05 0.1 0.15 0.2 0.25 0.3
0 0.05 0.1 0.15 0.2 0.25 0.3
X2
X3
0.1 0.2 0.3 0.4
0 0.05 0.1 0.15 0.2
X4
S
0 0.05 0.1 0.15
2 4 6 8

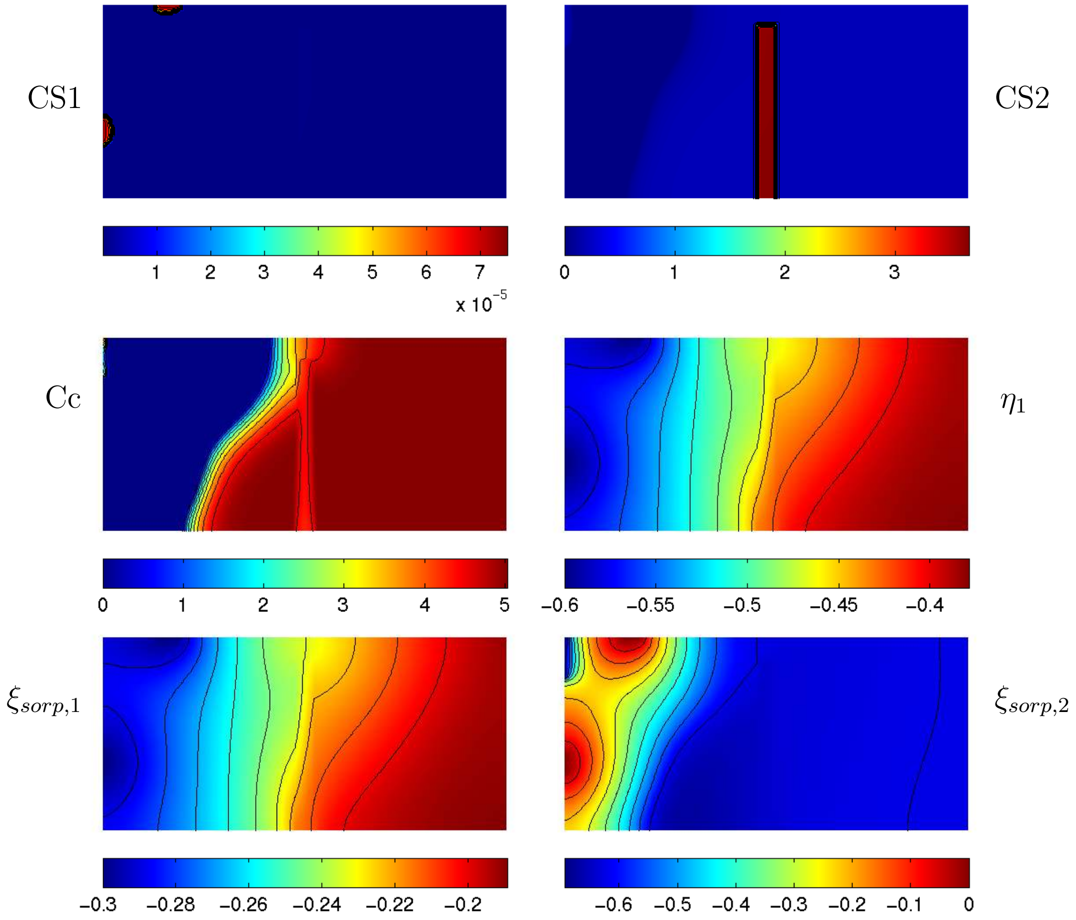


B.2.4 Diffusive Medium 2D

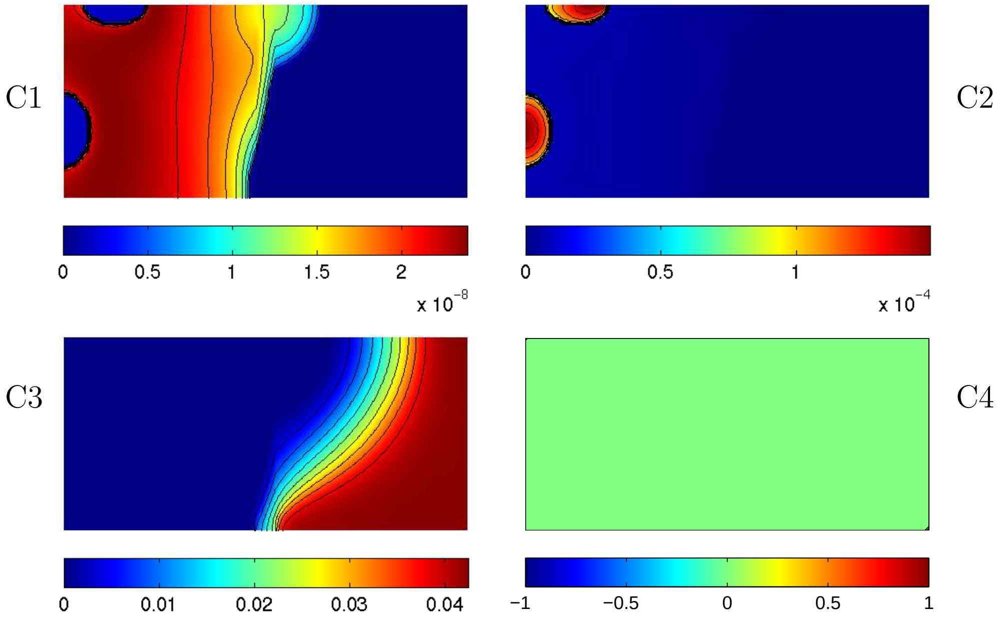
Concentration profile at $t = 10$



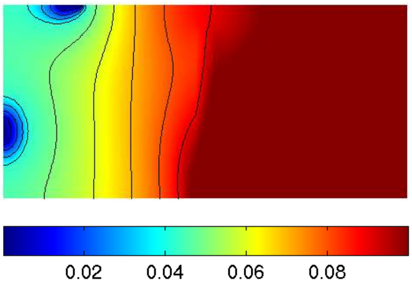

C3
0 0.01 0.02 0.03 0.04 0.05
C4
−1 −0.5 0 0.5 1
C5
0 0.02 0.04 0.06 0.08
C6
0 0.005 0.01 0.015 0.02 0.025
C7
0 0.05 0.1 0.15 0.2 0.25
X1
0.2 0.22 0.24 0.26 0.28 0.3
X2
0.02 0.04 0.06 0.08 0.1
X3
0 0.05 0.1 0.15 0.2
X4
0 1 2 3 4 5 6
x 10−3
S
1 1.5 2 2.5



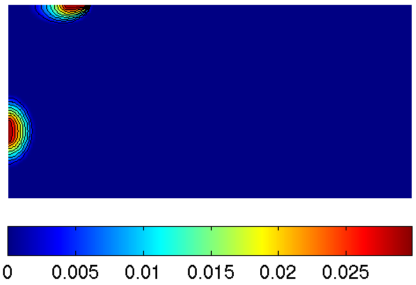
Concentration profile at $t = 50$



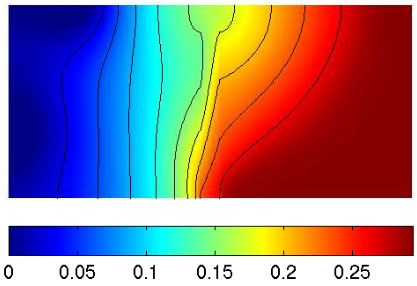
C5



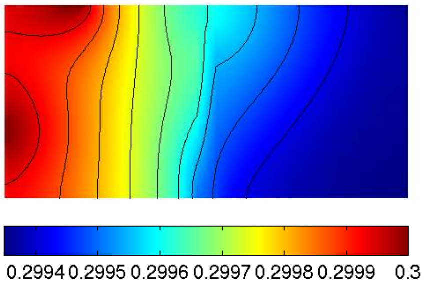
C6



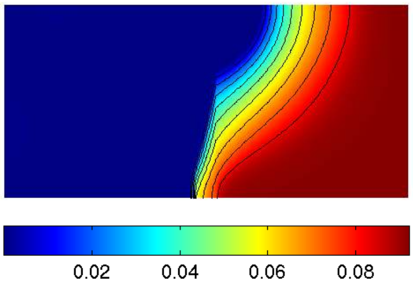
C7



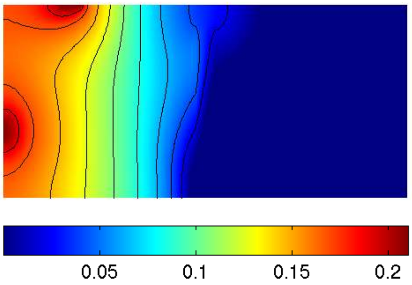
X1



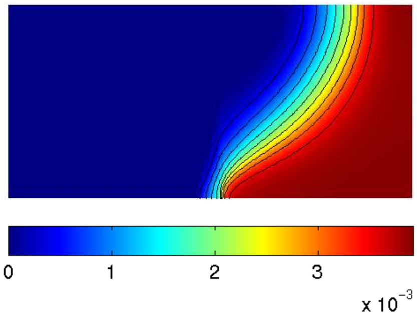
X2



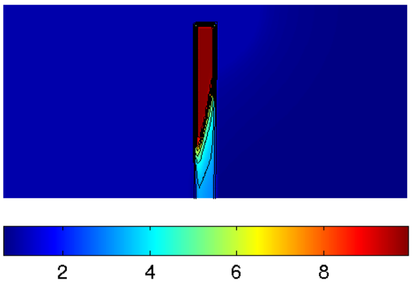
X3



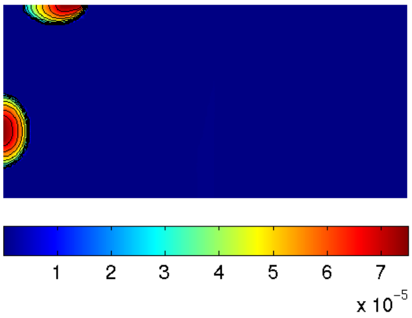
X4



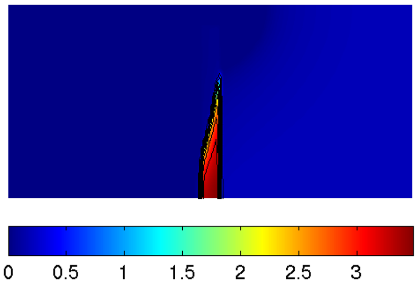
S

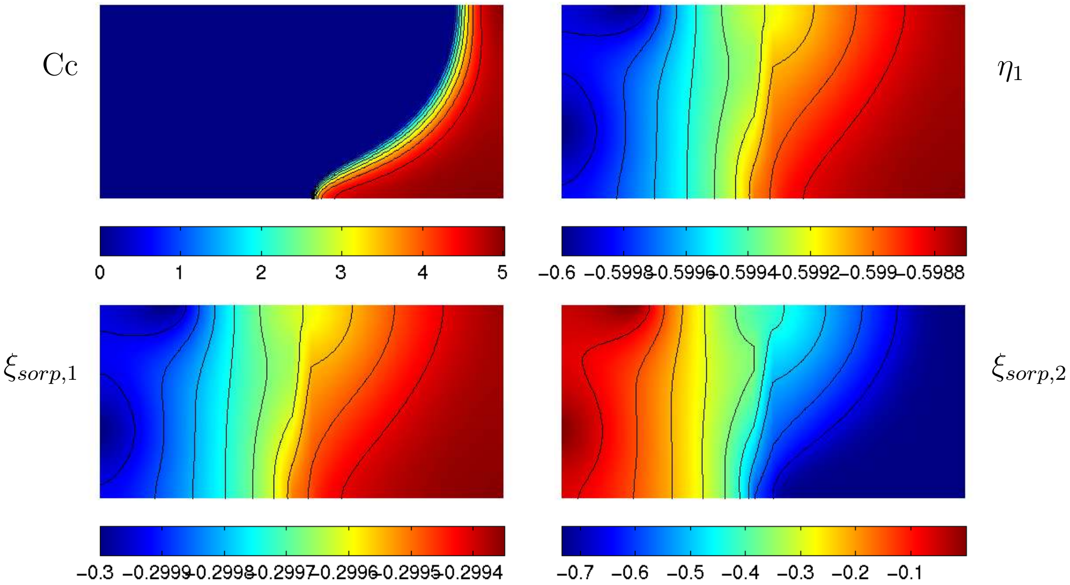


CS1

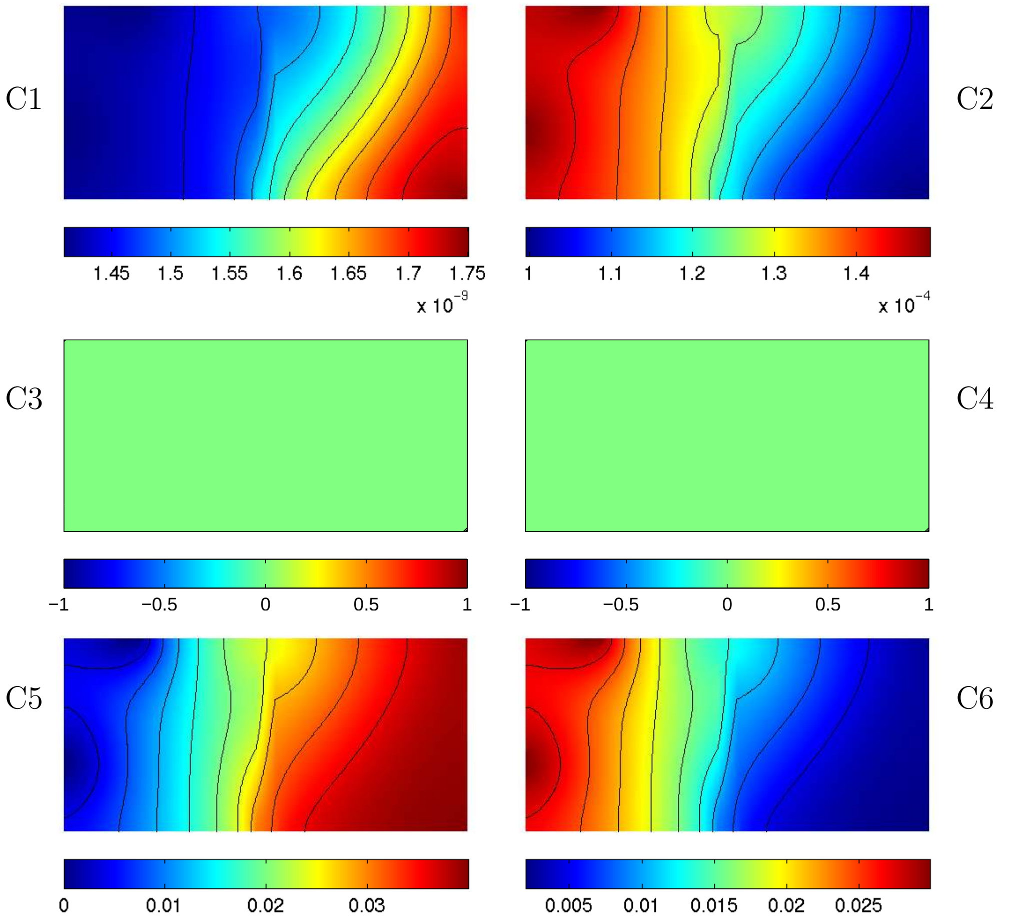


CS2

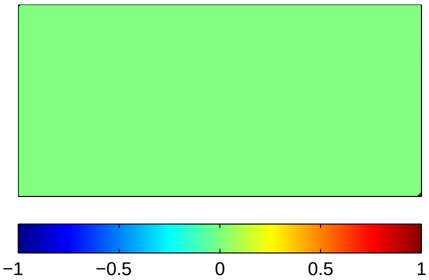




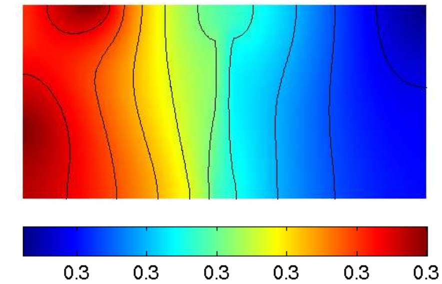
Concentration profile at $t = 100$



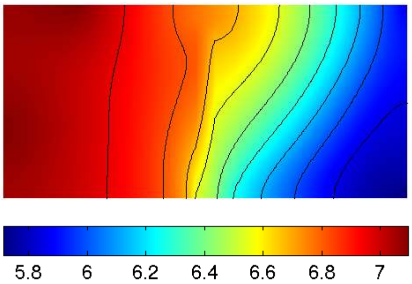
C7



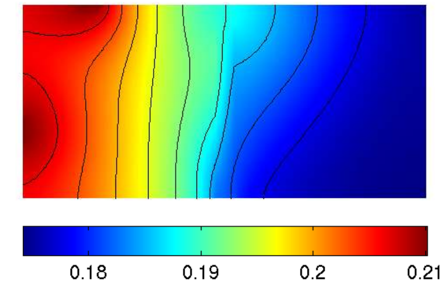
X1



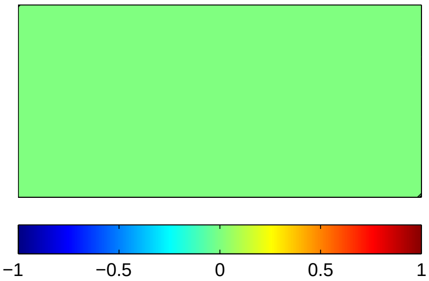
X2



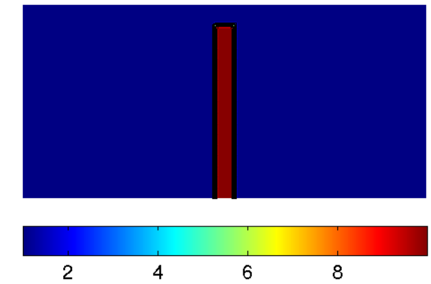
X3



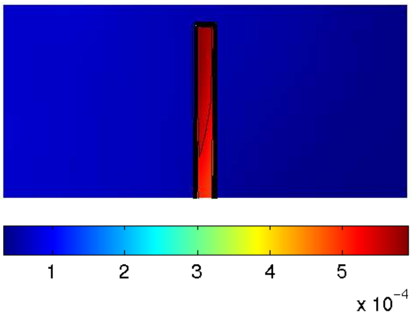
X4



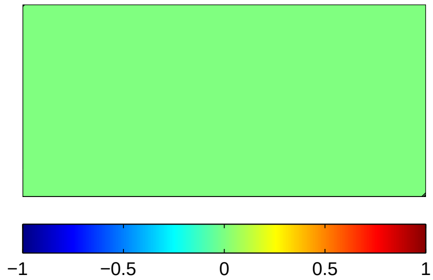
S



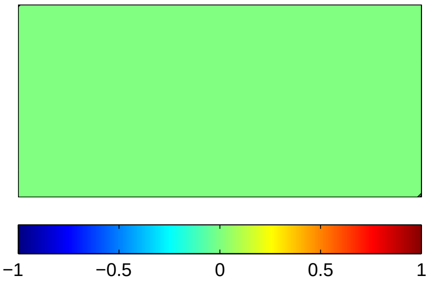
CS1



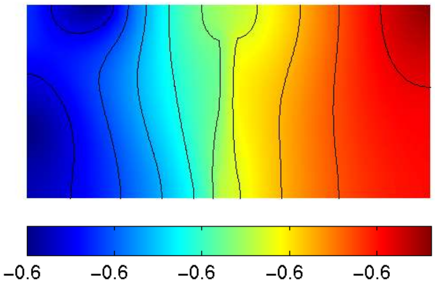
CS2



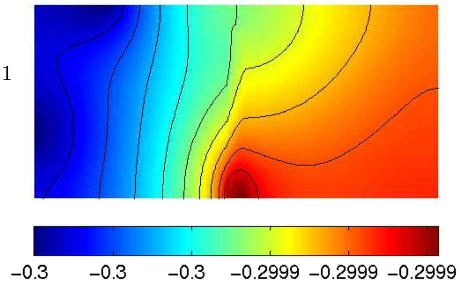
Cc

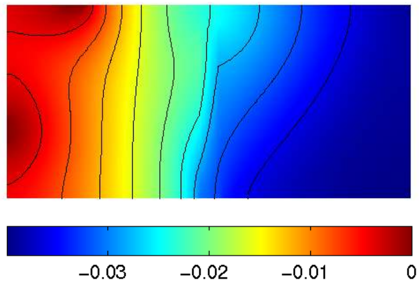


η_1



$\xi_{sorp,1}$

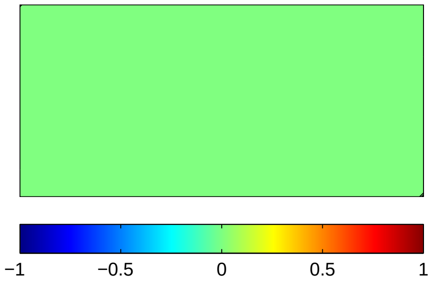


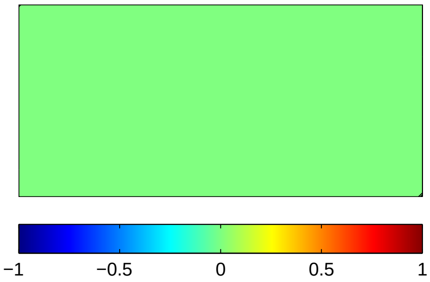


$\xi_{sorp,2}$

Concentration profile at $t = 5010$

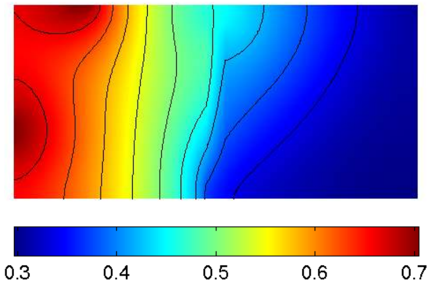
C1

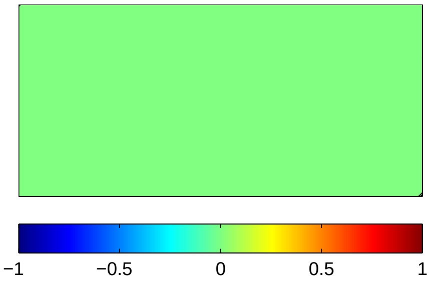




C2

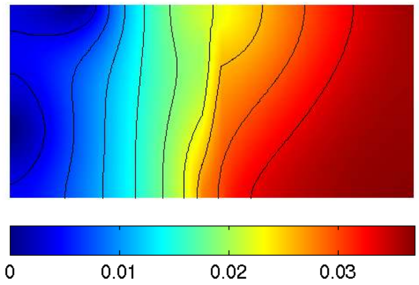
C3

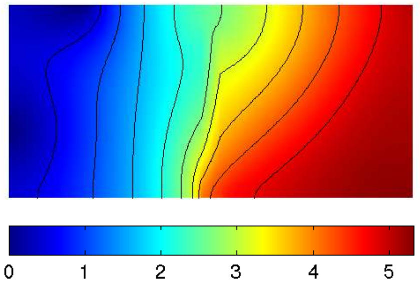




C4

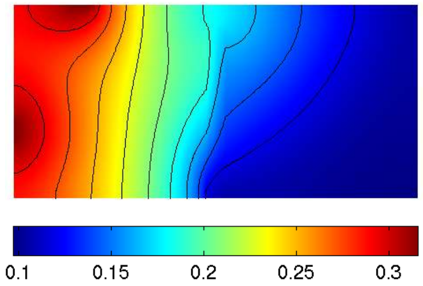
C5

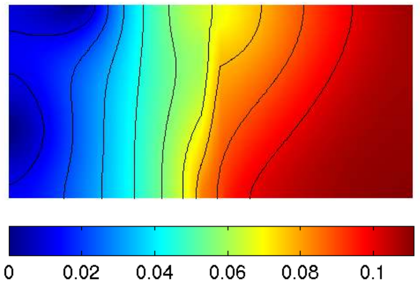




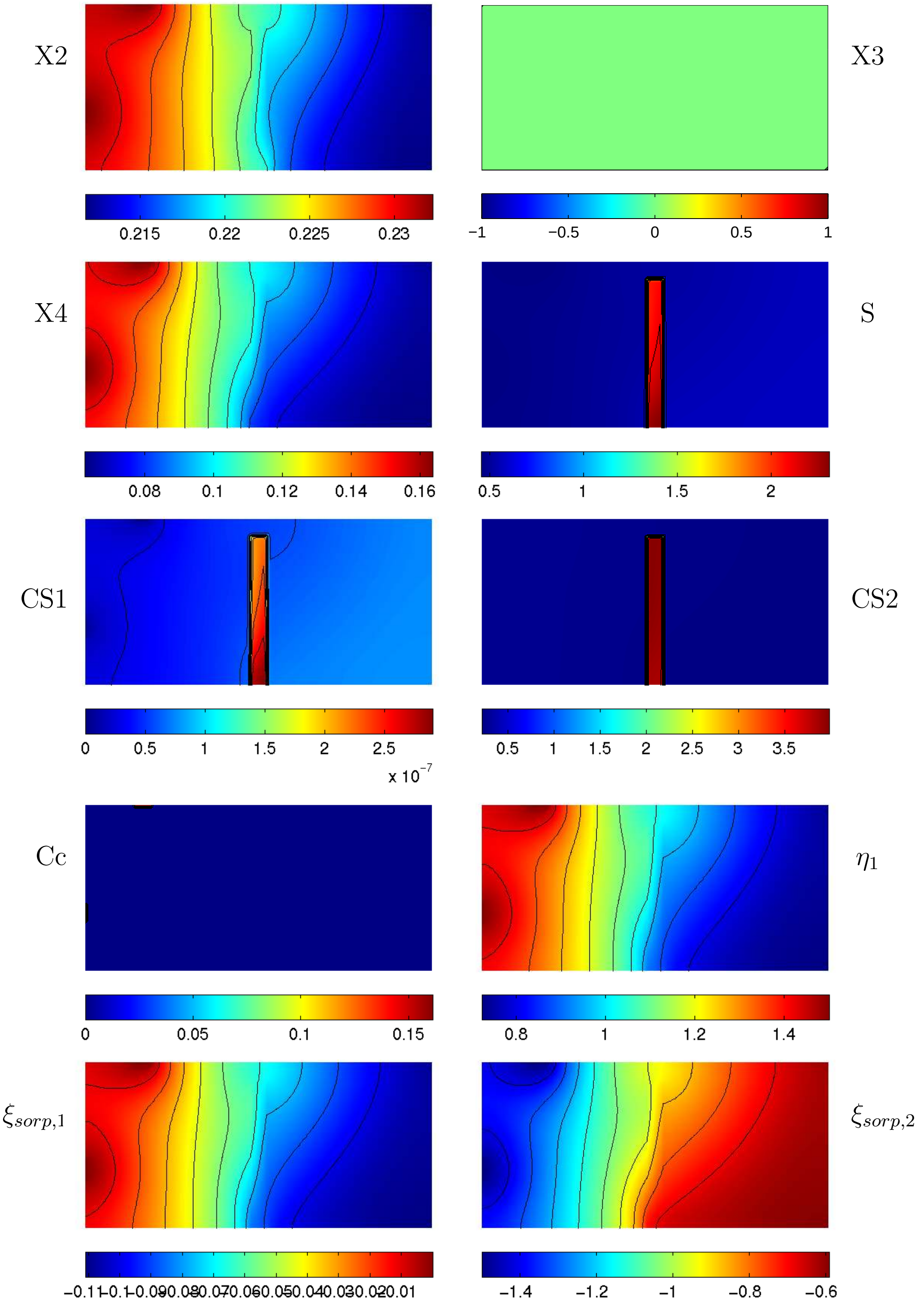
C6

C7

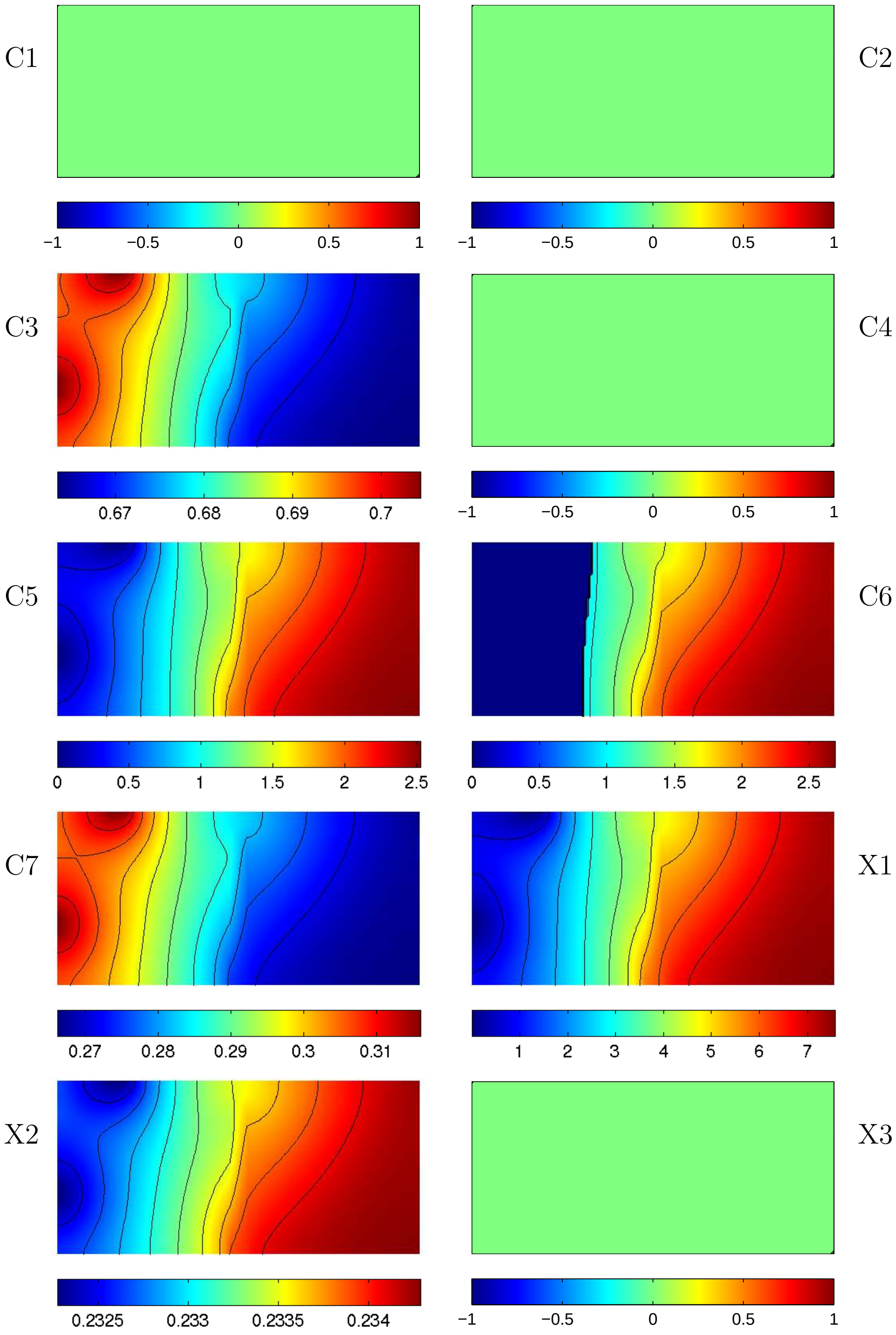


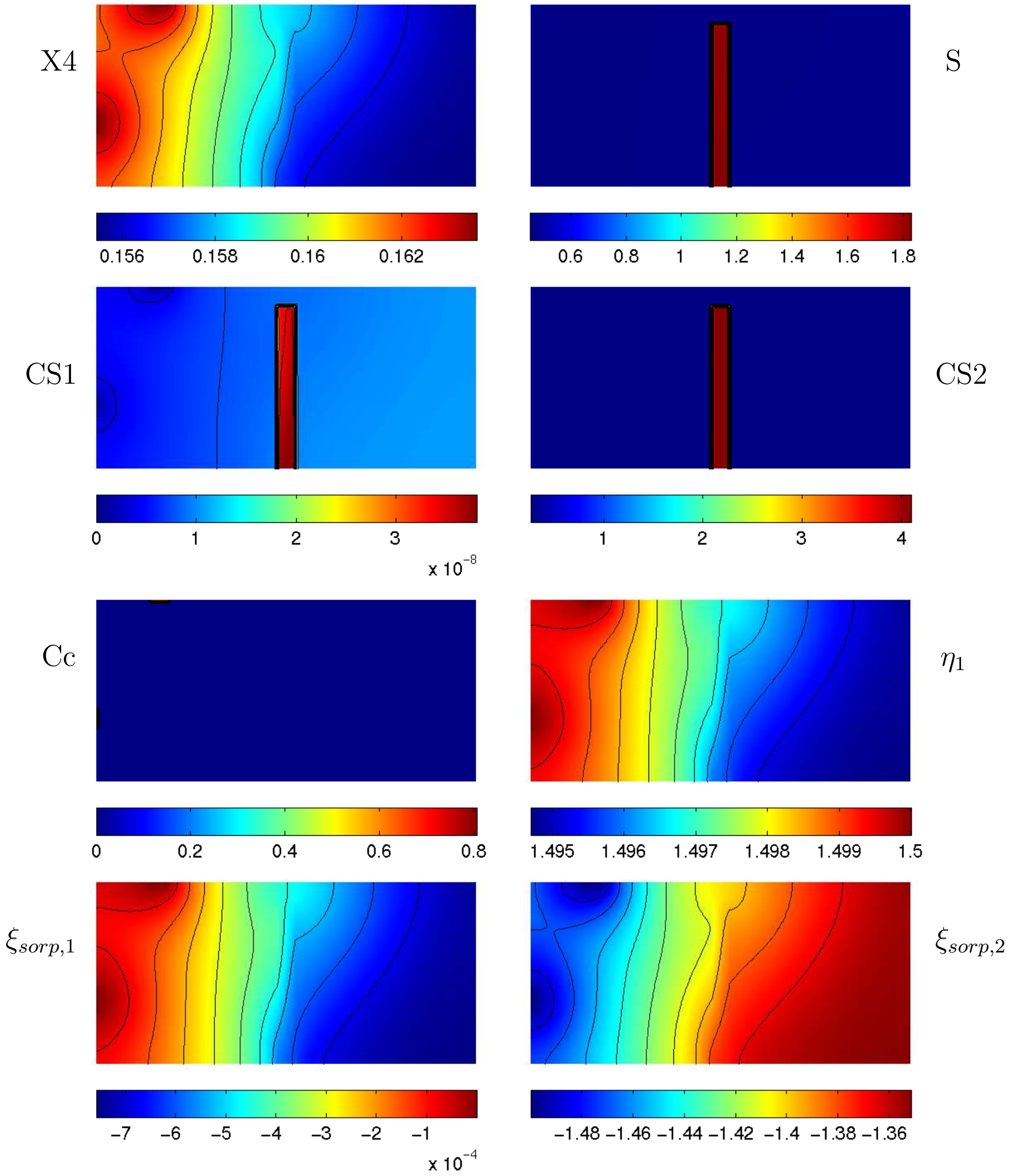


X1


X2
X3
0.215
0.22
0.225
0.23
−1
−0.5
0
0.5
1
X4
S
0.08
0.1
0.12
0.14
0.16
0.5
1
1.5
2
CS1
CS2
0
0.5
1
1.5
2
2.5
x 10^{-7}
0.5
1
1.5
2
2.5
3
3.5
Cc
η_1
0
0.05
0.1
0.15
0.8
1
1.2
1.4
$\xi_{sorp,1}$
$\xi_{sorp,2}$
−0.11
−0.1
−0.09
−0.08
−0.07
−0.06
−0.05
−0.04
−0.03
−0.02
−0.01
−1.4
−1.2
−1
−0.8
−0.6

Concentration profile at $t = 5050$

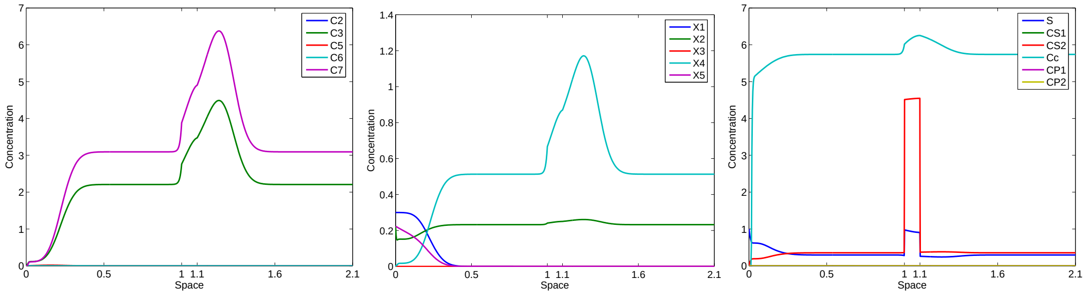



X4
0.156
0.158
0.16
0.162
S
0.6
0.8
1
1.2
1.4
1.6
1.8
CS1
0
1
2
3
x 10^{-8}
CS2
1
2
3
4
Cc
0
0.2
0.4
0.6
0.8
η_1
1.495
1.496
1.497
1.498
1.499
1.5
$\xi_{sorp,1}$
−7
−6
−5
−4
−3
−2
−1
x 10^{-4}
$\xi_{sorp,2}$
−1.48
−1.46
−1.44
−1.42
−1.4
−1.38
−1.36

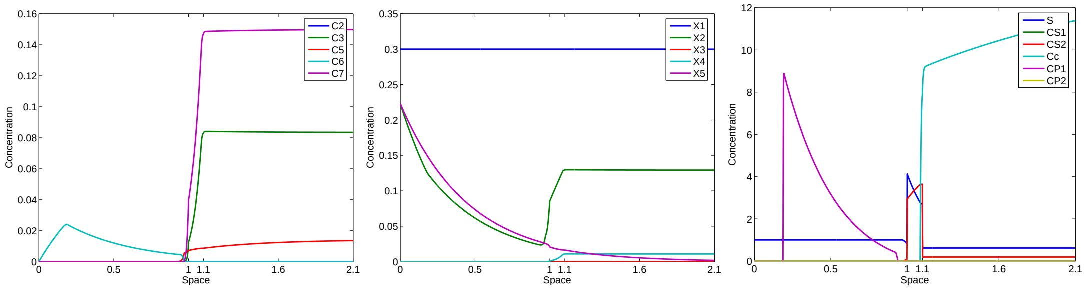
B.3 Hard Test Cases

B.3.1 Advective Hard 1D

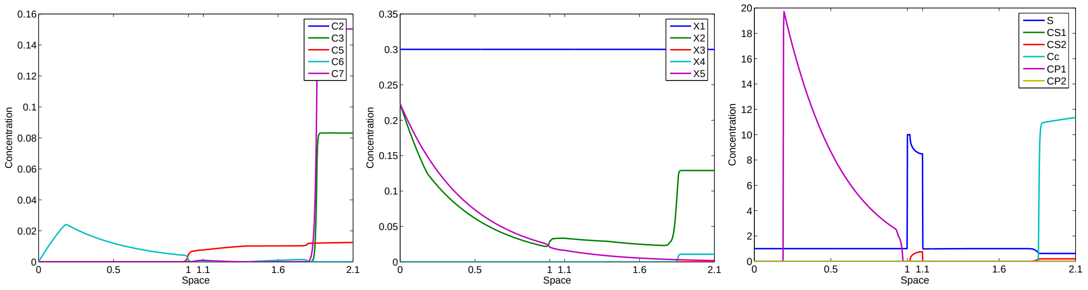
Concentration profile at $t = 10$



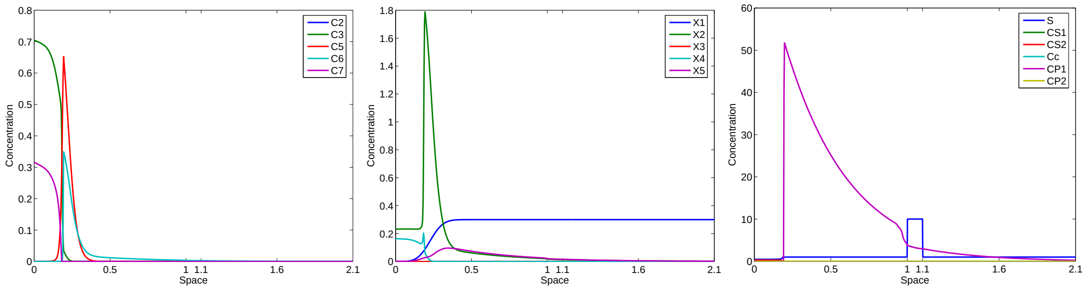
Concentration profile at $t = 1000$



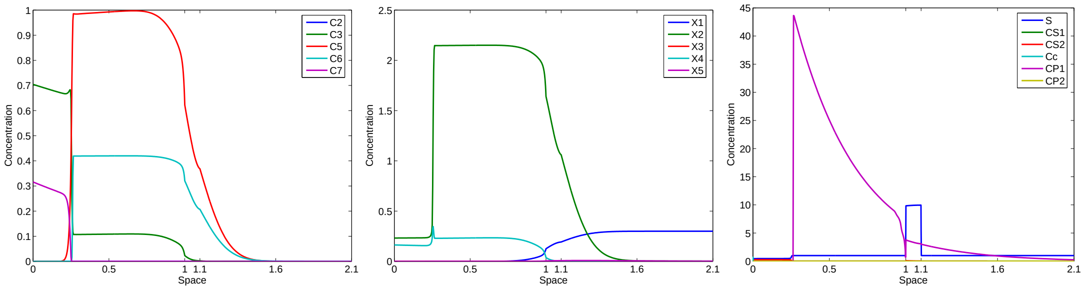
Concentration profile at $t = 2000$



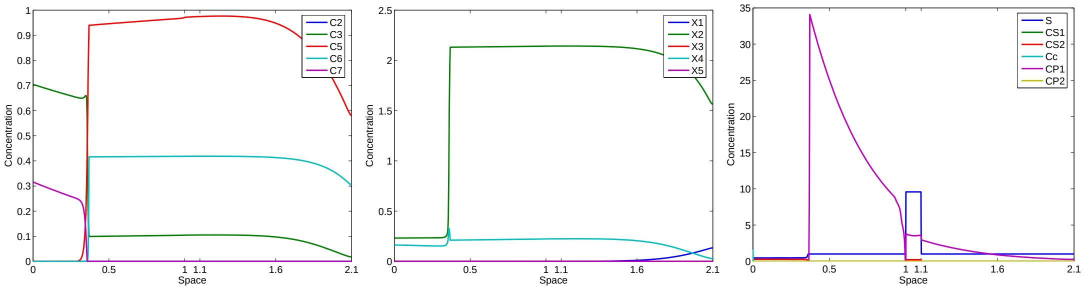
Concentration profile at $t = 5010$



Concentration profile at $t = 5050$

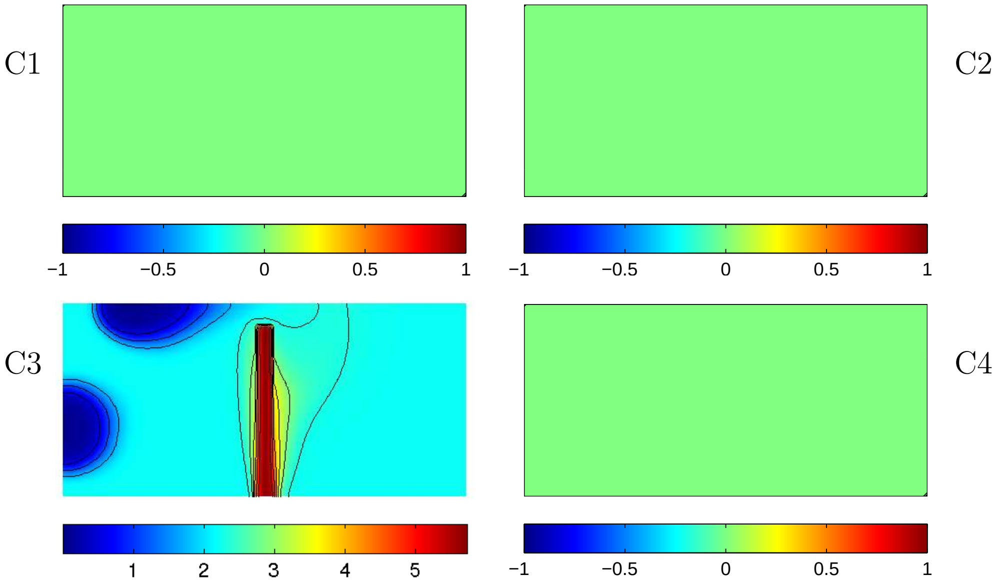


Concentration profile at $t = 5100$

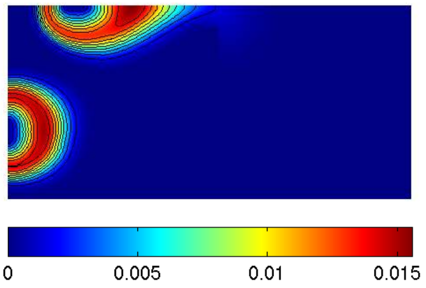


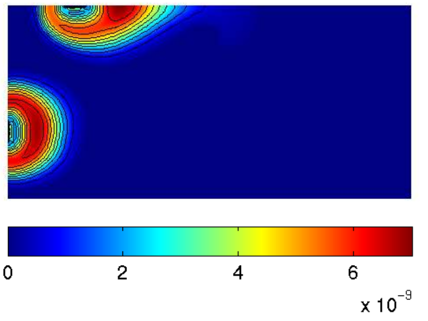
B.3.2 Advective Hard 2D

Concentration profile at $t = 10$



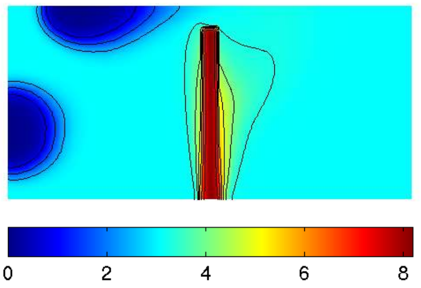
C5

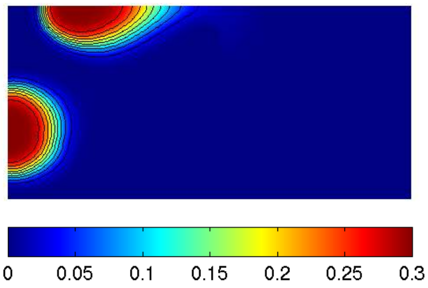




C6

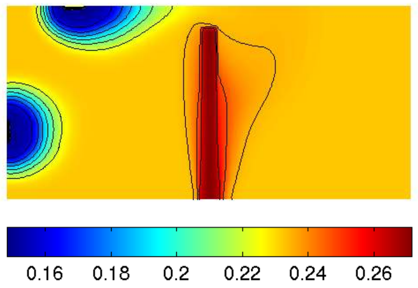
C7

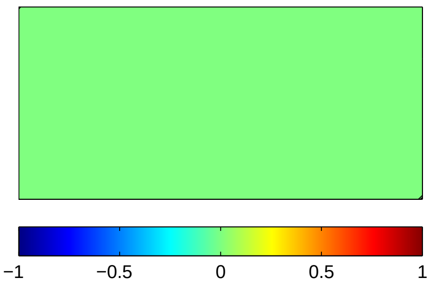




X1

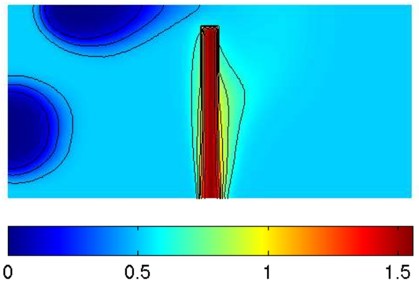
X2

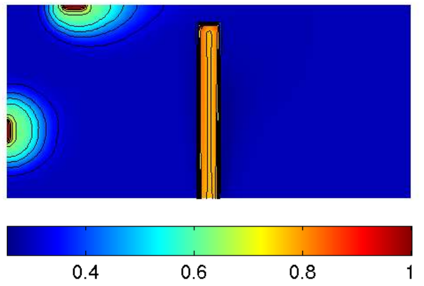




X3

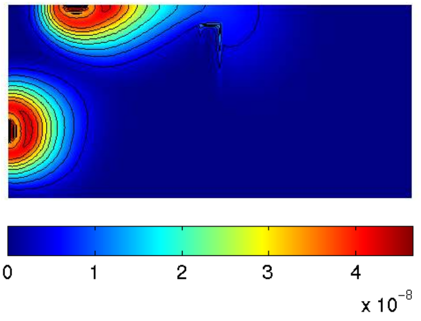
X4

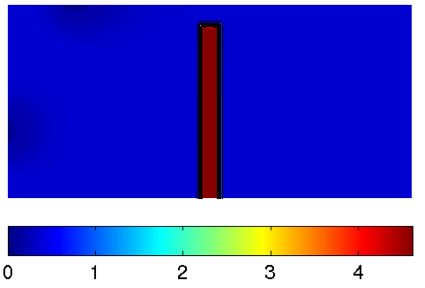




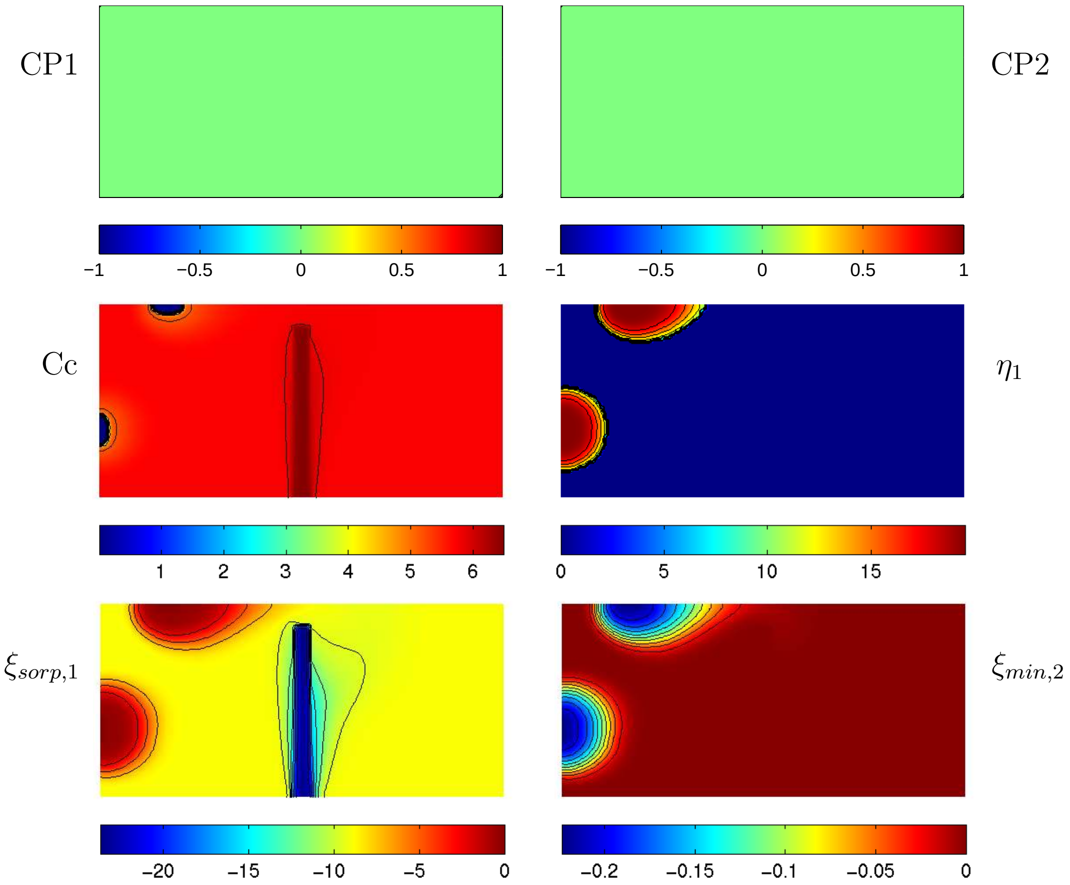
S

CS1

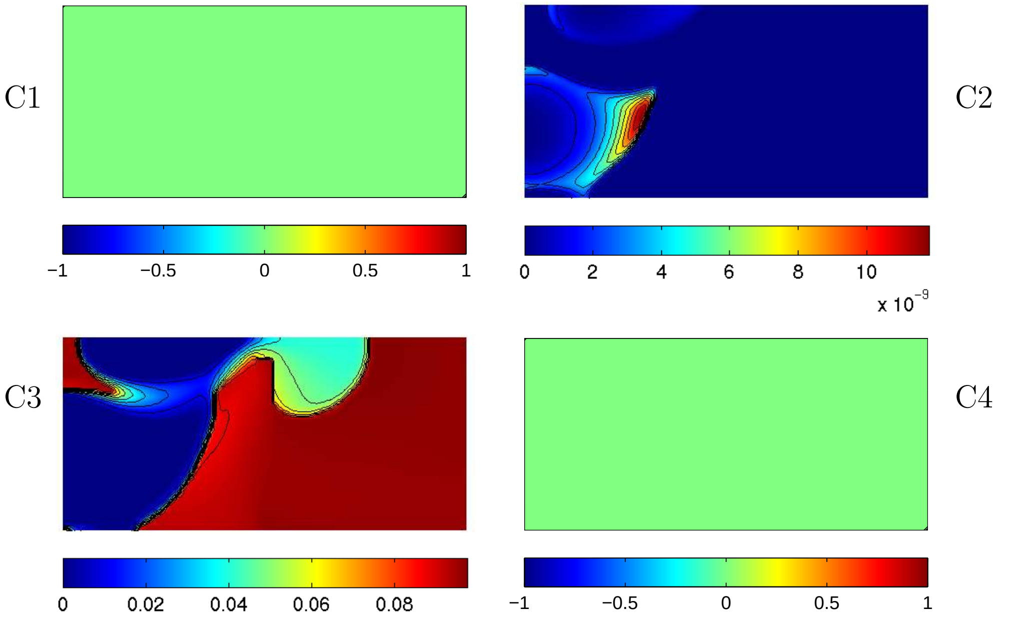




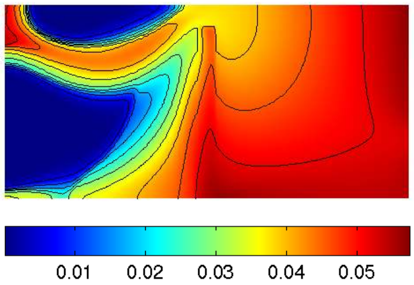
CS2

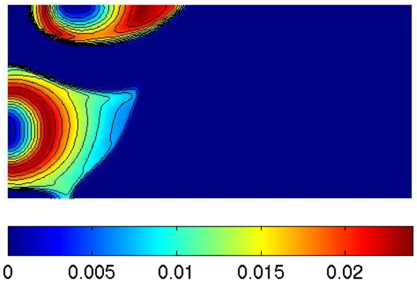


Concentration profile at $t = 1000$



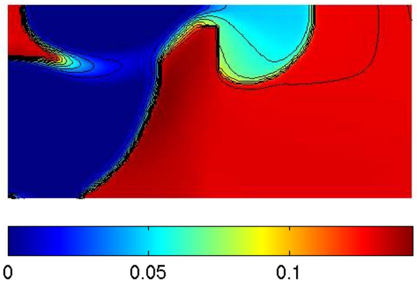
C5

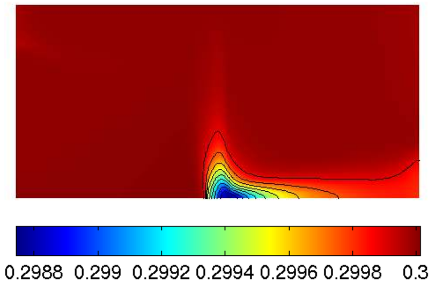




C6

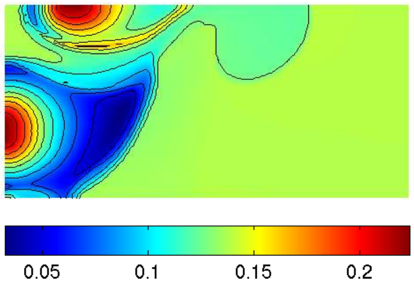
C7

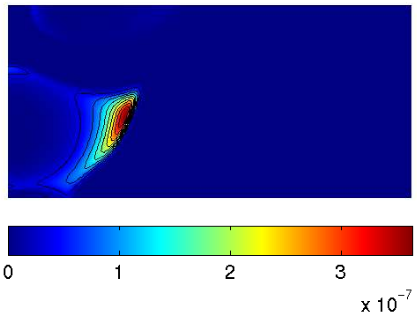




X1

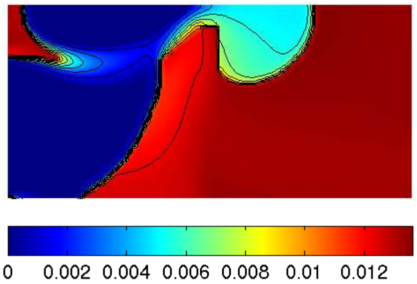
X2

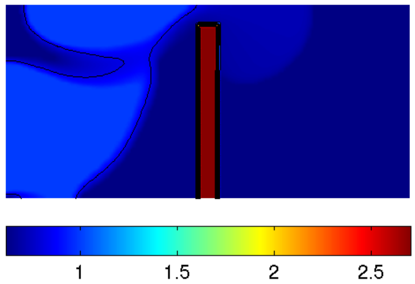




X3

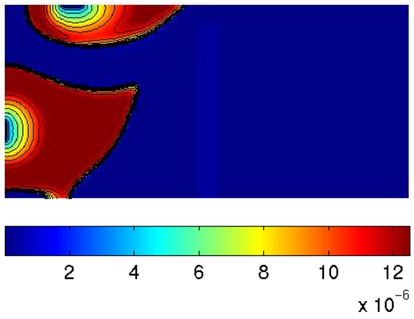
X4

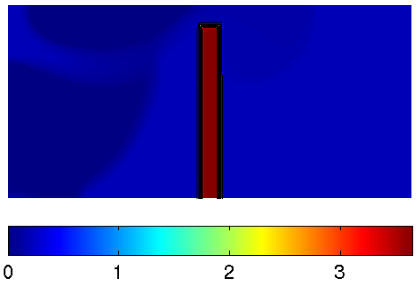




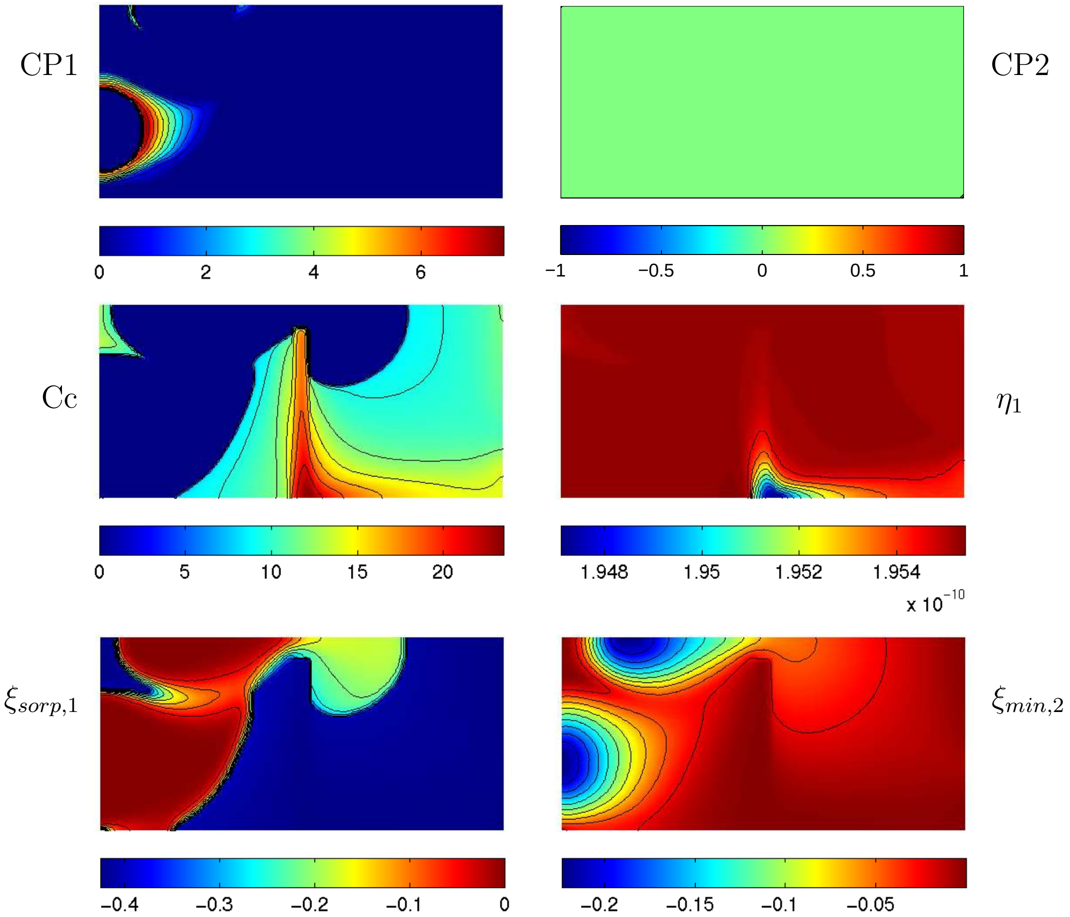
S

CS1

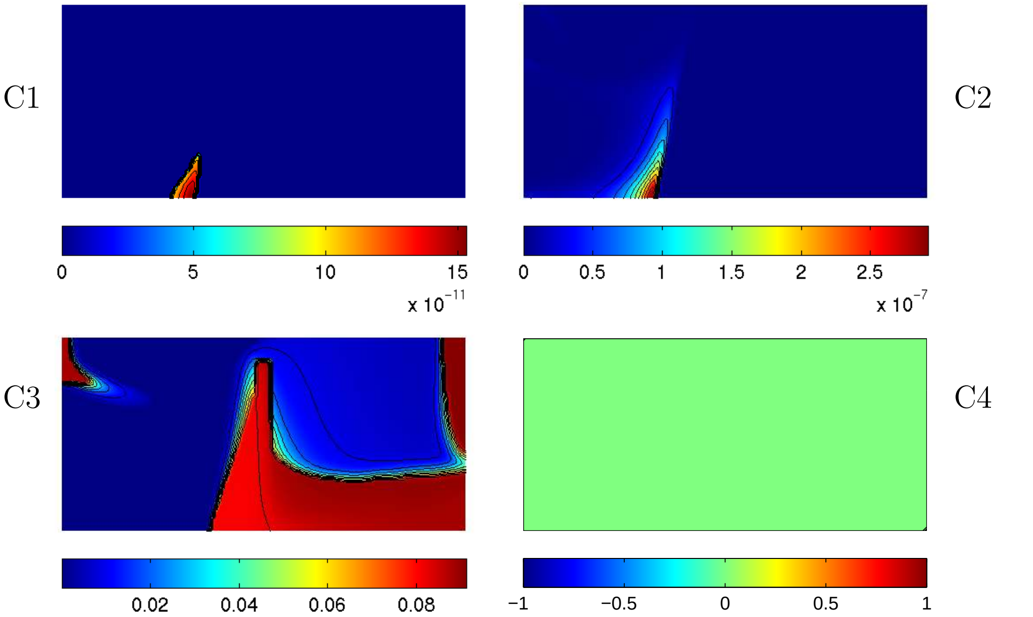




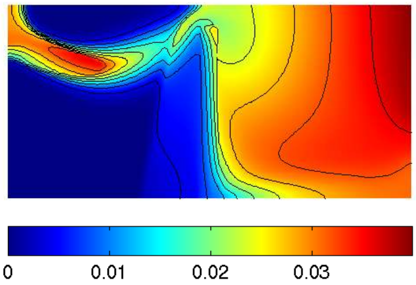
CS2

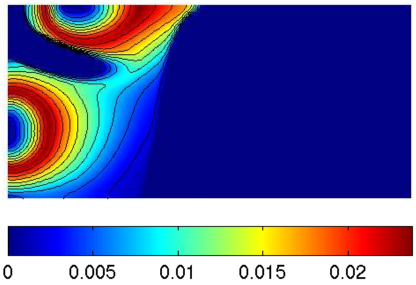


Concentration profile at $t = 2000$



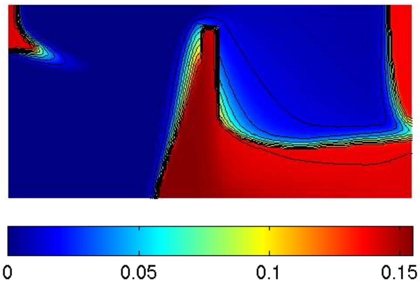
C5

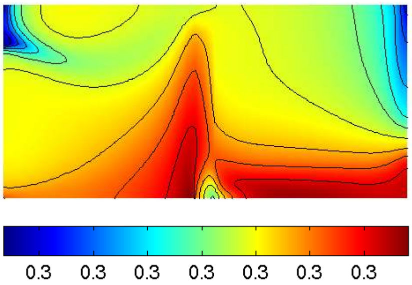




C6

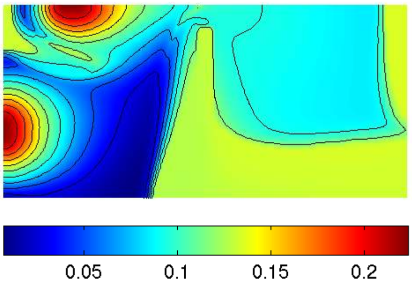
C7

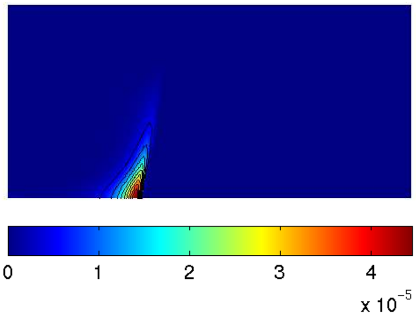




X1

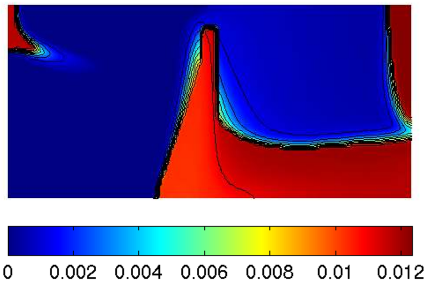
X2

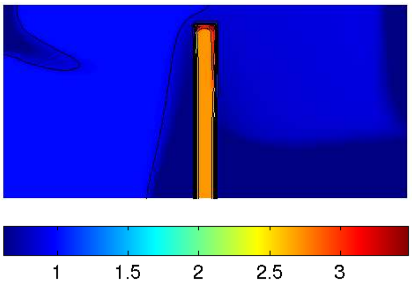




X3

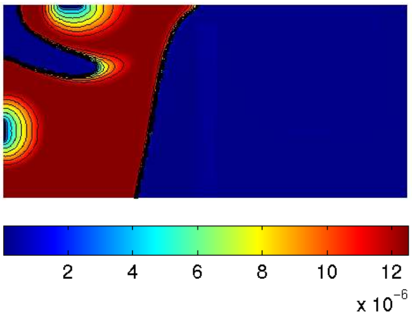
X4

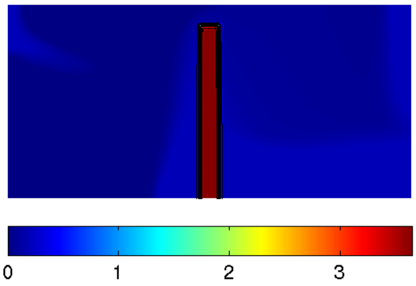




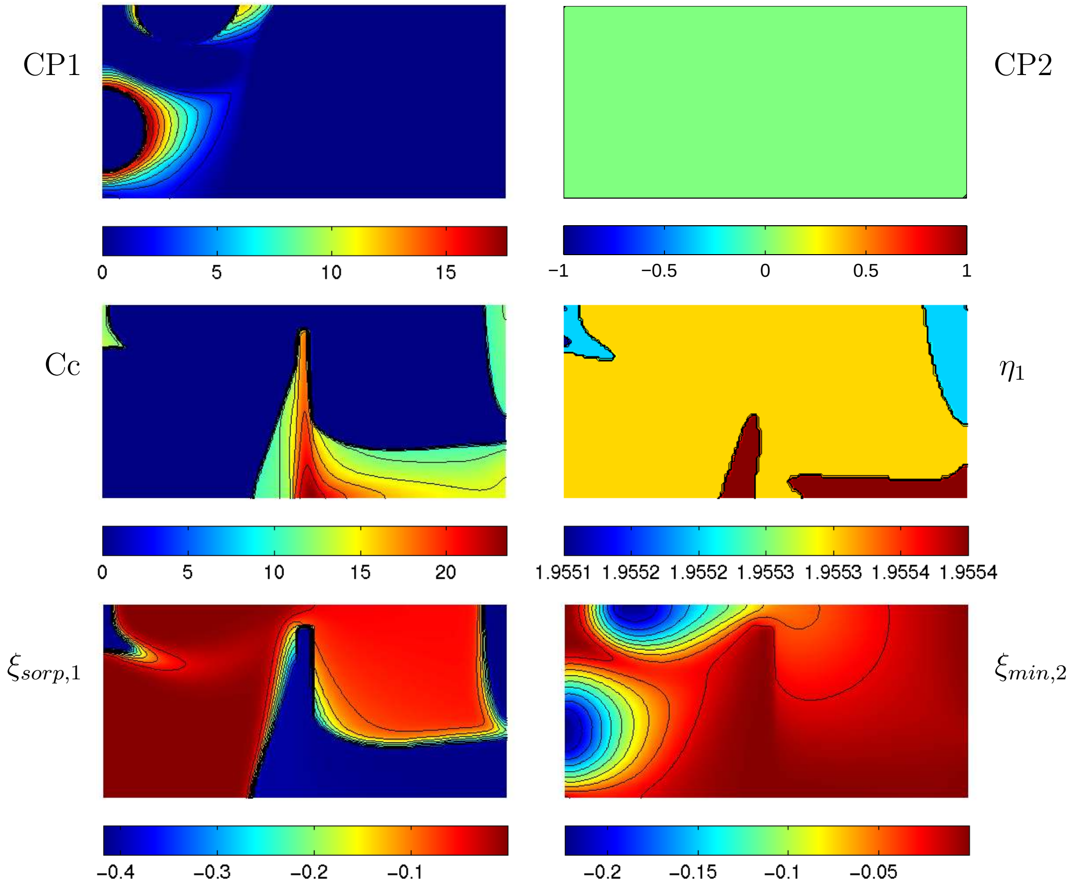
S

CS1

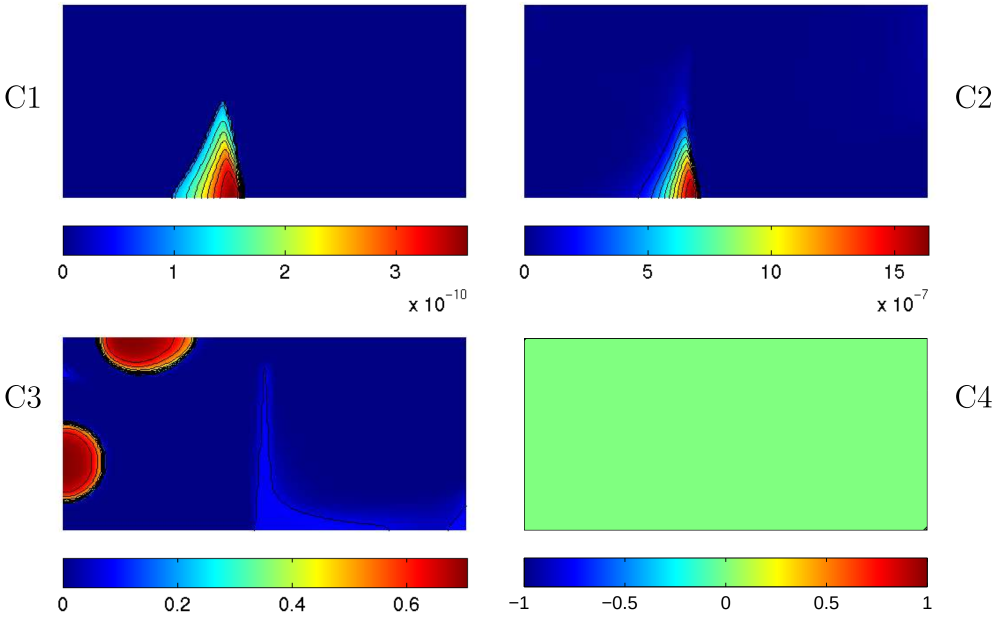




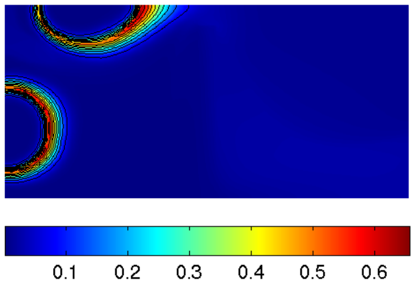
CS2



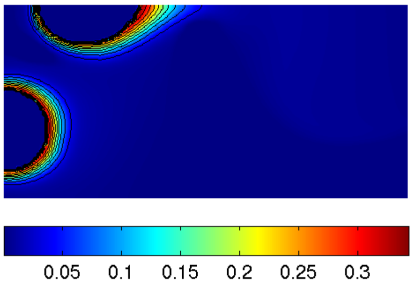
Concentration profile at $t = 5010$



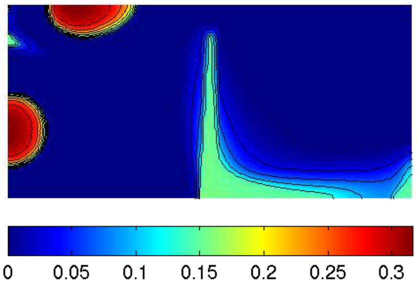
C5



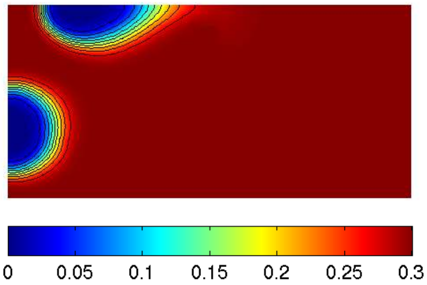
C6



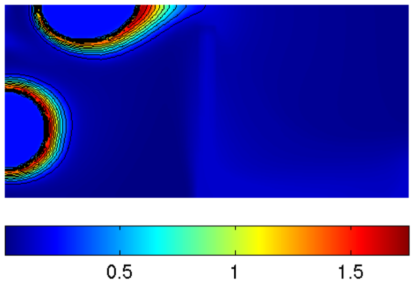
C7



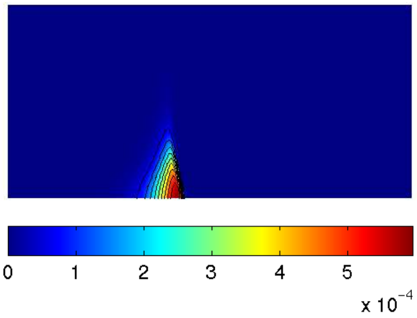
X1



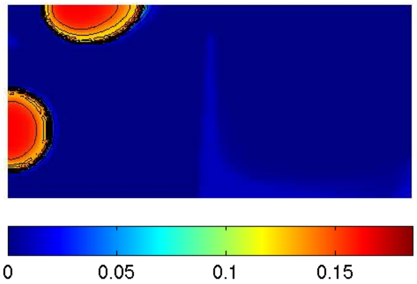
X2



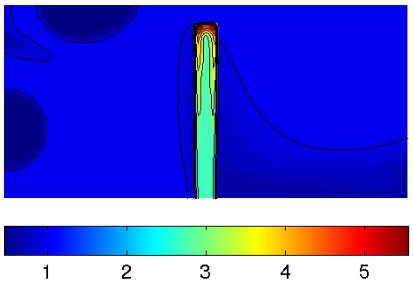
X3



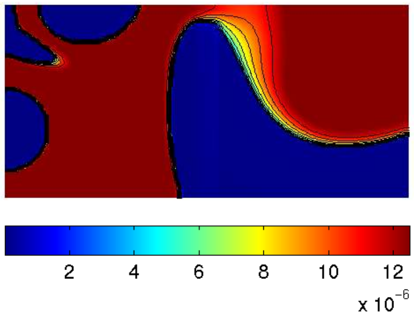
X4



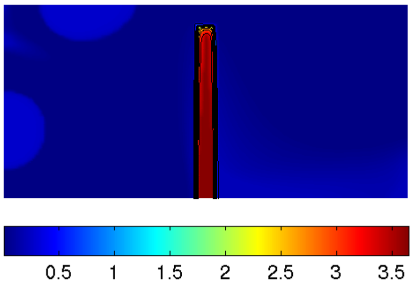
S

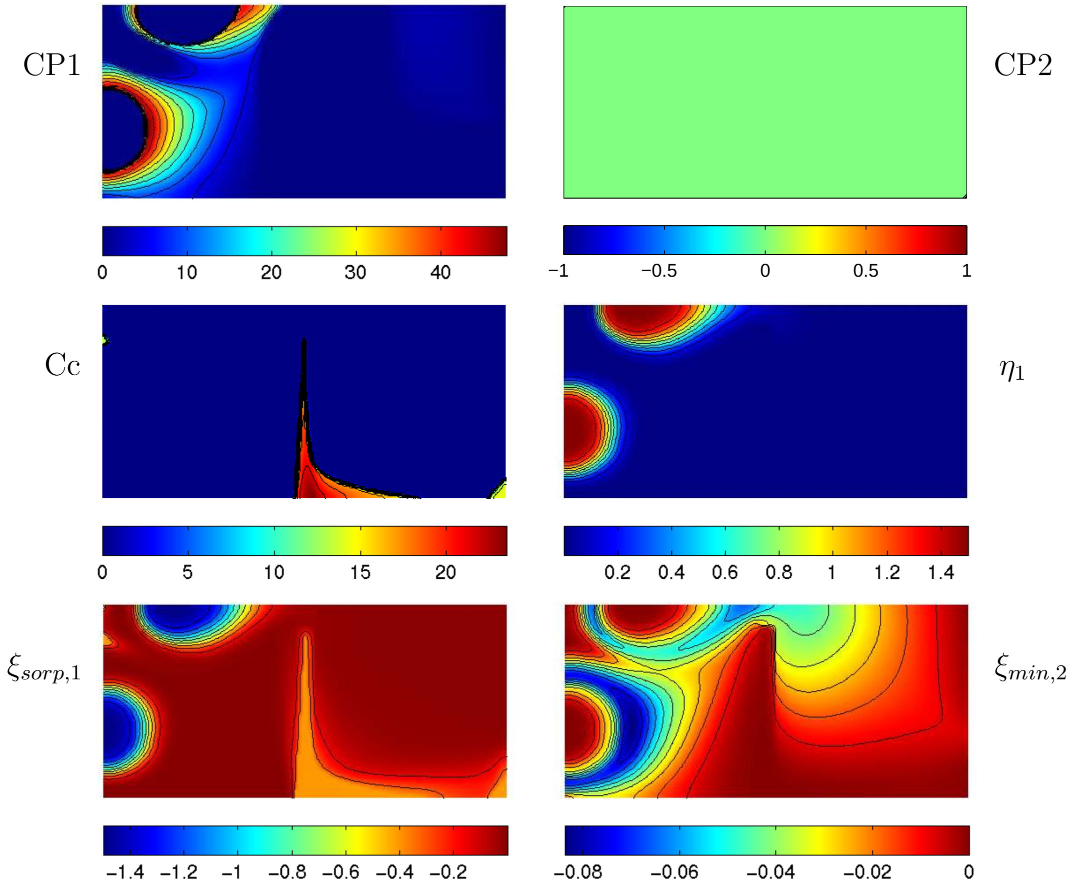


CS1

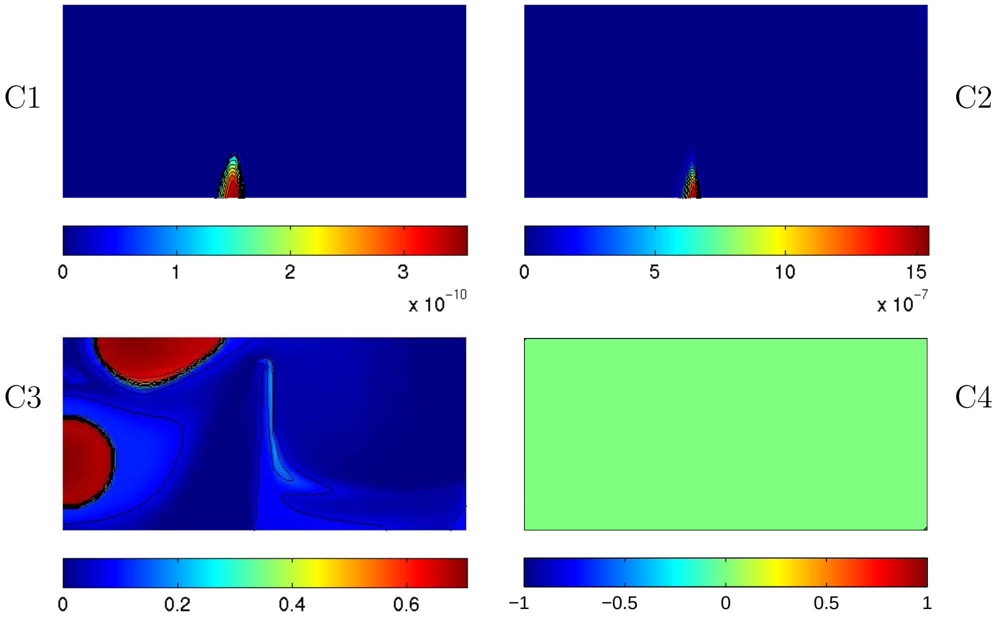


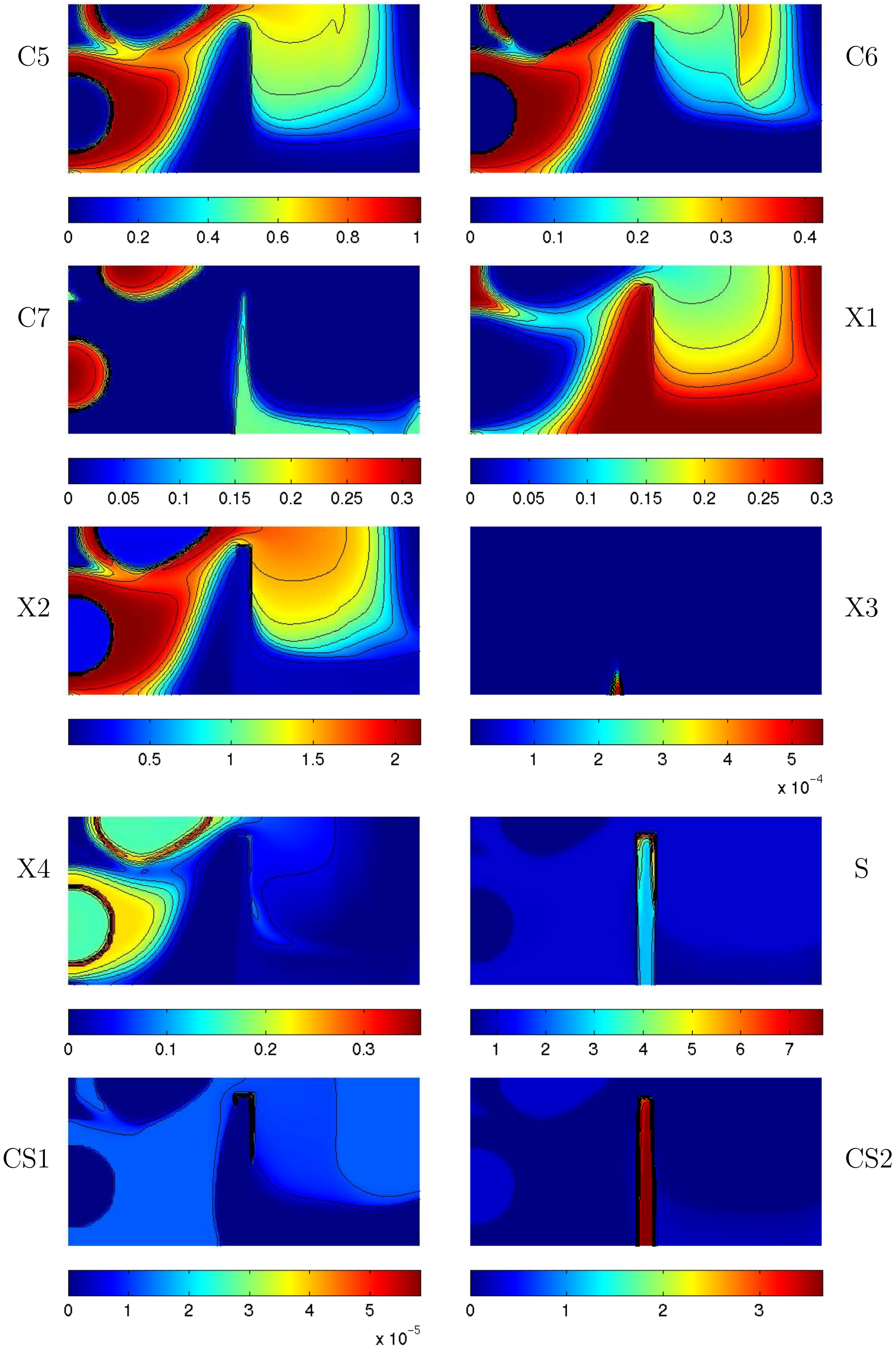
CS2

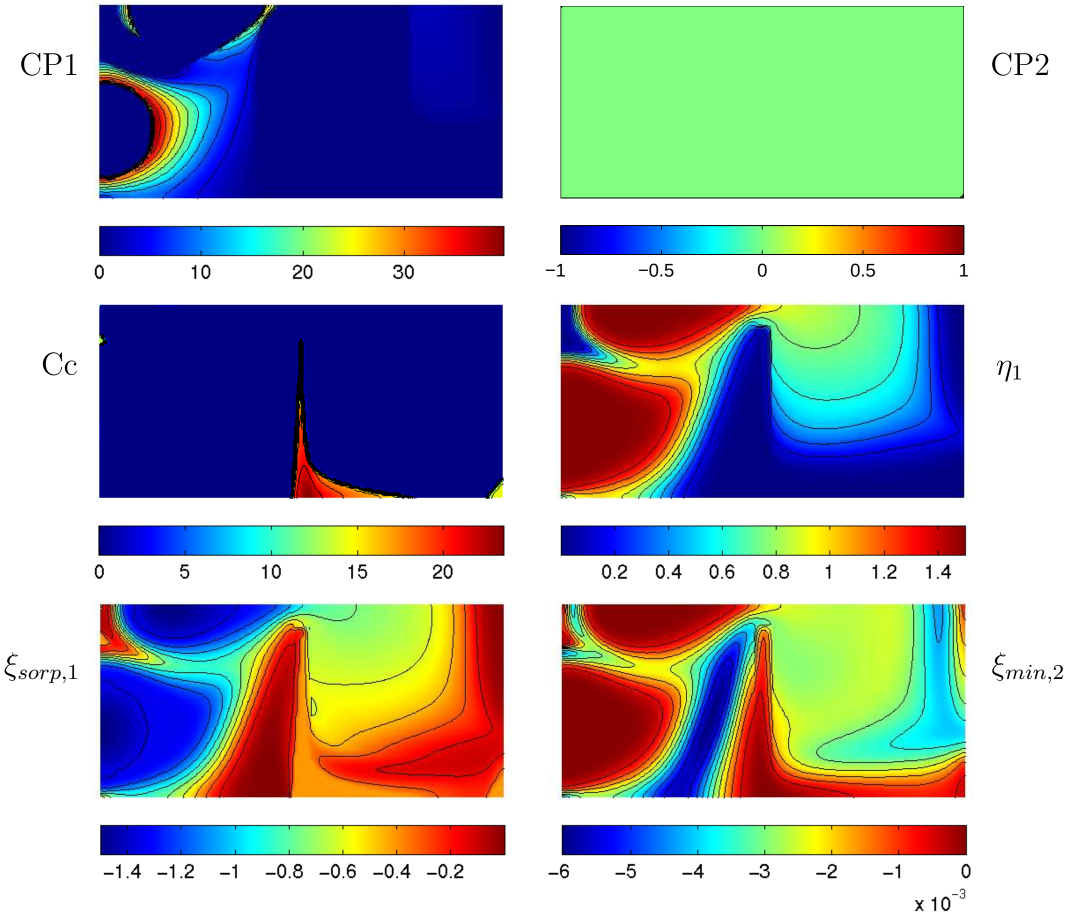




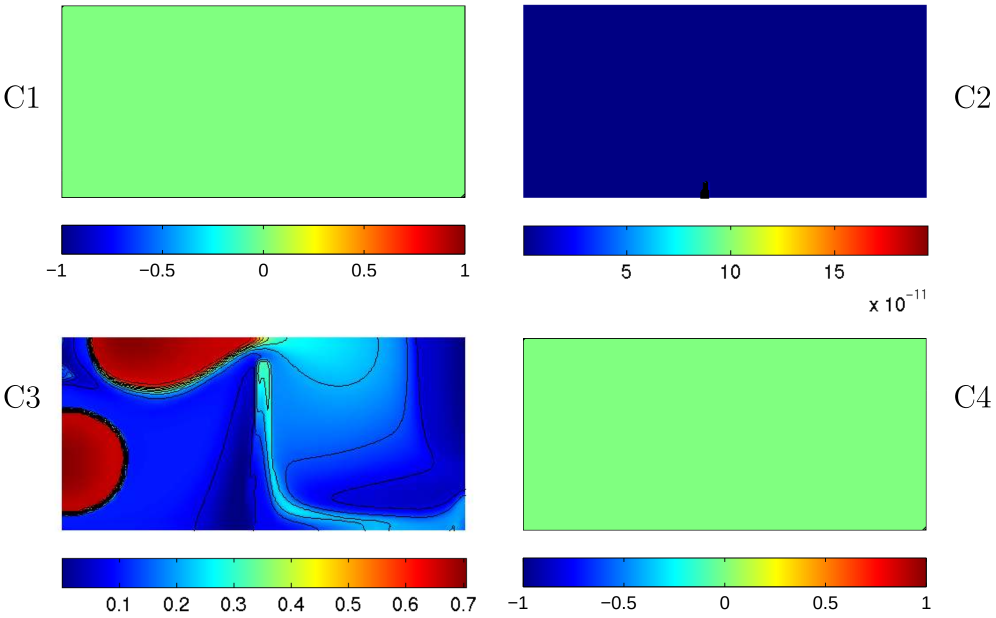
Concentration profile at $t = 5050$

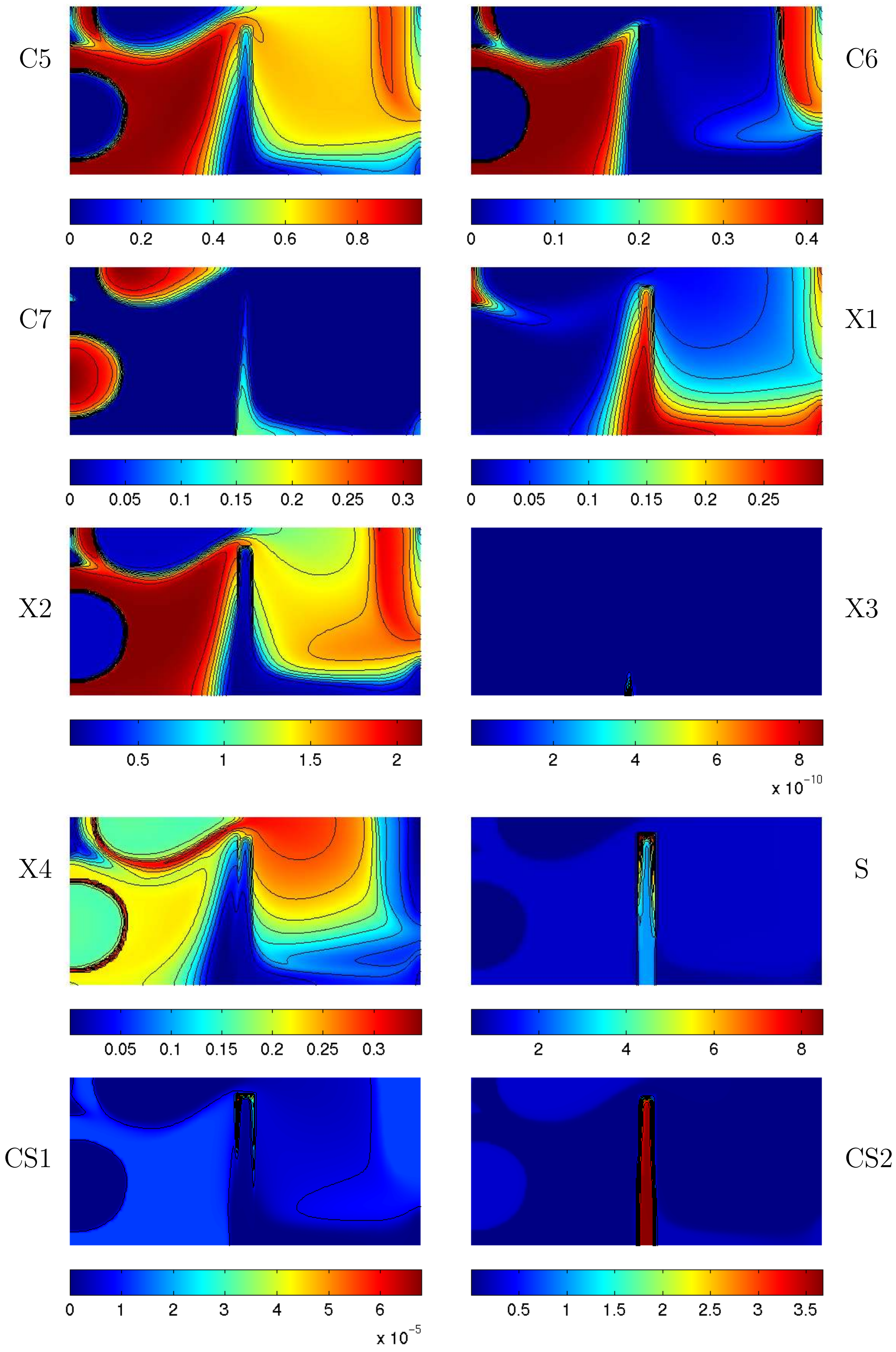



C5
C6
0 0.2 0.4 0.6 0.8 1
0 0.1 0.2 0.3 0.4
C7
X1
0 0.05 0.1 0.15 0.2 0.25 0.3
0 0.05 0.1 0.15 0.2 0.25 0.3
X2
X3
0.5 1 1.5 2
1 2 3 4 5
x 10^{-4}
X4
S
0 0.1 0.2 0.3
1 2 3 4 5 6 7
CS1
CS2
0 1 2 3 4 5
x 10^{-5}
0 1 2 3

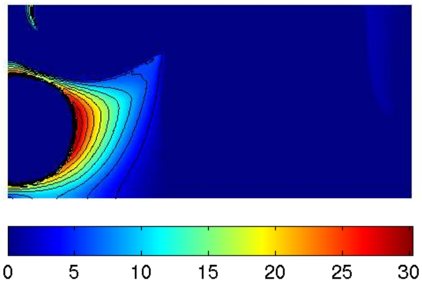


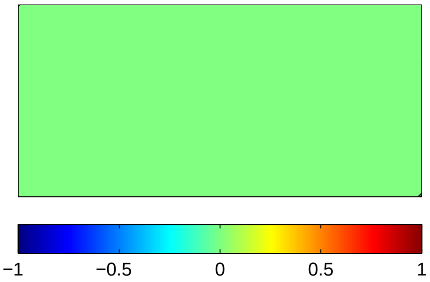
Concentration profile at $t = 5100$




C5
0 0.2 0.4 0.6 0.8
C6
0 0.1 0.2 0.3 0.4
C7
0 0.05 0.1 0.15 0.2 0.25 0.3
X1
0 0.05 0.1 0.15 0.2 0.25
X2
0.5 1 1.5 2
X3
2 4 6 8
x 10^{-10}
X4
0.05 0.1 0.15 0.2 0.25 0.3
S
2 4 6 8
CS1
0 1 2 3 4 5 6
x 10^{-5}
CS2
0.5 1 1.5 2 2.5 3 3.5

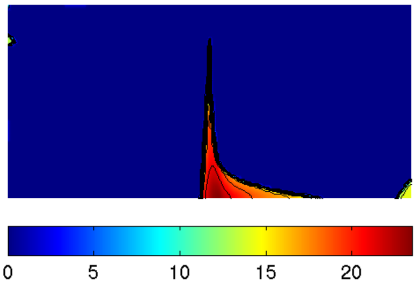
CP1

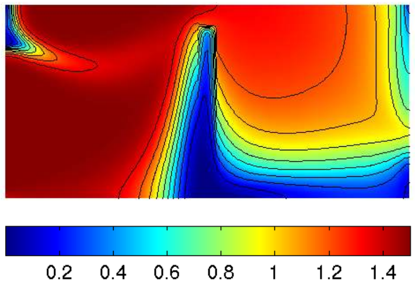




CP2

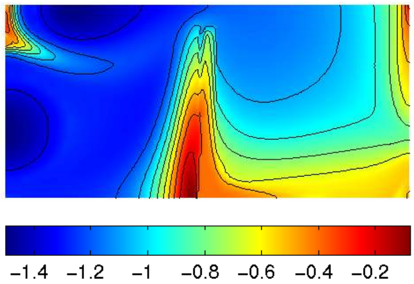
Cc

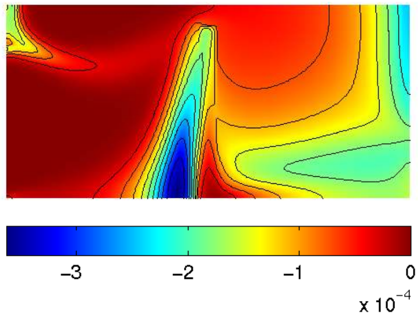




η_1

$\xi_{sorp,1}$



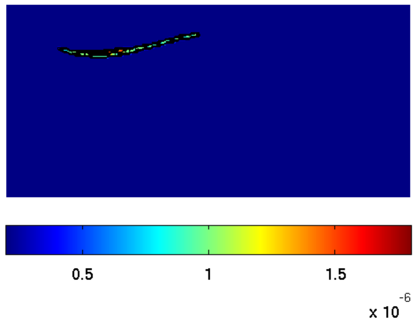


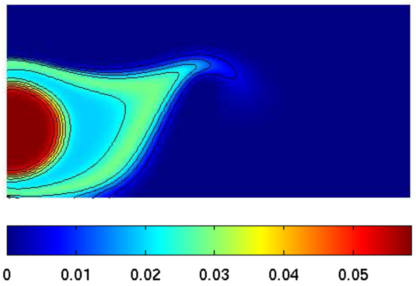
$\xi_{min,2}$

B.4 Modified Scenario

Concentration profile at $t = 50$

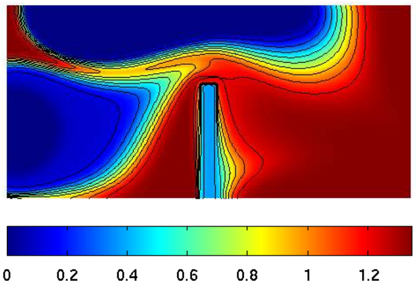
C1



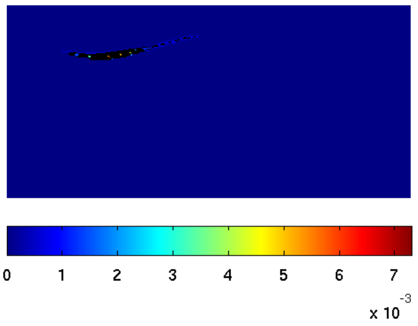


C2

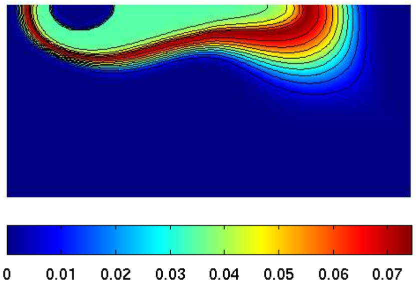
C3



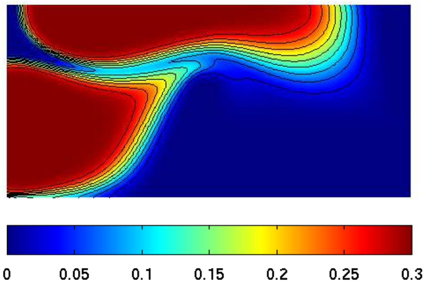
C4



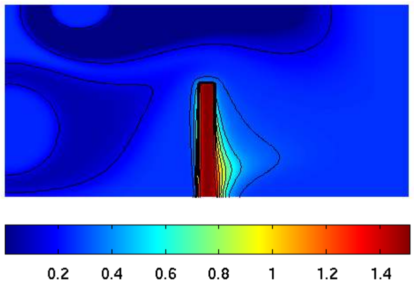
C5



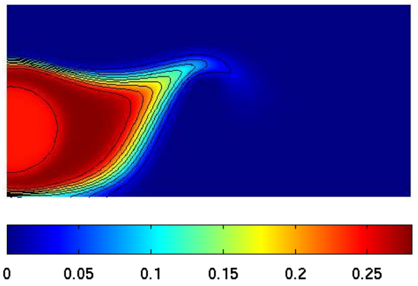
X1



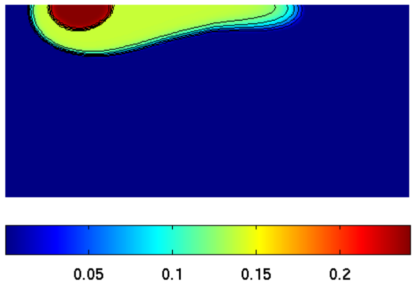
X2



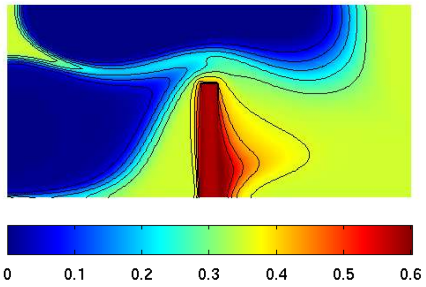
X3a



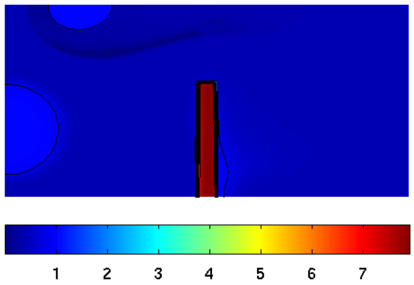
X3b



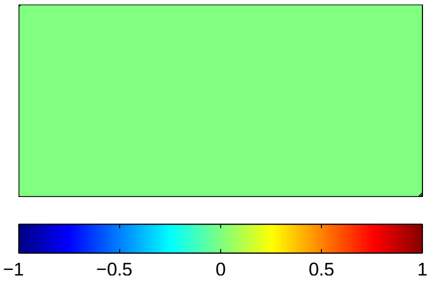
X4

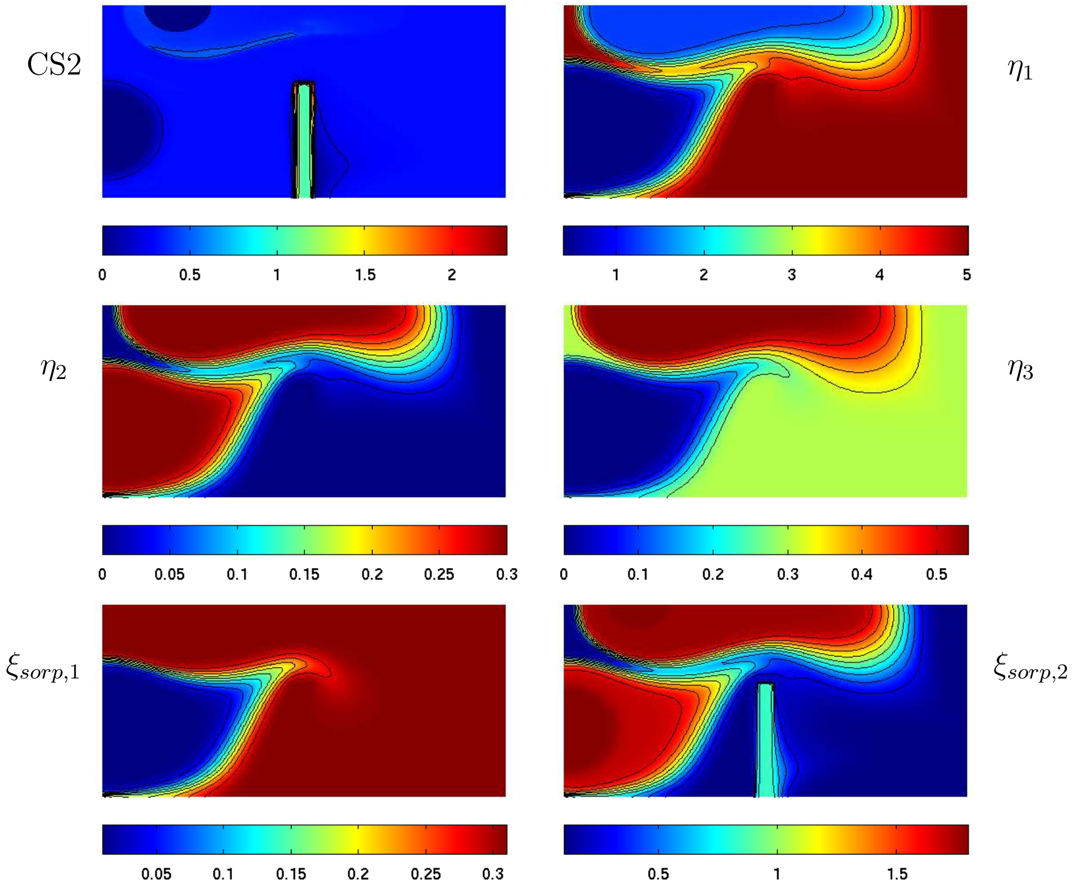


S

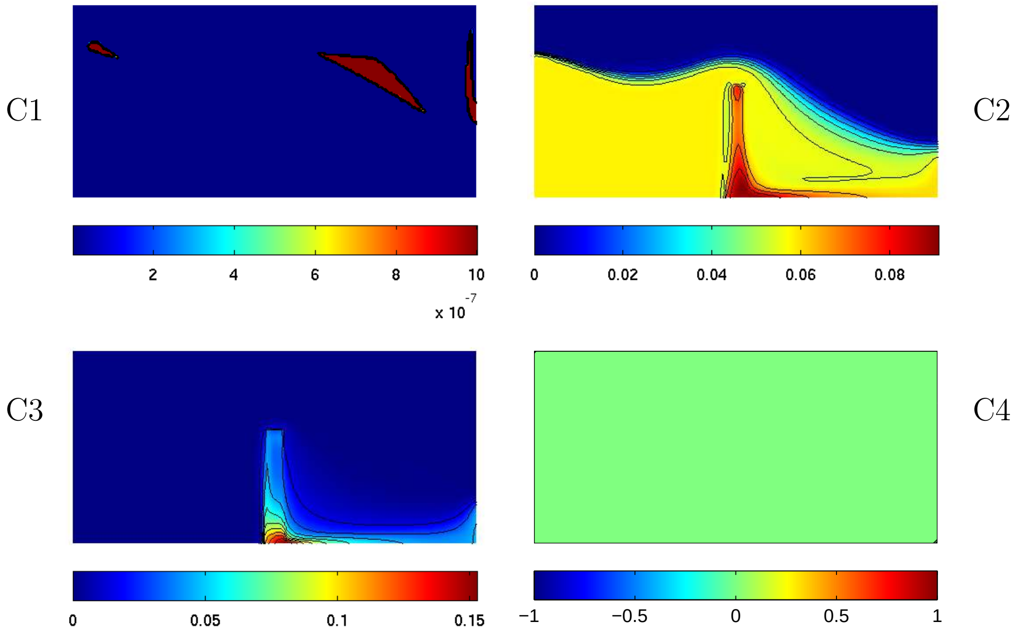


CS1

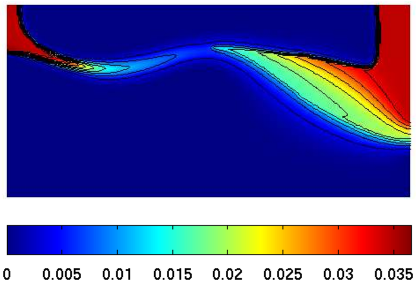




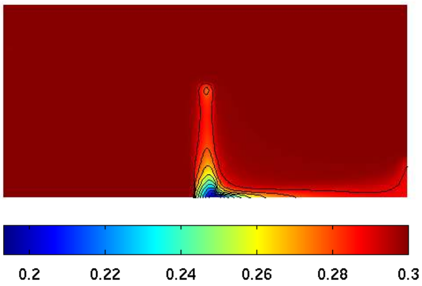
Concentration profile at $t = 750$



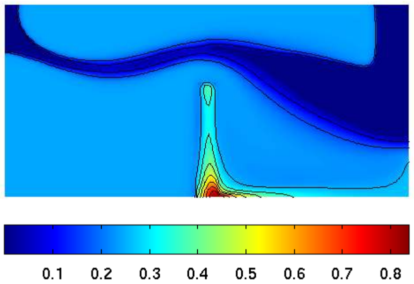
C5



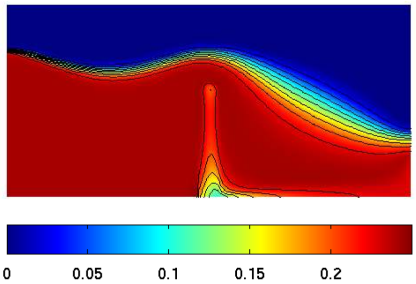
X1



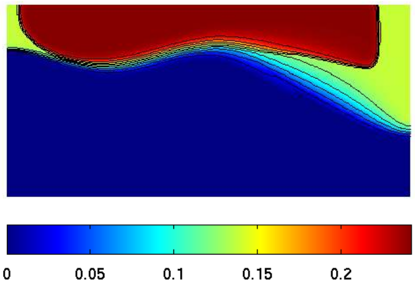
X2



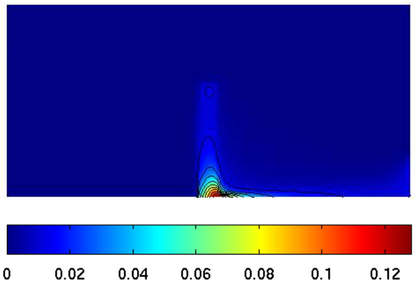
X3a



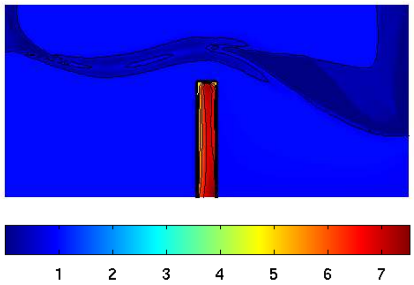
X3b



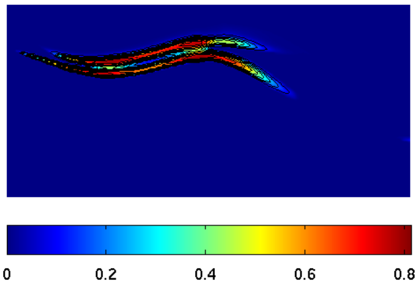
X4



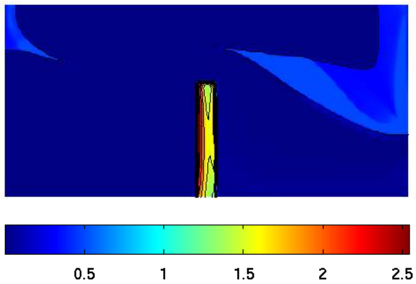
S



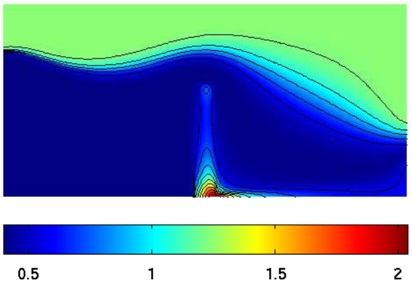
CS1

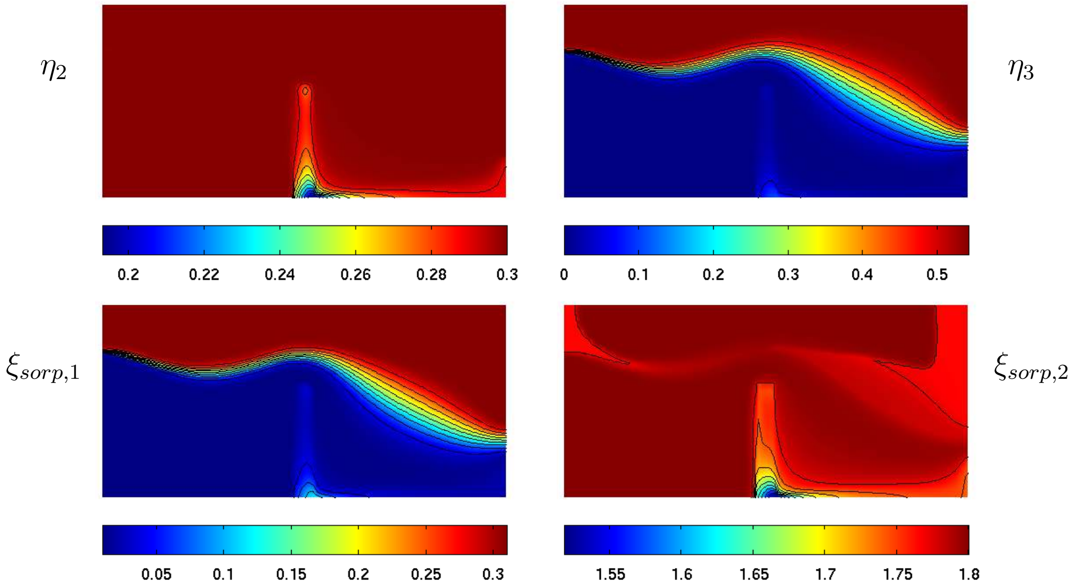


CS2

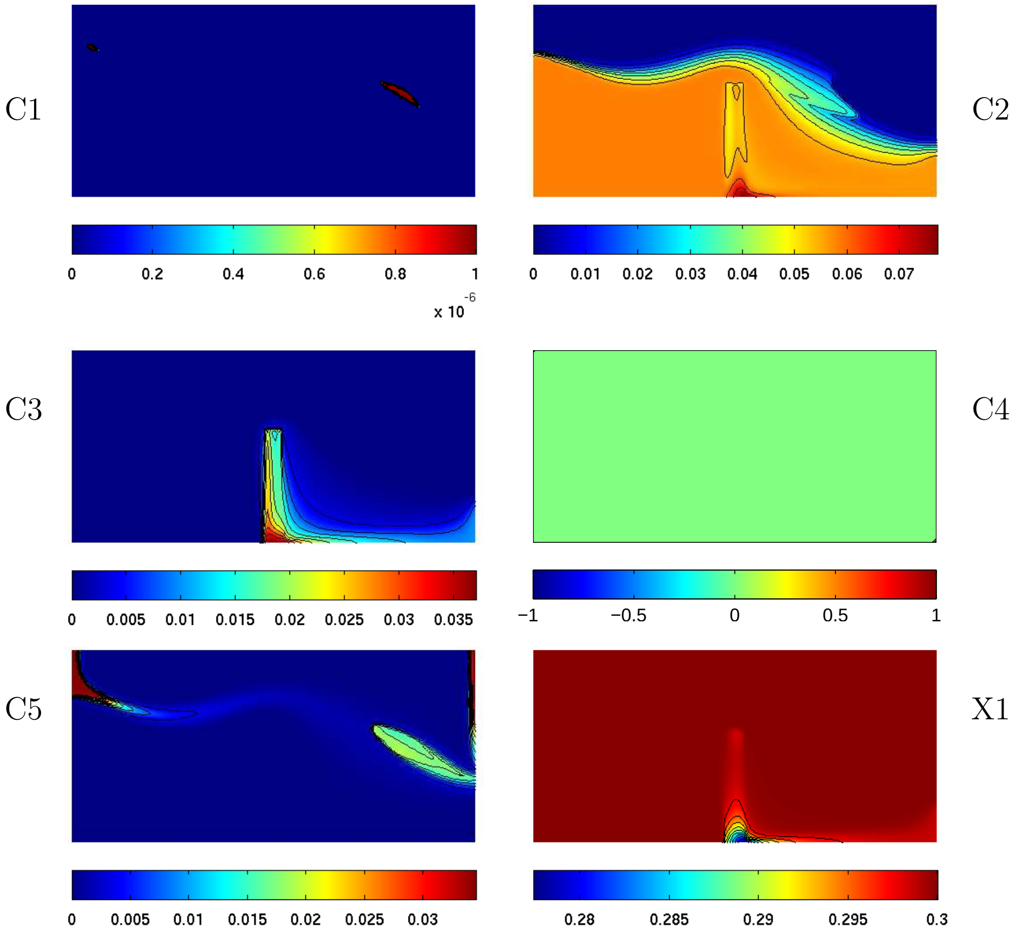


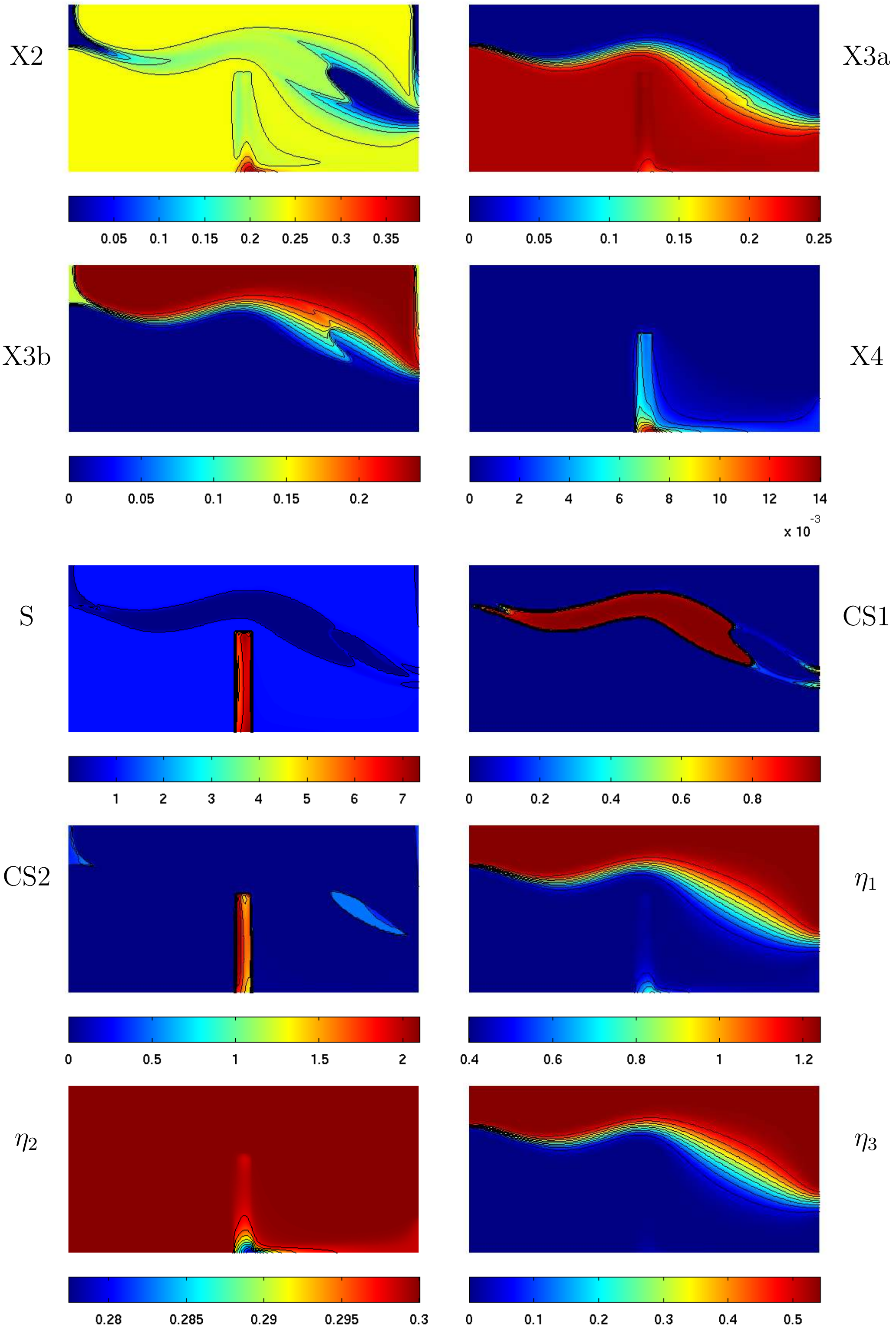
η_1



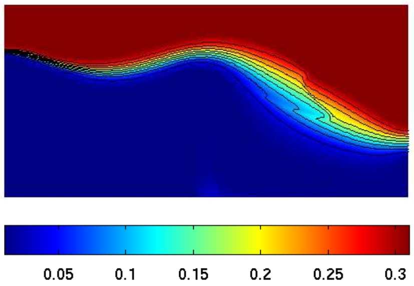


Concentration profile at $t = 1300$

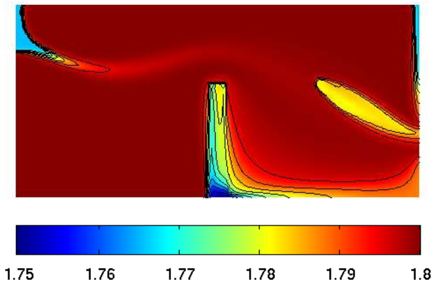



X2
0.05 0.1 0.15 0.2 0.25 0.3 0.35
X3a
0 0.05 0.1 0.15 0.2 0.25
X3b
0 0.05 0.1 0.15 0.2
X4
0 2 4 6 8 10 12 14
x 10^{-3}
S
1 2 3 4 5 6 7
CS1
0 0.2 0.4 0.6 0.8
CS2
0 0.5 1 1.5 2
η_1
0.4 0.6 0.8 1 1.2
η_2
0.28 0.285 0.29 0.295 0.3
η_3
0 0.1 0.2 0.3 0.4 0.5

$\xi_{sorp,1}$

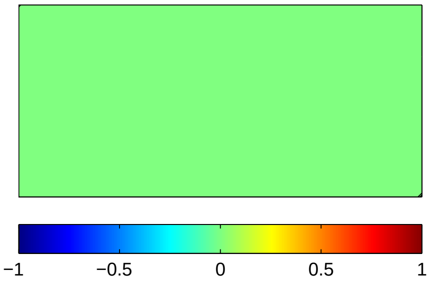


$\xi_{sorp,2}$

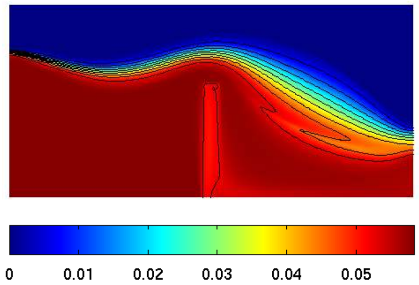


Concentration profile at $t = 2500$

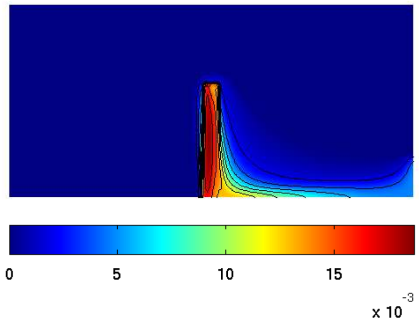
C1



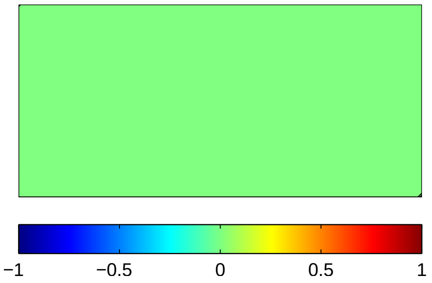
C2



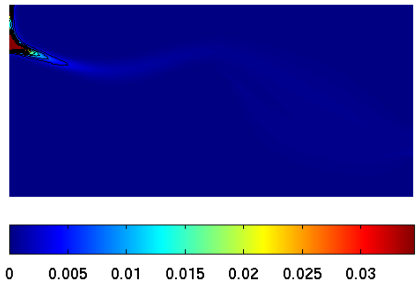
C3



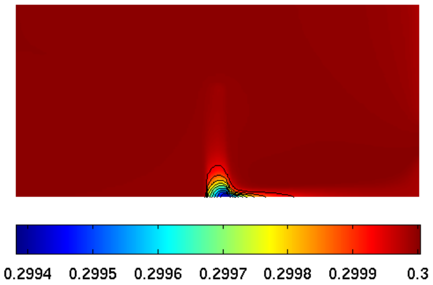
C4



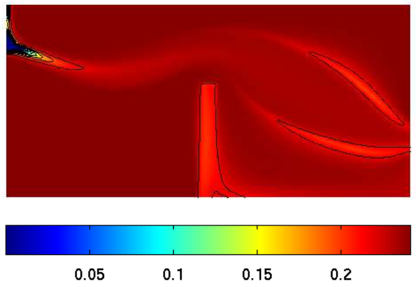
C5



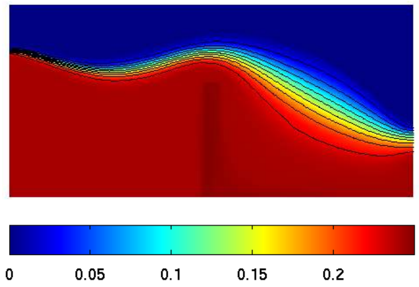
X1



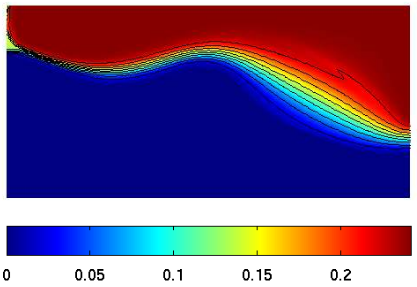
X2



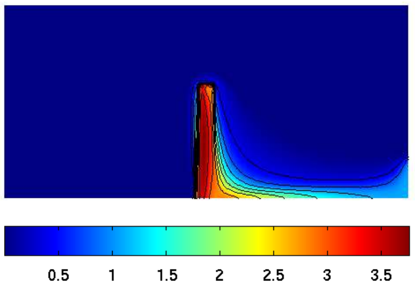
X3a



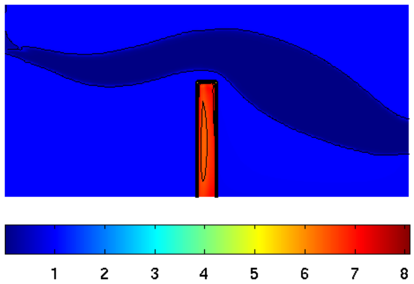
X3b



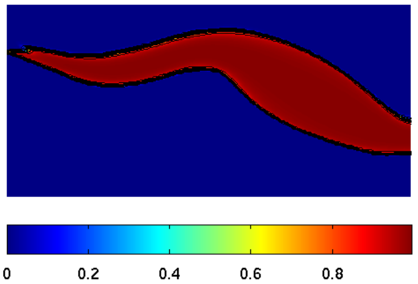
X4



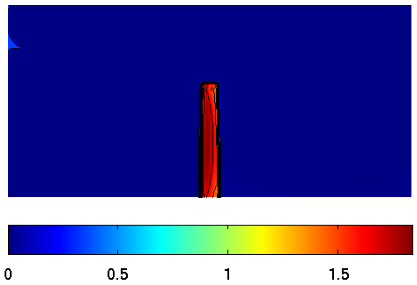
S



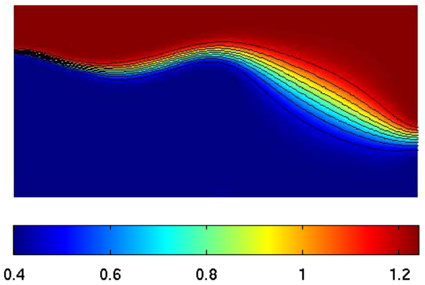
CS1



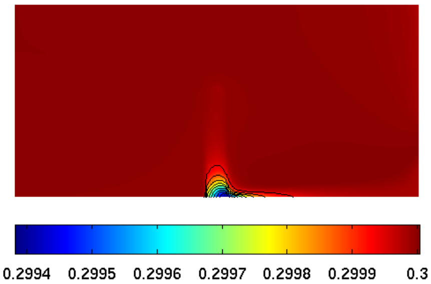
CS2



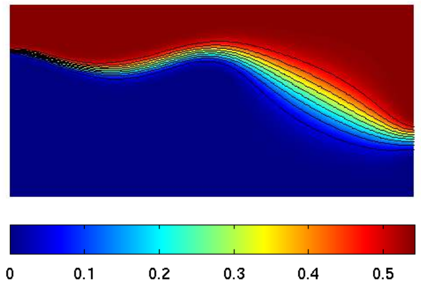
η_1



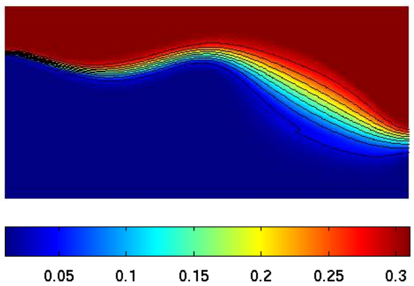
η_2



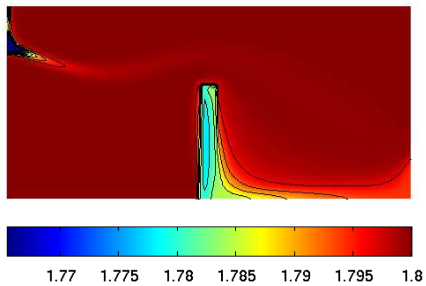
η_3



$\xi_{sorp,1}$



$\xi_{sorp,2}$



List of Symbols

$\boldsymbol{A}$ submatrix of $\boldsymbol{S}$ in standard form, size: $(I - J_{mob}) \times J_{sorp}$, 83

$\boldsymbol{A}_1$ matrix from relation (3.9), size: $(J_{mob}+J_{sorp,li}+J_{min}+J^*_{1,kin})\times J$, 23

$\boldsymbol{A}_{1,kin}$ $J^*_{1,kin} \times J_{kin}$-submatrix of $\boldsymbol{A}_1$, 23

$\boldsymbol{A}_{1,min}$ $J_{min} \times J_{kin}$-submatrix of $\boldsymbol{A}_1$, 23

$\boldsymbol{A}_{1,mob}$ $J_{mob} \times J_{kin}$-submatrix of $\boldsymbol{A}_1$, 23

$\boldsymbol{A}_{1,sorp}$ $J_{sorp,li} \times J_{kin}$-submatrix of $\boldsymbol{A}_1$, 23

$\boldsymbol{A}_2$ matrix from relation (3.9), size: $(J_{sorp} + J_{min} + J^*_{2,kin}) \times J$, 23

$\boldsymbol{A}_{2,kin}$ $J^*_{2,kin} \times J_{kin}$-submatrix of $\boldsymbol{A}_2$, 23

$\boldsymbol{A}_{2,sorp}$ $J_{sorp} \times J_{kin}$-submatrix of $\boldsymbol{A}_2$, 23

$\boldsymbol{A}_{2,sorp,ld}$ $(J_{sorp} - J_{sorp,li}) \times J_{kin}$-submat. of $\boldsymbol{A}_{2,sorp}$ with the last $(J_{sorp} - J_{sorp,li})$ rows, 27

$\boldsymbol{A}_{2,sorp,li}$ $J_{sorp,li} \times J_{kin}$-submat. of $\boldsymbol{A}_{2,sorp}$ with the first $J_{sorp,li}$ rows, 27

$\boldsymbol{A}_{ld}$ matrix from relation (3.2), size : $J_{min} \times (J_{sorp} - J_{sorp,li})$, 21

$\hat{\boldsymbol{B}}$ submatrix of $\boldsymbol{S}$ in standard form, size: $(\bar{I}_{nmin} - J_{sorp}) \times J_{sorp}$, 83

$\boldsymbol{B}_1$ matrix for variable transformation, size: $I \times (J_{mob} + J_{sorp,li} + J_{min} + J^*_{1,kin})$, 23

$\boldsymbol{B}_1^{\perp}$ matrix with max. system orthogonal to $\boldsymbol{B}_1$, size: $I \times (I - J_{mob} - J_{sorp,li} - J_{min} - J^*_{1,kin})$, 23

$\boldsymbol{B}_2$ matrix for variable transformation, size: $\bar{I} \times (J_{sorp} + J_{min} + J^*_{2,kin})$, 23

$\boldsymbol{B}_2^{\perp}$ matrix with max. system orthogonal to $\boldsymbol{B}_2$, size: $\bar{I} \times (\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin})$, 23

$\tilde{\boldsymbol{B}}_2^{\perp}$ $\bar{I}_{nmin} \times (\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin})$-submatrix of $\boldsymbol{B}_2^{\perp}$, 25

β_l longitudinal dispersion coefficient, 17

β_t transversal dispersion coefficient, 17

$\boldsymbol{C}$ submatrix of $\boldsymbol{S}$ in standard form, size: $(I - J_{mob}) \times J_{mob}$, 83

$\boldsymbol{C}_1$ transformation matrix of generalized reduction scheme, 94

$\boldsymbol{c}$ concentration vector of mobile species, length: I, 16

$\boldsymbol{c}_{prim}$ primary mobile concentrations, vector length: $I - J_{mob}$, 86

$\boldsymbol{c}_{sec}$ secondary mobile concentrations, vector length: J_{mob}, 86

$\bar{\boldsymbol{c}}$ concentration vector of immobile species, length: $\bar{I}$, 16

$\bar{\boldsymbol{c}}_{min}$ concentration vector of minerals, length: $\bar{I}_{min}$, 21

$\bar{\boldsymbol{c}}_{nmin}$ concentration vector of nonminerals, length: $\bar{I}_{nmin}$, 21

$\bar{\boldsymbol{c}}_{nmin,prim}$ primary nonmineral concentrations, vector length: $\bar{I}_{nmin} - J_{sorp}$, 86

$\bar{\boldsymbol{c}}_{nmin,sec}$ secondary nonmineral concentrations, vector length: J_{sorp}, 86

$\boldsymbol{D}$ submarix of $\boldsymbol{S}$ in standard form, size: $(I - J_{mob}) \times J_{min}$, 83

$d_{diff,i}$ diffusion coefficient of the i-th species, 17

$\boldsymbol{D}_i$ diffusion/dispersion tensor of the i-th species, 17

$\boldsymbol{\eta}$ transformed variables, number of var.: $I - J_{mob} - J_{sorp,li} - J_{min} - J^*_{1,kin}$, 24

$\bar{\boldsymbol{\eta}}$ transformed variables, number of var.: $\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin}$, 24

I number of mobile species, 16

$\bar{I}$ number of immobile species, 16

$\bar{I}_{min}$ number of minerals, 21

$\bar{I}_{nmin}$ number of nonminerals, 21

$\boldsymbol{I}_n$ identity matrix of size n, 21

J number of chemical reactions, 17

J_{eq} number of all equilibrium reactions, 19

J_{kin} number of kinetic reactions, 18

$J^*_{1,kin}$ number of columns of $\boldsymbol{S}^*_{1,kin}$, 23

$J^*_{2,kin}$ number of columns of $\boldsymbol{S}^*_{2,kin}$, 23

J_{min} number equilibrium mineral reactions, 21

J_{mob} number of equilibrium reactions with only mobile species, 21

J_{sorp} number of equilibrium sorption reactions, 21

$J_{sorp,li}$ number of columns of $\boldsymbol{S}_{1,sorp,li}$, 22

K_j equilibrium constant of the j-th equilibrium reaction, 18

$k_{b,j}$ backward coefficient of the j-th kinetic reaction, 18

$k_{f,j}$ forward coefficient of the j-th kinetic reaction, 18

L linear transport operator, 20

$\boldsymbol{l}$ logarithms of the mobile concentrations, vector length: I, 43

$\boldsymbol{l}_{nmin}$ logarithms of the nonmineral concentrations, vector length: I_{nmin}, 43

$\boldsymbol{\Lambda}$ diagonal matrix with reciprocal values of mobile concentrations, 29

$\bar{\boldsymbol{\Lambda}}_{nmin}$ diagonal matrix with reciprocal values of nonmineral concentrations, 29

$\tilde{\boldsymbol{\Lambda}}$ diagonal matrix with reciprocal values of all concentrations except minerals, 39

$\boldsymbol{\nu}$ outer normal, 33

Ω computational domain, 33

ϕ_j j-th equilibrium condition, 18

$\boldsymbol{\phi}$ vector of all equilibrium conditions, length: J_{eq}, 20

$\boldsymbol{\phi}_{min}$ equilibrium conditions of mineral reactions, number: J_{min}, 25

$\boldsymbol{\phi}_{mob}$ equilibrium conditions of reactions with only mobile species, number: J_{mob}, 25

$\boldsymbol{\phi}_{sorp}$ equilibrium conditions of sorption reactions, number: J_{sorp}, 25

ψ_j equilibrium condition of the j-th mineral reaction in the case mineral is present, 18

$\boldsymbol{q}$ Darcy flow, 16

$\boldsymbol{r}$ reaction rate vector, length: J, 17

$\boldsymbol{r}_{eq}$ reaction rates of equilibrium reactions, number: J_{eq}, 18

$\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})$ reaction rates of kinetic reactions, number: J_{kin}, 18

$\boldsymbol{S}$ stoichiometric matrix, size: $(I + \bar{I}) \times J$, 17

$\boldsymbol{S}_1$ $I \times J$-submatrix of $\boldsymbol{S}$ with entries of mobile species, 19

$\boldsymbol{S}_{1,eq}$ $I \times J_{eq}$-submatrix of $\boldsymbol{S}_{eq}$ with entries of mobile species, 19

$\boldsymbol{S}_{1,kin}$ $I \times J_{kin}$-submatrix of $\boldsymbol{S}_{kin}$ with entries of mobile species, 19

$\boldsymbol{S}^*_{1,kin}$ $I \times J^*_{1,kin}$-submatrix of $\boldsymbol{S}^*_1$ with entries of kinetic reactions, 22

$\boldsymbol{S}_{1,min}$ $I \times J_{min}$-submatrix of $\boldsymbol{S}_{1,eq}$ with entries of mineral reactions, 21

$\boldsymbol{S}_{1,min,\mathcal{A}}$ submatrix of $\boldsymbol{S}_{1,min}$ with active reactions (mineral not present), 29

$\boldsymbol{S}_{1,min,\mathcal{I}}$ submatrix of $\boldsymbol{S}_{1,min}$ with inactive reactions (mineral present), 29

$\boldsymbol{S}_{1,mob}$ $I \times J_{mob}$-submatrix of $\boldsymbol{S}_{1,eq}$ containing entries of reactions with only mobile species, 21

$\boldsymbol{S}^{\perp}_1$ matrix with max. system orthogonal to $\boldsymbol{S}^*_1$, size: $I \times (I - J_{mob} - J_{sorp,li} - J_{min} - J^*_{1,kin})$, 23

$\boldsymbol{S}^*_1$ matrix with max. system of linear independent columns of $\boldsymbol{S}_1$, size: $I \times (J_{mob} + J_{sorp,li} + J_{min} + J^*_{1,kin})$, 22

$\boldsymbol{S}_{1,sorp}$ $I \times J_{sorp}$-submatrix of $\boldsymbol{S}_{1,eq}$ with entries of sorption reactions, 21

$\boldsymbol{S}_{1,sorp,li}$ $I \times J_{sorp,li}$-submatrix of $\boldsymbol{S}_{1,sorp}$ s.t. $(\boldsymbol{S}_{1,mob}, \boldsymbol{S}_{1,sorp,li}, \boldsymbol{S}_{1,min})$ is linear independent, 21

$\boldsymbol{S}_2$ $\bar{I} \times J$-submatrix of $\boldsymbol{S}$ with entries of immobile species, 19

$\boldsymbol{S}_{2,eq}$ $\bar{I} \times J_{eq}$-submatrix of $\boldsymbol{S}_{eq}$ with entries of immobile species, 19

$\boldsymbol{S}_{2,kin}$ $\bar{I} \times J_{kin}$-submatrix of $\boldsymbol{S}_{kin}$ with entries of immobile species, 19

$\boldsymbol{S}^*_{2,kin}$ $\bar{I}_{nmin} \times J^*_{2,kin}$-submatrix of $\boldsymbol{S}^*_2$ with entries of nonminerals and kinetic reactions, 22

$\tilde{\boldsymbol{S}}_{2,kin}$ $\bar{I}_{nmin} \times J_{kin}$-submatrix of $\boldsymbol{S}_{2,kin}$ with entries of nonminerals, 22

$\boldsymbol{S}^{\perp}_2$ matrix with max. system orthogonal to $\boldsymbol{S}^*_2$, size: $\bar{I} \times (\bar{I} - J_{sorp} - J_{min} - J^*_{2,kin})$, 23

$\boldsymbol{S}^*_2$ matrix with max. system of linear independent columns of $\boldsymbol{S}_2$, size: $\bar{I} \times (J_{sorp} + J_{min} + J^*_{2,kin})$, 22

$\boldsymbol{S}_{2,sorp}$ $\bar{I}_{nmin} \times J_{sorp}$-submatrix of $\boldsymbol{S}_{2,eq}$ with entries of nonminerals and sorption reactions, 21

$\boldsymbol{S}_{eq}$ $(I + \bar{I}) \times J_{eq}$-submatrix of $\boldsymbol{S}$ with entries of equilibrium reactions, 19

$\boldsymbol{S}_{kin}$ $(I + \bar{I}) \times J_{kin}$-submatrix of $\boldsymbol{S}$ with entries of kinetic reactions, 19

$\tilde{\boldsymbol{S}}^{\perp T}_1$ transformation matrix of generalized reduction scheme, 94

T end time, 33

Δt time step size, 35

$\boldsymbol{T}$ total concentrations, number of variables: $I - J_{mob}$, 86

$\boldsymbol{T}_F$ fixed part of total concentrations, number of var.: $I - J_{mob}$, 86

$\boldsymbol{T}_M$ mobile part of total concentrations, number of var.: $I - J_{mob}$, 86

$\mathcal{T}_h$ triangulation, 34

θ water content, 16

$\boldsymbol{W}$ total fixed concentrations, number of variables: $\bar{I}_{nmin} - J_{sorp}$, 86

$\boldsymbol{\xi}$ transformed variables, number of var.: $J_{mob} + J_{sorp,li} + J_{min} + J^*_{1,kin}$, 24

$\bar{\boldsymbol{\xi}}$ transformed variables, number of var.: $J_{sorp} + J_{min} + J^*_{2,kin}$, 24

$\bar{\boldsymbol{\xi}}_{kin}$ transformed variables, number of variables: $J^*_{2,kin}$, 25

$\bar{\boldsymbol{\xi}}_{min}$ transformed variables, number of variables: J_{min}, 25

$\bar{\boldsymbol{\xi}}_{sorp}$ transformed variables, number of variables: J_{sorp}, 25

$\bar{\boldsymbol{\xi}}_{sorp,ld}$ transformed variables, number of variables: $J_{sorp,ld}$, 26

$\bar{\boldsymbol{\xi}}_{sorp,li}$ transformed variables, number of variables: $J_{sorp,li}$, 26

$\boldsymbol{\xi}_{glob}$ global variables, number of var.: $2J_{sorp,li} + 2J_{min} + J^*_{1,kin}$, 28

$\boldsymbol{\xi}_{kin}$ transformed variables, number of variables: $J^*_{1,kin}$, 25

$\boldsymbol{\xi}_{loc}$ local variables, number of variables: $J_{mob} + J_{sorp} + J_{min} + J^*_{2,kin}$, 28

$\boldsymbol{\xi}_{min}$ transformed variables, number of variables: J_{min}, 25

$\boldsymbol{\xi}_{mob}$ transformed variables, number of variables: J_{mob}, 25

$\boldsymbol{\xi}_{sorp}$ transformed variables, number of variables: $J_{sorp,li}$, 25

$\tilde{\boldsymbol{\xi}}$ transformed variables, number of variables: $J_{sorp,li} + J_{min}$, 28

$\tilde{\boldsymbol{\xi}}_{min}$ transformed variables, number of variables: J_{min}, 28

$\tilde{\boldsymbol{\xi}}_{sorp}$ transformed variables, number of variables: $J_{sorp,li}$, 28

Summary

The work in hand deals with the efficient numerical solving of multi-species reactive transport problems in porous media. The goal was that in doing so a existing reduction scheme is enhanced such that it is applicable to realistic, numerical challenging scenarios and to participate with it successfully in international benchmark computations. In the process the gain in CPU time should be carried out without any loss of accuracy.

The essential step of the advancement of the reduction scheme is the introduction of additional variables that are reaction invariant regarding equilibrium reactions. These additional variables are mandatory to get a well conditioned global problem. In addition a new solver for the local problem and a starting value search to avoid negative concentrations were developed. Likewise the cutting-off of the Newton iterate conduces to the avoiding of negative concentrations. For convection dominated problems a Finite Volume stabilization was integrated. To handle anisotropic dispersion tensors the mesh is adapted to the tensor because the standard method leads to negative concentrations.

All nonlinear systems of equations are solved with Newton's method. The space discretization is carried out with conformal Finite Elements using mass lumping, the time discretization with the implicit Euler method using adaptive time stepping, where the time step size depends on the number of Newton steps in the last time step. The global linear system is solved by an iterative solver (e.g. BiCGStab, QMRCGStab) with SSOR preconditioner.

It was shown that there is a connection between the reduction scheme and the widely used Morel formulation. That way the global problem is a kind of transport problem and the local problem a kind of chemical problem. Furthermore a generalized formulation of the reduction scheme was developed, such that the normal formulation of the reduction scheme and the Morel formulation are special cases of the generalized formulation. Hence comparative computations with the same source code are possible. Also a method between the reduction scheme and the Morel formulation, where only a part of the equations decouple, is possible.

Nine test cases of the numerically challenging MoMaS–benchmark were computed successfully with the new reduction scheme. In the 2D "easy advective test case" of the MoMaS–benchmark the reduction scheme is more than five times faster than the software of the second fastest group, which operates with iterative splitting. Comparative computations done with the help of the generalized formulation provide the same speed advantage for the reduction scheme, in doing so a weaker stopping criteria had to be used for the iterative splitting. Also with the help of the generalized formulation it was shown that the "global ODE approach", which was used by no one of the participants of the benchmark, requires double CPU time compared to the reduction scheme. Thus this method is the fasted one after the reduction scheme.

A suggestion for an additional test case was given, in which the transversal dispersion is crucial for the results and hence the numerical diffusion of the used method has a large influence on the numerical solution. For two methods with different numerical diffusion numerical results are given, which are clearly visibly different as it is expected.

For the kinetic mineral problem the different formulations (set-valued rate function, complementarity condition, discontinuous rate function) were compared. It turned out that for weak solutions all three formulations are equivalent. Afterwards the existence of a global solution was proven with the help of a regularization of the set-valued rate function and application of the fix point theorem of Schaefer.

Deutscher Titel und Zusammenfassung

Reaktiver Transport und Minerallösung/-fällung in porösen Medien: Effiziente Lösungsalgorithmen, Benchmarkrechnungen und Existenz einer globalen Lösung

Die vorliegende Arbeit beschäftigt sich mit dem effizienten numerischen Lösen von reaktiven Mehrkomponenten-Transportproblemen in porösen Medien. Ziel war es dabei, ein vorhandenes Reduktionsverfahren so weiterzuentwickeln, dass es auf realistische, numerisch herausfordernde Szenarien anwendbar ist und damit an einer internationalen Benchmark-Rechung erfolgreich teilzunehmen. Dabei sollte die Reduzierung der Rechenzeit ohne Einbußen bei der Genauigkeit erfolgen.

Der wesentliche Schritt bei der Weiterentwicklung des Reduktionsverfahrens ist die Einführung von zusätzlichen Variablen, die reaktionsinvariant bezüglich Gleichgewichtsreaktionen sind. Diese zusätzlichen Variablen sind zwingend notwendig, um ein gut konditioniertes globales Problem zu erhalten. Zusätzlich wurden ein neuer Löser für das lokale Problem und eine Startwertsuche zur Vermeidung negativer Konzentrationen entwickelt. Ebenfalls zur Vermeidung negativer Konzentrationen dient das Abschneiden der Newton–Iterierten. Für konvektionsdominate Probleme wurde eine Finite Volumen Stabilisierung eingebaut. Zur Handhabung von anisotropen Dispersionstensoren wird das Gitter an den Tensor angepasst, da die Standardmethode zu negativen Konzentrationen führt.

Zum Lösen aller nichtlinearen Gleichungssysteme wird das Newton–Verfahren eingesetzt. Die Ortsdiskretisierung erfolgt durch konforme Finite Elemente mit Mass Lumping, die Zeitdiskretisierung durch das implizite Euler Verfahren mit adaptiver Zeitschrittweite, wobei die Zeitschrittweite von der Anzahl der Newton–Schritte im letzten Zeitschritt abhängig ist. Zur Lösung des globalen linearen Gleichungssystems wird ein iterativer Löser (z.B. BiCGStab, QMRCGStab) mit

SSOR Vorkonditionierer verwendet.

Es wurde gezeigt, dass eine Beziehung zwischen dem Reduktionsverfahren und der meist verwendeten Morel–Formulierung besteht. So ist das globale Problem eine Art Transportproblem und das lokale Problem eine Art chemisches Problem. Weiter wurde eine verallgemeinerte Formulierung des Reduktionsverfahrens entwickelt, so dass das ursprüngliche Reduktionsverfahren und die Morel–Formulierung Spezialfälle der verallgemeinerten Formulierung sind. Somit sind Vergleichsrechnungen mit dem selben Programmcode möglich. Auch ein Methode zwischen dem Reduktionsverfahren und der Morel–Formulierung, bei der nur ein Teil der Gleichungen entkoppelt ist möglich.

Mit dem neuen Reduktionsverfahren wurden erfolgreich 9 Testfälle des numerisch herausfordernden MoMaS–Benchmarks gerechnet. Beim 2D “easy advective test case” des MoMaS–Benchmarks ist das Reduktionsverfahren mehr als fünfmal schneller als die Software der nächstplazierten Gruppe, die mit iterativen Splitting arbeitet. Vergleichsrechnungen mit Hilfe der verallgemeinerten Formulierung liefern den selben Geschwindigkeitsvorteil für das Reduktionsverfahren, wobei beim iterativen Splitting ein schwächeres Abbruchkriterium verwendet werden musste. Ebenfalls durch Vergleichsrechnungen mit Hilfe der verallgemeinerten Formulierung wurde gezeigt, dass der “global ODE approach”, der von keinem Teilnehmer des Benchmarks verwendet wurde, die doppelte Rechenzeit wie das Reduktionsverfahren benötigt. Somit ist dieser Ansatz die schnellste Methode nach dem Reduktionsverfahren.

Es wurde ein Vorschlag für einen zusätzlichen Testfall gegeben, bei dem die Querdispersion für das Ergebnis entscheidend ist und somit die numerische Diffusion des verwendeten Verfahrens einen großen Einfluss auf die numerische Lösung hat. Für zwei unterschiedlich diffusive Verfahren werden numerische Ergebnisse angegeben, die sich wie erwartet deutlich sichtbar unterscheiden.

Für das kinetisches Mineralproblem wurden die unterschiedlichen Formulierungen (mengenwertige Ratenfunktion, Komplementaritätsbedingung, unstetige Ratenfunktion) verglichen. Es stellte sich heraus, dass für schwache Lösungen alle drei Formulierungen äquivalent sind. Anschließend wurde die Existenz einer globalen Lösung mit Hilfe einer Regularisierung der mengenwertigen Ratenfunktion und durch Anwendung des Fixpunktsatzes von Schaefer bewiesen.

Bibliography

[AK09] L. Amir and M. Kern. Global method for coupling transport with chemistry in heterogeneous porous media. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9162-x.

[BBC$^+$] A. Bourgeat, S. Bryant, J. Carrayrou, A. Dimier, C.J. Van Duijn, M. Kern, P. Knabner, and N. Leterrier. GDR MoMaS benchmark reactive transport. http://www.gdrmomas.org/Ex_qualif/Geochimie/Documents/Benchmark-MoMAS.pdf.

[BEHM07] Nicolas Bouillard, Robert Eymard, Raphaele Herbin, and Philippe Montarnal. Diffusion with dissolution and precipitation in a porous medium: Mathematical analysis and numerical approximation of a simplified model. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(6):975–1000, 2007.

[Bet96] C. M. Bethke. *Geochemical Reaction Modeling.* Oxford University Press, New York, 1996.

[BK04] M. Bause and P. Knabner. Numerical simulation of contaminant biodegradation by higher order methods and adaptive time stepping. *Computing and Visualization in Science*, 39:61–78, 2004.

[BMCB97] D. Barry, C. Miller, P. Culligan, and K. Bajracharya. Analysis of split operator methods for nonlinear and multispecies groundwater chemical transport models. *Mathematics and Computers in Simulation*, 43:331–341, 1997.

[Car01] J. Carrayrou. *Modelisation du Transport de Solutes Reactifs en Milieu Poreux Sature.* PhD thesis, Universite Louis Pasteur, Straßburg, Frankreich, 2001.

[Car09] J. Carrayrou. Looking for some reference solutions for the reactive transport benchmark of MoMaS with SPECY. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9161-y.

[CHK$^+$10] J. Carrayrou, J. Hoffmann, P. Knabner, S. Kräutle, C. de Dieuleveult, J. Erhel, J. Van der Lee, V. Lagneau, K.U. Mayer, and K.T.B. McQuarrie. Comparison of numerical methods for simulating strongly non-linear and heterogeneous reactive transport problems — the MoMaS benchmark case. *Computational Geosciences*, 2010. published online, doi: 10.1007/s10596-010-9178-2.

[CJRT01] G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM Journal on Numerical Analysis*, 38(6):2047–2078, 2001.

[CKK09] J. Carrayrou, M. Kern, and P. Knabner. Reactive transport benchmark of MoMaS. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9157-7.

[CMB02] J. Carrayrou, R. Mosé, and P. Behra. New efficient algorithm for solving thermodynamic chemistry. *AIChE Journal*, 48(4):894–904, 2002.

[CMB04] J. Carrayrou, R. Mosé, and P. Behra. Operator-splitting procedures for reactive transport and comparison of mass balance errors. *Journal of Contaminant Hydrology*, 68:239–268, 2004.

[dD08] C. de Dieuleveult. *Un modèle numérique global et performant pour le couplage géochimie-transport*. PhD thesis, Université de Rennes 1, 2008.

[dDE09] C. de Dieuleveult and J. Erhel. A global approach for reactive transport: application to the benchmark easy test case of MoMaS. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9163-9.

[dDEK09] C. de Dieuleveult, J. Erhel, and M. Kern. A global strategy for solving reactive transport equations. *Journal of Computational Physics*, 228:6395–6410, 2009.

[DPvDC08] V. M. Devigne, I. S. Pop, C. J. van Duijn, and T. Clopeau. A numerical scheme for the pore scale simulation of crystal dissolution and precipitation in porous media. *SIAM Journal on Numerical Analysis*, 46(2):895–919, 2008.

[Eva98] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, 1998.

[FR92] J. C. Friedly and J. Rubin. Solute transport with multiple equilibrium-controlled or kinetically controlled reactions. *Water Resources Research*, 28(6):1935–1953, 1992.

[Fri64] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.

[Fri91] J. C. Friedly. Extent of reaction in open systems with multiple heterogeneous reactions. *AIChE Journal*, 37(5):687–693, 1991.

[Heu89] H. Heuser. *Gewöhnliche Differentialgleichungen*. Teubner, Stuttgart, 1989.

[HKK09] J. Hoffmann, S. Kräutle, and P. Knabner. A parallel global-implicit 2-d solver for reactive transport problems in porous media based on a reduction scheme and its application to the MoMaS benchmark problem. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9173-7.

[Hof05] J. Hoffmann. Ein Entkopplungsverfahren für Systeme von Transportreaktionsgleichungen in porösen Medien: Algorithmische Realisierung und Simulation realistischer 2D-Szenarien. Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Angewandte Mathematik I, Deutschland, 2005.

[KA03] P. Knabner and L. Angermann. *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*, volume 44 of *Texts in Applied Mathematics*. Springer, New York, 2003.

[Kan04] C. Kanzow. Inexakt semismooth newton methods for large-scale complementarity problems. *Optimization Methods and Software*, 19(3-4):309–325, 2004.

[KK05] S. Kräutle and P. Knabner. A new numerical reduction scheme for fully coupled multicomponent transport-reaction problems in porous media. *Water Resources Research*, 41:W09414, 2005. doi: 10.1029/2004WR003624.

[KK07] S. Kräutle and P. Knabner. A new numerical reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions. *Water Resources Research*, 43:W03429, 2007. doi: 10.1029/2005WR004465.

[Kna86] P. Knabner. A free boundary problem arising from the leaching of saine soils. *SIAM Journal on Mathematical Analysis*, 17(3):610–625, 1986.

[Kna02] P. Knabner. Numerische Mathematik I. http://www.am.uni-erlangen.de/am1/de/scripts/knabner/num1_2002.ps, 2002. Vorlesungsskript.

[Koh05] C. Kohlhepp. Gemischte Finite-Elemente-Methoden für elliptische und parabolische Differentialgleichungen mit Lösungen geringer Regularität: Konvergenzordnung und parallele Implementierung. Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Angewandte Mathematik I, Deutschland, 2005.

[Krä08] S. Kräutle. *General Multi-Species Reactive Transport Problems in Porous Media: Efficient Numerical Approaches and Existence of Global Solutions.* Habilitation thesis, University Erlangen–Nuremberg, 2008. http://www.am.uni-erlangen.de/am1/en/theses.html.

[KvDH95] P. Knabner, C.J. van Duijn, and S. Hengst. An analysis of crystal dissolutions fronts of flows through porous media. part 1: Compatible boundary conditions. *Advances in Water Resources*, 18:171–185, 1995.

[Lic85] P. C. Lichtner. Continuum model for simultaneous chemical reactions and mass transport in hydrothermal systems. *Geochimica et Cosmochimica Acta*, 49(3):779–800, 1985.

[Lic96] P. C. Lichtner. *Reactive Transport in Porous Media (P. C. Lichtner and C. I. Steefel and E. H. Oelkers, eds.)*, chapter Continuum formulation of multicomponent-multiphase reactive transport, pages 1–81. Reviews in Mineralogy 34. Mineralogical Society of America, Washington, 1996.

[LSU68] O.A. Ladyzenskaja, V.A. Solonnikov, and N.N. Uralceva. *Linear and Quasi-linear Equations of Parabolic Type*, volume 23 of *Translations of Mathematical Monographs*. American Mathematical Society, 1968.

[LvdL09] V. Lagneau and J. van der Lee. HYTEC results of the MoMaS reactive transport benchmark. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9159-5.

[May99] K. U. Mayer. *A numerical model for multicomponent reactive transport in variably saturated porous media*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1999.

[MCAS04] S. Molins, J. Carrera, C. Ayora, and M. W. Saaltink. A formulation for decoupling components in reactive transport problems. *Water Resources Research*, 40(10):W10301, 2004. doi: 10.1029/2003WR002970.

[MFB02] K. U. Mayer, E. O. Frind, and D. W. Blowes. Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions. *Water Resources Research*, 38(9):1174, 2002. doi: 10.1029/2001WR000862.

[MM09] K.U. Mayer and K.T.B. MacQuarrie. Formulation of the multicomponent reactive transport code MIN3P and implementation of MoMaS benchmark problems. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9158-6.

[PHKK06] A. Prechtel, J. Hoffmann, S. Kräutle, and P. Knabner. Reaktive Mehrkomponentenprobleme: Sicherung von Effizienz und Zuverlässigkeit. In *Modellierung und Prognose von Natural Attenuation-Prozessen im Untergrund. Statusseminar des KORA–TV 7, 08.06.2006, Gemeinsame Mitteilungen des Dresdner Grundwasserforschungszentrums e.V.*, pages 75–90. Eigenverlag des Hrsg., Dresden, 2006.

[Saa96] F. Saaf. *A Study of Reactive Transport Phenomena in Porous Media*. PhD thesis, Rice University, Houston, USA, 1996.

[SAC98] M. W. Saaltink, C. Ayora, and J. Carrera. A mathematical formulation for reactive transport that eliminates mineral concentrations. *Water Resources Research*, 34(7):1649–1656, 1998.

[SCA00] M. W. Saaltink, J. Carrera, and C. Ayora. A comparison of two approaches for reactive transport modelling. *Journal of Geochemical Exploration*, 69-70:97–101, 2000.

[Sch61] A.E. Scheidegger. General theory of dispersion in porous media. *Journal of Geophysical Research*, 66(10):3273–3278, 1961.

[SM96a] C. I. Steefel and K. T. B. MacQuarrie. *Reactive Transport in Porous Media (P. C. Lichtner and C. I. Steefel and E. H. Oelkers, eds.)*, chapter Approaches to modeling of reactive transport in porous media, pages 83–129. Reviews in Mineralogy 34. Mineralogical Society of America, Washington, 1996.

[SM96b] W. Stumm and J. J. Morgan. *Aquatic Chemistry.* John Wiley & Sons, New York, 1996.

[Sol64] V.A. Solonnikov. A priori estimates for second-order parabolic equations. *Trudy Mat. Inst. Steklov*, 70:133–212, 1964. *Amer. Math. Soc. Transl. Ser. 2*, 65:51-138, 1967.

[vDK97] C.J. van Duijn and P. Knabner. Travelling wave behaviour of crystal dissolution in porous media flow. *European Journal on Applied Mathematics*, 8:455 ff., 1997.

[vDKS98] C.J. van Duijn, P. Knabner, and R.J. Schotting. An analysis of crystal dissolution fronts in flows through porous media part 2: Incompatible boundary conditions. *Advances in Water Resources*, 22(1):1–16, 1998.

[vdLWLG03] J. van der Lee, L. De Windt, V. Lagneau, and P. Goblet. Modul-oriented modeling of reactive transport with HYTEC. *Computers & Geosciences*, 29:265–275, 2003.

[vDP04] C.J. van Duijn and I.S. Pop. Crystal dissolution and precipitation in porous media: Pore scale analysis. *J. Reine Angew. Math.*, 577:171–211, 2004.

[VM92] A. Valocchi and M. Malmstead. Accuracy of operator splitting for advection-dispersion-reaction problems. *Water Resources Research*, 28(5):1471–1476, 1992.

[Wie05] C. Wieners. Distributed Point Objects. A New Concept for Parallel Finite Elements. In *Domain decomposition methods in science and* *engineering, Lecture notes in computational science and engineering*, volume 40, pages 175–183. R. Kornhuber, R. Hoppe, J. Priaux, O. Pironneau, O. Widlund, J. Xu (editors), Springer, 2005.

[WYW06] Zhuoqun Wu, Jingxue Yin, and Chunpeng Wang. *Elliptic & Prabolic Equations.* World Scientific, Singapore, 2006.

[YT89] 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(1):93–108, 1989.

[Zie97] H. Zieschang. *Lineare Algebra und Geometrie.* Teubner, Stuttgart, 1997.

Curriculum vitae

Persönliche Daten:

Name:	Hoffmann
Vorname:	Joachim
Geburtsdatum:	22. November 1979
Geburtsort:	Erlangen, Deutschland
Anschrift:	Sachsenstraße 15 B 91074 Herzogenaurach Deutschland
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

Ausbildungsdaten:

Schulausbildung:	9/1986 – 7/1990	Grundschule Niederndorf
	9/1990 – 6/1999	Gymnasium Herzogenaurach
	6/1999	Abitur
Zivildienst:	8/1999 – 6/2000	Liebfrauenhaus Altenheim, Herzogenaurach
Studium:	9/2000 – 1/2006	Studium in Technomathematik an der FAU Erlangen–Nürnberg
	1/2006	Diplom, Thema der Diplomarbeit (Prof. Dr. P. Knabner): Transportreaktionsgleichungen in porösen Medien: Algorithmische Realisierung und Simulation realistischer 2D–Szenarien
Promotionsstudium:	seit 2/2006	Promotionsstudium in Mathematik am Department Mathematik der FAU Erlangen–Nürnberg

Berufliche Tätigkeiten:

Doktorand:	Seit Feb. 2006	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Angewandte Mathematik I der FAU Erlangen–Nürnberg

Beschäftigung in folgenden Drittmittelprojekten:

- BMBF–Förderschwerpunkt “Kontrollierter natürlicher Rückhalt und Abbau von Schadstoffen bei der Sanierung kontaminierter Böden und Grundwässer” KORA

- DFG–Projekt “Effiziente numerische Lösung von großen Komplementaritätsproblemen gekoppelt an PDEs bei reaktivem Mehrkomponenten–Transport mit Mineralien in porösen Medien”

Lehrveranstaltungen:

- Betreuung der Übung zu “Numerik I für Ingenieure”

- Betreuung der Übung (einschließlich MATLAB–Programmieraufgaben) zur Vorlesung “Numerics of Partial Differential Equations”

- Betreuung des Programmierpraktikums “Problemlösung mit Finiten Elementen”

Veröffentlichungen:

M. Bause and J. Hoffmann. Numerical study of mixed finite element and multi point flux approximation of flow in porous media. In *Proceedings of ENUMATH 2007, the 7th European Conference on Numerical Mathematics and Advanced Applications*, Graz, Austria, 2007.

M. Bause, J. Hoffmann, and P. Knabner. First-order convergence of multi-point flux approximation on triangular grids and comparison with mixed finite element methods. *Numerische Mathematik*, 2010. accepted.

J. Carrayrou, J. Hoffmann, P. Knabner, S. Kräutle, C. de Dieuleveult, J. Erhel, J. Van der Lee, V. Lagneau, K.U. Mayer, and K.T.B. McQuarrie. Comparison of numerical methods for simulating strongly non-linear and heterogeneous reactive transport problems — the MoMaS benchmark case. *Computational Geosciences*, 2010. published online, doi: 10.1007/s10596-010-9178-2.

Joachim Hoffmann, Serge Kräutle, and Peter Knabner. Computation of the MoMaS benchmark problem with a parallel global-implicit 2-d solver based on a reformulation of the PDE–ODE system. http://www-imfs.u-strasbg.fr/colloques/mrtpm2008/papers/Bench_Hoffamnn_et_al.pdf, 2008. International Workshop on Modelling Reactive Transport in Porous Media, Strasbourg.

J. Hoffmann, S. Kräutle, and P. Knabner. A parallel global-implicit 2-d solver for reactive transport problems in porous media based on a reduction scheme and its application to the MoMaS benchmark problem. *Computational Geosciences*, 2009. published online, doi: 10.1007/s10596-009-9173-7.

J. Hoffmann. Results of the GdR MoMaS reactive transport benchmark with RICHY2D. Preprint No. 326, Department of Mathematics, University of Erlangen–Nuremberg, Erlangen, Germany, 2008. ISSN 1435-5833:
http://www.am.uni-erlangen.de/papers/pr326.pdf.

P. Knabner, F. Frank, J. Hoffmann, S. Kräutle, St. Oßmann, and A. Prechtel. Entwicklung, Zuverlässigkeit und Effizienz reaktiver Mehrkomponententransportmodelle. In *Systemanalyse, Modellierung und Prognose von Natural Attenuation-Prozessen um Untergrund, Synopse des KORA–Themenverbund 7*, pages 197–234. DGFZ, Dresden, 2008.

S. Kräutle, P. Knabner, and J. Hoffmann. Efficient and accurate simulation of large general reactive multicomponent transport processes in porous media by model-preserving a priori decoupling techniques. In *Proceedings of the International Conference on Computational Methods in Water Resources (CMWR) XVI*, Copenhagen, 2006.

A. Prechtel, J. Hoffmann, S. Kräutle, and P. Knabner. Reaktive Mehrkomponentenprobleme: Sicherung von Effizienz und Zuverlässigkeit. In *Modellierung und Prognose von Natural Attenuation-Prozessen im Untergrund. Statusseminar des KORA–TV 7, 8.6.2006, Gemeinsame Mitteilungen des Dresdner Grundwasserforschungszentrums e.V.*, pages 75–90. Eigenverlag des Hrsg., Dresden, 2006.