Mathematical Modelling and Analysis
Volume 25, Issue 4, 546–568, 2020
https://doi.org/10.3846/mma.2020.10626

http://mma.vgtu.lt
ISSN: 1392-6292
eISSN: 1648-3510

Preconditioned Iterative Method for Reactive Transport with Sorption in Porous Media

Michel Kern[a,b], **Abdelaziz Taakili**[c] **and Mohamed M. Zarrouk**[d]

[a] *INRIA, Paris Research Center*
2 rue Simone Iff, 75589 Paris Cedex 12, France
[b] *Université Paris–Est. CERMICS (ENPC)*
77455 Marne–la–Vallé, France
[c] *ENSAM–Meknès*
Département de Mathématiques et Informatique
Marjane 2, B. P. 15250 Al–Mansor, Meknès, Maroc
[d] *FST–Errachidia*
Département de Mathématiques, 52000 Errachidia, Maroc
E-mail(*corresp.*): ataakili@gmail.com
E-mail: michel.kern@inria.fr
E-mail: ms.zarrouk@gmail.com

Received July 2, 2019; revised July 10, 2020; accepted July 24, 2020

Abstract. This work deals with the numerical solution of a nonlinear degenerate parabolic equation arising in a model of reactive solute transport in porous media, including equilibrium sorption. The model is a simplified, yet representative, version of multicomponents reactive transport models. The numerical scheme is based on an operator splitting method, the advection and diffusion operators are solved separately using the upwind finite volume method and the mixed finite element method (MFEM) respectively. The discrete nonlinear system is solved by the Newton–Krylov method, where the linear system at each Newton step is itself solved by a Krylov type method, avoiding the storage of the full Jacobian matrix. A critical aspect of the method is an efficient matrix-free preconditioner. Our aim is, on the one hand to analyze the convergence of fixed-point algorithms. On the other hand we introduce preconditioning techniques for this system, respecting its block structure then we propose an alternative formulation based on the elimination of one of the unknowns. In both cases, we prove that the eigenvalues of the preconditioned Jacobian matrices are bounded independently of the mesh size, so that the number of outer Newton iterations, as well as the number of inner GMRES iterations, are independent of the

mesh size. These results are illustrated by some numerical experiments comparing the performance of the methods.

Keywords: reactive transport, Newton–Krylov method, preconditioning.

AMS Subject Classification: 35K10; 35K65; 65F08; 65F10; 65N08; 65N30.

1 Introduction and mathematical model

Reactive transport models lead to a set of coupled partial differential equations, with algebraic equations in the case of equilibrium reactions or ordinary differential equations in the case of kinetic reactions. The system may be very large, as the number of unknowns is the number of grid points times the number of chemical species. In Amir and Kern [1] a method was introduced where the chemical equations are eliminated, and a set of transport equations are solved, with a source term implicitly representing the effect of chemistry. The resulting problem is solved by the Newton–Krylov method, where the linear system is solved by an iterative method. It was seen that an efficient preconditioning was a crucial component of the method. However, finding a preconditioner is difficult, as no matrix is constructed in the Newton–Krylov method, and one would like to preserve the decoupling between transport and chemistry that is the main advantage of the formulation in Amir and Kern [1].

In this work, we consider a simplified model with one species undergoing a sorption reaction, given by a known equilibrium isotherm. This choice is motivated by the facts that the resulting mathematical problem has the same structure as that considered in the more general multi-component model, that it is amenable to a more complete analysis, and that it can still be seen as representative of a physically relevant model, see for example Logan [19, Chapter 2] or [5, Sec. 7.3.3]. We denote by c the aqueous concentration (in mol/L) of the species, and by $\bar{c}$ that of the solid part (in mol/kg). The mathematical model given by writing the mass balance equation, and the adsorption relation is:

$$\begin{cases} \partial_t c + \rho_\omega \partial_t \bar{c} - \nabla.(D\nabla c) + \beta.\nabla c = 0 \text{ in } Q_T, \\ \bar{c} = \psi(c) \text{ in } Q_T, \\ c = 0 \text{ in } \partial\Omega, \\ c(0,.) = c_0 \text{ in } \Omega, \end{cases} \tag{1.1}$$

where Ω is a bounded domain in $\mathbf{R}^d$, $1 \leq d \leq 3$, $[0,T]$ a fixed time interval, $Q_T = \Omega \times (0,T]$. The parameters ω, ρ, and D are respectively the porosity, the diffusion-dispersion tensor, β is such that $\omega\beta$ is the Darcy velocity, and we let $\rho_\omega := \frac{\rho(1-\omega)}{\omega}$. For more details on the derivation of the model we refer to Logan [19, Chapter 2] or Bear–Cheng [5, Sec. 7.3.3].

In most of the paper we shall assume that the sorption isotherm is Lipschitz continuous (see (A3) below), as for instance the Langmuir isotherm,

$$\psi(c) = \sigma\, c/(K_L + c), \tag{1.2}$$

where σ and K_L are given constants, though in section 6.2 we will study an

example with the Freundlich isotherm:

$$\psi(c) = K_F c^\alpha, \text{ with } \alpha \in (0,1).$$

In this last case the derivative is singular at $c = 0$, so ψ is not Lipschitz.

The one species sorption model used here, and several of its generalization, has been studied both for its physical insight, as for example in [27], and for its rich mathematical structure, see [28, 29, 30] where the emphasis is put on the analysis of travelling waves, and [3,4] where error estimates for equilibrium adsorption processes for finite element approximation is considered with generalized adsorption isotherm, especially for the Freundlich adsorption isotherm. Note that in the purely hyperbolic case, a semi-analytical solution can be found, see [31], while the advection-diffusion case is treated in [16]. In this work, we concentrate on the interaction of reaction and diffusion in heterogeneous media.

We make the following assumptions on the model data:

(A1) The diffusivity D is a tensor valued function such that there exists $0 < d_0 < d_1$ satisfying:

$$d_0\|\xi\|^2 \leq \xi^T D(x)\xi \leq d_1\|\xi\|^2, \forall \xi \in \mathbf{R}^d, \ \forall x \in \Omega.$$

(A2) The advective velocity field β is a divergence free vector-valued function such that $\|\beta\|_{L^\infty(\Omega)} \leq \beta_0$.

(A3) The sorption isotherm ψ is non-decreasing, non-negative and Lipschitz continuous function with Lipschitz constant L_ψ, such that $\alpha_0 \leq \psi'(x) \leq \alpha_1$, $\forall\, x \in \mathbf{R}$ with $0 < \alpha_0 < \alpha_1$.

The remainder of this paper is organized as follows: the following section recalls how we discretize our model problem. Section 3 concerns iterative methods for solving the nonlinear system: we present and analyze a fixed point method as well as a Newton–Krylov method. In Section 4, we discuss different techniques for linear and nonlinear preconditioning of the Jacobian matrix so as to accelerate the convergence of the Newton–Krylov method. In Section 5, we present a spectral analysis of the preconditioned Jacobian matrix, showing that the eigenvalues of the preconditioned Jacobian matrix are bounded independently of the mesh size h, so that the inner-outer iterations are independent of h. The last section is devoted to some numerical experiments.

2 Discretization

In this section, we describe briefly our discretization method. Our approach is an operator splitting method (cf. [9,23]), the advection and diffusion operators are solved separately using different schemes in time and space. This has the advantage that each physical phenomenon is solved with an appropriate method. The drawbacks are that the method is (formally) first order and that it may be difficult to implement boundary conditions.

Throughout this paper, we use common notation in functional analysis. By $\|\ \|$, we mean the norm in $L^2(\Omega)$. We denote by $H(\mathrm{div};\Omega)$ the space of vector valued function having an L^2 divergence. Its norm will be denoted by

$\| \ \|_{H(\mathrm{div};\Omega)}$. For the space discretization, we assume that Ω is a polygonal (or polyhedral in 3D) domain, and we denote by $\mathcal{T}_h$ a regular decomposition of Ω into closed d-simplices; h stands for the mesh diameter. For the time discretization, we consider a partition of $[0,T]$ into N sub-intervals (t_n, t_{n+1}), for $n = 0, \ldots, N-1$. We denote by $\Delta t = t_{n+1} - t_n$, the time-step.

2.1 Operator splitting method and a semi-discretized problem

The splitting method for the system (1.1) is defined in two steps: for $n = 0, 1, \ldots, N-1$.

Advection step:

$$\begin{cases} \partial_t c + \beta.\nabla c = 0 \text{ in } \Omega \times (t_n, t_{n+1}], \\ c = 0 \text{ on } \partial\Omega \times (t_n, t_{n+1}], \\ c(0,.) = c_0 \text{ in } \Omega. \end{cases} \tag{2.1}$$

Reaction-diffusion step:

$$\begin{cases} \partial_t c + \rho_\omega \partial_t \bar{c} - \nabla.(D\nabla c) = 0 \text{ in } \Omega \times (t_n, t_{n+1}], \\ \bar{c} = \psi(c) \text{ in } \Omega \times (t_n, t_{n+1}], \\ c = 0 \text{ on } \partial\Omega \times (t_n, t_{n+1}], \\ c(0,.) = c^*(t_{n+1},.) \text{ in } \Omega, \end{cases} \tag{2.2}$$

where $c^*(t_{n+1},.)$ is the solution of (2.1).

For the time derivative, we restrict to a simple Euler method. The advective part is treated explicitly, and the diffusive part is treated implicitly.

We denote by c^n an approximation of $c(t_n)$. Each interval (t_n, t_{n+1}) is subdivided into sub-intervals $(t_{n,m}, t_{n,m+1})$, for $m = 0, ..., M-1$, with $t_{n,0} = t_n$ and $t_{n,M} = t_{n+1}$. We denote by $\Delta t_c = t_{n,m+1} - t_{n,m}$ the advection time-step such that $\Delta t = M \Delta t_c$. The semi-discretized problem is given by:

$$\begin{cases} \dfrac{c^{n,m+1} - c^{n,m}}{\Delta t_c} + \beta.\nabla c^{n,m} = 0 \text{ in } \Omega, \ m = 0, ..., M-1, \\ c^{n,m+1} = 0 \text{ on } \partial\Omega, \\ c^{n,0} = c^n, \end{cases} \tag{2.3}$$

and for a given $c^{n,M}$ solution (2.3), we find c^{n+1} and $\bar{c}^{n+1}$ solution of the following problem:

$$\begin{cases} c^{n+1} + \rho_\omega \bar{c}^{n+1} - \Delta t \nabla.(D\nabla c^{n+1}) = f^n \text{ in } \Omega, \\ \bar{c}^{n+1} = \psi(c^{n+1}) \text{ in } \Omega, \\ c^{n+1} = 0 \text{ on } \partial\Omega, \end{cases} \tag{2.4}$$

with $f^n = c^{n,M} + \rho_\omega \bar{c}^n$.

2.2 The fully discrete problem

For the space discretization, we use an upwind cell centered finite volume method for the advection equation (2.1), and a mixed finite element method for the diffusion equation (2.2).

We denote by $|T|$ the volume of mesh element T and by $\mathcal{F}_h$ the set of the faces of the mesh. An interior face $F \in \mathcal{F}_h$ is shared by two mesh elements T^+ and T^-, we arbitrarily chose a normal n_F pointing from T^+ to T^-.

We start by discretizing the advection equation (2.1). We denote by c^m the approximation of the concentration at time $t_{0,m}$ with $c^{0,0} = c^0$. The upwind Euler scheme for the advection equation is obtained by integrating equation (2.3) over a mesh element T

$$\int_T \frac{c^{m+1} - c^m}{\Delta t_c}\, dx - \int_{\partial T} c^{m*}\, \beta.n_T\, ds = 0, \quad m = 0, ..., M-1, \tag{2.5}$$

where c^{m*} denotes the upstream concentration, and taking a piece-wise constant approximation for c^m. We denote by c_T^m the approximation of c^m over T. Equation (2.5) becomes

$$|T|\frac{c_T^{m+1} - c_T^m}{\Delta t_c} - \sum_{F \subset \partial T} c^{m*} \xi_{T,F} \beta_F = 0,$$

where $\xi_{T,F}$ and β_F are given by

$$\xi_{T,F} = n_T.n_F = \begin{cases} 1, \text{ if } n_S \text{ outgoing from } T, \\ -1, \text{ otherwise.} \end{cases}, \quad \beta_F = \int_F \beta.n_F\, ds.$$

The time-step Δt_c is chosen such that $\Delta t = M \Delta t_c$ with $M \geq 1$. It is controlled by the following CFL condition (2.6) that ensures that the scheme is stable and determines the value of M

$$\Delta t_c < \min_{T \in \mathcal{T}_h} |T| / \int_{F^- \subset \partial T} \beta.n_{F^-} ds. \tag{2.6}$$

We now describe the space discretization of problem (2.4), which will be achieved by a mixed finite element method. We let $V_h \subset H(\mathrm{div}; \Omega)$ and $L_h \subset L^2(\Omega)$ denote finite element spaces of dimension n_{V_h} and m_{L_h}, respectively.

For a given $c_h^M = \sum_{T \in \mathcal{T}_h} c_T^M \mathbf{1}_T$, a mixed finite element discretization of (2.4) seeks c_h^{n+1}, $\bar{c}_h^{n+1} \in L_h$ and $\mathbf{q}_h^{n+1} \in V_h$, $n \geq 0$ satisfying

$$\begin{cases} a(\mathbf{q}_h^{n+1}, \mathbf{v_h}) + b(c_h^{n+1}, \mathbf{v}_h) = 0, \quad \forall \mathbf{v}_h \in V_h, \\ m(c_h^{n+1}, p_h) + \rho_\omega m(\bar{c}_h^{n+1}, p_h) - \Delta t\, b(p_h, \mathbf{q}_h^{n+1}) = l(p_h), \ \forall p_h \in L_h, \\ m(\bar{c}_h^{n+1}, q_h) = m(\psi(c_h^{n+1}), q_h), \ \forall q_h \in L_h, \end{cases} \tag{2.7}$$

where the bi-linear forms a, b and m and the linear form l are given by

$$a(\mathbf{u}_h, \mathbf{v}_h) := \int_\Omega \mathbf{v}_h^T D^{-1} \mathbf{u}_h\, dx, \quad b(s_h, \mathbf{v}_h) := -\int_\Omega s_h \nabla.\mathbf{v}_h\, dx,$$
$$m(p_h, q_h) := \int_\Omega p_h q_h\, dx, \quad l(p_h) := \int_\Omega f_h^n p_h\, dx.$$

To ensure that the discretization (2.7) of (2.4) is stable, the finite element spaces V_h and L_h will be assumed to satisfy the coercivity and uniform inf-sup conditions:

$$\text{coercivity: } \exists\, \alpha > 0,\ a(\mathbf{v}_h, \mathbf{v}_h) \geq \alpha \|\mathbf{v}_h\|^2_{H(\mathrm{div};\Omega)},\ \forall \mathbf{v}_h \in \mathcal{K}_0^h, \tag{2.8}$$

$$\text{inf-sup: } \exists\, \gamma > 0,\ \inf_{q_h \in L_h} \sup_{\mathbf{v}_h \in V_h} \frac{b(q_h, \mathbf{v}_h)}{\|q_h\|_{0,\Omega} |||\mathbf{v}_h|||} \geq \gamma, \tag{2.9}$$

where $\mathcal{K}_0^h = \{\mathbf{v}_h \in V_h\ : b(q_h, \mathbf{v}_h) = 0,\ \forall q_h \in L_h\}$.

In the following, we will restrict the discussion to the case where V_h and L_h are the lowest order Raviart–Thomas–Nédélec spaces for velocity and pressure respectively, which are known to satisfy the above two conditions, see [7]. For this choice, the discrete velocities within each element are determined uniquely by the fluxes on the faces (when $d = 3$) or edges (when $d = 2$) of the elements.

We recall the definition of the corresponding spaces in two dimensions:

$$L_h = \{p \in L^2(\Omega) :\ p_{|_T} \in \mathbb{P}_0,\ \forall T \in \mathcal{T}_h\},$$
$$V_h = \left\{\mathbf{v} \in H(\mathrm{div};\Omega) :\ \mathbf{v}_{|_T} \in \mathbb{R}T_0,\ \forall T \in \mathcal{T}_h\right\},$$

$\mathbb{P}_0$ is the space of constant functions and $\mathbb{R}T_0$ is the lowest-order Raviart–Thomas space,

$$\mathbb{R}T_0 = \left\{ \begin{pmatrix} ax + b \\ az + c \end{pmatrix}, a,\, b,\, c \in \mathbb{R} \right\}.$$

A non-linear system corresponding to (2.7) can be constructed as follows. Let $\{\phi_i(x)\}_{1 \leq i \leq n_{V_h}}$ denote the basis for V_h built from the standard choice of degrees of freedom (integral of the normal fluxes on each edge), and let $\{\varphi_1(x), \varphi_2(x), ..., \varphi_{m_{L_h}}(x)\}$ be a basis for L_h, where $\varphi_i(x)$ are the characteristic functions of an element T_i. Expand $\mathbf{q}_h^{n+1}$, c_h^{n+1} and $\bar{c}_h^{n+1}$ using this basis:

$$\mathbf{q}_h^{n+1}(x) = \sum_{i=1}^{n_{V_h}} (\mathbf{q}_h^{n+1})_i \phi_i(x),$$
$$c_h^{n+1}(x) = \sum_{i=1}^{m_{L_h}} (c_h^{n+1})_i \varphi_i(x), \quad \bar{c}_h^{n+1}(x) = \sum_{i=1}^{m_{L_h}} (\bar{c}_h^{n+1})_i \varphi_i(x),$$

where with some abuse of notation, we have used $\mathbf{q}_h(x)$ to denote a finite element function and $\mathbf{q}_h$ to denote its vector representation relative to the given basis. Substituting $\mathbf{v}_h(x) = \phi_i(x)$ for $i = 1, ..., n_{V_h}$ and $p_h(x) = \varphi_i(x)$ for $i = 1, ..., m_{L_h}$ into the above yields the following algebraic system:

For $n = 0, 1, \ldots$, solve

$$\begin{cases} A\mathbf{q}_h^{n+1} + B^T c_h^{n+1} = 0, \\ M c_h^{n+1} + \rho_\omega M \bar{c}_h^{n+1} - \Delta t\, B\mathbf{q}_h^{n+1} = F^n, \\ M\bar{c}_h^{n+1} = M\Psi(c_h^{n+1}), \end{cases} \tag{2.10}$$

where we define the matrices A, B, M as

$$(A)_{ij} = a(\phi_i, \phi_j),\ (B)_{ij} = b(\phi_j, \varphi_i), \text{ and } (M)_{ij} = m(\varphi_j, \varphi_i),$$

the vector F^n as $(F^n)_i = l(\varphi_i)$, and the function Ψ as $(\Psi(c_h^{n+1}))_i = \psi(c_{h,i}^{n+1})$. By eliminating the unknowns $\mathbf{q}_h^{n+1}$ in (2.10), we obtain the following nonlinear system:

$$\begin{cases} Sc_h^{n+1} + \rho_\omega M\bar{c}_h^{n+1} = F^{n+1}, \\ M\bar{c}_h^{n+1} = M\Psi(c_h^{n+1}), \end{cases} \tag{2.11}$$

where S is defined as $S := M + \Delta t\, BA^{-1}B^T$.

Existence for the solution of the discrete system (2.11) will be proved in Proposition 1 below.

2.3 Properties of A and B

In this section, we summarize properties of matrices A and B in system (2.10). The matrix A is symmetric positive definite of size n_{V_h}, it is sparse for the choice of the basis ϕ_i indicated above and it corresponds to a mass matrix when $D = I$. When $D \neq I$, we obtain that:

$$\frac{1}{d_1}\mathbf{v}_h^T G\mathbf{v}_h \leq \mathbf{v}_h^T A\mathbf{v}_h \leq \frac{1}{d_0}\mathbf{v}_h^T G\mathbf{v}_h,\ \mathbf{v}_h \in V_h, \tag{2.12}$$

where G denotes the velocity mass matrix.

In addition to the above, the following property will hold for matrix B [20, lem 10.74, p. 478].

$$|q_h^T B\mathbf{v}_h| \leq c^* h^{-1}\|\mathbf{v}_h\|_A \|q_h\|_M,\ \mathbf{v}_h \in V_h \text{ and } q_h \in L_h, \tag{2.13}$$

where M denotes the mass matrix for the concentration $\|q_h\|_M^2 = q_h^T M q_h = \|q_h\|_{L^2(\Omega)}^2$ and c^* is positive constant independent of the mesh size h.

We now recall the following bounds for the Schur complement S.

Lemma 1. *Let the coercivity and the inf-sup conditions* (2.8) *and* (2.9) *hold. Then there exists $c^* > 0$ independent of h, such that*

$$(1 + d_0\gamma^2\Delta t)(q_h^T M q_h) \leq q_h^T S q_h \leq (1 + c^* h^{-2}\Delta t d_1)(q_h^T M q_h). \tag{2.14}$$

Proof. Follows from bounds (2.12) and (2.13). See [20, lem 10.74, p. 479]. □

3 Iterative methods for nonlinear systems

In this section, we discuss iterative algorithms applied to the nonlinear system (2.11). We first consider an iterative fixed point algorithm, then we present a Newton–Krylov algorithm. In both case, we establish the convergence result.

3.1 Fixed point iteration

The fixed point iteration for the nonlinear system (2.11) reads: Let $\bar{c}_h^{n+1,0} \in L_h$ be some initial starting point and define $c_h^{n+1,k+1}$ and $\bar{c}_h^{n+1,k+1}$ iteratively as the solution of

$$\begin{cases} Sc_h^{n+1,k+1} + \rho_\omega M\bar{c}_h^{n+1,k} = F^{n+1}, \\ M\bar{c}_h^{n+1,k+1} = M\Psi(c_h^{n+1,k+1}). \end{cases} \tag{3.1}$$

Proposition 1. *Assume (A1)–(A3) hold. Under the condition*

$$\rho_\omega L_\psi/(1+\gamma^2 d_0 \Delta t) < 1, \tag{3.2}$$

the nonlinear system (2.11) *admits a unique solution that is the limit of the fixed-point iteration* (3.1).

Proof. To simplify notation, we will remove the superscript $n+1$ throughout the proof. We can also eliminate the unknown $\bar{c}$, and the proposition will follow if we prove that (if condition (3.2) is satisfied) the function $G : \mathbf{R}^{m_{\mathcal{L}_h}} \to \mathbf{R}^{m_{\mathcal{L}_h}}$ defined implicitly by $SG(c_h) = F - \rho_\omega M \Psi(c_h)$ is a contraction. We compute, for two given vectors c_h and c_h':

$$S(G(c_h) - G(c_h')) = -\rho_\omega M(\Psi(c_h) - \Psi(c_h')),$$

and take the inner product with $\Delta G := G(c_h) - G(c_h')$. On the left hand side, using Lemma 1, we obtain (with the notation $\|c\|_M = \sqrt{c^T M c}$)

$$(1 + d_0\gamma^2 \Delta t)\, \|\Delta G\|_M^2 \le \Delta G^T S \Delta G,$$

while on the right hand side, denoting $\Delta\Psi := \Psi(c_h) - \Psi(c_h')$, there holds, using the Cauchy–Schwarz inequality

$$|\rho_\omega \Delta G^T M \Delta \Psi| \le \rho_\omega \|\Delta G\|_M \, \|\Delta \Psi\|_M.$$

Now we observe that

$$\|\Delta\Psi\|_M^2 = \int_\Omega \left(\psi(c_h) - \psi(c_h')\right)^2 \, dx \le L_\psi^2 \int_\Omega (c_h - c_h')^2 \, dx = L_\psi^2 \|c_h - c_h'\|_M^2.$$

The conclusion of the Proposition follows. □

Remark 1. Condition (3.2) and the conclusion of Proposition 1 is surprising as it says that the iterative method converges provided the time-step is *large enough*. Indeed, condition (3.2) splits into two cases:

- $\rho_\omega L_\psi < 1$: in that case, there are non restrictions on Δt;
- $\rho_\omega L_\psi \ge 1$: in that case, condition (3.2) becomes $\Delta t > (\rho_\omega L_\psi - 1)/(\gamma^2 d_0)$.

In the second case, the condition may be too restrictive, given that we use an implicit method, which has no stability condition. For this reason, it may be interesting to turn to Newton's method.

3.2 A Newton–Krylov algorithm

In this section, we present the Newton–Krylov algorithm for the nonlinear system (2.11), where a linear system is solved by a Krylov iterative method at each Newton step. See [17,18] for general references or [15] in a related context. The non-linear system to be solved at each time-step is

$$F\begin{pmatrix} c_h \\ \bar{c}_h \end{pmatrix} = 0, \tag{3.3}$$

where F is given by

$$F\begin{pmatrix} c_h \\ \bar{c}_h \end{pmatrix} = \begin{pmatrix} Sc_h + \rho_\omega M\bar{c}_h - F \\ M\bar{c}_h - M\Psi(c_h) \end{pmatrix}.$$

Let us set $X = (c_h, \bar{c}_h)$, the Newton–Krylov algorithm for the system (3.3) reads: Given an initial iterate $X^{(0)}$ and letting $X^{(k)}$ be the current approximate solution, the next approximate solution $X^{(k+1)}$ is obtained through the following steps:

1. Inexactly solve (using a Krylov subspace method such as GMRES) the linear system
$$J_k \delta X = -F(X^{(k)}), \tag{3.4}$$
and obtain an approximate Newton direction $d^{(k)}$ such that
$$\|F(X^{(k)}) + J_k d^{(k)}\| \leq \eta_k \|F(X^{(k)})\|. \tag{3.5}$$

2. Compute the new approximate solution
$$X^{(k+1)} = X^{(k)} + \lambda d^{(k)}, \tag{3.6}$$
where $\lambda \in (0, 1]$ is chosen through a line-search to ensure global convergence (see [17]).

In the description above, J_k denotes the Jacobian matrix, the first derivative of F with respect to the degrees of freedom, it is given by

$$J_k = \begin{pmatrix} S & \rho_\omega M \\ -D^{(k)} & M \end{pmatrix} \tag{3.7}$$

with $D_{ij}^{(k)} = \int_\Omega \psi'(c_h)\varphi_j\varphi_i dx$. Notice that, because of our choice of basis for L_h, the matrix $D^{(k)}$ is diagonal.

Additionally, $\eta_k \in [0, 1)$ is a forcing term that controls how accurately system (3.4) should be solved. Several possible strategies for choosing η_k have been proposed in [12] and are reviewed in [17]. Typically η_k has to go to zero as the non-linear iteration converges, and it is chosen to strike a balance between over-solving the linear system (3.5) at the beginning of the iterations and keeping the fast convergence of the (exact) Newton method.

Let us now recall the local convergence result of the inexact Newton iteration (3.6). The Newton iteration for the nonlinear system (2.11) reads: Let c_h^0, $\bar{c}_h^0 \in L_h$ be some initial starting point and define c_h^{k+1} and $\bar{c}_h^{k+1}$ iteratively as the solution of

$$\begin{cases} Sc_h^{k+1} + \rho_\omega M\bar{c}_h^{k+1} = F, \\ M\bar{c}_h^{k+1} - D^{(k)}c_h^{k+1} = M\Psi(c_h^k) - D^{(k)}c_h^k. \end{cases}$$

We quote here part of the basic convergence result from C. T. Kelley's book, Theorem 6.1.2 in [17]:

Lemma 2. *Assume* (3.2) *holds, so that* (3.3) *has a unique solution* X^**, and that* $J_k(X^*)$ *is non-singular. Assume also that* ψ' *is Lipschitz continuous with Lipschitz constant* $L_{\psi'}$*,* $\lambda = 1$ *(no line search is performed). Let* $\mathcal{B}(\delta) = \{X | \, \|X - X^*\| < \delta\}$ *Then there exists* $\delta > 0$ *such that, if* $X^{(0)} \in \mathcal{B}(\delta)$ *and if the sequence* η_k *tends to 0 as* $k \to \infty$*, then the inexact Newton iteration* (3.6) *converges q-superlinearly to* X^**.*

4 Linear and nonlinear preconditioning

In the Newton–Krylov method, the linear system (3.4) is solved using an iterative Krylov solver as GMRES, the Jacobian J_k is only used once per solver iteration in matrix-vector product between the Jacobian and the Krylov vector w, as in

$$v = J_k w. \tag{4.1}$$

The Jacobian-vector product can be approximated by finite differencing of the residual or computed exactly as in (3.7).

The efficiency of Newton–Krylov method depends on the choice of an adequate preconditioner. The purpose of preconditioning the system in (3.4) is to reduce the number of iterations, and thus accelerate the convergence rate of iterative solver. In this section, we distinguish between two types of preconditioning: linear preconditioning where the linear system is preconditioned respecting the block structure of the Jacobian matrix and nonlinear preconditioning where the original nonlinear system (3.3) is replaced by new system $\tilde{F}(\bar{c}_h) = 0$ by eliminating the unknown c_h such that the two systems lead to the same solution.

4.1 Linear preconditioning

Using right preconditioning, the system (3.4) can be rewritten as

$$\left(J_k P^{-1}\right) P \begin{pmatrix} \delta c_h \\ \delta \bar{c}_h \end{pmatrix} = -F \begin{pmatrix} c_h^k \\ \bar{c}_h^k \end{pmatrix},$$

where P represents the preconditioning matrix, and we denote by $X = (\delta c_h, \delta \bar{c}_h)$. The solution of this system can be divided into two steps, we first solve

$$J_k P^{-1} Y = -F,$$

and then

$$PX = Y. \tag{4.2}$$

The preconditioner matrix P should be a good approximation of J_k such that the cost of constructing P should be minimal and solving the linear system (4.2) should be easier than solving the original system.

When GMRES is applied to solve this preconditioned system, the Jacobian-vector product in (4.1) takes the form

$$v = J_k P^{-1} w,$$

which decomposes into solving the linear system $PZ = w$, for Z and then computing the Jacobian-vector product $v = J_k Z$.

4.1.1 Block Jacobi

Our first strategy is to approximate the Jacobian system with

$$P = \begin{pmatrix} S & 0 \\ 0 & M \end{pmatrix}.$$

We neglect the coupling between the transport and the chemistry fields. In this case, we have

$$J_k P^{-1} = \begin{pmatrix} \mathcal{I} & \rho_\omega \mathcal{I} \\ -D^{(k)} S^{-1} & \mathcal{I} \end{pmatrix}.$$

The Schur complement of the preconditioned matrix $J_k P^{-1}$ is given by

$$\tilde{S} = \mathcal{I} + \rho_\omega D^{(k)} S^{-1}.$$

4.1.2 Block Gauss–Seidel

Our second strategy is to approximate the Jacobian system with

$$P = \begin{pmatrix} S & 0 \\ -D^{(k)} & M \end{pmatrix}.$$

In this case, we have

$$J_k P^{-1} = \begin{pmatrix} \mathcal{I} + \rho_\omega D^{(k)} S^{-1} & \rho_\omega \mathcal{I} \\ 0 & \mathcal{I} \end{pmatrix}.$$

4.2 Nonlinear preconditioning

In this section, we propose an alternative formulation for the original nonlinear system (3.3) by eliminating the unknown c_h. Since S is positive definite, it follows from the first equation of (3.3) that

$$c_h = S^{-1}(F - \rho_\omega M \bar{c}_h).$$

Substituting this into the second equation of (3.3), we obtain

$$\tilde{F}(\bar{c}_h) := M\bar{c}_h - M\Psi(S^{-1}(F - \rho_\omega M\bar{c}_h)) = 0. \tag{4.3}$$

The Jacobian matrix for the nonlinear system (4.3) can be computed exactly as

$$\tilde{J}_k = M + \rho_\omega D^{(k)} S^{-1} M. \tag{4.4}$$

This formulation looks complicated because of the presence of S^{-1}, but is actually fairly easy to implement. As usual, the inverse of S is not actually computed. Rather, when one needs to evaluate the residual, one simply solves a linear system with the matrix S, and this turns out be the building block singled out before, namely the solution of one transport step, with some given source term.

We remark that $\tilde{J}_k$ is the Schur complement of the block Jacobi preconditioned Jacobian matrix $\tilde{S}$ multiplied by the mass matrix, but the preconditioner is nonlinear, it acts on the nonlinear system.

5 Spectral analysis of the preconditioned linear system

In this section, we present an analysis of the spectrum of the preconditioned Jacobian matrix showing that the spectra of the preconditioned systems are bounded independently of the discretization mesh size h. We denote by μ_j, $j = 1, ..., N$ the eigenvalues of the generalized eigenproblem

$$Su = \mu D^{(k)} u.$$

The eigenvalues μ_j are real positive numbers because S is symmetric positive and $D^{(k)}$ is symmetric positive definite, if ψ is an increasing function.

Let us first present a spectral analysis of the matrix S.

Lemma 3. *Let assumptions (A1)–(A3), as well as the coercivity and inf-sup conditions hold. Then, with the notation as in Section 2.12, the eigenvalues μ_j, $j = 1, ..., N$ satisfy*

$$\alpha_1^{-1}(1 + d_0\gamma^2\Delta t) \le \mu_j \le \alpha_0^{-1}(1 + c^* h^{-2} d_1 \Delta t).$$

Proof. As matrix S is symmetric and $D^{(k)}$ is diagonal, we use the characterization of the eigenvalues as a Rayleigh quotient: $\mu = u^T S u / u^T D^{(k)} u$ for an eigenvector u. We bound the numerator using (2.14), and the denominator using assumption (A3), so that $\alpha_0\, u^T M u \le u^T D^{(k)} u \le \alpha_1\, u^T M u$. □

Now, we are able to give a bound on the eigenvalues of the preconditioned Jacobian matrix.

Proposition 2. *The eigenvalues $\{\lambda_i\}_{i=1}^{2N}$ of the generalized eigenvalue problem*

$$\begin{pmatrix} S & \rho_\omega M \\ -D^{(k)} & M \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix} = \lambda \begin{pmatrix} S & 0 \\ 0 & M \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix}$$

lie on the union of two vertical segments in the complex plane, $\{1\} \times [\beta_0, \beta_1] \cup \{1\} \times [-\beta_1, -\beta_0]$, where β_0 and β_1 are given by

$$\beta_0 = h\sqrt{\rho_\omega \alpha_0/(h^2 + c^* d_1 \Delta t)}, \quad \beta_1 = \sqrt{\rho_\omega \alpha_1/(1 + d_0\gamma^2\Delta t)}.$$

Proof. The eigenvalues $\{\lambda_i\}_{i=1}^{2N}$ satisfy

$$S\underline{u} + \rho_\omega M\underline{v} = \lambda S\underline{u} \text{ and } D^{(k)}\underline{u} = (1 - \lambda) M\underline{v}.$$

We first note that 1 cannot be an eigenvalue. Indeed, if that were the case, we would get $Mv = 0$ from the first equation and $D^{(k)}u = 0$ from the second. Since both matrices are invertible, we have a contradiction. We can then eliminate $\underline{v}$ to obtain

$$S\underline{u} = -\frac{\rho_\omega}{(\lambda - 1)^2} D^{(k)}\underline{u}.$$

From this equality, we obtain

$$\rho_\omega/(\lambda - 1)^2 = -\mu,$$

where μ is an eigenvalue of the generalized problem

$$S\underline{u} = \mu D^{(k)}\underline{u}.$$

Equivalently, as Lemma 3 shows that the μs are all positive,

$$\lambda^{\pm} = 1 \pm i\sqrt{\rho_\omega/\mu}.$$

We conclude using the bounds obtained in Lemma 3 again. □

We have a similar result for the Schur complement matrix, which corresponds to the Jacobian matrix (4.4). This time, all eigenvalues are real and positive.

Proposition 3.

- *The N eigenvalues $\{\lambda_i\}_{i=1}^N$ of the generalized eigenvalues problem*

$$\tilde{J}_k v = \lambda M v \tag{5.1}$$

lie in the interval $[\sigma_0, \sigma_1]$ with $\sigma_0 = 1 + h^2\rho_\omega\alpha_0/(h^2 + c^* d_1 \Delta t)$ and $\sigma_1 = 1 + \rho_\omega\alpha_1/(1 + d_0\gamma^2\Delta t)$.

- *The $2N$ eigenvalues $\{\lambda_i\}_{i=1}^{2N}$ of the generalized eigenvalues problem*

$$\begin{pmatrix} S & \rho_\omega M \\ -D^{(k)} & M \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix} = \lambda \begin{pmatrix} S & 0 \\ -D^{(k)} & M \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{v} \end{pmatrix} \tag{5.2}$$

consist of $\lambda = 1$ together with the generalized eigenvalues of $\tilde{J}_k$.

Therefore, the eigenvalues are bounded independently of h.

Proof. Let λ be an eigenvalue of the generalized eigenvalues problem $\tilde{J}_k v = \lambda M v$. Equivalently,

$$\rho_\omega D^{(k)} S^{-1} M v = (\lambda - 1) M v.$$

As in the previous proposition, we see that $\lambda = 1$ cannot be an eigenvalue. We then put $\underline{v} = S^{-1}Mv$ to see that $\frac{\rho_\omega}{\lambda - 1}$ is a eigenvalue of the generalized problem

$$S\underline{u} = \mu D^{(k)}\underline{u}.$$

Thus, $\lambda = 1 + \rho_\omega/\mu$ and, using Lemma 3 one more time, we obtain

$$1 + h^2 \frac{\rho_\omega \alpha_0}{h^2 + c^* d_1 \Delta t} \leq \lambda \leq 1 + \frac{\rho_\omega \alpha_1}{1 + d_0 \gamma^2 \Delta t}.$$

This time, $\lambda = 1$ is an eigenvalue of the larger matrix in (5.2), the eigenspace being generated by vectors of the form $(u, 0)^T$, $u \in \mathbf{R}^N$. The other eigenvalues are the same as those determined in the first part of the proof. For each eigenvector v of (5.1), $\begin{pmatrix} {D^{(k)}}^{-1} M v \\ v \end{pmatrix}$ is an eigenvector of (5.2). □

This result shows that the eigenvalues of the preconditioned Jacobian matrix are bounded independently of h, and cluster near 1 as h goes to zero. However, this is not sufficient to show that the rate of convergence is independent of the mesh parameter. Indeed a beautiful result by Greenbaum et al. [14] states that one can prescribe both the eigenvalues and the sequence of residuals, and there exists a matrix with the given eigenvalues such that GMRES applied to this matrix will converge with the given residuals. On the other hand, it has been pointed out by Wathen [32, sec. 7] (quoting a previous result of Taussky) that because the eigenvalues of $\tilde{J}_k$ are all real and simple, this matrix must be self-adjoint, albeit for a *non-standard* inner product.

We summarize some well-known results concerning the convergence of GMRES. The most general result states that, for a given system of linear equations $Ax = b$, with $A \in \mathbb{C}^{n\times n}$, and $x, b \in \mathbb{C}^n$, the residuals induced by the GMRES iterates, $r^{(l)} = b - Ax^{(l)}$ satisfy the minimum residual property

$$\|r^{(l)}\|_2 \leq \min_{p\in\mathcal{P}_l^*} \|p(A)\|_2 \|r^{(0)}\|_2,$$

where $\mathcal{P}_l^* = \{p \in \mathcal{P}_l \,:\, p(0) = 1\}$, $\mathcal{P}_l$ is the set of polynomials of degree l or less.

The first convergence bound suggested for GMRES predicts convergence at a rate determined by $\Lambda(A)$, the set of eigenvalues of A. If A is normal, $\Lambda(A)$ determines convergence. This is not the case for non normal matrices. Assuming that A is diagonalizable, $A = V\Lambda V^{-1}$, we have

$$\frac{\|r^{(l)}\|_2}{\|r^{(0)}\|_2} \leq \kappa(V) \min_{p\in\mathcal{P}_l^*} \max_{\lambda\in\Lambda(A)} |p(\lambda)|, \tag{5.3}$$

where, $\kappa(V) = \|V\|_2\|V^{-1}\|_2$ is the 2-norm condition number of the eigenvector matrix V.

One approach avoiding this difficulty is due to Trefethen [24, 26], who has derived residual bounds based on pseudospectra of the matrix A. For a positive number, ϵ, the associated ϵ-pseudospectrum of A is the set in the complex plane defined by $\Lambda_\epsilon(A) = \{z \,:\, \|(zId - A)^{-1}\| \geq 1/\epsilon\}$. This set contains the spectrum of A. This results in the bound

$$\|p_l(A)\|_2 \leq \frac{L(\Gamma_\epsilon)}{2\pi\epsilon} \|p_l\|_{\Gamma_\epsilon},$$

where Γ_ϵ the boundary of $\Lambda_\epsilon(A)$ and $L(\Gamma_\epsilon)$ the length of the curve Γ_ϵ, which implies the bound

$$\frac{\|r^{(l)}\|_2}{\|r^{(0)}\|_2} \leq \frac{L(\Gamma_\epsilon)}{2\pi\epsilon} \min_{p\in\mathcal{P}_l^*} \max_{\lambda\in\Gamma_\epsilon} |p(\lambda)|$$

for the residual reduction.

Pseudospectra can sometimes result in much more realistic bounds than (5.3) but are expensive to compute. Moreover, it is not always clear which value of ϵ leads to the most useful information. In this paper, we consider another set associated with the matrix $\tilde{J}_k$ for predicting the convergence rate

of minimum residual methods, namely the field of values of $\tilde{J}_k$

$$W(\tilde{J}_k) \equiv \left\{ \frac{x^* \tilde{J}_k x}{x^* x} | x \in \mathbb{C}^n, x \neq 0 \right\},$$

sometimes also called its numerical range. The field of values of a matrix is known to be a convex and compact set in the complex plane that contains the eigenvalues (see [21]).

Bounds for GMRES convergence have recently been developed, starting with work by Eiermann [10], and an older bound from [11] can also be interpreted in terms of the distance of the field of values to the origin, see [6, 13].

The bound can be stated in terms of the angle $\beta \in [0, \pi/2)$:

$$\|r_k\| \leq \sin^k(\beta)\|r_0\|, \qquad \text{where } \cos\beta = \operatorname{dist}(0, W(\tilde{J}_k))/\|\tilde{J}_k\|, \tag{5.4}$$

provided $0 \notin W(\tilde{J}_k)$, which unfortunately we've been unable to prove.

6 Numerical experiments

In this section, we first perform a number of studies concerning the three different preconditioners for Lipschitz and non-Lipschitz isotherm cases. Then, we look at the dependence of the field of values with respect to the mesh size. We finish this section by some numerical simulations comparing the fixed-point and Newton methods. The velocity β and the diffusion tensor $D > 0$ are assumed to be constants. For the all numerical tests (except 2D-example), the domain $\Omega =]0, L[$ with $(L = 5)$, the mesh size is $h = 0.05$, $\rho = 1$, $\beta = 1.$, $D = 0.05$ and the initial condition is $c_0(x) = 0$ for $0 < x < L$, and the boundary conditions are $c(0, t) = 1$ and a zero diffusive flux at $x = L$.

In the tables below, we denote by BJ the Block Jacobi preconditioner, BGS the Block Gauss–Seidel preconditioner, NNI the average number (rounded to integer) of nonlinear iterations per time-step, NLI the average number (rounded to integer) of linear iterations per time-step and RT, the run time in seconds.

6.1 Preconditioner comparison

We consider a 1-dimensional model for single-species nonlinear adsorption with a Langmuir isotherm, cf (1.2). For numerical runs, we choose $T = 0.5$, the porosity $\omega = 0.1$, the time-step is $\Delta t = 0.0135$.

Our first study looks at the effects of changing the value of the density σ by setting K_L constant then investigating preconditioner responses to increasing the value of σ.

Tables 1–2 represent the average over time of the number of nonlinear and linear iterations with respect to σ for the three formulations. We conclude from these tables that the NLI increases when we increase the value of σ. This is consistent with Propositions 2 and 3, as increasing σ has the effect of increasing both α_0 and α_1 in Assumption (A3), and this in turn increases both σ_0 and σ_1 in Proposition 3. The number of nonlinear iterations in the $c\bar{c}$-formulation case also increases with σ. Table 1 shows also the good performance of the Block of Gauss–Seidel preconditioner with respect to the other preconditioners.

Table 1. Linear and nonlinear iterations with respect to σ (Linear preconditioning).

Preconditioner (PC)	σ	NNI	NLI	RT	σ	NNI	NLI	RT
None	0.025	7	145	16.07	0.5	15	175	21.21
BJ	0.025	5	11	2.92	0.5	12	35	7.33
BGS	0.025	4	5	1.72	0.5	13	20	5.87
None	1.	24	255	31.61	2.5	36	415	52.01
BJ	1.	21	78	14.21	2.5	33	173	29.12
BGS	1.	22	47	11.22	2.5	35	104	22.22

Table 2. Linear and nonlinear iterations with respect to σ (Nonlinear preconditioning).

σ	0.025	0.5	2.5	5.	10.	20.	40.
NNI	4	5	6	6	7	7	9
NLI	5	11	25	32	56	75	127

In our second study, we choose $K_L = 1$ and $\sigma = 1.5$, the other parameters are the same as the first study. Then we look at the effects of mesh size on the convergence rate of the preconditioned linear system.

Table 3. Average over time of the number of nonlinear and linear iterations for the three preconditioners (Exact Newton method).

Mesh/PC	h		$h/2$		$h/4$		$h/8$		$h/16$	
	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI
None	3	104	3	167	3	275	3	453	—	—
BJ	3	68	3	67	3	63	3	60	3	62
BGS	3	48	3	48	3	47	3	45	3	44
Elimi. of c_h	3	41	3	41	3	41	3	40	3	40

Tables 3–4 represent the average over time of the number of nonlinear and linear iterations with respect to the mesh size. These tables show that the number of nonlinear iterations does not increase when the mesh is refined. The number of linear iterations for the unpreconditioned method increases, whereas it remains stable for the two linear precondtioners as well as for the non-linear elimination method, as predicted in Propositions 2 and 3. The tables show also a good performance of the nonlinear preconditioner.

6.2 An example with non-Lipschitz isotherm

In this section, we discuss the case of non-Lipschitz sorption. We restrict the discussion to the case of a Freundlich isotherm: $\psi(c) = c^\alpha$, $\alpha \in (0, 1]$ ($K_F = 1$). The case $\alpha = 1$ can be included in the previous section, therefore we consider

Table 4. Average over time of the number of nonlinear and linear iterations for the three preconditioners (Inexact Newton method).

Mesh/PC	h		$h/2$		$h/4$		$h/8$		$h/16$	
	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI
None	8	42	8	76	10	105	10	177	—	—
BJ	7	27	7	27	7	26	7	26	7	26
BGS	8	23	7	24	7	22	8	25	8	24
Elimi. of c_h	5	15	5	15	5	15	5	15	5	15

here $\alpha \in (0,1)$. Clearly, the derivative is singular at $c = 0$, so ψ is not Lipschitz. In this case a regularization step is needed to use fixed point or Newton method. For a given $\epsilon > 0$ we define

$$\psi_\epsilon(c) = \begin{cases} \psi(c), & \text{if } c \notin [0,\epsilon], \\ \alpha\epsilon^{\alpha-1}c + (1-\alpha)\epsilon^\alpha, & \text{if } c \in [0,\epsilon], \end{cases}$$

and we recall the following lemma:

Lemma 4. *The regularized sorption isotherm is non-decreasing. Further, $\psi_\epsilon(.)$ and $\psi'_\epsilon(.)$ are Lipschitz continuous on $[0,\infty)$ with the Lipschitz constants $L_{\psi_\epsilon(.)} = \alpha\epsilon^{\alpha-1}$, respectively $L_{\psi'_\epsilon(.)} = \alpha(\alpha-1)\epsilon^{\alpha-2}$. Finally, we have*

$$0 \le \psi_\epsilon(c) - \psi(c) \le (1-\alpha)\epsilon^\alpha,$$

if $c \in (0,\epsilon)$, whereas $\psi(c) = \psi_\epsilon(c)$ whenever $c \notin (0,\epsilon)$.

Proof. See [22, lem 3.1]. □

As a first study, we choose $T = 2.$, the porosity $\omega = 1.$ and the time-step is $\Delta t = 0.1$. Then, we look at the dependence of the convergence rate of the preconditioned system with respect to the parameter ϵ for different values of α.

Table 5. The average over time of the number of nonlinear and linear iterations for the three preconditioners (Inexact Newton method) with $\alpha = 0.8$.

Epsilon/PC	0.5		0.1		0.05		0.01		0.001	
	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI
None	5	123	5	107	5	102	5	89	5	74
BJ	5	37	5	42	5	45	5	53	5	69
BGS	5	18	5	21	5	22	5	26	5	34
Elimi. of c_h	6	20	6	25	6	28	6	36	6	46

Table 6. The average over time of the number of nonlinear and linear iterations for the three preconditioners (Inexact Newton method) with $\alpha = 0.5$.

Epsilon/PC	0.5		0.1		0.05		0.01		0.001	
	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI	NNI	NLI
BJ	5	28	5	38	5	43	5	62	5	108
BGS	5	14	5	19	5	22	5	31	5	58
Elimi. of c_h	5	14	5	20	6	26	6	40	6	60

Tables 5 and 6 show that the number of the linear and the nonlinear iterations increase when the parameter ϵ tends to zero for the three preconditioners. The tables show also a good performance of BGS preconditioner.

As a second study, we are interested in the convergence of the fixed point method according to the inequality (3.2). We choose $\alpha = 0.8$, $\omega = 0.8$, $\epsilon = 0.5$ and the nonlinear tolerance is set to 10^{-12}. In this case $\rho_\omega L_{\psi_\epsilon} < 1$ with $L_{\psi_\epsilon(.)} = \alpha\epsilon^{\alpha-1}$. Table 7 represents the number of non-linear iterations with respect to the time-step. As predicted by the analytical result (Proposition 1), Table 7 shows the convergence of fixed point method without any restriction on Δt, albeit the convergence deteriorates when Δt is reduced.

Table 7. Nonlinear iterations with respect to Δt (fixed point method).

Δt	0.5	0.4	0.3	0.2	0.1	0.05	0.01
NNI	45	52	62	80	125	192	296

Table 8. Nonlinear iterations with respect to Δt (fixed point method).

Δt	0.5	0.4	0.3	0.2	0.15	0.1	0.09
NNI	64	77	100	143	—	—	—

Now, we choose $\alpha = 0.8$, $\omega = 0.5$ and $\epsilon = 0.1$, then $\rho_\omega L_{\psi_\epsilon} > 1$. The maximum number of the fixed point iterations is set to 1000.

Table 8 shows that the fixed point method converges only for Δt large enough, again confirming the condition from Proposition 1.

6.3 Field of values

In this section, we consider the same problem as in the last section, we are in particular interested in the effect of the mesh on the numerical radius $\mu(A) := \max\{|\xi| : \xi \in W(A)\}$ a measure for the size of $W(A)$. Figure 1 represents the eigenvalues, isolines of the pseudospectra and the field of values for the

Jacobian matrix $\tilde{J}_1$ with different mesh size. The figure was generated with the help of the EigTool software[1] [25].

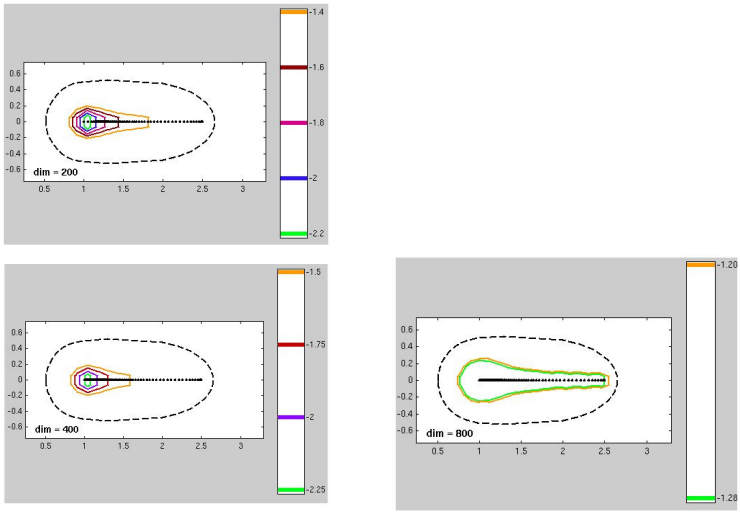

Figure 1. Pseudospectra, eigenvalues and field of values for different mesh resolutions. Top figure: 200 grid cells, bottom figures 400 and 800 grid cells. On each figure, the eigenvalues are the dots on the real axis, the colored lines are iso-lines of the pseudospectra and the black line is the boundary of the field of values.

These figures show that the eigenvalues are bounded independently of the mesh size, as was proved in Proposition 3, and that the field of values of the Jacobian matrix J_1 does not change, and stays well away from the origin, when the mesh is refined. This is an indication that the right hand side of inequality (5.4) is bounded independently of the mesh, as consequence of this the convergence rate of the GMRES method applied to the linearized system after elimination of the unknown c_h is also independent of the mesh. We emphasize that this is a numerical observation, and that we currently have no theoretical result that proves this observation.

6.4 Comparison of fixed point and Newton: a 2D example

In this section, we consider the geometry of the 2-dimensional benchmark MoMas problem (see [8] where a full statement of the flow and transport problems is described, including the boundary and initial conditions), but we keep the idealized chemistry studied in this paper, with Langmuir adsorption. The domain Ω is the benchmark geometry (see Figure 2), we choose $T = 100$, the time-step $\Delta t = 1.$, the velocity β is obtained by solving the incompressible Darcy flow problem.

The domain Ω is heterogeneous and is comprised of two media. The porosity

[1] Thomas G. Wright. EigTool. `http://www.comlab.ox.ac.uk/pseudospectra/eigtool/`, 2002.

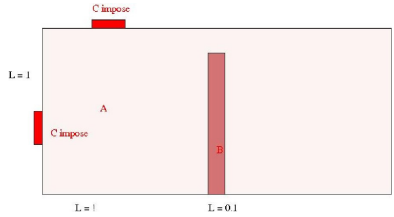

Figure 2. Geometry of benchmark MoMaS.

and the transverse and longitudinal dispersion coefficients are respectively given by: Medium A : $\omega = 0.25$, $\alpha_L = 10^{-2}$, $\alpha_T = 10^{-3}$; Medium B : $\omega = 0.5$, $\alpha_L = 6.10^{-2}$, $\alpha_T = 6.10^{-3}$. We recall that the Scheidegger dispersion model is [5]

$$D_{ij} = \left(\alpha_T \delta_{ij} + (\alpha_L - \alpha_T)\beta_i\beta_j/\|\beta\|^2\right)\|\beta\|.$$

The dispersion coefficients α_L and α_T have dimensions of length, but note that all data in the benchmark are non-dimensional. We choose $\sigma = 1.$, $K_L = 0.25$, and we fix the residual tolerance at 10^{-12}. We first study the convergence of the Newton algorithm and we compare it to the Fixed Point algorithm at the fourth time iteration (see Figure 3).

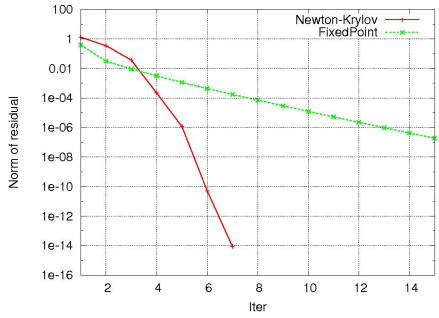

Figure 3. Convergence of Newton–Krylov and Fixed Point algorithms.

This figure shows a linear (resp. quadratic) convergence for Fixed Point (resp. Newton–Krylov) algorithms, which corresponds to the theoretical one, the total number of linear iteration for Newton–Krylov algorithm is 29 iterations.

7 Conclusions and perspectives

In this paper, we have introduced and analyzed different preconditioners for the linearized two species reactive transport equation. Specifically, we have focused on the dependence of the GMRES convergence rate with respect to the discretization parameter. We have proven that the eigenvalues of the preconditioned Jacobian matrix are bounded independently of the mesh size, this is confirmed by numerical experiments that also show a good performance of the

nonlinear elimination formulation. As the preconditioned Jacobian matrix is non symmetric, we have not been able to prove that the GMRES convergence rate is independent of mesh size. It has been observed numerically that the number of linear iterations is bounded independently of the mesh size. We have also observed numerically that the field of values of the Jacobian matrix remains stable, and away from the origin, when we refine the mesh, so the convergence rate of the GMRES method applied to the linearized system is independent of the mesh. We note that this study has already been generalized to a multi-components reactive transport system (see [2]). Our aim in a future work is to study theoretically the independence of the GMRES convergence rate applied to the linearized preconditioned system with respect to the mesh size.

Acknowledgements

We thank the two anonymous reviewers whose thorough reading and extensive comments hugely contributed to improve this paper. In particular, the idea of using the fixed point method in Proposition 1 to prove the existence of the solution is due to one of the reviewers.

References

[1] L. Amir and M. Kern. A global method for coupling transport with chemistry in heterogeneous porous media. *Computational Geosciences*, **14**:465–481, 2010. https://doi.org/10.1007/s10596-009-9162-x.

[2] L. Amir and M. Kern. Preconditioning a coupled model for reactive transport in porous media. *Int. J. Numer. Anal. Model.*, **16**(1):18–48, 2019. ISSN 1705-5105.

[3] J.W. Barrett and P. Knabner. Finite element approximation of the transport of reactive solutes in porous media. I. Error estimates for nonequilibrium adsorption processes. *SIAM J. Numer. Anal.*, **34**(1):201–227, 1997. ISSN 0036-1429. https://doi.org/10.1137/S0036142993249024.

[4] J.W. Barrett and P. Knabner. Finite element approximation of the transport of reactive solutes in porous media. II. Error estimates for equilibrium adsorption processes. *SIAM J. Numer. Anal.*, **34**(2):455–479, 1997. ISSN 0036-1429. https://doi.org/10.1137/S0036142993258191.

[5] J. Bear and A.H.-D. Cheng. *Modeling Groundwater Flow and Contaminant Transport*. Number 23 in Theory and Applications of Transport in Porous Media. Springer, New-York, 2010. https://doi.org/10.1007/978-1-4020-6682-5.

[6] B. Beckermann, S.A. Goreinov and E.E. Tyrtyshnikov. Some remarks on the Elman estimate for GMRES. *SIAM J. Matrix Anal. Appl.*, **27**(3):772–778, 2005. ISSN 0895-4798. https://doi.org/10.1137/040618849.

[7] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991. https://doi.org/10.1007/978-1-4612-3172-1.

[8] J. Carrayrou, M. Kern and P. Knabner. Reactive transport benchmark of MoMaS. *Computational Geosciences*, **14**:385–392, 2010. https://doi.org/10.1007/s10596-009-9157-7.

[9] C.N. Dawson. Godunov-mixed methods for advective flow problems in one space dimension. *SIAM Journal on Numerical Analysis*, **28**(5):1282–1309, 1991. https://doi.org/10.1137/0728068.

[10] M. Eiermann. Field of values and iterative methods. *Lin. Alg. Applic.*, **180**, 1993. https://doi.org/10.1016/0024-3795(93)90530-2.

[11] S.C. Eisenstat, H.C. Elman and M.H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, **20**(2):345–357, 1983. ISSN 0036-1429. https://doi.org/10.1137/0720023.

[12] S.C. Eisenstat and H.F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*, **17**(1):16–32, 1996. https://doi.org/10.1137/0917003.

[13] A. Greenbaum. *Iterative methods for solving linear systems*, volume 17 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. ISBN 0-89871-396-X. https://doi.org/10.1137/1.9781611970937.

[14] A. Greenbaum, V. Pták and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, **17**(3):465–469, 1996. ISSN 0895-4798. https://doi.org/10.1137/S0895479894275030.

[15] G.E. Hammond, A.J. Valocchi and P.C. Lichtner. Application of Jacobian-free Newton–Krylov with physics-based preconditioning to biogeochemical transport. *Adv. Water Resour.*, **28**:359–376, 2005. https://doi.org/10.1016/j.advwatres.2004.12.001.

[16] J. Kacur, B. Malengier and M. Remesikova. Solution of contaminant transport with equilibrium and non-equilibrium adsorption. *Computer Methods in Applied Mechancs and Engineering*, **194**(2-5):479–489, 2005. ISSN 0045-7825. https://doi.org/10.1016/j.cma.2004.05.017.

[17] C.T. Kelley. *Iterative methods for linear and nonlinear equations*. Society for Industrial and Applied Mathematics, 1995. https://doi.org/10.1137/1.9781611970944.

[18] D.A. Knoll and D.E. Keyes. Application of Jacobian-free Newton-Krylov with physics-based preconditioning to biogeochemical transport. *Journal of Computational Physics*, **28**:359–376, 2005. https://doi.org/10.1016/j.advwatres.2004.12.001.

[19] J.D. Logan. *Transport Modeling in Hydrogeochemical Systems*. Springer-Verlag, 2001. https://doi.org/10.1007/978-1-4757-3518-5.

[20] T.P.A. Mathew. *Domain Decomposition Methods for the Numerical Solution of Partial Differential Equations*. Springer-Verlag, 2008. https://doi.org/10.1007/978-3-540-77209-5.

[21] R.A. and Ch.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991. ISBN 0-521-30587-X. https://doi.org/10.1017/CBO9780511840371.

[22] F.A. Radu and I.S. Pop. Newton method for reactive solute transport with equilibrium sorption in porous media. *Journal of Computational and Applied Mathematics*, **234**(7):2118–2127, 2010. https://doi.org/10.1016/j.cam.2009.08.070.

[23] P. Siegel, R. Mosé, Ph. Ackerer and J. Jaffré. Solution of the advection–diffusion equation using a combination of discontinuous and mixed finite elements. *International Journal for Numerical Methods in Fluids*, **24**(6):595–613, 1997. https://doi.org/10.1002/(SICI)1097-0363(19970330)24:6<595::AID-FLD512>3.0.CO;2-I.

[24] L.N. Trefethen. Pseudospectra of matrices. In D.F. Griffiths and G.A. Watson(Eds.), *Numerical Analysis 1991*, pp. 234–266, Dundee, 1991.

[25] L.N. Trefethen. Computation of pseudospectra. *Acta Numerica*, **8**:247–295, 1999. https://doi.org/10.1017/S0962492900002932.

[26] L.N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, 2005.

[27] A.J. Valocchi, R.L. Street and P.V. Roberts. Transport of ion-exchanging solutes in groundwater: Chromatographic theory and field simulation. *Water Resources Research*, **17**(5):1517–1527, 1981. ISSN 1944-7973. https://doi.org/10.1029/WR017i005p01517.

[28] C.J. van Duijn and P. Knabner. Solute transport in porous media with equilibrium and nonequilibrium multiple-site adsorption: travelling waves. *J. Reine Angew. Math.*, **415**:1–49, 1991. ISSN 0075-4102. https://doi.org/10.1515/crll.1991.415.1.

[29] C.J. van Duijn and P. Knabner. Travelling waves in the transport of reactive solutes through porous media: Adsorption and binary ion exchange – Part 1. *Transport in Porous Media*, **8**(2):167–194, Jun 1992. ISSN 1573-1634. https://doi.org/10.1007/BF00617116.

[30] C.J. van Duijn and P. Knabner. Travelling waves in the transport of reactive solutes through porous media: Adsorption and binary ion exchange – Part 2. *Transport in Porous Media*, **8**(3):199–225, Jul 1992. ISSN 1573-1634. https://doi.org/10.1007/BF00618542.

[31] M. van Veldhuizen, J.A. Hendriks and C.A.J. Appelo. Numerical computation in heterovalent chromatography. *Applied Numerical Mathematics*, **28**(1):69–89, 1998. https://doi.org/10.1016/S0168-9274(98)00016-6.

[32] A.J. Wathen. Preconditioning. *Acta Numer.*, **24**:329–376, 2015. ISSN 0962-4929. https://doi.org/10.1017/S0962492915000021.