FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
INSTITUT FÜR ANGEWANDTE MATHEMATIK

A general reduction scheme for reactive transport in porous media

by

J. Hoffmann, S. Kräutle & P. Knabner

No. 353 2012

Institut für Angewandte Mathematik
Martensstraße 3
D-91058 Erlangen

Lehrstuhl AM I:
Tel.: ++49/9131/85-27015
Fax: ++49/9131/85-27670

Lehrstuhl AM II:
Tel.: ++49/9131/85-27510
Fax: ++49/9131/85-28126

Lehrstuhl AM III:
Tel.: ++49/9131/85-25226
Fax: ++49/9131/85-25228

internet: `http://www.am.uni-erlangen.de`

ISSN 1435-5833, 2012, No. 353

A general reduction scheme for reactive transport in porous media

Joachim Hoffmann, Serge Kräutle, Peter Knabner[1]

Abstract We present a method to transform the governing equations of multi-species reactive transport in porous media. The reformulation leads to a smaller problem size by decoupling of equations and by elimination of unknowns, which increases the efficiency of numerical simulations. The reformulation presented here is a generalization of earlier works. In fact, a whole class of transformations is now presented. This class is parametrized by the choice of certain transformation matrices. For specific choices, some known formulations of reactive transport can be retrieved. Hence, the software based on the presented transformation can be used to obtain efficiency comparisons of different solution approaches. For our efficiency tests we use the MoMaS Benchmark Problem on Reactive Transport.

Keywords reactive transport, porous media, reduction of problem size, numerical simulation, global implicit approach

MSC(2010) 35K57, 35K58, 65Y20

1 Introduction

This work is devoted to efficient numerical simulations of large-scale reactive transport problems in porous media. We present a reformulation of the given system of partial differential equations (PDEs), ordinary differential equations (ODEs), algebraic equilibrium equations (AEs) and inequalities that govern multicomponent reactive transport problems. The reformulation uses linear transformations of the equations and variables. The size of the govering system of equations is reduced by decoupling of some linear PDEs and by implicit elimination of local equations (i.e., AEs, ODEs) and unknowns. The work presented here is based on earlier versions of the reduction scheme which are proposed in [13, 14, 11].

Let us emphasize that our method is a global implicit approach (GIA) and does neither use operator splitting nor simplifications of the model. Operator splitting schemes usually decompose a time step into a transport problem and a reaction problem. There are two main types, the sequential non-iterative approach (SNIA) and the sequential iterative approach (SIA).

[1]Email: kraeutle/knabner@am.uni-erlangen.de
University of Erlangen-Nuremberg, Department of Mathematics, Cauerstr. 11, 91058 Erlangen, Germany
Submitted: Feb 21, 2012.

For example the software SPECY uses a non-iterative operator splitting scheme (see [3]). The drawback of SNIA is the introduction of splitting errors (see e.g. [6]). To circumvent this problem, iterative operator splitting schemes can be used. For example the software HYTEC uses an iterative operator splitting scheme (see [22]). However, iterative operator splitting schemes may need many iteration steps and small time steps to converge in chemically difficult cases which can impact the effectiency (see e.g. [20]). An advantage of splitting schemes seems that the implementation might be easier compared to a reformulation scheme. One GIA method, which is known for several years, is the direct substitutional approach (DSA). For example the code MIN3P (see [17]) uses this method. One disadvantage of DSA is that it leads to a nonlinear system which might be difficult to solve numerically. A recently proposed global implicit approach is to use a differential algebraic equation (DAE) solver (see [7, 8]). 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 seems to lead to larger computation times (see [4]).

Besides the presentation of the reformulation technique, the other aim of this paper is to give a efficiency comparison to other methods, based on a demanding benchmark problem [5], which is characterized by strongly nonlinear reactions, variation of concentrations and scales of many magnitudes, and two space dimensions.

Compared to earlier presentations [14, 9, 10], the reformulation presented here is more general. In fact, a whole class of transformations is described. In this class of transformations, not only [14, 9, 10], but also the Morel formulation [5, 18], which is a common standard for such problems, is found as a representative.

Hence, with one implementation comparisons between the reduction scheme and the Morel formulation are possible. That excludes differences in the CPU time due to the specific implementation and specific computer systems, etc. We hope that this contributes to unbiased comparisons of the different methods.

The paper is organized as follows. In Sec. 2 we present the model. It includes mobile species and immobile species such as sorbed species and minerals, and it includes both kinetic and equilibrium reactions. In Sec. 3 we develop the reformulation and size reduction technique. Sec. 4 is devoted to the question if the substitution of local equations into the remaining equations, which is proposed in Sec. 3, is always possible. Sec. 5 shows that both the formulation [14] and the Morel formulation [5] are special cases of the presented general scheme. In Sec. 6 we give numerical results for the

schemes for the MoMaS benchmark [5]. These results can be seen as an extension of the results presented in [4, 10].

2 Mathematical model

Our model of multicomponent reactive transport covers both kinetic and equilibrium reactions, and both mobile and immobile species. The details are as follows.

2.1 Chemical reactions

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. [2])

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

of the chemical reaction. The number of reactions of this type is denoted by J_{kin}. The number of mobile and immobile species is denoted by I, $\bar{I}$, respectively. In the following we will frequently mark immobile species by a bar ('$\bar{c}_i$').

Another kind of kinetic reactions are biodegradation reactions. They can be described with help of the Monod model. A presentation of the model and how to apply the reduction mechanism to that type of reactions can be found in [13, Sec. 5].

Equilibrium reactions according to law of mass action

If the j-th equilibrium reaction can be described by the law of mass action then the j-th equilibrium condition reads

$$\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}) := -\ln(K_j) + \sum_{i=1}^{I+\bar{I}} s_{ij} \ln(c_i) = 0. \tag{2}$$

In case of equilibrium reactions the corresponding reaction rate $r_{eq,j}$ is not known. So $r_{eq,j}$ gets an additional unknown and the equilibrium condition (2) is added to the system of PDEs and ODEs as an additional equation.

Equilibrium minerals

If the j-th equilibrium reaction is a mineral reaction we define

$$\psi_j(\boldsymbol{c}) := -\ln(K_j) + \sum_{i=1}^{I} s_{ij} \ln(c_i). \tag{3}$$

For a mineral reaction the equilibrium condition consisting of equations and inequalities reads (e.g. [12])

$$\Big(\psi_j(\boldsymbol{c}) = 0 \wedge \bar{c}_{min,j} \geq 0\Big) \vee \Big(\psi_j(\boldsymbol{c}) > 0 \wedge \bar{c}_{min,j} = 0\Big) \tag{4}$$

where $\bar{c}_{min,j}$ denotes the concentration of that mineral taking part in the j-th equilibrium reaction. The left part of (4) describes saturation, the right part describes undersaturation of the fluid with respect to mineral j.

The equilibrium condition (4) can be expressed by

$$\phi_j(\boldsymbol{c}, \bar{\boldsymbol{c}}_{min}) := \min\{\psi_j(\boldsymbol{c}),\, \bar{c}_{min,j}\} = 0\,. \tag{5}$$

The equations (2) and (5) are subsumed to $\boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) = \boldsymbol{0}$ and added to the system of PDEs and ODEs, see next section.

2.2 Reactive transport

In matrix notation the system of equations reads

$$\begin{aligned} \partial_t(\theta \boldsymbol{c}) + L\boldsymbol{c} &= \theta \boldsymbol{S}_{1,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \theta \boldsymbol{S}_{1,eq} \boldsymbol{r}_{eq} \\ \partial_t(\theta \bar{\boldsymbol{c}}) &= \theta \boldsymbol{S}_{2,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) + \theta \boldsymbol{S}_{2,eq} \boldsymbol{r}_{eq} \\ \boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) &= \boldsymbol{0} \end{aligned} \tag{6}$$

with the Darcy velocity $\boldsymbol{q}$, the linear transport operator $L_i u_i := -\nabla \cdot (\boldsymbol{D}_i \nabla u_i - \boldsymbol{q} u_i)$, $L\boldsymbol{u} = \mathrm{diag}(L_i u_i)$, and the stoichiometric matrix

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

Some assumptions on the shape of the blocks of this matrix are given in the next section.

The Scheidegger diffusion/dispersion tensor $\boldsymbol{D}_i$ (see [21]) is used to describe diffusion and dispersion

$$\boldsymbol{D}_i = -(\theta d_{diff,i} + \beta_t |\boldsymbol{q}|)\boldsymbol{I} + (\beta_l - \beta_t)\frac{\boldsymbol{q}\,\boldsymbol{q}^T}{|\boldsymbol{q}|}. \tag{7}$$

This system of equations is widely used for modelling reactive transport (see, e.g., [19, eq. (19),(1)]).

3 The general reduction mechanism

The reduction scheme presented here is an extension to the reduction scheme described in [13], [14], [11] and [10]. 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 usually small compared to the mechanic dispersion.

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

$$\begin{aligned} \boldsymbol{S}_{1,eq} &= \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}, \\ \boldsymbol{S}_{2,eq} &= \begin{pmatrix} \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{pmatrix} \end{aligned} \tag{8}$$

$\boldsymbol{I}_{J_{min}}$ denotes the identity matrix of the size J_{min}. The number of the reactions of each type is denoted by J_{mob}, J_{sorp} and J_{min}, respectively. Then we define the abbreviation $\boldsymbol{S}_{1,het} := \begin{pmatrix} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}$. Hence, we get

$$\boldsymbol{S}_{1,eq} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,het} \end{pmatrix} \tag{9}$$

Also the immobile species are arranged in a specific order. We take the nonminerals first and then the minerals $\bar{\boldsymbol{c}} = (\bar{\boldsymbol{c}}_{nmin}, \bar{\boldsymbol{c}}_{min})$. The number of nonminerals[2] is named $\bar{I}_{nmin}$.

We assume that both the columns of

$$\begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,min} \end{pmatrix} \text{ and of } \boldsymbol{S}_{2,sorp} \tag{10}$$

are linear independent. Furthermore we assume[3] that there is a subset of columns $\boldsymbol{S}^*_{1,sorp}$ of the columns of $\boldsymbol{S}_{1,sorp}$ and a subset of columns $\boldsymbol{S}^*_{1,kin}$ of the columns of $\boldsymbol{S}_{1,kin}$ in such a way that the columns of the matrix $(\boldsymbol{S}_{1,mob}\ \boldsymbol{S}^*_{1,sorp}\ \boldsymbol{S}_{1,min}\ \boldsymbol{S}^*_{1,kin})$ are a maximal system of linear independent columns of matrix $\boldsymbol{S}_1$ and that all the remaining columns of block $\boldsymbol{S}_{1,sorp}$ are linear combinations of the columns of $\boldsymbol{S}^*_{1,sorp}$, $\boldsymbol{S}_{1,min}$, $\boldsymbol{S}^*_{1,kin}$ (but not of $\boldsymbol{S}_{1,mob}$).

[2] In this work nonminerals always denote all *immobile* species that are not a mineral.

[3] This assumption is rather technical. It could be replaced by the simpler assumption that all columns of matrix $(\boldsymbol{S}_{1,mob}\ \boldsymbol{S}_{1,sorp}\ \boldsymbol{S}_{1,min})$ are linear independent, as it is done in [11, p.131]. Hence, the assumption after (10) is a relaxation of the assumption made in [11].

Now let us incorporate also the kinetic reactions. 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:

$$\begin{aligned} \boldsymbol{S}_1 &= \begin{pmatrix} \boldsymbol{S}_{1,eq} & \boldsymbol{S}_{1,kin} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,het} & \boldsymbol{S}_{1,kin} \end{pmatrix} \\ \boldsymbol{S}_2 &= \begin{pmatrix} \boldsymbol{S}_{2,eq} & \boldsymbol{S}_{2,kin} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} & \tilde{\boldsymbol{S}}_{2,kin} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix} \end{aligned} \tag{11}$$

Note that this implies that we assume that an equilibrium mineral does not participate in any kinetic reaction. So the stoichiometric coefficients in $\boldsymbol{S}_{2,kin}$ connected to minerals are zero, and it is possible to express $\boldsymbol{S}_{2,kin}$ by $\tilde{\boldsymbol{S}}_{2,kin}$ and a zero block.

Recall that our system (6) reads

$$\partial_t(\theta \boldsymbol{c}) + L\boldsymbol{c} = \theta \boldsymbol{S}_1 \begin{pmatrix} \boldsymbol{r}_{eq} \\ \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \end{pmatrix} \tag{12}$$

$$\partial_t(\theta \bar{\boldsymbol{c}}) = \theta \boldsymbol{S}_2 \begin{pmatrix} \boldsymbol{r}_{eq} \\ \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \end{pmatrix} \tag{13}$$

$$\boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) = \boldsymbol{0} \tag{14}$$

with the block structure (11) and equilibrium conditions (14) from (2), (5).

In order to prepare the reformulation, we choose 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 (10)) it is always possible to choose $\boldsymbol{S}_1^*$ and $\boldsymbol{S}_2^*$ such that the matrices have the form

$$\begin{aligned} \boldsymbol{S}_1^* &= \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,het}^* & \boldsymbol{S}_{1,kin}^* \end{pmatrix} \in \mathbb{R}^{I \times (J_{mob} + J_{1,het}^* + J_{1,kin}^*)} \\ \boldsymbol{S}_2^* &= \begin{pmatrix} \boldsymbol{S}_{2,sorp} & \boldsymbol{0} & \boldsymbol{S}_{2,kin}^* \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{\bar{I} \times (J_{sorp} + J_{min} + J_{2,kin}^*)} \end{aligned} \tag{15}$$

where $\boldsymbol{S}_{1,het}^*$, $\boldsymbol{S}_{1,kin}^*$ and $\boldsymbol{S}_{2,kin}^*$ consist of some of the columns of $\boldsymbol{S}_{1,het}$, $\boldsymbol{S}_{1,kin}$ and $\tilde{\boldsymbol{S}}_{2,kin}$, respectively. The number of columns of $\boldsymbol{S}_{1,het}^*$, $\boldsymbol{S}_{1,kin}^*$ and $\boldsymbol{S}_{2,kin}^*$ is denoted $J_{1,het}^*$, $J_{1,kin}^*$ and $J_{2,kin}^*$, respectively.

Since matrix $\boldsymbol{S}_i^*$ consists of linear independent columns, there is always a matrix $\boldsymbol{A}_i$ such that

$$\boldsymbol{S}_i = \boldsymbol{S}_i^* \boldsymbol{A}_i \qquad i = 1, 2\,. \tag{16}$$

With the block structure from (11) for $\boldsymbol{S}_1$, $\boldsymbol{S}_2$ and the block structure from (15) for $\boldsymbol{S}_1^*$, $\boldsymbol{S}_2^*$, and with the assumptions made in (10) and thereafter, we get (similar as in [14]) for $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ the block structure

$$\begin{aligned} \boldsymbol{A}_1 &= \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{A}_{1,mob} \\ \boldsymbol{0} & \boldsymbol{A}_{1,het} & \tilde{\boldsymbol{A}}_{1,kin} \end{pmatrix} \\ &\in \mathbb{R}^{(J_{mob}+J_{1,het}^*+J_{1,kin}^*)\times(J_{mob}+J_{het}+J_{kin})} \\ \boldsymbol{A}_2 &= \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} & \boldsymbol{A}_{2,sorp} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{2,kin} \end{pmatrix} \\ &\in \mathbb{R}^{(J_{sorp}+J_{min}+J_{2,kin}^*)\times(J_{mob}+J_{sorp}+J_{min}+J_{kin})}, \end{aligned} \tag{17}$$

where $J_{mob}+J_{sorp}+J_{min}+J_{kin} = J_{mob}+J_{het}+J_{kin} = J$ is the total number of reactions.

Transformation of the PDE block

To derive the reduction scheme the block of the PDEs (12) is multiplied with some matrices ${\tilde{\boldsymbol{S}}_1^{\perp}}^T$ and $\boldsymbol{C}_1^T$. In order to allow the user some freedom of choice, we do not prescribe specific matrices $\tilde{\boldsymbol{S}}_1^{\perp}$ and $\boldsymbol{C}_1$. Instead, we give the following five conditions which have to be met by the matrices $\tilde{\boldsymbol{S}}_1^{\perp}$, $\boldsymbol{C}_1$ to be chosen. The conditions are formulated as general as possible. Note that two different specific choices of the matrices are presented in Sec. 5. The conditions are:

1. The number of rows of both matrices $\tilde{\boldsymbol{S}}_1^{\perp}, \boldsymbol{C}_1$ is I
2. $\operatorname{span}\{\tilde{\boldsymbol{S}}_1^{\perp}, \boldsymbol{C}_1\} = \mathbb{R}^I$
3. All columns of $\tilde{\boldsymbol{S}}_1^{\perp}, \boldsymbol{C}_1$ are linear independent
4. 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 $\boldsymbol{S}_1^*$): ${\tilde{\boldsymbol{S}}_1^{\perp}}^T \boldsymbol{S}_1^* = \boldsymbol{0}$
5. There is a matrix $\boldsymbol{D}_1$ such that

$$\boldsymbol{C}_1^T \boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_1 \end{pmatrix} \tag{18}$$

(the block $\boldsymbol{D}_1$ may be rectangular)[4]

Since the matrix $\boldsymbol{S}_1^*$ consists of $J_{mob} + J_{1,het}^* + J_{1,kin}^*$ columns, and 4. holds, the number of columns of $\tilde{\boldsymbol{S}}_1^\perp$ is less than or equal to $I-(J_{mob}+J_{1,het}^*+J_{1,kin}^*)$. Let N_* be defined by postulating that the number of columns of $\tilde{\boldsymbol{S}}_1^\perp$ is $I - J_{mob} - N_*$. It immediatly follows that $J_{1,het}^* + J_{1,kin}^* \leq N_* \leq I - J_{mob}$. From 1., 2. and 3. it follows that the size of the matrix $\boldsymbol{C}_1$ is $I \times (J_{mob}+N_*)$. Hence, the product (18) has the size $(J_{mob} + N_*) \times (J_{mob} + J_{1,het}^* + J_{1,kin}^*)$. As a consequence, $\boldsymbol{D}_1$ has the size $N_* \times (J_{1,het}^* + J_{1,kin}^*)$.

Multiplying the block of the PDEs (12) by ${\tilde{\boldsymbol{S}}_1^\perp}^T$ and by $\boldsymbol{C}_1^T$, and exploiting condition 4, we obtain the two equations

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

Using condition 5. and the structure of $\boldsymbol{A}_1$ (see (17)) it follows that

$$\begin{aligned} &\partial_t\Big(\theta {\tilde{\boldsymbol{S}}_1^\perp}^T \boldsymbol{c}\Big) + L {\tilde{\boldsymbol{S}}_1^\perp}^T \boldsymbol{c} = \boldsymbol{0}\,, \\ &\partial_t(\theta \boldsymbol{C}_1^T \boldsymbol{c}) + L \boldsymbol{C}_1^T \boldsymbol{c} \\ &\quad = \theta \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{A}_{1,mob} \\ \boldsymbol{0} & \boldsymbol{D}_1 \boldsymbol{A}_{1,het} & \boldsymbol{D}_1 \tilde{\boldsymbol{A}}_{1,kin} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_{eq} \\ \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \end{pmatrix} \\ &\quad = \theta \begin{pmatrix} \boldsymbol{r}_{mob} + \boldsymbol{A}_{1,mob} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \\ \boldsymbol{R} \begin{pmatrix} \boldsymbol{r}_{sorp} \\ \boldsymbol{r}_{min} \end{pmatrix} + \boldsymbol{A}_{1,*} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) \end{pmatrix}, \end{aligned} \tag{19}$$

where

$$\boldsymbol{R} := \boldsymbol{D}_1 \boldsymbol{A}_{1,het}, \qquad \boldsymbol{A}_{1,*} := \boldsymbol{D}_1 \tilde{\boldsymbol{A}}_{1,kin} \tag{20}$$

and where $\boldsymbol{r}_{eq} = (\boldsymbol{r}_{mob}, \boldsymbol{r}_{sorp}, \boldsymbol{r}_{min})$ analogously to the decomposition of $\boldsymbol{S}_{1,eq}$ in (8). The size of $\boldsymbol{R}$ is $N_* \times J_{het}$ and the size of $\boldsymbol{A}_{1,*}$ is $N_* \times J_{kin}$.

The formulation (19) motivates to introduce the following transformed variables

$$\boldsymbol{\eta} := {\tilde{\boldsymbol{S}}_1^\perp}^T \boldsymbol{c}\,, \qquad \boldsymbol{\xi} = (\boldsymbol{\xi}_{mob}, \boldsymbol{\xi}_*) := \boldsymbol{C}_1^T \boldsymbol{c}\,. \tag{21}$$

[4]The simplest choice is the following: Let the columns of $\tilde{\boldsymbol{S}}_1^\perp$ be a basis of the orthogonal complement of the space spanned by the columns of $\boldsymbol{S}_1^*$, and let $\boldsymbol{C}_1 = \boldsymbol{S}_1^* ({\boldsymbol{S}_1^*}^T \boldsymbol{S}_1^*)^{-T}$; then $\boldsymbol{C}_1^T \boldsymbol{S}_1^* = \boldsymbol{I}_{J_{mob}+J_{1,het}^*+J_{1,kin}^*}$, i.e., $\boldsymbol{D}_1 = \boldsymbol{I}_{N_*}$ ($N_* = J_{1,het}^* + J_{1,kin}^*$). This choice is underlying to the method in [14].

The number of the variables $\boldsymbol{\eta}$ (reaction invariants) 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 (19) reads

$$\begin{aligned} \partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} &= \mathbf{0}\,, \\ \partial_t(\theta\boldsymbol{\xi}_*) + L\boldsymbol{\xi}_* &= \theta\boldsymbol{R}\begin{pmatrix} \boldsymbol{r}_{sorp} \\ \boldsymbol{r}_{min} \end{pmatrix} + \theta\boldsymbol{A}_{1,*}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \end{aligned} \tag{22}$$

and

$$\partial_t(\theta\boldsymbol{\xi}_{mob}) + L\boldsymbol{\xi}_{mob} = \theta\,\boldsymbol{r}_{mob} + \theta\boldsymbol{A}_{1,mob}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}). \tag{23}$$

This last equation is usually dropped together with the unknowns $\boldsymbol{r}_{mob}$, which reduces the size of the problem. If desired, this equation can be used for an a posteriori computation of the rates $\boldsymbol{r}_{mob}$.

The transformation inverse to (21) obviously reads

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{C}_1 & \tilde{\boldsymbol{S}}_1^{\perp} \end{pmatrix}^{-T} \begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}_* \\ \boldsymbol{\eta} \end{pmatrix}. \tag{24}$$

Because of the conditions 2. and 3. the inverse exists. Partitioning the columns of the inverse analogously to the entries of the vector $(\boldsymbol{\xi}_{mob}, \boldsymbol{\xi}_*, \boldsymbol{\eta})$ we write

$$\begin{pmatrix} \boldsymbol{C}_1 & \tilde{\boldsymbol{S}}_1^{\perp} \end{pmatrix}^{-T} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y} & \boldsymbol{Z} \end{pmatrix}. \tag{25}$$

where the size of the blocks $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ is $I \times J_{mob}$, $I \times N_*$, $I \times (I - J_{mob} - N_*)$, respectively. With help of the conditions 4. and 5. one gets

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

Multiplying this equation by (25) from the left we get

$$\boldsymbol{S}_1^* = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}\boldsymbol{D}_1 \end{pmatrix}.$$

We multiply by $\boldsymbol{A}_1$ and obtain (cf. (16))

$$\boldsymbol{S}_1 = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}\boldsymbol{D}_1 \end{pmatrix} \boldsymbol{A}_1.$$

Using the block structure of $\boldsymbol{A}_1$ (see (17)) and the definition of the matrices $\boldsymbol{R}$ and $\boldsymbol{A}_{1,*}$ (see (20)) we get

$$\boldsymbol{S}_1 = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y}\boldsymbol{R} & \boldsymbol{X}\boldsymbol{A}_{1,mob} + \boldsymbol{Y}\boldsymbol{A}_{1,*} \end{pmatrix}.$$

In particular (see (11) for the block structure of $\boldsymbol{S}_1$),

$$\boldsymbol{X} = \boldsymbol{S}_{1,mob}\,, \quad \boldsymbol{YR} = \boldsymbol{S}_{1,het} = \begin{pmatrix} \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \end{pmatrix}. \tag{26}$$

With the left equality and with (25) the inverse transformation (24) can be rewritten as

$$\boldsymbol{c} = \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\boldsymbol{\xi}_* + \boldsymbol{Z}\boldsymbol{\eta}. \tag{27}$$

Transformation of the ODE block

We construct a matrix $\boldsymbol{S}_2^\perp$ and a matrix $\boldsymbol{B}_2$. The matrix $\boldsymbol{S}_2^\perp$ consists of a maximal system of linear independent vectors that are orthogonal to all columns of $\boldsymbol{S}_2^*$. We choose a matrix $\boldsymbol{B}_2$ being of the same size as $\boldsymbol{S}_2^*$, that fulfills the condition that the columns of $\boldsymbol{B}_2$, $\boldsymbol{S}_2^\perp$ form a basis of the whole space. Furthermore $\boldsymbol{B}_2$ should have a block structure like $\boldsymbol{S}_2^*$ (c.f. (15)), i.e.,

$$\boldsymbol{B}_2 = \begin{pmatrix} * & \boldsymbol{0} & * \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} & \boldsymbol{0} \end{pmatrix}. \tag{28}$$

Note that one possible choice obviously is $\boldsymbol{B}_2 = \boldsymbol{S}_2^*$. Analogously to $\boldsymbol{S}_2^\perp$ a matrix $\boldsymbol{B}_2^\perp$ is constructed from $\boldsymbol{B}_2$, such that the columns of $\boldsymbol{B}_2^\perp$ form a maximal system of columns orthogonal to the columns of $\boldsymbol{B}_2$. The orthogonality relations imply

$$\boldsymbol{S}_2^{\perp\,T}\boldsymbol{S}_2^* = \boldsymbol{0}, \qquad \boldsymbol{B}_2^T\boldsymbol{B}_2^\perp = \boldsymbol{0}\,. \tag{29}$$

Multiplying the block of the ODEs (13) by[5]

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

and applying (16), (29) leads to

$$\begin{aligned} \partial_t\Big(\theta(\boldsymbol{S}_2^{\perp\,T}\boldsymbol{B}_2^\perp)^{-1}\boldsymbol{S}_2^{\perp\,T}\bar{\boldsymbol{c}}\Big) &= \boldsymbol{0} \\ \partial_t(\theta(\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}}) &= \theta\boldsymbol{A}_2\begin{pmatrix} \boldsymbol{r}_{eq} \\ \boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \end{pmatrix} \end{aligned} \tag{30}$$

This motivates the following definition of new variables:

$$\bar{\boldsymbol{\eta}} := (\boldsymbol{S}_2^{\perp\,T}\boldsymbol{B}_2^\perp)^{-1}\boldsymbol{S}_2^{\perp\,T}\bar{\boldsymbol{c}}, \quad \bar{\boldsymbol{\xi}} := (\boldsymbol{B}_2^T\boldsymbol{S}_2^*)^{-1}\boldsymbol{B}_2^T\bar{\boldsymbol{c}} \tag{31}$$

[5]For a proof that the two inverses exist see [9, Sec. 3.1].

The variables $\bar{\boldsymbol{\xi}}$ are partitioned analogously to the partitioning of the columns of $\boldsymbol{S}_2^*$ in (15) into

$$\bar{\boldsymbol{\xi}} = (\bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin}) =: (\bar{\boldsymbol{\xi}}_{het}, \bar{\boldsymbol{\xi}}_{kin}). \tag{32}$$

Then (30) reads

$$\begin{aligned}
\partial_t(\theta\bar{\boldsymbol{\eta}}) &= \mathbf{0} && (33)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{sorp}) &= \theta(\boldsymbol{r}_{sorp} + \boldsymbol{A}_{2,sorp}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})) && (34)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{min}) &= \theta\boldsymbol{r}_{min} && (35)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{kin}) &= \theta\boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}})\,. && (36)
\end{aligned}$$

Hence, we have found some more reaction invariants (33). The variables $\bar{\boldsymbol{\eta}}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin}$ have the sizes $\bar{I} - J_{sorp} - J_{kin} - J^*_{2,kin}$, J_{sorp}, J_{min}, $J^*_{2,kin}$.

For the transformation inverse to (31) we get

$$\bar{\boldsymbol{c}} = \boldsymbol{S}_2^*\bar{\boldsymbol{\xi}} + \boldsymbol{B}_2^{\perp}\bar{\boldsymbol{\eta}}\,. \tag{37}$$

To verify that (37) and (31) are inverse, just substitute the one into the other and use the orthogonality relations (29).

Since the matrices $\boldsymbol{S}_2^*$ and $\boldsymbol{B}_2$ are of the form (28), the transformed variables $\bar{\boldsymbol{\xi}}_{sorp}$ and $\bar{\boldsymbol{\xi}}_{kin}$ in (31) only depend on the upper part $\bar{\boldsymbol{c}}_{nmin}$ of $\bar{\boldsymbol{c}}$, but not on the lower part $\bar{\boldsymbol{c}}_{min}$. Furthermore, since the second column block of (28) consists of unit vectors, the orthogonality relations (29) lead to the fact that the last J_{min} entries in every column of $\boldsymbol{S}_2^{\perp}$ and $\boldsymbol{B}_2^{\perp}$ are always zero. Hence, also $\bar{\boldsymbol{\eta}}$ in (31) only depends on $\bar{\boldsymbol{c}}_{nmin}$, but not on $\bar{\boldsymbol{c}}_{min}$, and we can decompose $\boldsymbol{B}_2^{\perp}$ into an upper block $\tilde{\boldsymbol{B}}_2^{\perp}$ and a lower zero block. Using this and the partitionings (15), (32) we can rewrite the inverse transformation (37) as

$$\bar{\boldsymbol{c}} = \begin{pmatrix} \boldsymbol{S}_{2,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \tilde{\boldsymbol{B}}_2^{\perp}\bar{\boldsymbol{\eta}} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix}. \tag{38}$$

By now, we have transformed system (12)-(14) into (22), (33)-(36), (14), for the unknowns $\boldsymbol{\eta}$, $\bar{\boldsymbol{\eta}}$, $\boldsymbol{\xi}_{mob}$, $\boldsymbol{\xi}_*$, $\bar{\boldsymbol{\xi}}_{het} = (\bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})$, $\bar{\boldsymbol{\xi}}_{kin}$, where $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$ are expressed by (24)/(27), (38). The equations for $\boldsymbol{\eta}$ and for $\bar{\boldsymbol{\eta}}$ are decoupled and linear, reducing the size of the remaining nonlinear system. Note that, while the equilibrium rates $\boldsymbol{r}_{mob}$ are already eliminated (cf. (23)), our aim will be to eliminate the equilibrium rates $\boldsymbol{r}_{sorp}, \boldsymbol{r}_{min}$ as well.

Additional variables

Before we present the final transformed system of equations, let us introduce additional variables $\tilde{\boldsymbol{\xi}}$ defined by

$$\tilde{\boldsymbol{\xi}} := \boldsymbol{\xi}_* - \boldsymbol{R}\bar{\boldsymbol{\xi}}_{het}. \tag{39}$$

The advantage of these additional variables is discussed in Sec. 4.5.

If we solve equation (39) for $\boldsymbol{\xi}_*$ and substitute it into the inverse transformation (27), then using (26), the inverse transformation reads

$$\boldsymbol{c} = \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\tilde{\boldsymbol{\xi}} + \boldsymbol{S}_{1,het}\bar{\boldsymbol{\xi}}_{het} + \boldsymbol{Z}\boldsymbol{\eta}. \tag{40}$$

Solving the ODEs (34) and (35) for $\boldsymbol{r}_{sorp}$ and $\boldsymbol{r}_{min}$ and plugging this in the PDEs (22) gives

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

which can be rewritten as

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}) + L\boldsymbol{\xi}_* = \theta\boldsymbol{A}_*\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) \tag{41}$$

with

$$\boldsymbol{A}_* := \boldsymbol{A}_{1,*} - \boldsymbol{R}\begin{pmatrix} \boldsymbol{A}_{2,sorp} \\ \boldsymbol{0} \end{pmatrix}. \tag{42}$$

Note that all the equilibrium rates are eliminated now from the system. Altogether we obtain the equations

$$\begin{aligned}
\partial_t(\theta\boldsymbol{\eta}) + L\boldsymbol{\eta} &= \boldsymbol{0} && (43)\\
\partial_t(\theta\bar{\boldsymbol{\eta}}) &= \boldsymbol{0} && (44)\\
\tilde{\boldsymbol{\xi}} &= \boldsymbol{\xi}_* - \boldsymbol{R}\bar{\boldsymbol{\xi}}_{het} && (45)\\
\partial_t(\theta\tilde{\boldsymbol{\xi}}) + L\boldsymbol{\xi}_* &= \theta\boldsymbol{A}_*\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) && (46)\\
\partial_t(\theta\bar{\boldsymbol{\xi}}_{kin}) &= \theta\boldsymbol{A}_{2,kin}\boldsymbol{r}_{kin}(\boldsymbol{c},\bar{\boldsymbol{c}}) && (47)\\
\boldsymbol{\phi}(\boldsymbol{c},\bar{\boldsymbol{c}}) &= \boldsymbol{0} && (48)
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{c} &= \boldsymbol{S}_{1,mob}\boldsymbol{\xi}_{mob} + \boldsymbol{Y}\tilde{\boldsymbol{\xi}} + \boldsymbol{S}_{1,het}\bar{\boldsymbol{\xi}}_{het} + \boldsymbol{Z}\boldsymbol{\eta} \\
\bar{\boldsymbol{c}} &= \begin{pmatrix} \boldsymbol{S}_{2,sorp}\bar{\boldsymbol{\xi}}_{sorp} + \boldsymbol{S}^*_{2,kin}\bar{\boldsymbol{\xi}}_{kin} + \tilde{\boldsymbol{B}}_2^{\perp}\bar{\boldsymbol{\eta}} \\ \bar{\boldsymbol{\xi}}_{min} \end{pmatrix}.
\end{aligned} \tag{49}$$

Compared to (12)-(13), but also to other reformulations used in the literature, the advantage of system (43)-(48) is that the equations (43) and (44) decouple from the rest of the system. So the coupled nonlinear system, which must be solved numerically, is smaller. Counting the PDEs in the nonlinear system (the number of PDEs severely influences the number of nonzero entries in the Jacobian), we find N_* many. As pointed out in Footnote 4, a proper choice of $\boldsymbol{S}_1^\perp$ leads to $N_* = J_{1,het}^* + J_{1,kin}^* \leq J_{sorp} + J_{min} + J_{kin} = J - J_{mob}$.

4 Implicit elimination

4.1 The main idea

In order to further reduce the size of the system, we are formally solving some 'local' equations (equations without spatial couplings) for some 'local' variables, and substitute these local variables into the remaining 'global' equations (PDEs). Hence, we perceive the system (45)-(48) as

$$\begin{aligned} \boldsymbol{f}(\boldsymbol{u},\boldsymbol{v}) &= \boldsymbol{0} \\ \boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) &= \boldsymbol{0} \end{aligned} \tag{50}$$

with global equations $\boldsymbol{f}$, local equations $\boldsymbol{g}$, global variables $\boldsymbol{u}$, local variables $\boldsymbol{v}$. The number of global (local) variables must equal the number of global (local) equations.

If the second block has a resolution function $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{u})$, i.e., the function $\boldsymbol{v}(\boldsymbol{u})$ fulfills

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{u})) = \boldsymbol{0} \text{ for all } \boldsymbol{u}\,, \tag{51}$$

then we can rewrite system (50) by the smaller system

$$\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{u})) = \boldsymbol{0}\,. \tag{52}$$

The existence of the resolution function $\boldsymbol{v}(\boldsymbol{u})$ will be checked in Sec. 4.2. To solve this condensed system (52) for $\boldsymbol{u}$ with Newton's method, the evaluation of the mapping $\boldsymbol{u} \mapsto \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{v}(\boldsymbol{u}))$, i.e., of $\boldsymbol{v}(\boldsymbol{u})$, and of the Jacobian of (52),i.e.,[6]

$$\boldsymbol{J} = \partial_1 \boldsymbol{f} + \partial_2 \boldsymbol{f} \boldsymbol{v}', \tag{53}$$

are needed for a given Newton iterate $\boldsymbol{u}$. The evaluation of $\boldsymbol{v}(\boldsymbol{u})$ can be done by solving the local problem

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{0} \quad \text{for given } \boldsymbol{u}$$

[6] $\partial_{1/2}$ denote the partial derivative with respect to the first/second block of variables

with Newton's method, i.e., by a nested Newton iteration. Note that due to the lack of spatial couplings within $\boldsymbol{g}$, there is a decoupling in the sense that the nested Newton iteration can be done mesh point by mesh point. The derivative $\boldsymbol{v}'$ is obtained by differentiating (51) with respect to $\boldsymbol{u}$. This yields $\partial_1 \boldsymbol{g} + \partial_2 \boldsymbol{g}\, \boldsymbol{v}' = \boldsymbol{0}$, i.e.,

$$\boldsymbol{v}' = -(\partial_2 \boldsymbol{g})^{-1}\, \partial_1 \boldsymbol{g}\,. \tag{54}$$

See also Sec. 4.4 for some details on the computation of (54). The invertibility of matrix $\partial_2 \boldsymbol{g}$ is directly related to the existence of the resolution function, cf. Sec. 4.2. Note that $\partial_2 \boldsymbol{g}$ is also the Jacobian of the local problem.

4.2 Existence of the resolution function

We consider equations (47)-(48) to be local and (45)-(46) to be global in the sense of Sec. 4.1. The unknowns are split into local unknowns

$$\boldsymbol{v} = \boldsymbol{\xi}_{loc} := (\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin}) \tag{55}$$

and global unknowns

$$\boldsymbol{u} = \boldsymbol{\xi}_{glob} := (\tilde{\boldsymbol{\xi}}, \boldsymbol{\xi}_*)\,. \tag{56}$$

Now we want to prove the existence of a resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ defined by the equations (47)-(48) (after time discretization of (47) with the implicit Euler method) and $(\boldsymbol{c}, \bar{\boldsymbol{c}})$ given by (49). Note that the resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ does not depend on the variables $\boldsymbol{\xi}_*$ because these variables do not appear in the equations (47)-(48) and in the inverse transformation (49). Hence, we want to show that the equations (48) and

$$\boldsymbol{\phi}_{kin} := \bar{\boldsymbol{\xi}}_{kin} - \bar{\boldsymbol{\xi}}_{kin,old} - \Delta t\, \boldsymbol{A}_{2,kin} \boldsymbol{r}_{kin}(\boldsymbol{c}, \bar{\boldsymbol{c}}) = 0,$$

which is the time-discretized version of (47), have a resolution function

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

We assume that a positive lower bound for the concentrations exists. The following proof is adapted from [14, Appendix]. To apply the implicit function theorem, we have to check that the matrix

$$\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min}, \boldsymbol{\phi}_{kin})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})}\,, \tag{57}$$

which is $\partial_2 \boldsymbol{g}$, is invertible. For that purpose the equilibrium mineral reactions are split into 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.

In a first step, we consider matrix (57) in the limit case $\Delta t = 0$. For this case we obtain that the matrix (57) reads as in Table 1, where we have used the diagonal matrices

$$\begin{aligned} \boldsymbol{\Lambda} &= \mathrm{diag}(1/c_1, \ldots, 1/c_I)\,, \\ \bar{\boldsymbol{\Lambda}}_{nmin} &= \mathrm{diag}(1/\bar{c}_{I+1}, \ldots, /\bar{c}_{I+\bar{I}_{nmin}})\,. \end{aligned}$$

We can rewrite that matrix from Table 1 as

Table 1: Jacobian (57) of the local problem in the limit case $\Delta t = 0$.

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

$$\frac{\partial(\boldsymbol{\phi}_{mob},\boldsymbol{\phi}_{sorp},\boldsymbol{\phi}_{min},\boldsymbol{\phi}_{kin})}{\partial(\boldsymbol{\xi}_{mob},\bar{\boldsymbol{\xi}}_{sorp},\bar{\boldsymbol{\xi}}_{min},\bar{\boldsymbol{\xi}}_{kin})} = \begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\mathcal{A}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{J_{2,kin}^*} \end{pmatrix} \tag{58}$$

with

$$\begin{aligned} \boldsymbol{M}_1 &= \boldsymbol{Q}^T\tilde{\boldsymbol{\Lambda}}\boldsymbol{Q}\,, \quad \tilde{\boldsymbol{\Lambda}} = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{\Lambda}}_{nmin} \end{pmatrix}, \\ \boldsymbol{Q} &= \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min,\mathcal{I}} \\ \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}, \end{aligned} \tag{59}$$

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

Due to the linear independence assumption (see after (10)) the columns of the matrix $\boldsymbol{Q}$ are linear independent. Hence, matrix $\boldsymbol{M}_1$ is symmetric positive definite because the diagonal matrix $\tilde{\boldsymbol{\Lambda}}_{nmin}$ only has positive entries. Hence, matrix (58) is invertible.

For sufficiently small Δt we use a continuity argument, exactly as in [14], to finish the proof.

Using this resolution function and assuming that the decoupled equations (43) and (44) are solved, (43)-(48) can be reduced to the condensed system (cf. (52))

$$\tilde{\boldsymbol{\xi}} \;=\; \boldsymbol{\xi}_* - \boldsymbol{R}\bar{\boldsymbol{\xi}}_{het}(\tilde{\boldsymbol{\xi}}) \tag{60}$$

$$\partial_t(\theta\tilde{\boldsymbol{\xi}}) + L\boldsymbol{\xi}_* \;=\; \theta\boldsymbol{A}_*\tilde{\boldsymbol{r}}_{kin}(\tilde{\boldsymbol{\xi}}). \tag{61}$$

The reaction rates $\boldsymbol{r}_{kin}$ are written as a function of $\tilde{\boldsymbol{\xi}}$. This can be achieved by (49) and the resolution function. Comparing with (43)-(48) one can see that the number of coupled equations is reduced drastically. This nonlinear system will be used for the numerical computations. Contrary to DSA there is no nonlinearity under the transport operator. That is an advantage because it is known that a nonlinearity under the differential operator may cause numerical problems; at least it produces many nonzero entries in the Jacobian.

4.3 The local problem seen as a minimization problem

It is well known that mass action reactions are closely related to the minimization of the so-called Gibbs free energy. We want to show that in our formulation the local problem can be regarded as a convex minimization problem, provided that there are no variables $\bar{\boldsymbol{\xi}}_{kin}$ (i.e., $J^*_{2,kin} = 0$). In [11, Sec. 2.4.4] this consideration is done for the special case that there are no equilibrium minerals and for a formulation that comes without the additional 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

$$\begin{aligned}\mu_i(c_i) &:= \mu_{0,i} - 1 + \ln(c_i)\\ \bar{\mu}_i(\bar{c}_i) &:= \begin{cases} \bar{\mu}_{0,i} - 1 + \ln(\bar{c}_{nmin,i}), & i = I+1, ..., I+\bar{I}_{nmin} \\ \bar{\mu}_{0,i}, & i = I+\bar{I}_{nmin}+1, ..., I+\bar{I}\,, \end{cases}\end{aligned}$$

where the vector $(\boldsymbol{\mu}_0, \bar{\boldsymbol{\mu}}_0) \in \mathbb{R}^{I+\bar{I}}$ is a solution of the linear system

$$\boldsymbol{S}_{eq}^T \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_0 \end{pmatrix} = -\ln(\boldsymbol{K})\,. \tag{62}$$

We calculate

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

We see that the Hessian $D^2G = \mathrm{diag}(\frac{1}{c_1}, ..., \frac{1}{\bar{c}_{I+\bar{I}_{nmin}}}, 0, ..., 0)$ is positive semidefinite. So it follows that G as a function of $(\boldsymbol{c}, \bar{\boldsymbol{c}})$ is a convex functional.

With help of the inverse transformation (49) and for given values for the variables $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\xi}}$, $\bar{\boldsymbol{\eta}}$ we can write G as a function of the local variables $\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}$. Remember 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 function. 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 (49), we compute

$$\begin{aligned}\nabla_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} G &= \boldsymbol{S}_{eq}^T \nabla_{(\boldsymbol{c}, \bar{\boldsymbol{c}})} G \\ &= \boldsymbol{S}_{eq}^T \begin{pmatrix} \boldsymbol{\mu}_0 \\ \bar{\boldsymbol{\mu}}_{0,nmin} \\ \bar{\boldsymbol{\mu}}_{0,min} \end{pmatrix} + \boldsymbol{S}_{eq}^T \begin{pmatrix} \ln(\boldsymbol{c}) \\ \ln(\bar{\boldsymbol{c}}_{nmin}) \\ \boldsymbol{0} \end{pmatrix}.\end{aligned} \tag{63}$$

For the Hessian of G we get

$$D^2_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} G = \boldsymbol{S}_{eq}^T \boldsymbol{\Lambda} \boldsymbol{S}_{eq} \tag{64}$$

with $\boldsymbol{\Lambda} = \mathrm{diag}(\frac{1}{c_1}, ..., \frac{1}{c_I}, \frac{1}{\bar{c}_{I+1}}, ..., \frac{1}{\bar{c}_{I+\bar{I}_{nmin}}}, 0, ..., 0)$. 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 constrained minimization problem

$$\begin{aligned}&\min\ G(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}) \\ &s.t.\ \ \bar{\boldsymbol{\xi}}_{min} \geq \boldsymbol{0}\,.\end{aligned} \tag{65}$$

The Lagrangian 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 (63) and (62) we get for the associated KKT system

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

where the entries of $\boldsymbol{\psi}_{min}$ are defined according to (3). Hence the minimization problem (65) is equivalent to solving the equilibrium conditions (48). So the resolution function $\boldsymbol{\xi}_{loc}(\boldsymbol{\xi}_{glob})$ of the previous subsection can also be defined as the solution of the constrained minimization problem (65) if there are no variables $\bar{\boldsymbol{\xi}}_{kin}$. The difference to [11, 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 in [11] there are two coupled minimization problems are needed.

Note that the Hessian (64) can be written as

$$\begin{aligned}
&D^2_{(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min})} G = \hat{\boldsymbol{S}}_{eq}^T\, \mathrm{diag}\Big(\tfrac{1}{c_1}, ..., \tfrac{1}{\bar{c}_{I+\bar{I}_{nmin}}}\Big)\, \hat{\boldsymbol{S}}_{eq}\,, \\
&\hat{\boldsymbol{S}}_{eq} = \begin{pmatrix} \boldsymbol{S}_{1,mob} & \boldsymbol{S}_{1,sorp} & \boldsymbol{S}_{1,min} \\ \boldsymbol{0} & \boldsymbol{S}_{2,sorp} & \boldsymbol{0} \end{pmatrix}.
\end{aligned}$$

From the assumption (10) we know that the columns of $\hat{\boldsymbol{S}}_{eq}$ are linear independent. As a consequence, the Hessian ist strictly positive definit, i.e., the constrained minimization problem is strictly convex. From this we can derive that the solution of the local problem is unique.

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

In order to use the reformulation of Sec. 3 with the implicit elimination of Sec. 4.1-4.2, we have to compute the matrix $\boldsymbol{v}'$ (see (54)) to apply a Newton

step for (52). Instead of computing $\boldsymbol{v}'$ exactly, we will use an approximation. Note that the usage of an approximate $\boldsymbol{v}'$ does not affect the accuracy of the solution of the time step; it might only affect the convergence of the Newton iteration, if the approximation is poor. We assume that kinetic rate terms, since they contain a factor Δt, are sufficiently small to be neglected for the computation of $\boldsymbol{v}'$. This means that within the computation of $\boldsymbol{v}'$, we replace any occurence of Δt by zero.

To compute $\boldsymbol{v}'$, i.e., $D_{\tilde{\boldsymbol{\xi}}}(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})$, by formula (54) we need the matrices

$$\partial_2 \boldsymbol{g} = \frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min}, \boldsymbol{\phi}_{kin})}{\partial(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min}, \bar{\boldsymbol{\xi}}_{kin})} ,$$
$$\partial_1 \boldsymbol{g} = \frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min}, \boldsymbol{\phi}_{kin})}{\partial \tilde{\boldsymbol{\xi}}} .$$

With the approximation described above, we obtain $\partial_2 \boldsymbol{g}$ by the formulas (58)-(59).

Analogously we can calculate that

$$\begin{aligned} -\partial_1 \boldsymbol{g} &= -\frac{\partial(\boldsymbol{\phi}_{mob}, \boldsymbol{\phi}_{sorp}, \boldsymbol{\phi}_{min,\mathcal{I}}, \boldsymbol{\phi}_{min,\mathcal{A}}, \boldsymbol{\phi}_{kin})}{\partial \tilde{\boldsymbol{\xi}}} \\ &= \begin{pmatrix} \boldsymbol{Q}^T \tilde{\boldsymbol{\Lambda}} \boldsymbol{C} \\ \boldsymbol{0} \\ \boldsymbol{I}_{J^*_{2,kin}} \end{pmatrix} \quad \text{with } \boldsymbol{C} = \begin{pmatrix} -\boldsymbol{Y} \\ \boldsymbol{0} \end{pmatrix} . \end{aligned}$$

Because of the block structure of $\partial_1 \boldsymbol{g}$ and of $\partial_2 \boldsymbol{g}$ we obtain that the matrix $\boldsymbol{v}'$ has a block structure $\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{0} \\ \boldsymbol{I}_{J^*_{2,kin}} \end{pmatrix}$ where $\boldsymbol{U}$ is the solution of the linear systems

$$\boldsymbol{Q}^T \tilde{\boldsymbol{\Lambda}} \boldsymbol{Q} \boldsymbol{U} = \boldsymbol{Q}^T \tilde{\boldsymbol{\Lambda}} \boldsymbol{C} , \tag{66}$$

to be solved at every mesh point. $\boldsymbol{U}$ contains the derivatives $D_{\tilde{\boldsymbol{\xi}}}(\boldsymbol{\xi}_{mob}, \bar{\boldsymbol{\xi}}_{sorp}, \bar{\boldsymbol{\xi}}_{min,\mathcal{I}})$. The matrix of the linear systems $\boldsymbol{Q}^T \tilde{\boldsymbol{\Lambda}} \boldsymbol{Q}$ is symmetric and positive definite and hence it is always invertible.

4.5 Benefit of the Additional Variables

In the end of Sec. 3 we introduced the additional variables $\tilde{\boldsymbol{\xi}}$, which were not present in earlier versions of the reformulation method. Let us motivate this.

We need the matrix $\boldsymbol{v}' = D_{\boldsymbol{\xi}_{glob}} \boldsymbol{\xi}_{loc}$ to solve the condensed system (60)-(61). With introduction of the additional variables, it turns out that the entries of this matrix, i.e., the entries of $\boldsymbol{U}$ (cf. Sec. 4.4) are bounded independently of concentration values and reaction constants (see [9, Sec. 3.5.1] for a proof). This would not be the case for matrix $\boldsymbol{v}'$ without the additional variables $\tilde{\boldsymbol{\xi}}$. As a consequence of the boundedness, for the MoMaS model problem, one equilibrium sorption reaction and in the limit case $\Delta t = 0$, one can prove for the condition number of the Jacobian matrix (53) that $\kappa(\boldsymbol{J}) \leq (1+\sqrt{5})^2/4$ (see [9, Sec. 3.5.2]). Without the additional variables $\tilde{\boldsymbol{\xi}}$ the condition number for large K (K: nonlogarithmized reaction constant) is $\kappa(\boldsymbol{J}) \approx 4K$ (see [9, Sec. 3.6.1]). In realistic problems reaction constants can have very large values, such as $K = 10^{35}$. So the additional variables are mandatory for many realistic problems.

5 Choice of the Transformation Matrices

In the following we will give two specific choices for the transformation matrices $\boldsymbol{C}_1, \tilde{\boldsymbol{S}}_1^{\perp}, \boldsymbol{B}_2, \boldsymbol{B}_2^{\perp}$ of Sec. 3.

The first choice allows to establish a close connection to the well known Morel formulation. Hence, the computer code for the reaformulation technique can be used to simulate Morel-based algorithms.

The second method minimizes the number of remaining coupled PDEs and coincides with the reformulation presented in [9] and is similar to the version given in [14].

5.1 Morel formulation as special case of general reduction scheme

If there are no kinetic reactions, then in the Morel formulation the stoichiometric matrix has the form (see [10] for the notation used in the Morel formulation)

$$\boldsymbol{S} = \left(\begin{array}{ccc} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \\ \hline \boldsymbol{0} & \hat{\boldsymbol{B}} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{I}_{J_{min}} \end{array} \right),$$

and some variables $\boldsymbol{T}, \boldsymbol{T}_M, \boldsymbol{T}_F, \boldsymbol{W}$ denoting the so-called total, total mobile, total fixed and constant concentrations. It is possible to achieve that

$$\tilde{\boldsymbol{\xi}} = \boldsymbol{T}, \quad \boldsymbol{\xi}_* = \boldsymbol{T}_M, \quad \boldsymbol{R}\bar{\boldsymbol{\xi}}_{het} = -\boldsymbol{T}_F, \quad \bar{\boldsymbol{\eta}} = \boldsymbol{W},$$

giving our variables $\tilde{\boldsymbol{\xi}}$, $\boldsymbol{\xi}_*$, $\bar{\boldsymbol{\eta}}$ the concrete meaning of the total, the total mobile, and the constant concentrations.

For this purpose one has to choose the transformation matrices

$$\boldsymbol{C}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{N_*} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{C}^T \end{pmatrix}, \quad \tilde{\boldsymbol{S}}_1^{\perp} \text{ empty matrix} \tag{67}$$

$$\boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I}_{J_{sorp}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{J_{min}} \end{pmatrix}, \quad \boldsymbol{B}_2^{\perp} = \begin{pmatrix} \boldsymbol{I}_{\bar{I}-J_{sorp}-J_{min}} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \tag{68}$$

with $N_* = I - J_{mob}$. Note that the matrix $\tilde{\boldsymbol{S}}_1^{\perp}$ consists of zero columns, and so there will be no decoupled PDEs. This is of course a disadvantage for efficient numerical computations. For the variables $\boldsymbol{\xi}$ defined in (21) it holds

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\xi}_{mob} \\ \boldsymbol{\xi}_* \end{pmatrix} &= \boldsymbol{C}_1^T \boldsymbol{c} = \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_{J_{mob}} \\ \boldsymbol{I}_{N_*} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \boldsymbol{c}_{prim} \\ \boldsymbol{c}_{sec} \end{pmatrix} \\ &= \begin{pmatrix} -\boldsymbol{c}_{sec} \\ \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} \end{pmatrix}. \end{aligned}$$

In particular, one gets the coincidence $\boldsymbol{\xi}_* = \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} = \boldsymbol{T}_M$. The above choice of $\boldsymbol{B}_2$, $\boldsymbol{B}_2^{\perp}$ leads to

$$\bar{\boldsymbol{\xi}}_{sorp} = -\bar{\boldsymbol{c}}_{nmin,sec}\,, \quad \bar{\boldsymbol{\xi}}_{min} = -\bar{\boldsymbol{c}}_{min}\,, \quad \bar{\boldsymbol{\eta}} = \boldsymbol{W}.$$

Since all columns of $\boldsymbol{S}_1$ are linear independent there is $\boldsymbol{S}_1^* = \boldsymbol{S}_1$ and $\boldsymbol{A}_1 = \boldsymbol{I}_{J_{eq}}$. So by computing the matrix product

$$\begin{aligned} \boldsymbol{C}_1^T \boldsymbol{S}_1^* &= \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_{J_{mob}} \\ \boldsymbol{I}_{N_*} & \boldsymbol{C} \end{pmatrix} \begin{pmatrix} \boldsymbol{C} & \boldsymbol{A} & \boldsymbol{D} \\ -\boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_{J_{mob}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} & \boldsymbol{D} \end{pmatrix} \end{aligned}$$

ones sees with the definitions of the matrices $\boldsymbol{D}_1$ (18) and $\boldsymbol{R}$ (20) that

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{D} \end{pmatrix}.$$

Using this we obtain for $\tilde{\boldsymbol{\xi}}$ defined in (39) that

$$\begin{aligned}\tilde{\boldsymbol{\xi}} &= \boldsymbol{\xi}_* - \boldsymbol{R}\bar{\boldsymbol{\xi}}_{het} = \boldsymbol{\xi}_* - \begin{pmatrix}\boldsymbol{A} & \boldsymbol{D}\end{pmatrix}\bar{\boldsymbol{\xi}}_{het} \\ &= \boldsymbol{c}_{prim} + \boldsymbol{C}\boldsymbol{c}_{sec} + \boldsymbol{A}\bar{\boldsymbol{c}}_{nmin,sec} + \boldsymbol{D}\bar{\boldsymbol{c}}_{min}.\end{aligned}$$

Therefore $\tilde{\boldsymbol{\xi}} = \boldsymbol{T}$ and $\boldsymbol{R}\bar{\boldsymbol{\xi}}_{het} = -\boldsymbol{T}_F$ hold.

With help of these identities it is easy to see that also the Morel equations

$$\begin{aligned}\partial_t(\theta\boldsymbol{T}) + L\boldsymbol{T}_M &= \boldsymbol{0} \\ \boldsymbol{T} &= \boldsymbol{T}_M + \boldsymbol{T}_F(\bar{\boldsymbol{c}}_{nmin,sec}, \bar{\boldsymbol{c}}_{min}) \\ \partial_t(\theta\boldsymbol{W}) &= \boldsymbol{0} \\ \boldsymbol{\phi}(\boldsymbol{c}, \bar{\boldsymbol{c}}) &= \boldsymbol{0}\end{aligned}$$

coincide with ours. Using the reduction scheme one has a resolution function $\bar{\boldsymbol{\xi}}_{het}(\tilde{\boldsymbol{\xi}})$. 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

$$\begin{aligned}\partial_t(\theta\boldsymbol{T}) + L\boldsymbol{T}_M &= \boldsymbol{0} \\ \boldsymbol{T} &= \boldsymbol{T}_M + \boldsymbol{T}_F(\boldsymbol{T})\,.\end{aligned}$$

Solving this system is called *global-ODE approach* (see [8]). The formulation used here is very similar to the formulation in [1]. The only difference is that the equation $\boldsymbol{T}_F = \boldsymbol{\psi}(\boldsymbol{T})$, appearing in the formulation of [1], is plugged into the other equations of the formulation.

5.2 The standard version of the reduction scheme as special case of the general reduction scheme

To get the standard version of the reduction scheme, as it is presented in [9, 10], and similar in [13, 11], proceed as follows. Choose a matrix $\boldsymbol{S}_1^{\perp}$ such that its columns form a basis of the orthogonal complement of the range of $\boldsymbol{S}_1^*$. Then let $\boldsymbol{B}_1$ be a matrix such that the columns of $\boldsymbol{B}_1$ and of $\boldsymbol{S}_1^{\perp}$ form a basis of $\mathbb{R}^I$ (so $\boldsymbol{B}_1 = \boldsymbol{S}_1^*$ is a possible choice)[7]. Then construct $\boldsymbol{B}_1^{\perp}$ from $\boldsymbol{B}_1$ in the same way $\boldsymbol{S}_1^{\perp}$ is constructed from $\boldsymbol{S}_1^*$. Then set

$$\tilde{\boldsymbol{S}}_1^{\perp^T} = \left(\boldsymbol{S}_1^{\perp^T}\boldsymbol{B}_1^{\perp}\right)^{-1}\boldsymbol{S}_1^{\perp^T}, \quad \boldsymbol{C}_1^T = \left(\boldsymbol{B}_1^T\boldsymbol{S}_1^*\right)^{-1}\boldsymbol{B}_1^T. \tag{69}$$

For this choice $\tilde{\boldsymbol{S}}_1^{\perp}$ is a *maximal* system of linear independent vectors orthogonal to the range of $\boldsymbol{S}_1^*$. Hence, the size of $\boldsymbol{\eta}$, i.e., the number of decoupling

[7] In [13, 14], $\boldsymbol{B}_1 = \boldsymbol{S}_1^*$ and $\boldsymbol{B}_1^{\perp} = \boldsymbol{S}_1^{\perp}$ is taken.

PDEs, becomes maximal for this choice. We further get $\boldsymbol{D}_1 = \boldsymbol{I}_{N_*}$ (see (18)) and, using (20),

$$\boldsymbol{R} = \boldsymbol{A}_{1,het}. \tag{70}$$

Comparing the definitions of the transformed variables for the generalized formulation (21) and the standard formulation in [10, 9, 11] we have the equality

$$\boldsymbol{\xi}_* = (\boldsymbol{\xi}_{sorp}, \boldsymbol{\xi}_{min}, \boldsymbol{\xi}_{kin}). \tag{71}$$

Using all this and the definition of $\boldsymbol{A}_{1,*}$ (cf. (20)) one sees that the PDEs (22)

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

are identical to the equations

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

in [9, Eq. (3.19)-(3.21)].

Furthermore, for $\boldsymbol{Y}, \boldsymbol{Z}$ defined in (25) the equations

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{S}^*_{1,het} & \boldsymbol{S}^*_{1,kin} \end{pmatrix}, \qquad \boldsymbol{Z} = \boldsymbol{B}_1^{\perp}$$

hold. Plugging $\boldsymbol{R}$ (70) and the partitioning of $\boldsymbol{\xi}_*$ (71) in the definition of the additional variables (39) of the generalized reduction scheme and using the fact that

$$\boldsymbol{A}_{1,het} = \begin{pmatrix} \boldsymbol{I}_{J_{sorp,li}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{ld} & \boldsymbol{I}_{J_{min}} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

in the case of the standard formulation of the reduction scheme (compare (17) with [9, Eq. (3.10)]) gives

$$\begin{pmatrix} \tilde{\boldsymbol{\xi}}_{sorp} \\ \tilde{\boldsymbol{\xi}}_{min} \\ \tilde{\boldsymbol{\xi}}_{kin} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_{sorp} \\ \boldsymbol{\xi}_{min} \\ \boldsymbol{\xi}_{kin} \end{pmatrix} - \begin{pmatrix} \bar{\boldsymbol{\xi}}_{sorp,li} \\ \boldsymbol{A}_{ld}\bar{\boldsymbol{\xi}}_{sorp,ld} + \bar{\boldsymbol{\xi}}_{min} \\ \boldsymbol{0} \end{pmatrix}.$$

Here we find a tiny difference to the standard formulation of the reduction scheme as it is presented in [9, (3.39)] and [10, (18)] where the additional kinetic variables $\tilde{\boldsymbol{\xi}}_{kin}$ do not appear. However, in the general formulation we just doubled these variables $\tilde{\boldsymbol{\xi}}_{kin} = \boldsymbol{\xi}_{kin}$.

6 Numerical Results

In this section we are going to compare the performance of the reformulation in the version which is called the 'standard formulation' in the previous section, to other numerical methods.

For all numerical computations we use the so-called "2-D advective easy test case" of the numerically challenging MoMaS benchmark. See [5] for a description of the benchmark.

In Sec. 6.1 we compare our performance results with those given by other groups. In Sec. 6.2 we try to simulate other numerical methods by our code in order to present results which were obtained using the same numerical kernel for all methods.

6.1 Comparison of the CPU time with other groups

A detailed comparison of the easy test case results of all participants of the benchmark can be found in a synthesis article [4]. So here only a short summary will be given.

For the "2D advective easy test case", three groups presented results. Lagneau and van der Lee presented results computed with the code HYTEC (see [15]), which uses iterative operator splitting (SIA). Mayer and MacQuarrie presented results computed with the code MIN3P (see [16]), which uses the direct substitutional approach (DSA). Hoffmann, Kräutle and Knabner presented results using the presented reformulation method, Sec. 5.2 (for the specific matrices $\boldsymbol{B}_1, \boldsymbol{B}_1^{\perp}, \boldsymbol{B}_2, \boldsymbol{B}_2^{\perp}$ see [10]). In Fig. 1 the normalized CPU times of the three participants in dependence on the number of cells can be seen. This figure is taken from [4].

The other groups give results only for coarser grids (in our opinion too coarse grids, see Sec. 6.3). By extrapolating the lines in Fig. 1 one can estimate that the reduction scheme is faster by a factor greater than five compared with the second fastest code.

The main drawback of DSA is that it leads to a nonlinear system which is difficult to solve numerically because the reaction constants appear in the global system. This is avoided by using the resolution function from Sec. 4. Hence using the resolution function seems to be the preferable method to reduce the number of equations in the global system. Another drawback of DSA is that the decoupling of linear PDEs is not possible. A discussion of the drawbacks of SIA can be found in the following section.

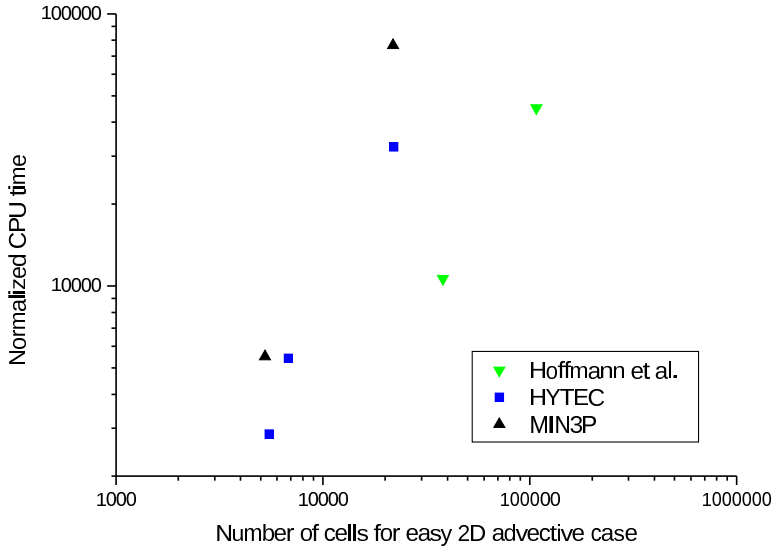

Figure 1: CPU time in the 2D "advective easy test case"

6.2 Comparison of the CPU time with other methods

In this section the universality of the implementation of the reduction scheme (see Sec. 3) is used to compare the CPU times of the reduction scheme with those of the global–ODE approach and those of iterative operator splitting. The time step size is chosen adaptively, based on the number of Newton steps.

In a first comparison the variables and equations of the reduction scheme are used, but the $\boldsymbol{\eta}$-equations are added to the global problem and so are solved simultaneously with the global problem, i.e., the decoupling of the $\boldsymbol{\eta}$-equations is not exploited by the numerical solver. By doing so one can see the saving of CPU time gained by the decoupling of the $\boldsymbol{\eta}$-equations. In Table 2 these results are named "no decoupling of $\boldsymbol{\eta}$-equations".

To obtain the global–ODE approach with the implementation of the generalized reduction scheme one has to choose $\boldsymbol{C}_1$ according to (67).

To do SIA with 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 Jacobian and to plug equation (45) into (46).

Moreover a stopping criterion for the SIA iteration has to be chosen. For SIA we use the same type of stopping criterion as we use for the reduction

scheme and the global-ODE approach, i.e.,

$$|\boldsymbol{r}| < \max\{Eps, |\boldsymbol{r}_0|Red\} \tag{72}$$

where $\boldsymbol{r}_0$ is the initial residual and Eps, Red are user specified parameters. However, in order to get a result in finite time, we had to relax the criterion for SIA by multiplying Red by a factor of ten and multipying Eps by a factor of 100 or 200, i.e., we tried two different values $Eps = 2 \cdot 10^{-8}, 10^{-8}$.

All computations presented are done with the same implementation based on M++ (see [23]), and a FV stabilization with exponential upwinding is used because of the convection dominance.

In Table 2 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.

Table 2: CPU time, time steps, Newton steps (for SIA: iteration steps) for different methods

	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

Comparing the results of the reduction scheme and the case "no decoupling of $\boldsymbol{\eta}$-equations" one can see that the gain of CPU time by the decoupling of the $\boldsymbol{\eta}$-equations is a 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 for 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 feasible to choose the same stopping criterion. Using SIA it is necessary to enlarge the stopping parameter *Eps*, i.e., the error, by a factor of about one 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 is displayed in Table 2.

Analyzing the SIA method one sees that it has a linear convergence behavior like it is expected for a fixed point method. After some iteration steps (about five) in each iteration the residual is reduced by an approximately constant factor. The reason why a severe stopping criterion is practically unusable is that sometimes this factor is only 1.3 depending on the problem and the time step size. Then it would take 18 iteration steps to reduce the residual by a factor of one hundred.

6.3 Grid convergence

The implementation of the reduction scheme also runs on parallel computers. So the use of very fine grids is possible. Furthermore the implementation is done for unstructured grids. So pre-adapted grids can be used. A suitable pre-adapted grid is shown in Fig. 2.

To check the convergence of the method a reference solution on a fine, pre-adapted grid with 608256 cells is computed and the differences between the solutions and the reference solution measured in the discrete L^2-norm are computed (see Table 3). We observe a numerical convergence order of $h^{2/3}$. Since the given permeability is discontinuous, the water flow has only low regularity. The dispersion tensor depends on this flow. So the solution of the transport problem has very low regularity. Hence, such a low numerical convergence order is not really surprising.

Table 3: L^2-error for different grids with respect to the reference solution

number of cells	pre-adapted	error in discrete L^2-norm
26880	no	0.0120
107520	no	0.0076
38016	yes	0.0100

Most streamlines go through the high velocity zone ($1 \leq x \leq 1.1$, $0.9 \leq$

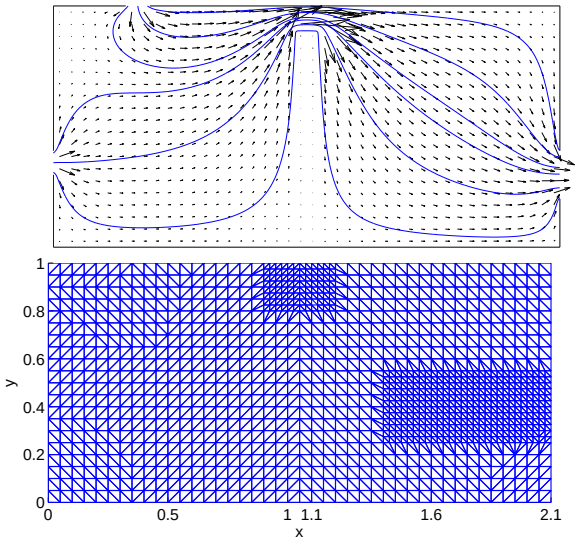

Figure 2: Water flow (top) and grid refined in the high velocity zone and near the outflow (bottom)

$y \leq 1.0$) (see Fig. 2). Hence, the number of cells in this part of the grid is crucial for the accuracy of the solution in the right half of the domain. The numbers of cells in y-direction in the high velocity zone are 64 for the reference solution, 8 for the grid with 26880, 7 for the grid used by HYTEC, 10 for the grid used by MIN3P (in the case 'without CPU time limitation'; in the case 'with CPU time limitation' a coarser grid was used; see [4] for the grids used by HYTEC and MIN3P), 4 for the grid used by GDAE2D (see [7]). The grid used to compute the reference solution is the only pre-adapted grid in this list. All other grids are regular grids.

The grid with 26880 cells is comparable with the grids used by HYTEC and MIN3P. There are considerable differences between the solution on the grid with 26880 cells and the reference solution, especially between x=1 and 1.2 at the bottom, and concerning the extent of the red zone on top right (see Fig. 3). Hence, a mesh with this seems too coarse to give good results on the right-hand side of the domain.

The possibility to use a sufficiently fine mesh (our reference solution) we owe to the high computational efficiency of the reduction method.

So in the "2D advective test case" of the MoMaS benchmark the implementation of the reduction scheme was the only code able to compute a

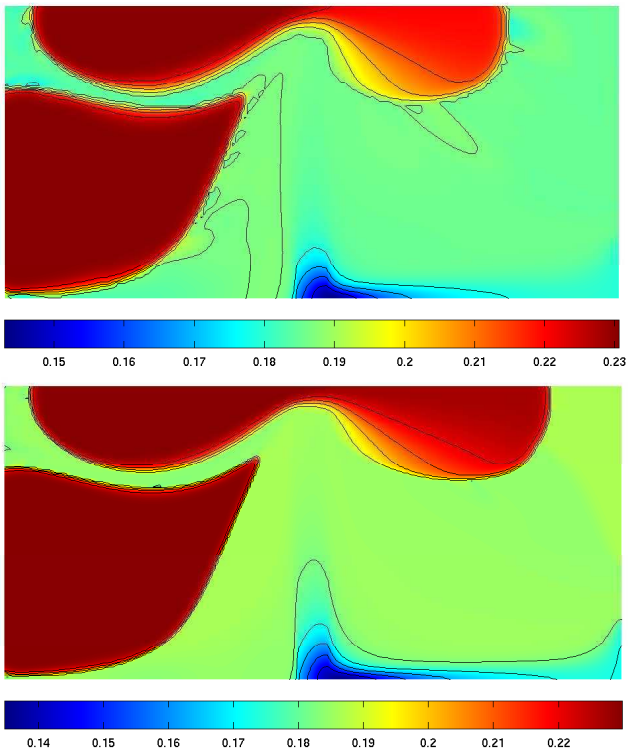

Figure 3: Component X_3 at time 1000. Grid with 26880 cells (top), reference solution with 608256 cells (bottom).

solution on a sufficiently fine mesh, up to now.

7 Conclusions

The reduction scheme decreases the number of coupled equations and with it the CPU time significantly. It is possible to apply the method to realistic problems with large reaction constants and concentrations ranging over many orders of magnitude. The numerically challenging MoMaS benchmark was solved successfully. In this benchmark the reduction scheme is more than five times faster than the software of the second fastest group. Compared with the standard methods SIA and DSA, the reduction scheme is more efficient for this type of problem. Due to the efficiency of the reduction scheme, simulations on finer grids are accomplishable even on standard PCs.

For realistic 2D problems such fine grids are necessary to get an accurate solution.

References

[1] Amir, L., Kern, M.: A global method for coupling transport with chemistry in heterogeneous porous media, Comp. Geosci. 14, 465481 (2010), doi:10.1007/s10596-009-9162-x

[2] Bethke, C.M.: Geochemical reaction modeling, Oxford University Press, New York (1996)

[3] Carrayrou, J.: Modélisation du transport de solutés réactifs en milieu poreux saturé, doctoral thesis, Université Louis Pasteur Straßburg (2001)

[4] Carrayrou, J., Hoffmann, J., Knabner, P., Kräutle, S., de Dieuleveult, C., Erhel, J., Van der Lee, J., Lagneau, V., Mayer, K.U., McQuarrie, K.T.B.: Comparison of numerical methods for simulating strongly non-linear and heterogeneous reactive transport problems — the MoMaS benchmark case, Comp. Geosci. 14, 483-502 (2010), doi:10.1007/s10596-010-9178-2

[5] Carrayrou, J., Kern, M., Knabner, P.: Reactive transport benchmark of MoMaS, Comp. Geosci. 14, 385-392 (2010), doi:10.1007/s10596-009-9157-7

[6] Carrayrou, J., Mosé, R., Behra, P.: Operator-splitting procedures for reactive transport and comparison of mass balance errors, J. Cont. Hydr. 68, 239-268 (2004)

[7] de Dieuleveult, C.: Un modèle numérique global et performant pour le couplage géochimie-transport, doctoral thesis, Université de Rennes 1 (2008)

[8] de Dieuleveult, C., Erhel, J., Kern, M.: A global strategy for solving reactive transport equations, J. Comput. Phys. 228, 6395-6410 (2009)

[9] Hoffmann, J.: Reactive transport and mineral dissolution/precipitation in porous media: efficient solution algorithms, benchmark computations and existence of global solutions, doctoral thesis, University Erlangen-Nuremberg (2010)

[10] Hoffmann, J., Kräutle, S., Knabner, P.: 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, Comp. Geosci. 14, 421-433 (2010), doi:10.1007/s10596-009-9173-7

[11] Kräutle, S.: General multi-species reactive transport problems in porous media: efficient numerical approaches and existence of global solutions, habilitation thesis, University Erlangen–Nuremberg (2008)

[12] Kräutle, S.: The semismooth Newton method for multicomponent reactive transport with minerals, Advances Water Res. 34, 137-151 (2011), doi:10.1016/j.advwatres.2010.10.004

[13] Kräutle, S., Knabner, P.: A new numerical reduction scheme for fully coupled multicomponent transport-reaction problems in porous media, Water Resour. Res. 41 (2005), W09414, doi:10.1029/2004WR003624

[14] Kräutle, S., Knabner, P.: A reduction scheme for coupled multicomponent transport-reaction problems in porous media: Generalization to problems with heterogeneous equilibrium reactions, Water Resour. Res. 43 (2007), W03429, doi:10.1029/2005WR004465

[15] Lagneau, V., van der Lee, J.: HYTEC results of the MoMaS reactive transport benchmark, Comp. Geosci. 14, 435-449 (2010), doi:10.1007/s10596-009-9159-5

[16] Mayer, K.U., MacQuarrie, K.T.B.: Formulation of the multicomponent reactive transport code MIN3P and implementation of MoMaS benchmark problems, Comp. Geosci. 14, 405-419 (2010), doi:10.1007/s10596-009-9158-6

[17] Mayer, K.U., Frind, E.O., Blowes, D.W.: Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions, Water Resour. Res. 38, 1174-1194 (2002), doi:10.1029/2001WR000862

[18] Morel, F., Hering, J.: Principles and applications of aquatic chemistry, Wiley, New York (1993)

[19] Saaltink, M.W., Ayora, C., Carrera, J.: A mathematical formulation for reactive transport that eliminates mineral concentrations, Water Resour. Res. 34, 1649-1656 (1998)

[20] Saaltink, M.W., Carrera, J., Ayora, C.: A comparison of two approaches for reactive transport modelling, J. Geochem. Explo. 69-70, 97-101 (2000)

[21] Scheidegger, A.E.: General theory of dispersion in porous media, J. Geophys. Res. 66, 3273-3278 (1961)

[22] van der Lee, J., De Windt, L., Lagneau, V., Goblet, P.: Modul-oriented modeling of reactive transport with HYTEC, Computers & Geosciences 29, 265-275 (2003)

[23] Wieners, C.: 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, vol. 40, R. Kornhuber, R. Hoppe, J. Priaux, O. Pironneau, O. Widlund, J. Xu (editors), Springer, 175-183 (2004)

PREPRINTS

DES INSTITUTS FÜR ANGEWANDTE MATHEMATIK DER UNIVERSITÄT ERLANGEN-NÜRNBERG

ZULETZT ERSCHIENENE BEITRÄGE:

330 J. Haslinger , G. Leugering , M. Kočvara , M. Stingl: *Multidisciplinary Free Material Optimization*

331 B. Schmidt , M. Stingl , D. A. Berry , M. Döllinger: *Material parameter optimization in a multi-layered vocal fold model*

332 M. Kaiser, A. Thekale: *Solving nonlinear feasibility problems with expensive functions*

333 I. Bomze , G. Eichfelder: *Copositivity detection by difference-of-convex decomposition and ω-subdivision*

334 M. Prechtel, G. Leugering, P. Steinmann, M. Stingl: *Towards optimization of crack resistance of composite materials by adjustment of fiber shapes*

335 A. M. Khludnev , G. Leugering: *Optimal control of cracks in elastic bodies with thin rigid inclusions*

336 M. Prechtel, P. Leiva Ronda, R. Janisch, A. Hartmaier, G. Leugering, P. Steinmann, M. Stingl: *Simulation of fracture in heterogeneous elastic materials with cohesive zone models*

337 E. Marchand: *Combined Deterministic-Stochastic Sensitivity Analysis; Application to Uncertainty Analysis.*

338 G. Eichfelder , T.X.D. Ha: *Optimality conditions for vector optimization problems with variable ordering structures*

339 F.A. Radu, N. Suciu, J. Hoffmann, A. Vogel, O. Kolditz, C-H. Park , S. Attinger: *Accuracy of numerical simulations of contaminant transport in heterogeneous aquifers: a comparative study*

340 M. Kaiser, K. Klamroth, A. Thekale: *Test examples for nonlinear feasibility problems with expensive functions*

341 J. Jahn, T.X.D. Ha: *New Order Relations in Set Optimization*

342 G. Eichfelder , J. Povh: *On reformulations of nonconvex quadratic programs over convex cones by set-semidefinite constraints*

343 T.X.D. Ha and J. Jahn: *Properties of Bishop-Phelps Cones*

344 N. Ray, Ch. Eck, A. Muntean, P. Knabner: *Variable Choices of Scaling in the Homogenization of a Nernst-Planck-Poisson Problem*

345 A. Muntean , T. L. Van Noorden: *Corrector estimates for the homogenizationof a locally-periodic medium with areas oflow and high diffusivity*

346 N.Suciu , S.Attinger , F.A.Radu , C.Vamoş , J.Vanderborght , H.Vereecken , P. Knabner: *Solute transport in aquifers with evolving scale heterogeneity*

347 F. Brunner, F. A. Radu, M. Bause, P. Knabner : *Optimal order convergence of a modified BDM_1 mixed finite element scheme for reactive transport in porous media*

348 G. Eichfelder: *Cone-valued maps in optimization*

349 G. Eichfelder , J. Povh: *On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets*

350 S. Kräutle: *Existence of global solutions of multicomponent reactive transport problems with mass action kinetics in porous media*

351 N. Ray, T. van Noorden, F. Frank, P. Knabner: *Colloid and Fluid Dynamics in Porous Media including an Evolving Microstructure*

352 F. Frank, N. Ray, P. Knabner: *Numerical Investigation of a Homogenized Stokes-Nernst-Planck-Poisson Problem*