Journal of Computational and Applied Mathematics 366 (2020) 112338

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Numerical analysis of one dimensional motion of magma without mass forces

Miglena N. Koleva [a,*], Lubin G. Vulkov [b]

[a] Department of Mathematics, Ruse University, 8 Studentska St., 7017 Ruse, Bulgaria
[b] Department of Applied Mathematics and Statistics, Ruse University, 8 Studentska St., 7017 Ruse, Bulgaria

ARTICLE INFO

Article history:
Received 29 December 2018
Received in revised form 21 May 2019

Keywords:
Darcy flow magma model
Degenerate parabolic equation
Finite difference method
Implicit–explicit scheme
Iterative implicit scheme
Positivity
Conservation
Convergence

ABSTRACT

In this paper, we present numerical treatment for a compaction-driven Darcian flow viscoelastic rock magma model. This problem is a strongly coupled system of one quasi-linear parabolic equation and one integro-differential equation for the density and the porosity of the flow. The numerical discretization uses cell-centered finite difference method, combined with semi-implicit and implicit time stepping. Implicit–explicit schemes, as well as implicit–explicit iterative algorithms have been developed to solve the corresponding discrete problems. Some properties (positivity, boundness, conservation) of the numerical solutions are investigated. Convergence study of the iteration processes is also presented. The efficiency and the accuracy of the proposed methods are illustrated numerically by test examples with near-real data.

1. Introduction

We consider a two-phase model for isothermal motion of magma in porous rock. This process is described by the laws of conservation of mass for each phase, Darcy's law for the fluid phase (taking into account the motion of a solid skeleton), the rheological law and the equation of the conservation of momentum for the system [1–3]. For posing the differential problem, we follow A. Papin and M. Tokareva [2]. The authors consider the following quasi-linear system

$$\frac{\partial(1-\phi)\rho_s}{\partial t}+\frac{\partial}{\partial x}((1-\phi)\rho_s v_s)=0, \tag{1}$$

$$\frac{\partial(\rho_f\phi)}{\partial t}+\frac{\partial}{\partial x}(\rho_f\phi v_f)=0, \tag{2}$$

$$\phi(v_f-v_s)=-k(\phi)\left(\frac{\partial p_f}{\partial x}-\rho_f g_m\right), \tag{3}$$

$$\frac{\partial v_s}{\partial x}=-\frac{1}{\xi(\phi)}p_e,\quad p_e=p_{tot}-p_f, \tag{4}$$

$$\frac{\partial p_{tot}}{\partial x}=-\rho_{tot}g_m,\qquad p_{tot}=\phi p_f+(1-\phi)p_s,\qquad p_{tot}=\phi\rho_f+(1-\phi)\rho_s, \tag{5}$$

* Corresponding author.
E-mail addresses: mkoleva@un-ruse.bg (M.N. Koleva), lvalkov@un-ruse.bg (L.G. Vulkov).

https://doi.org/10.1016/j.cam.2019.07.003
0377-0427/

with a solution, defined in the domain $(x,t) \in Q_T = \Omega \times (0,T)$, $\Omega = (0,1)$. To complete the model, the following boundary and initial conditions are imposed

$$v_s(0,t) = v_s(1,t) = 0, \quad v_f(0,t) = v_f(1,t) = 0, \tag{6}$$

$$\phi(x,0) = \phi^0(x), \quad \rho_f(x,0) = \rho^0(x). \tag{7}$$

The quantities in (1)–(5) have the following physical meaning: ρ_f, ρ_s, v_f, v_s are real density and velocity of fluid and solid phases, respectively; ϕ is the porosity, p_f and p_s are pressures of the fluid and solid phases; p_e is the efficiency pressure; p_{tot} is the total pressure; ρ_{tot} is the density of the two-phase medium; g_m is the density of the mass forces; $k(\phi)$ is the coefficient of filtration; $\xi(\phi)$ is the coefficient of the rock shear viscosity (specific function).

The problem (1)–(7) is written in Eulerian coordinates (x,t). The unknown functions are ϕ, ρ_f, v_f, v_s, p_f, p_s. The real density of the solid particles ρ_s is assumed to be a constant. In the system (1)–(5) we use the equation of state of the fluid $p_f = p(\rho_f)$. An often used relation in the applications is

$$\frac{dp_f}{d\rho_f} = \frac{1}{\beta_f \rho_f}, \tag{8}$$

or in the case of slightly compressible flow [4, p. 15]

$$\frac{dp_f}{d\rho_f} = \frac{1}{\beta_f \rho_f^0}, \tag{9}$$

where ρ_f is the fluid compressibility, β_f is the compressibility coefficient of the pore-fluid and ρ_f^0 is the density at the reference pressure p^0 [1,3–5].

In [2,6], the following dependencies of the functional parameters of the problem are used:

$$k(\phi) = \frac{\bar{k}}{\mu}\phi^n, \quad \xi(\phi) = v\phi^{-r}, \tag{10}$$

where $r \in [0,2]$, $n = 3$, v, μ, $\bar{k}$ are positive environmental settings [6].

Numerical methods for standard two-phase models have been subject of extensive research in the last decades. The problems are solved numerically by finite difference and finite element schemes [7], mixed finite element method [8,9], cell-centered finite difference method, based on lowest order Raviart–Thomas elements [10,11], finite volume based discretization [12–14].

Initial boundary-value problems for similar structured system of equations as (1)–(5) are numerically investigated by Crank–Nicolson and alternating direction implicit finite difference schemes [6] and finite element method [15].

Finite volume method for 2D Darcy fluid non-linear non-local reaction–advection–diffusion problem is developed in [16].

In all these papers the authors treat *directly* the corresponding model problems.

In this work we study the model system (1)–(7) without mass forces, i.e. $g_m = 0$ in Eqs. (3), (5). The formulation of such compact equations is mathematically simple, but the resulting system (see (12)–(16)) consists of one strongly non-linear parabolic equation and one non-linear integro-differential equation. Analytical solutions to this strongly coupled system are cumbersome even in particular cases. In this paper, we therefore present numerical solutions.

The rest of the paper is organized as follows. In the next section we formulate the transformed differential problem. In Section 3, we develop different finite difference discretizations of the model problem. In Section 4 we discuss the realization of the corresponding numerical schemes. Section 5 is devoted to the investigation of the properties of the numerical solutions. Illustrative numerical examples are given and discussed in Section 6. Finally, we wind up the paper with some concluding remarks.

2. The differential problem

In this section we follow [2,17], where the case $g_m = 0$ is studied and in view of (5), $p_{tot} = p^*(t)$. To be self-contained, we describe the derivation of the model, suggested in [17].

Suppose that $\bar{x} = \bar{x}(\tau, x, t)$ is a solution of the Cauchy problem

$$\frac{\partial x}{\partial \tau} = v_s(\bar{x}, \tau), \quad \bar{x}|_{\tau = t} = x.$$

We set $\widehat{x} = \bar{x}(0, x, t)$, take $\widehat{x}$ and t for new independent variables, taking into account that $1 - \phi(\hat{x}, t) = (1 - \phi^0(\hat{x}))\frac{\partial \hat{x}}{\partial x}(\hat{x}, t)$ and pass to mass Lagrangian variables (y, t) by the rule

$$(1 - \phi^0(\widehat{x}))d\widehat{x} = dy, \qquad y(\widehat{x}) = \int_0^{\widehat{x}} (1 - \phi^0(\eta))d\eta \in [0,1].$$

Now, preserving the notation x for the variable y and with essential use of the zero mass forces, the authors of [17] obtain from (1)–(5) the system

$$\frac{\partial(1-\phi)}{\partial\phi}+(1-\phi)^2\frac{\partial v_s}{\partial x}=0,\quad \frac{\partial}{\partial t}\left(\rho_f\frac{\phi}{1-\phi}\right)+\frac{\partial}{\partial x}\left(\rho_f\phi(v_f-v_s)\right)=0,$$

$$\phi(v_s-v_f)=k(\phi)(1-\phi)\frac{\partial p(\rho_f)}{\partial x},\quad (1-\phi)\frac{\partial v_s}{\partial x}=-a_1(\phi)p_e,\quad p_e=p^*(t)-p(\rho_f).$$

Next, we introduce dimensionless variables

$$\begin{aligned}&t'=\frac{t}{t_1},\quad x'=\frac{x}{L},\quad v_s'=\frac{v_s}{v_1},\quad v_f'=\frac{v_f}{v_1},\quad \rho_f'=\frac{\rho_f}{\rho_s},\\ &p_f'=\frac{p_f}{p_1},\quad p_s'=\frac{p_s}{p_1},\quad p_e'=\frac{p_e}{p_1},\quad p_{tot}'=\frac{p_{tot}}{p_1},\quad a_1'(\phi)=\frac{a_1(\phi)}{a^0},\quad k'(\phi)=\frac{k(\phi)}{k_1},\end{aligned}\tag{11}$$

where $L=\int_0^1\left(1-\phi^0(\eta)\right)d\eta$, $t_1=\dfrac{L}{v_1}$, $a^0=\dfrac{v_1}{Lp_1}$, $k_1=\dfrac{v_1L}{p_1}$, v_1, p_1 are fixed positive constants, having dimension of velocity and pressure, accordingly.

Further, taking into account that the domain of x' is the interval [0,1] and omit the dashed notation, we derive the following dimensionless parabolic-ordinary differential equations system for finding functions ρ_f and ϕ:

$$\frac{\partial}{\partial t}\left(a(\phi)\rho_f\right)-\frac{\partial}{\partial x}\left(K(\phi)b(\rho_f)\frac{\partial\rho_f}{\partial x}\right)=0,\tag{12}$$

$$\frac{\partial G(\phi)}{\partial t}=p(\rho_f)-p^*(t),\qquad p^*(t)=\int_0^1\frac{a_1(\phi)}{1-\phi}p_f dx\left[\int_0^1\frac{a_1(\phi)}{1-\phi}dx\right]^{-1}\equiv P^*(\phi,\rho_f).\tag{13}$$

Here, $a_1(\phi)=1/\xi(\phi)$, the function $G(\phi)$ is defined by the equation

$$\frac{dG(\phi)}{d\phi}=\frac{1}{(1-\phi)a_1(\phi)}\tag{14}$$

and

$$a(\phi)=\frac{\phi}{1-\phi},\qquad K(\phi)=k(\phi)(1-\phi),\qquad b(\rho_f)=\rho_f\frac{dp(\rho_f)}{d\rho_f}.\tag{15}$$

The system (12)–(15) is subjected to the initial conditions (7) and boundary conditions

$$\frac{\partial\rho_f}{\partial x}(0,t)=0,\quad \frac{\partial\rho_f}{\partial x}(1,t)=0.\tag{16}$$

In [2] is proved a local solvability of the problem (12)–(16) in $\overline{Q}_{t_*}$ for $\phi^0\in C^{2+\alpha}(\overline{\Omega})$, $\rho^0\in C^{2+\alpha}(\overline{\Omega})$, i.e. there exist t_*, such that $(\phi(x,t),\rho_f(x,t))\in C^{2+\alpha,1+\alpha/2}(\overline{Q}_{t_*})$, $\alpha\in(0,1]$. Moreover, authors show that if

$$\frac{dp_f(\rho^0)}{dx}\Big|_{x=0}=\frac{dp_f(\rho^0)}{dx}\Big|_{x=1}=0,\quad 0<m_0\le\phi^0(x)\le M_0<1,\quad 0<m_1\le\rho^0(x)\le M_1<\infty,\quad x\in\overline{\Omega},$$

for given positive constants m_0, M_0, m_1 and M_1, then $0<\phi<1$ and $\rho_f>0$ for $(x,t)\in\overline{Q}_{t_*}$.

The problem (12)–(16) is challenging for a numerical investigation because of the several difficulties: non-linearity in the diffusion and evolution terms, spatial non-local nature of p^*, the degeneration of the coefficient functions in (12)–(14) and their derivatives.

The aim of the present work is to develop and analyze efficient numerical methods that preserve qualitative properties of the differential problem, for solving the integro-differential initial boundary value problem (7), (12)–(16).

3. Difference schemes approximations

In this section we propose implicit and implicit–explicit finite difference discretizations of (7), (12)–(16), treating the time derivative in (13) by two different ways.

In the space–time domain $[0,1]\times[0,T]$ we define a uniform mesh $\overline{w}_{h\tau}=\overline{w}_h\times\overline{w}_\tau$:

$$\overline{w}_h=\{x_i=ih,\quad i=0,1,\dots,N,\quad Nh=1\},$$

$$\overline{w}_\tau=\{t_j=t_{j-1}+\tau_j,\quad j=1,2,\dots,J,\quad t_0=0,\quad t_J=T\}.$$

The numerical solutions at grid nodes (x_i,t_j) are denoted by $\rho_i^j=\rho_f(x_i,t_j)$ and $\phi_i^j=\phi(x_i,t_j)$. Further, for a mesh functions y (defined on $\overline{w}_{h\tau}$) and the derivative of the function $G(\phi)$, we use the following notations

$$y_i:=y_i^j=y(x_i,t_j),\qquad \widehat{y}_i:=y_i^{j+1}=y(x_i,t_{j+1}),\qquad g(v)=\frac{dG(\phi)}{d\phi}\Big|_{\phi=v}.$$

Following [18], we approximate the system (12), (13) by two different finite difference schemes:
IMEX schemes (IMEX 1 and IMEX 2):

$$\begin{aligned}
&\frac{a(\widehat{\phi}_i)\widehat{\rho}_i - a(\phi_i)\rho_i}{\tau_j} = \frac{1}{h}\left[\mathcal{K}(\widehat{\phi}_{i+1})\mathcal{B}(\rho_{i+1})\frac{\widehat{\rho}_{i+1}-\widehat{\rho}_i}{h} - \mathcal{K}(\widehat{\phi}_i)\mathcal{B}(\rho_i)\frac{\widehat{\rho}_i-\widehat{\rho}_{i-1}}{h}\right], \quad i = 1, 2, \ldots, N-1,\\
&\frac{a(\widehat{\phi}_0)\widehat{\rho}_0 - a(\phi_0)\rho_0}{\tau_j} = \frac{2}{h}\left[\mathcal{K}(\widehat{\phi}_1)\mathcal{B}(\rho_1)\frac{\widehat{\rho}_1-\widehat{\rho}_0}{h}\right],\\
&\frac{a(\widehat{\phi}_N)\widehat{\rho}_N - a(\phi_N)\rho_N}{\tau_j} = -\frac{2}{h}\left[\mathcal{K}(\widehat{\phi}_N)\mathcal{B}(\rho_N)\frac{\widehat{\rho}_N-\widehat{\rho}_{N-1}}{h}\right],
\end{aligned} \tag{17}$$

$$\begin{aligned}
&\frac{G^*(\widehat{\phi}_i) - G^*(\phi_i)}{\tau_j} = p(\rho_i) - p^*(t_j), \quad i = 0, 1, \ldots, N,\\
&p^*(t_j) = P_h^*(\phi, \rho) := \sum_{i=0}^{N}\alpha_i\frac{a_1(\phi_i)}{1-\phi_i}p(\rho_i)\left(\sum_{i=0}^{N}\alpha_i\frac{a_1(\phi_i)}{1-\phi_i}\right)^{-1},
\end{aligned} \tag{18}$$

where

$$G^*(\widehat{\phi}_i) = \begin{cases} g(\phi_i)\widehat{\phi}_i, & \text{IMEX1},\\ G(\widehat{\phi}_i), & \text{IMEX2}, \end{cases} \qquad G^*(\phi_i) = \begin{cases} g(\phi_i)\phi_i, & \text{IMEX1},\\ G(\phi_i), & \text{IMEX2}, \end{cases} \qquad \alpha_i = \begin{cases} \frac{1}{2}, & \text{if } i = \{0, N\},\\ 1, & \text{otherwise}, \end{cases}$$

$$\mathcal{B}_i(v) = \frac{1}{2}\left(b(v_{i-1}) + b(v_i)\right), \qquad \mathcal{K}_i(v) = \frac{1}{2}\left(K(v_{i-1}) + K(v_i)\right).$$

Implicit schemes (IS 1 and IS 2):

$$\begin{aligned}
&\frac{a(\widehat{\phi}_i)\widehat{\rho}_i - a(\phi_i)\rho_i}{\tau_j} = \frac{1}{h}\left[\mathcal{K}(\widehat{\phi}_{i+1})\mathcal{B}(\widehat{\rho}_{i+1})\frac{\widehat{\rho}_{i+1}-\widehat{\rho}_i}{h} - \mathcal{K}(\widehat{\phi}_i)\mathcal{B}(\widehat{\rho}_i)\frac{\widehat{\rho}_i-\widehat{\rho}_{i-1}}{h}\right], \quad i = 1, 2, \ldots, N-1,\\
&\frac{a(\widehat{\phi}_0)\widehat{\rho}_0 - a(\phi_0)\rho_0}{\tau_j} = \frac{2}{h}\left[\mathcal{K}(\widehat{\phi}_1)\mathcal{B}(\widehat{\rho}_1)\frac{\widehat{\rho}_1-\widehat{\rho}_0}{h}\right],\\
&\frac{a(\widehat{\phi}_N)\widehat{\rho}_N - a(\phi_N)\rho_N}{\tau_j} = -\frac{2}{h}\left[\mathcal{K}(\widehat{\phi}_N)\mathcal{B}(\widehat{\rho}_N)\frac{\widehat{\rho}_N-\widehat{\rho}_{N-1}}{h}\right],
\end{aligned} \tag{19}$$

$$\frac{G^*(\widehat{\phi}_i) - G^*(\phi_i)}{\tau_j} = p(\widehat{\rho}_i) - p^*(t_{j+1}), \quad p^*(t_{j+1}) = P_h^*(\widehat{\phi}, \widehat{\rho}), \quad i = 0, 1, \ldots, N, \tag{20}$$

where

$$G^*(\widehat{\phi}_i) = \begin{cases} g(\widehat{\phi}_i)\widehat{\phi}_i, & \text{IS1},\\ G(\widehat{\phi}_i), & \text{IS2}, \end{cases} \qquad G^*(\phi_i) = \begin{cases} g(\widehat{\phi}_i)\phi_i, & \text{IS1},\\ G(\phi_i), & \text{IS2}. \end{cases}$$

4. Realization of the schemes

Our aim is to implement the numerical discretizations in more efficient way, such that to save a computational time.

4.1. IMEX schemes

A natural way for the realization of the scheme (17), (18) is first to solve the (18) in order to find $\widehat{\phi} = \left(\widehat{\phi}_0, \widehat{\phi}_1, \ldots, \widehat{\phi}_N\right)$ and then to compute $\widehat{\rho} = \left(\widehat{\rho}_0, \widehat{\rho}_1, \ldots, \widehat{\rho}_1\right)$ from (17), for already known $\widehat{\phi}$. Thus, in general, at each time level, instead of solving one systems of $2(N+1)$ algebraic equations, we solve two systems of $N+1$ equations.

Applying IMEX 1, from (18) we find $\widehat{\phi}$ explicitly.

Regarding to IMEX 2, we compute $\widehat{\phi}$ on two stages. First, from (18) we find $G(\widehat{\phi})$. Next, to compute $\widehat{\phi}$ at each space grid node, we solve the non-linear equation

$$G(\widehat{\phi}) = \int_{\phi^0}^{\widehat{\phi}} \frac{1}{(1-v)a_1(v)}dv. \tag{21}$$

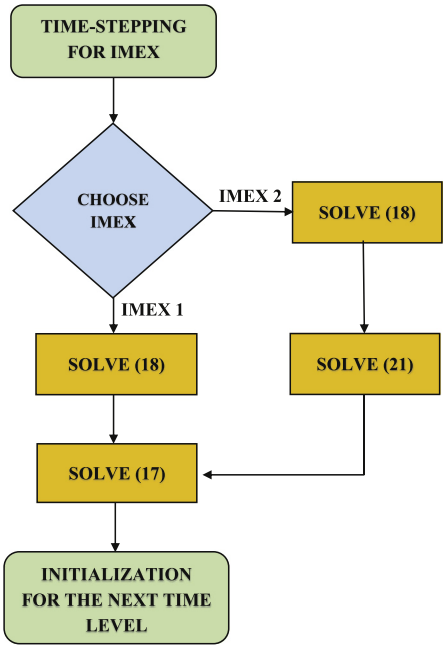

Fig. 1. Time-stepping with IMEX schemes for solving (12)–(16).

For example, in view of (10), (14), for some particular values of r, we have

$$G(\widehat{\phi}) = \nu \begin{cases} \ln\left(\widehat{\phi}(1-\phi^0)(\phi^0)^{-1}(1-\widehat{\phi})^{-1}\right) - (\widehat{\phi})^{-1} + (\phi^0)^{-1}, & r = 2, \\ 2\,\mathrm{arctanh}(\widehat{\phi}^{1/2}) - 2\widehat{\phi}^{-1/2} - 2\,\mathrm{arctanh}\left((\phi^0)^{1/2}\right) + 2(\phi^0)^{-1/2}, & r = 1.5, \\ \ln\left(\widehat{\phi}(1-\phi^0)(\phi^0)^{-1}(1-\widehat{\phi})^{-1}\right), & r = 1, \\ 2\,\mathrm{arctanh}(\widehat{\phi}^{1/2}) - 2\,\mathrm{arctanh}\left((\phi^0)^{1/2}\right), & r = 0.5. \end{cases}$$

Similarly, for p defined by (8) we determine

$$p(\rho) = \beta_f^{-1} \ln \frac{\rho}{\rho_f^0} + p^0 \tag{22}$$

and in the case of (9), we get

$$p(\rho) = \beta_f^{-1} \left(\frac{\rho}{\rho_f^0} - 1 \right) + p^0. \tag{23}$$

To solve (21) we require iteration method (for example, Newton method), using as initial guess ϕ - the solution at old time level.

On Fig. 1 we illustrate the realization of the IMEX schemes.

4.2. Implicit scheme

To find the solution $(\widehat{\rho}_i, \widehat{\phi}_i)$ of the non-linear system of algebraic equation (19)–(20), generated from the fully implicit discretization, we need iterative methods. To this aim, we initiate Picard-like iterative process.

Iteration schemes (ItS 1 and ItS 2). Let $m = 0, 1, \ldots$ be the number of iteration and denote the mth approximation of the solutions by the vectors

$$\phi^{(m)} = \left(\phi_1^{(m)}, \phi_2^{(m)}, \ldots, \phi_N^{(m)}\right), \quad \rho^{(m)} = \left(\rho_1^{(m)}, \rho_2^{(m)}, \ldots, \rho_N^{(m)}\right).$$

For each $m = 0, 1, \ldots$ we perform the following two stages.

First, from the (20) we compute $\phi^{(m+1)}$:

$$\frac{G^*\left(\phi_i^{(m+1)}\right) - G^*(\phi_i)}{\tau_j} = p\left(\rho_i^{(m)}\right) - p^*(t_{j+1}), \quad p^*(t_{j+1}) = P_h^*(\phi^{(m)}, \rho^{(m)}), \quad i = 0, 1, \ldots, N, \tag{24}$$

where

$$G^*\left(\phi_i^{(m+1)}\right)=\begin{cases} g\left(\phi_i^{(m)}\right)\phi_i^{(m+1)}, & \text{ItS1},\\ G\left(\phi_i^{(m+1)}\right), & \text{ItS2},\end{cases}\qquad G^*(\phi_i)=\begin{cases} g\left(\phi_i^{(m)}\right)\phi_i, & \text{ItS1},\\ G(\phi_i), & \text{ItS2}.\end{cases}$$

Next, for already known $\phi^{(m+1)}$, from (20) we find $\rho^{(m+1)}$:

$$\begin{aligned}\frac{a\left(\phi_i^{(m+1)}\right)\rho_i^{(m+1)}-a(\phi_i)\rho_i}{\tau_j}=&\frac{1}{h}\left[\mathcal{K}\left(\phi_i^{(m+1)}\right)\mathcal{B}\left(\rho_{i+1}^{(m)}\right)\frac{\rho_{i+1}^{(m+1)}-\rho_i^{(m+1)}}{h}\right.\\ &\left.-\mathcal{K}\left(\phi_i^{(m+1)}\right)\mathcal{B}\left(\rho_i^{(m)}\right)\frac{\rho_{i+1}^{(m+1)}-\rho_i^{(m+1)}}{h}\right],\quad i=1,\dots,N-1,\\ \frac{a\left(\phi_0^{(m+1)}\right)\rho_0^{(m+1)}-a(\phi_0)\rho_0}{\tau_j}=&\mathcal{K}\left(\phi_0^{(m+1)}\right)\mathcal{B}\left(\rho_1^{(m)}\right)\frac{\rho_1^{(m+1)}-\rho_0^{(m+1)}}{h^2},\\ \frac{a\left(\phi_N^{(m+1)}\right)\rho_N^{(m+1)}-a(\phi_N)\rho_N}{\tau_j}=&-\mathcal{K}\left(\phi_N^{(m+1)}\right)\mathcal{B}\left(\rho_N^{(m)}\right)\frac{\rho_N^{(m+1)}-\rho_{N-1}^{(m+1)}}{h^2}.\end{aligned}\tag{25}$$

As initial guess we use the solution value (ρ_i, ϕ_i) at previous time level i.e. $\rho_i^{(0)}=\rho_i$, $\phi_i^{(0)}=\phi_i$, $i=0,1,\dots,N$.

Note that the scheme (25) is a linear with respect to $\rho^{(m+1)}$ and the coefficient matrix is a tridiagonal. To solve the system of algebraic equations (25) we apply modified Thomas method.

Let us rewrite Eqs. (25) in the form

$$-A_i^{(m)}\rho_{i-1}^{(m+1)}+C_i^{(m)}\rho_i^{(m+1)}-B_i^{(m)}\rho_{i+1}^{(m+1)}=F_i,\quad i=0,1,\dots,N,\tag{26}$$

where

$$\begin{aligned}&A_i^{(m)}=\mathcal{K}\left(\phi_i^{(m+1)}\right)\mathcal{B}\left(\rho_i^{(m)}\right)\frac{\tau_j}{h^2},\quad B_i^{(m)}=\mathcal{K}\left(\phi_{i+1}^{(m+1)}\right)\mathcal{B}\left(\rho_{i+1}^{(m)}\right)\frac{\tau_j}{h^2},\quad i=1,2,\dots,N-1,\\ &A_0=0,\quad B_0=2\mathcal{K}\left(\phi_1^{(m+1)}\right)\mathcal{B}\left(\rho^{(m)}{}_1^{(m)}\right)\frac{\tau_j}{h^2},\quad A_N=2\mathcal{K}\left(\phi_{N-1}^{(m+1)}\right)\mathcal{B}\left(\rho^{(m)}{}_{N-1}\right)\frac{\tau_j}{h^2},\quad B_N=0,\\ &C_i^{(m)}=A_i^{(m)}+B_i^{(m)}+a(\widehat{\phi}_i^{(m+1)}),\quad F_i=a(\phi_i)\rho_i,\quad i=0,1,\dots,N.\end{aligned}$$

By recurrent formulas we compute three sweep coefficients (right sweep) [18]:

Forward elimination

$$\begin{aligned}&\alpha_{i+1}^{(m)}=\frac{B_i^{(m)}}{C_i^{(m)}-\alpha_i^{(m)}A_i^{(m)}},\quad i=1,2,\dots,N-1,\quad \alpha_1^{(m)}=\frac{B_0^{(m)}}{C_0^{(m)}},\\ &\beta_{i+1}^{(m)}=\frac{A_i^{(m)}\beta_i^{(m)}+F_i^{(m)}}{C_i^{(m)}-\alpha_i^{(m)}A_i^{(m)}},\quad i=1,2,\dots,N-1,\quad \beta_1^{(m)}=\frac{F_0}{C_0^{(m)}}.\end{aligned}\tag{27}$$

Backward substitution

$$\rho_N^{(m+1)}=\frac{F_N+A_N^{(m)}\beta_N^{(m)}}{C_N^{(m)}-A_N^{(m)}\alpha_N^{(m)}},\quad \rho_i^{(m+1)}=\alpha_{i+1}^{(m)}\rho_{i+1}^{(m+1)}+\beta_{i+1}^{(m)},\quad i=N-1,\dots,1,0.\tag{28}$$

We measure the distance between the vectors ρ^s and ρ^l by the strong norm

$$d(\rho^s,\rho^l)=\|\rho^s,\rho^l\|=\max_{i=1,\dots,N}|\rho_i^s-\rho_i^l|.\tag{29}$$

The convergence of the process is controlled by the difference between to consecutive iterations

$$\|\rho^{(m+1)}-\rho^{(m)}\|<\varepsilon,\tag{30}$$

where $\varepsilon>0$ is a small constant. The iteration process continues until the inequality (30) is satisfied.

If the inequality (30) does not hold even for one solution component $\rho_l^{(m+1)}$, $l\in\{0,1,\dots,N\}$, the procedure (27), (28), (30), implies post computation of all $\rho_i^{(m+1)}$, $i=0,1,\dots,N$. When the number of such components $\rho_l{}^{(m+1)}$ is small, this post computation is practically non-useful.

Improvement. In order to save a computational time, we make the following. On the first iteration the computations are fully performed by formulas (27), (28). At the checking of the convergence with the inequality (30) we store all

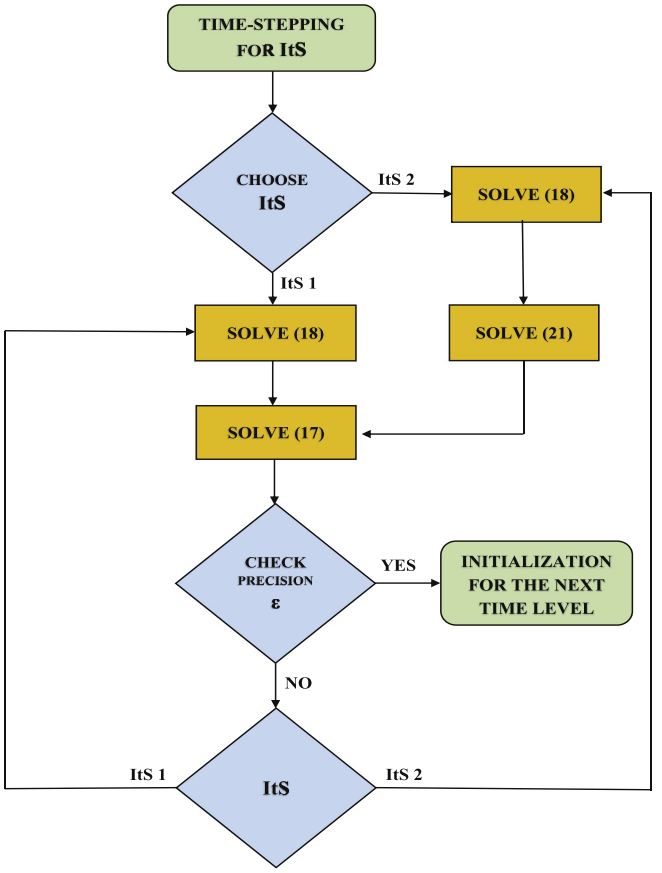

Fig. 2. Time-stepping with ItS schemes for solving (12)–(16).

numbers l for which this inequality fails. Then, on the next iteration only the coefficients $\alpha_{l+1}^{(m)}$, $\beta_{l+1}^{(m)}$ in (27) and the solution $\rho_l^{(m+1)}$ in (28) are post computed and the others are taken from the previous iteration.

On Fig. 2 we illustrate the realization of ItS 1 and ItS 2.

5. Properties of the numerical solution

In this section we discuss positivity, boundness and conservation properties of the numerical solution, obtained by the proposed numerical schemes. Also, we investigate convergence of the iteration process (24)–(25).

We observe conservation properties of the differential problem (12)–(16). Indeed, integrating the (12) over interval [0, 1] and taking into account the boundary conditions (16), we get

$$\int_0^1 a(\phi)\rho_f dx = \int_0^1 a(\phi^0)\rho_f^0 dx. \tag{31}$$

Further, integrating (13), first over interval $[0, t]$ and then over [0, 1], we derive

$$\int_0^1 \left(G(\phi) - H(\phi, \rho_f)\right) dx = \int_0^1 \left(G(\phi^0) - H(\phi^0, \rho_f^0)\right) dx, \tag{32}$$

where $H(\phi, \rho) = \mathcal{P}(\rho_f) - \mathcal{P}^*(\phi, \rho_f)$ and

$$\mathcal{P}(\rho_f) - \mathcal{P}(\rho_f^0) = \int_0^t p\Big(\rho_f(x,s)\Big)ds, \quad \mathcal{P}^*(\phi, \rho_f) - \mathcal{P}^*(\phi^0, \rho_f^0) = \int_0^t P^*\big(\phi(x,s), \rho_f(x,s)\big)ds.$$

Let us represent (13) in the form

$$\frac{d\phi}{dt} = [g(\phi)]^{-1}\left(p(\rho_f) - p^*(t)\right). \tag{33}$$

As before, integrating (33) over $[0, t]$ and then over $[0, 1]$, we obtain

$$\int_0^1 \left(\phi - \overline{H}(\phi, \rho_f)\right) dx = \int_0^1 \left(\phi^0 - \overline{H}(\phi^0, \rho_f^0)\right) dx, \tag{34}$$

where $\overline{H}(\phi, \rho) = \overline{\mathcal{P}}(\rho_f) - \overline{\mathcal{P}}^*(\phi, \rho_f)$ and

$$\overline{\mathcal{P}}(\rho_f) - \overline{\mathcal{P}}(\rho_f^0) = \int_0^t [g(\phi(x,s))]^{-1} p\Big(\rho_f(x,s)\Big) ds,$$

$$\overline{\mathcal{P}}^*(\phi, \rho_f) - \overline{\mathcal{P}}^*(\phi^0, \rho_f^0) = \int_0^t [g(\phi(x,s))]^{-1} P^*\big(\phi(x,s), \rho_f(x,s)\big) ds.$$

Further, we use the following notations

$$\|f(x,t)\|_\infty = \max_{(x,t)\in\Omega_T} |f(x,t)|, \quad v^+ = \max\{0, v\}, \quad v^- = \max\{0, -v\}, \quad v_{\max} = \max_{0\le i\le N} v_i, \quad v_{\min} = \min_{0\le i\le N} v_i.$$

and set the typical for the model assumptions:

$$\begin{gathered} \beta_f > 0; \quad \xi(\phi) > 0, \quad a_1(\phi) > 0 \ \text{ for } \ \phi > 0; \\ 0 < m_1 \le \rho^0(x) \le M_1 < \infty, \quad 0 < m_0 \le \phi^0(x) \le M_0 < 1 \ \text{ for } \ x \in [0,1] \ \text{ and} \\ p(\rho_f) > 0, \quad a(\phi) > 0, \quad K(\phi) \ge 0, \quad b(\rho) > 0, \quad g(\phi) > 0 \ \text{ for } \ \rho > 0, \quad 0 < \phi < 1. \end{gathered} \tag{35}$$

5.1. IMEX schemes (17), (18)

First, we establish that the implicit–explicit numerical schemes preserve the conservation properties (31), (32) or (34). Indeed, multiplying the first ($i = 0$) and the last ($i = N$) equation in (17) by $h/2$ and all other equations ($i = 1, 2, \dots, N-1$) by h, then summing up all resulting equations, we get

$$h\sum_{i=0}^{N} \alpha_i a(\widehat{\phi}_i)\widehat{\rho}_i = h\sum_{i=0}^{N} \alpha_i a(\phi_i^j)\rho_i^j. \tag{36}$$

Hence, applying (36) at each time level and returning to the notations $\phi^{j+1} = \widehat{\phi}$, $\rho^{j+1} = \widehat{\rho}$, we reach to the trapezoidal rule approximation of (31)

$$h\sum_{i=0}^{N} \alpha_i a(\phi_i^{j+1})\rho_i^{j+1} = h\sum_{i=0}^{N} \alpha_i a(\phi_i^j)\rho_i^j = h\sum_{i=0}^{N} \alpha_i a(\phi_i^{j-1})\rho_i^{j-1} = \cdots = h\sum_{i=0}^{N} \alpha_i a(\phi_i^0)\rho_i^0.$$

Consider (18), IMEX 2. As before, multiplying the first ($i = 0$) and the last ($i = N$) equation in (17) by $h/2$ and all other equations ($i = 1, 2, \dots, N-1$) by h, then summing up all resulting equations to obtain

$$h\sum_{i=0}^{N} \alpha_i G(\widehat{\phi}_i) - h\sum_{i=0}^{N} \alpha_i G(\phi_i) = \tau_j \left(h\sum_{i=0}^{N} \alpha_i p(\rho_i) - h\sum_{i=0}^{N} P_h^*(\phi_i, \rho_i) \right). \tag{37}$$

Write (37) for each time layer $j = 0, 1, \dots, J$ and summing up all these equation, we get

$$h\sum_{i=0}^{N} \alpha_i G(\phi_i^{j+1}) - h\sum_{i=0}^{N} \alpha_i G(\phi_i^0) = \sum_{l=0}^{j} \tau_l \left(h\sum_{i=1}^{N} \alpha_i p(\rho_i^l) - h\sum_{i=1}^{N} \alpha_i P_h^*(\phi_i^l, \rho_i^l) \right). \tag{38}$$

Note that the right-hand side of (38) is the approximation (by trapezoidal rule in space and rectangular rule in time) of

$$\int_0^1 \left(\mathcal{P}(\rho_f) - \mathcal{P}(\rho_f^0)\right) dx - \int_0^1 \left(\mathcal{P}^*(\phi, \rho_f) - \mathcal{P}^*(\phi^0, \rho_f^0)\right) dx,$$

while the left-hand side of (38) is the approximation (by trapezoidal rule) of

$$\int_0^1 G(\phi) dx - \int_0^1 G(\phi^0) dx.$$

Therefore, (38) is the discrete version of (32).

Treating (18), IMEX 1 similarly, we derive

$$h\sum_{i=0}^{N} \alpha_i \phi_i^{j+1} - h\sum_{i=0}^{N} \alpha_i \phi_i^0 = \sum_{l=0}^{j} \tau_l \left(h\sum_{i=1}^{N} \alpha_i [g(\phi_i^l)]^{-1} \left(p(\rho_i^l) - P_h^*(\phi_i^l, \rho_i^l) \right) \right), \tag{39}$$

which is a discretization of (34).

The numerical schemes IMEX 1 and IMEX 2 differ by different treating the (18). So, we will discuss them separately.

Theorem 1 (*Positivity and Boundness, IMEX 1*). *Let the assumptions (35) hold and the time step satisfy the inequality*

$$\tau_j < \min\left\{ \min_{0\le i\le N} \frac{\phi_i g(\phi_i)}{\left(p(\rho_i)-p^*(t_j)\right)^-}, \min_{0\le i\le N} \frac{(1-\phi_i) g(\phi_i)}{\left(p(\rho_i)-p^*(t_j)\right)^+} \right\}. \tag{40}$$

Then, at each time level, for the numerical solution of IMEX 1 we have $\widehat{\rho} > 0$, $0 < \widehat{\phi} < 1$.

Proof. Suppose that the statement of the theorem is fulfilled at time level t_j, namely $\rho > 0$, $0 < \phi < 1$. From (18), in view of (35), we have

$$\widehat{\phi}_i = \phi_i + \tau_j\left(p(\rho_i)-p^*(t_j)\right)[g(\phi_i)]^{-1} \ge \phi_i - \tau_j\left(p(\rho_i)-p^*(t_j)\right)^- [g(\phi_i)]^{-1}. \tag{41}$$

Therefore $\widehat{\phi}_i > 0$, $i = 0, 1, \ldots, N$, if

$$\tau_j < \min_{0\le i\le N} \frac{\phi_i g(\phi_i)}{\left(p(\rho_i)-p^*(t_j)\right)^-}. \tag{42}$$

Similarly, we obtain

$$\widehat{\phi}_i - 1 = \phi_i - 1 + \tau_j\left(p(\rho_i)-p^*(t_j)\right)[g(\phi_i)]^{-1} \le \phi_i - 1 + \tau_j\left(p(\rho_i)-p^*(t_j)\right)^+ [g(\phi_i)]^{-1}.$$

Now, it is clear that $\widehat{\phi}_i - 1 < 0$, $i = 0, 1, \ldots, N$ for

$$\tau_j < \min_{0\le i\le N} \frac{(1-\phi_i) g(\phi_i)}{\left(p(\rho_i)-p^0(t_j)\right)^+}. \tag{43}$$

Let us rewrite the equations in (17) in the form

$$-A_i\widehat{\rho}_{i-1} + C_i\widehat{\rho}_i - B_i\widehat{\rho}_{i+1} = F_i, \quad i = 0, 1, \ldots, N, \tag{44}$$

where

$$A_i = \mathcal{K}(\phi_i)\mathcal{B}(\rho_i)\frac{\tau_j}{h^2}, \quad \widehat{B}_i = \mathcal{K}(\widehat{\phi}_{i+1})\mathcal{B}(\rho_{i+1})\frac{\tau_j}{h^2}, \quad i = 1, 2, \ldots, N-1,$$

$$A_0 = 0, \quad B_0 = 2\mathcal{K}(\widehat{\phi}_1)\mathcal{B}(\rho_1)\frac{\tau_j}{h^2}, \quad A_N = 2\mathcal{K}(\widehat{\phi}_{N-1})\mathcal{B}(\rho_{N-1})\frac{\tau_j}{h^2}, \quad B_N = 0,$$

$$C_i = A_i + B_i + a(\widehat{\phi}_i), \quad F_i = a(\phi_i)\rho_i, \quad i = 0, 1, \ldots, N.$$

Taking into account that $\widehat{\phi}$ is known (see Section 4.1) and $0 < \widehat{\phi} < 1$, if the time step satisfies the conditions (42), (43), in view of (35), we deduce that the coefficient matrix of the system (44) is strictly diagonal dominant with positive main diagonal elements and non-positive off-diagonal entries. Therefore, being a tridiagonal M-matrix, the inverse of the coefficient matrix of (44) is a totally positive [19, Theorem 2.2]. Since F_i is positive, we conclude that $\widehat{\rho} > 0$ [20].

Collecting the conditions (42), (43), we obtain (40). The proof is completed, applying the same considerations at each time level. □

Theorem 2 (*Positivity and Boundness, IMEX 2*). *Let the function (14) is continuous in the interval (0,1), the assumptions (35) hold and the time step satisfy the inequality*

$$\tau_j < \min\left\{ \min_{0\le i\le N} \frac{G(\phi_i)-G(0+\epsilon)}{\left(p(\rho_i)-p^*(t_j)\right)^-}, \min_{0\le i\le N} \frac{G(1-\epsilon)-G(\phi_i)}{\left(p(\rho_i)-p^*(t_j)\right)^+} \right\}, \quad 0 < \epsilon \ll 1. \tag{45}$$

Then, at each time level, for the numerical solution of IMEX 2 we have $\widehat{\rho} > 0$, $0+\epsilon \le \widehat{\phi} \le 1-\epsilon$.

Proof. Suppose that at time level t_j, the statement of the theorem is fulfilled, namely $\rho > 0$, $0+\epsilon \le \phi \le 1-\epsilon$. Consider the (18), where $\widehat{\phi}$ is the solution of (21). Thus, we have

$$\begin{aligned}
\mathcal{F}(\widehat{\phi}_i) &:= G(\widehat{\phi}) - \int_{\phi_i^0}^{\widehat{\phi}_i} \frac{1}{(1-v)a_1(v)}dv \\
&= G(\phi_i) + \tau_j\left(p(\rho_i)-p^*(t_j)\right) - \int_{\phi_i^0}^{\widehat{\phi}_i} \frac{1}{(1-v)a_1(v)}dv \\
&= \int_{\phi_i^0}^{\phi_i} \frac{1}{(1-v)a_1(v)}dv + \tau_j\left(p(\rho_i)-p^*(t_j)\right) - \int_{\phi_i^0}^{\widehat{\phi}_i} \frac{1}{(1-v)a_1(v)}dv \\
&= \tau_j\left(p(\rho_i)-p^*(t_j)\right) - \int_{\phi_i}^{\widehat{\phi}_i} \frac{1}{(1-v)a_1(v)}dv = 0, \quad i = 0, 1, \ldots, N.
\end{aligned}$$

We investigate the conditions that guarantee the solvability of $\mathcal{F}(\widehat{\phi}_i)=0$, $i=0,1,\ldots,N$ in the interval $[0+\epsilon,1-\epsilon]$. Let $p(\rho_i)-p^*(t_j)>0$. In this case, we obtain

$$\mathcal{F}(1-\epsilon)=\tau_j\left(p(\rho_i)-p^*(t_j)\right)^+-\int_{\phi_i}^{1-\epsilon}\frac{1}{(1-v)a_1(v)}dv,$$
$$\mathcal{F}(0+\epsilon)=\tau_j\left(p(\rho_i)-p^*(t_j)\right)^++\int_{0+\epsilon}^{\phi_i}\frac{1}{(1-v)a_1(v)}dv>0.$$

Therefore, $\mathcal{F}(1-\epsilon)<0$, if

$$\tau_j<\min_{0\le i\le N}\frac{\int_{\phi_i}^{1-\epsilon}\frac{1}{(1-v)a_1(v)}dv}{\left(p(\rho_i)-p^*(t_j)\right)^+}=\min_{0\le i\le N}\frac{G(1-\epsilon)-G(\phi_i)}{\left(p(\rho_i)-p^*(t_j)\right)^+}. \tag{46}$$

Similarly, for $p(\rho_i)-p^*(t_j)<0$, we get

$$\mathcal{F}(1-\epsilon)=-\tau_j\left(p(\rho_i)-p^*(t_j)\right)^--\int_{\phi_i}^{1-\epsilon}\frac{1}{(1-v)a_1(v)}dv<0,$$
$$\mathcal{F}(0+\epsilon)=-\tau_j\left(p(\rho_i)-p^*(t_j)\right)^-+\int_{0+\epsilon}^{\phi_i}\frac{1}{(1-v)a_1(v)}dv.$$

Hence, to ensure that $\mathcal{F}(0+\epsilon)>0$, we have to restrict the time step by

$$\tau_j<\min_{0\le i\le N}\frac{\int_{0+\epsilon}^{\phi_i}\frac{1}{(1-v)a_1(v)}dv}{\left(p(\rho_i)-p^*(t_j)\right)^-}=\min_{0\le i\le N}\frac{G(\phi_i)-G(0+\epsilon)}{\left(p(\rho_i)-p^*(t_j)\right)^-}. \tag{47}$$

The conditions (46), (47) ensure that the equation $\mathcal{F}(\widehat{\phi}_i)=0$ has at least one root $\widehat{\phi}_i\in[0+\epsilon,1-\epsilon]$.

Observing that $\mathcal{F}(\widehat{\phi}_i)$ is increasing function for $\widehat{\phi}_i\in[0+\epsilon,1-\epsilon]$, because $\frac{d\mathcal{F}(\widehat{\phi}_i)}{d\widehat{\phi}_i}=\frac{dG(\widehat{\phi}_i)}{d\widehat{\phi}_i}>0$ (in view of (35)), we conclude that $\widehat{\phi}_i$ is the unique solution in $[0+\epsilon,1-\epsilon]$. Therefore, if the condition (45) is satisfied, we have $0+\epsilon\le\widehat{\phi}\le1-\epsilon$.

The proof that $\widehat{\rho}>0$ is the same as in Theorem 1. □

Taking into consideration the particular form of the function $g(\phi)=a^0\nu[(1-\phi)\phi^r]^{-1}$ (see (11), (14)), we obtain more precise time step restrictions. For example, consider IMEX 1. Observing that for $r>0$, $g(\phi)$ attains minimum value in $(0,1)$ at $\phi_{min}=r/(r+1)$ and $g(\phi_{min})<a^0\nu$, from (41) we deduce that $\widehat{\phi}_i>0$, $i=0,1,\ldots,N$, if

$$\tau_j<a^0\nu\begin{cases}\min\limits_{0\le i\le N}\dfrac{\phi_i}{\left(p(\rho_i)-p^*(t_j)\right)^-}, & 0\le r<1,\\ \|\left(p(\rho)-p^*(t_j)\right)^-\|^{-1}, & 1\le r\le 2.\end{cases} \tag{48}$$

Similarly, $\widehat{\phi}_i<1$, $i=0,1,\ldots,$ if the time step is restricted as follows

$$\tau_j<a^0\nu\|\left(p(\rho)-p^*(t_j)\right)^+\|^{-1}. \tag{49}$$

Moreover, taking into account that $p^j_{\min}\le p^*(t_j)\le p^j_{\max}$, $\|p(\rho^j)-p^*(t_j)\|\le p^j_{\max}$ and in view of (22), (23), (35) we obtain rough, but illustrative estimates for the solution of IMEX 1.

For clarity, let us restore the notations $\phi^{j+1}=\widehat{\phi}$, assume that at time layer t_j the solution satisfies the inequalities $0<m_0^j\le\phi_i^j\le M_0^j<1$, $0<m_1^j\le\rho_i^j\le M_1^j<1$ ($m_0^0=m_0$, $m_1^0=m_1$, $M_0^0=M_0$, $M_1^0=M_1$) and set $p^0=0$. Thus, from (11) and (41) we consequently derive

$$\phi_i^{j+1}\ge\phi_i^j-\tau_j\frac{p^j_{\max}}{a^0\nu p_1}\ge\begin{cases}m_0^j-\tau_j\mathcal{D}_1\ln\dfrac{M_1^j}{m_1^0}, & \text{for } p \text{ defined by (22)},\\ m_0^j-\tau_j\mathcal{D}_1\dfrac{M_1^j}{m_1^0}, & \text{for } p \text{ defined by (23)},\end{cases}\qquad \mathcal{D}_1=\frac{L}{\nu v_1\beta_f}.$$

Let $0<m_0^{j+1}<m_0^j$. Hence, $\phi_i^{j+1}\ge m_0^{j+1}$, $i=0,1,\ldots,N$, if

$$\tau_j\le\frac{1}{\mathcal{D}_1}\left(m_0^j-m_0^{j+1}\right)\begin{cases}\left[\ln\dfrac{M_1^j}{m_1^0}\right]^{-1}, & \text{for } p \text{ defined by (22)},\\ \dfrac{m_1^0}{M_1^j}, & \text{for } p \text{ defined by (23)}.\end{cases} \tag{50}$$

In the same manner, we get

$$\phi_i^{j+1} \le \phi_i^j + \tau_j \frac{p^j_{\max}}{a^0 \nu p_1} \le \begin{cases} M_0^j + \tau_j \mathcal{D}_1 \ln \dfrac{M_1^j}{m_1^0}, & \text{for } p \text{ defined by (22)}, \\ M_0^j + \tau_j \mathcal{D}_1 \dfrac{M_1^j}{m_1^0}, & \text{for } p \text{ defined by (23)}. \end{cases}$$

Suppose that there exists M_0^{j+1}, such that $M_0^j < M_0^{j+1} < \infty$. Then, $\phi_i^{j+1} \le M_0^{j+1}$, $i = 0, 1, \ldots, N$, if the time step satisfies the inequality

$$\tau_j \le \frac{1}{\mathcal{D}_1}\left(M_0^{j+1} - M_0^j\right) \begin{cases} \left[\ln \dfrac{M_1^j}{m_1^0}\right]^{-1}, & \text{for } p \text{ defined by (22)}, \\ \dfrac{m_1^0}{M_1^j}, & \text{for } p \text{ defined by (23)}. \end{cases} \tag{51}$$

Next, we consider Eqs. (44). Evidently, in view of (35) we have

$$C_i\widehat{\rho}_i \le A_i\widehat{\rho}_{\max} + B_i\widehat{\rho}_{\max} + a(\phi_i)\rho_{\max}, \quad i = 0, 1, \ldots, N,$$

$$C_i\widehat{\rho}_i \ge A_i\widehat{\rho}_{\min} + B_i\widehat{\rho}_{\min} + a(\phi_i)\rho_{\min}, \quad i = 0, 1, \ldots, N.$$

Since $a(\phi)$, $0 < \phi < 1$ is strictly increasing function, we obtain

$$\widehat{\rho}_{\max} \le M_1^{j+1}, \quad \widehat{\rho}_{\min} \ge m_1^{j+1}, \quad M_1^{j+1} = \frac{m_0^j M_1^j \left(1 - M_0^{j+1}\right)}{\left(1 - m_0^j\right) M_0^{j+1}}, \quad m_1^{j+1} = \frac{\left(1 - m_0^{j+1}\right) M_0^j m_1^j}{m_0^{j+1}\left(1 - M_0^j\right)}. \tag{52}$$

5.2. Implicit scheme (19), (20)

Similarly to the IMEX schemes, one can show that IS 1 and IS 2 preserve conservation properties (31) and (32) or (34), respectively. The only difference is that now in (38) and (39) the time layer summation $\sum_{l=0}^{j}$ is replaced by $\sum_{l=1}^{j+1}$.

In view of the realization of the implicit discretization by the iteration schemes ItS 1, ItS 2, we may consider positivity and boundness of the numerical solution at each time level and at each iteration. Applying similar considerations as for the IMEX schemes, we obtain the following results.

Theorem 3 (*Positivity and Boundness, ItS 1*). *Let the assumptions (35) hold and the time step satisfy the inequality*

$$\tau_j < \min\left\{ \min_{0\le i\le N} \frac{\phi_i g(\phi_i^{(m)})}{\left(p\left(\rho_i^{(m)}\right) - P_h^*(\phi^{(m)}, \rho^{(m)})\right)^-}, \min_{0\le i\le N} \frac{(1-\phi_i) g(\phi_i^{(m)})}{\left(p\left(\rho_i^{(m)}\right) - P_h^*(\phi^{(m)}, \rho^{(m)})\right)^+} \right\}.$$

Then, at each time level and at each iteration, for the numerical solution of IMEX 1 we have $\rho^{(m+1)} > 0$, $0 < \phi^{(m+1)} < 1$.

Theorem 4 (*Positivity and Boundness, ItS 2*). *Let the function (14) is continuous in the interval (0,1), the assumptions (35) hold and the time step satisfy the inequality*

$$\tau_j < \min\left\{ \min_{0\le i\le N} \frac{G(\phi_i) - G(0+\epsilon)}{\left(p(\rho_i) - P_h^*(\phi^{(m)}, \rho^{(m)})\right)^-}, \min_{0\le i\le N} \frac{G(1-\epsilon) - G(\phi_i)}{\left(p(\rho_i) - P_h^*(\phi^{(m)}, \rho^{(m)})\right)^+} \right\}, \quad 0 < \epsilon \ll 1.$$

Then, at each time level, and at each iteration, for the numerical solution of IMEX 2 we have $\rho^{(m+1)} > 0$, $0+\epsilon \le \phi^{(m+1)} \le 1-\epsilon$.

For the particular representation of the functions $g(\phi)$, $p(\rho)$, we may deduce similar to (48)–(52) results. On this base, further we assume:

(A1) There exists positive constants m_0^j, M_0^j, m_1^j, M_1^j, $j = 0, 1, \ldots, J$, such that for a sufficiently small time step τ_j, we have

$$0 < m_0^j \le \phi_i^j \le M_0^j < 1, \quad 0 < m_1^j \le \rho_i^j \le M_1^j < 1, \quad j = 0, 1, \ldots, J.$$

(A2) Function $a(\phi)$, $K(\phi)$, $[g(\phi)]^{-1}$, $f(\phi) = a_1(\phi)/(1-\phi)$, $p(\rho)$, $b(\rho)$ and their derivatives are bounded for $0 < m_{0*} \le \phi \le M_0^* < 1$, $0 < m_{1*} \le \rho \le M_1^* < 1$.

Theorem 5 (*Convergence of the Iteration Process, ItS 1*)**.** *Let the time step τ_j is sufficiently small, and the assumptions (A1), (A2) hold. Then, for the iteration process ItS 1, we have*

$$\lim_{m\to\infty}\left(\|\rho^{(m+1)}-\rho^{(m)}\|+\|\phi^{(m+1)}-\phi^{(m)}\|\right)=0.$$

Proof. From (A1) follows that there exists constants $m_{0*}=\min_j m_0^j$, $M_0^*=\max_j M_0^j$, $m_{1*}=\min_j m_1^j$, $M_1^*=\max_j M_1^j$, such that $0<m_{0*}\le\phi_i^j\le M_0^*<1$, $0<m_{1*}\le\rho_i^j\le M_1^*<1$, $i=0,1,\dots,N$, $j=0,1,\dots,J$.

Let

$$w_i^{(m+1)}=\rho_i^{(m+1)}-\rho_i^{(m)},\quad z_i^{(m+1)}=\phi_i^{(m+1)}-\phi_i^{(m)},\quad i=0,1,\dots,N \tag{53}$$

and for $0<\theta_i^s<1$, $s=\{1,2,3,4\}$ involve the notations

$$\widetilde{\widetilde{\phi}}_i=\phi_i^{(m+1)}+\theta_i^1\phi_i^{(m)},\quad \widetilde{\phi}_i=\phi_i^{(m)}+\theta_i^2\phi_i^{(m-1)},\quad \widetilde{\widetilde{\rho}}_i=\rho_i^{(m+1)}+\theta_i^3\rho_i^{(m)},\quad \widetilde{\rho}_i=\rho_i^{(m)}+\theta_i^4\rho_i^{(m-1)}. \tag{54}$$

Consider the iteration procedure ItS 1 at the time layer t_j. Subtracting the (24) at the mth iteration from (24) at the $m+1$-st iteration and applying Taylor series expansion around $(\phi_i^{(m-1)},\rho_i^{(m-1)})$, resulting in

$$\begin{aligned}
z_i^{(m+1)}=&\tau_j[g(\phi_i^{(m)})]^{-1}\left(p(\rho_i^{(m)})-P_h^*(\phi_i^{(m)},\rho_i^{(m)})\right)-\tau_j\left[g(\phi_i^{(m-1)})\right]^{-1}\left(p(\rho_i^{(m-1)})-P_h^*(\phi_i^{(m-1)},\rho_i^{(m-1)})\right)\\
=&\tau_j\left(\frac{d\left[g(\widetilde{\phi})\right]^{-1}}{d\phi}(\widetilde{\phi}_i)\right)\left(p(\widetilde{\rho}_i)-P_h^*(\widetilde{\phi}_i,\widetilde{\rho}_i)\right)z_i^{(m)}+\tau_j\left[g(\widetilde{\phi}_i)\right]^{-1}\frac{dp(\rho)}{d\rho}(\widetilde{\rho}_i)w_i^{(m)}\\
&-\tau_j\left[g(\widetilde{\phi})\right]^{-1}\left(P_h^*(\phi_i^{(m)},\rho_i^{(m)})-P_h^*(\phi_i^{(m-1)},\rho_i^{(m-1)})\right).
\end{aligned}$$

Therefore, in view of the conditions of the theorem, for $0<m_{0*}\le\phi\le M_0^*<1$, $0<m_{1*}\le\rho\le M_1^*<1$ we may estimate

$$\begin{aligned}
\|z^{(m+1)}\|\le&\tau_j\left\|\frac{d\left[g(\phi)\right]^{-1}}{d\phi}\right\|\|p(\rho)\|\|z^{(m)}\|+\tau_j\left\|\frac{dp(\rho)}{d\rho}\left[g(\phi)\right]^{-1}\right\|\|w^{(m)}\|\\
&+\tau_j\left\|\left[g(\phi)\right]^{-1}\left(P_h^*(\phi^{(m)},\rho^{(m)})-P_h^*(\phi^{(m-1)},\rho^{(m-1)})\right)\right\|.
\end{aligned} \tag{55}$$

Let us consider the term $P_h^*(\phi_i^{(m)},\rho_i^{(m)})-P_h^*(\phi_i^{(m-1)},\rho_i^{(m-1)})$. Applying Taylor series expansion for $P_h^*(\phi_i^{(m)},\rho_i^{(m)})$ around $(\phi_i^{(m-1)},\rho_i^{(m-1)})$ and using the notations (54), we obtain

$$\begin{aligned}
P_h^*(\phi_i^{(m)},\rho_i^{(m)})-P_h^*(\phi_i^{(m-1)},\rho_i^{(m-1)})=&\sum_{i=0}^{N}\frac{\frac{dp(\rho)}{d\rho}(\widetilde{\rho}_i)f(\widetilde{\phi}_i)}{\sum_{i=0}^{N}f(\widetilde{\phi}_i)}w_i^{(m)}\\
&+\sum_{i=0}^{N}\left(\frac{\frac{df(\phi)}{d\phi}(\widetilde{\phi}_i)p(\widetilde{\rho}_i)}{\sum_{i=0}^{N}f(\widetilde{\phi}_i)}-\frac{\frac{df(\phi)}{d\phi}(\widetilde{\phi}_i)\sum_{i=0}^{N}p(\widetilde{\rho}_i)f(\widetilde{\phi}_i)}{\left(\sum_{i=0}^{N}f(\widetilde{\phi}_i)\right)^2}\right)z_i^{(m)}.
\end{aligned}$$

Consequently,

$$\left\|P_h^*(\phi^{(m)},\rho^{(m)})-P_h^*(\phi^{(m-1)},\rho^{(m-1)})\right\|\le\left\|\frac{dp(\rho)}{d\rho}\right\|\|w^{(m)}\|+2\|p(\rho)\|\left\|\frac{df(\phi)}{d\phi}[f(\phi)]^{-1}\right\|\|z^{(m)}\|. \tag{56}$$

Taking into account (56), from (55) we get

$$\|z^{(m+1)}\|\le\tau_j\|p(\rho)\|\left(\left\|\frac{d[g(\phi)]^{-1}}{d\phi}\right\|+2\left\|\frac{df(\phi)}{d\phi}[f(\phi)g(\phi)]^{-1}\right\|\right)\|z^{(m)}\|+2\tau_j\left\|\frac{dp(\rho)}{d\rho}[g(\phi)]^{-1}\right\|\|w^{(m)}\|.$$

Hence,

$$\|z^{(m+1)}\|\le\tau_jC_1\|z^{(m)}\|+\tau_jC_2\|w^{(m)}\|, \tag{57}$$

where

$$C_1=C_{11}C_{12},\quad \|p(\rho)\|_\infty\le C_{11},\quad \left\|\frac{d[g(\phi)]^{-1}}{d\phi}\right\|_\infty+2\left\|\frac{df(\phi)}{d\phi}[f(\phi)g(\phi)]^{-1}\right\|_\infty\le C_{12},\quad 2\left\|\frac{dp(\rho)}{d\rho}[g(\phi)]^{-1}\right\|_\infty\le C_2.$$

Now, we consider equations (25), $1 \le i \le N-1$. Subtract the ith equation at mth iteration from the same equation, but corresponding to the $m+1$-st iteration. The resulting equation is

$$\frac{1}{\tau_j}\left(a\big(\phi_i^{(m+1)}\big)\rho_i^{(m+1)} - a\big(\phi_i^{(m)}\big)\rho_i^{(m)}\right) = \frac{1}{h^2}\left(\mathcal{K}\big(\phi_{i+1}^{(m+1)}\big)\mathcal{B}\big(\rho_{i+1}^{(m)}\big)\big(\rho_{i+1}^{(m+1)} - \rho_i^{(m+1)}\big)\right.$$
$$-\mathcal{K}\big(\phi_{i+1}^{(m)}\big)\mathcal{B}\big(\rho_{i+1}^{(m-1)}\big)\big(\rho_{i+1}^{(m)} - \rho_i^{(m)}\big) - \mathcal{K}\big(\phi_i^{(m+1)}\big)\mathcal{B}\big(\rho_i^{(m)}\big)\big(\rho_i^{(m+1)} - \rho_{i-1}^{(m+1)}\big)$$
$$\left.+\mathcal{K}\big(\phi_i^{(m)}\big)\mathcal{B}\big(\rho_i^{(m-1)}\big)\big(\rho_i^{(m)} - \rho_{i-1}^{(m)}\big)\right).$$

Applying Taylor series expansion, rearranging the expression and in view of the notations (53), (54), we derive

$$\left(\frac{a(\widetilde{\widetilde{\phi}}_i)}{\tau_j} + \frac{\mathcal{K}\big(\widetilde{\widetilde{\phi}}_{i+1}\big)\mathcal{B}\big(\widetilde{\rho}_{i+1}\big) + \mathcal{K}\big(\widetilde{\widetilde{\phi}}_i\big)\mathcal{B}\big(\widetilde{\rho}_i\big)}{h^2}\right) w_i^{(m+1)} - \frac{\mathcal{K}\big(\widetilde{\widetilde{\phi}}_{i+1}\big)\mathcal{B}\big(\widetilde{\rho}_{i+1}\big)}{h^2} w_{i+1}^{(m+1)} - \frac{\mathcal{K}\big(\widetilde{\widetilde{\phi}}_i\big)\mathcal{B}\big(\widetilde{\rho}_i\big)}{h^2} w_{i-1}^{(m+1)}$$
$$= -\frac{1}{\tau_j}\frac{da}{d\phi}(\widetilde{\widetilde{\phi}}_i)\widetilde{\widetilde{\rho}}_i z_i^{(m+1)}$$
$$+\left(\frac{1}{2h^2}\frac{dK}{d\phi}(\widetilde{\widetilde{\phi}}_{i+1})\mathcal{B}\big(\widetilde{\rho}_{i+1}\big)\big(\widetilde{\widetilde{\rho}}_{i+1} - \widetilde{\widetilde{\rho}}_i\big)\right) z_{i+1}^{(m+1)} - \left(\frac{1}{2h^2}\frac{dK}{d\phi}(\widetilde{\widetilde{\phi}}_{i-1})\mathcal{B}\big(\widetilde{\rho}_i\big)\big(\widetilde{\widetilde{\rho}}_i - \widetilde{\widetilde{\rho}}_{i-1}\big)\right) z_{i-1}^{(m+1)}$$
$$+\left(\frac{1}{2h^2}\frac{dK}{d\phi}(\widetilde{\widetilde{\phi}}_i)\mathcal{B}\big(\widetilde{\rho}_{i+1}\big)\big(\widetilde{\widetilde{\rho}}_{i+1} - \widetilde{\widetilde{\rho}}_i\big) - \frac{1}{2h^2}\frac{dK}{d\phi}(\widetilde{\widetilde{\phi}}_i)\mathcal{B}\big(\widetilde{\rho}_i\big)\big(\widetilde{\widetilde{\rho}}_i - \widetilde{\widetilde{\rho}}_{i-1}\big)\right) z_i^{(m+1)}$$
$$+\left(\frac{1}{2h^2}\frac{db}{d\rho}(\widetilde{\rho}_{i+1})\mathcal{K}\big(\widetilde{\widetilde{\phi}}_{i+1}\big)\big(\widetilde{\widetilde{\rho}}_{i+1} - \widetilde{\widetilde{\rho}}_i\big)\right) w_{i+1}^{(m)} - \left(\frac{1}{2h^2}\frac{db}{d\rho}(\widetilde{\rho}_{i-1})\mathcal{K}\big(\widetilde{\widetilde{\phi}}_i\big)\big(\widetilde{\widetilde{\rho}}_i - \widetilde{\widetilde{\rho}}_{i-1}\big)\right) w_{i-1}^{(m)}$$
$$+\left(\frac{1}{2h^2}\frac{db}{d\rho}(\widetilde{\rho}_i)\mathcal{K}\big(\widetilde{\widetilde{\phi}}_{i+1}\big)\big(\widetilde{\widetilde{\rho}}_{i+1} - \widetilde{\widetilde{\rho}}_i\big) - \frac{1}{2h^2}\frac{db}{d\rho}(\widetilde{\rho}_i)\mathcal{K}\big(\widetilde{\widetilde{\phi}}_i\big)\big(\widetilde{\widetilde{\rho}}_i - \widetilde{\widetilde{\rho}}_{i-1}\big)\right) w_i^{(m)}.$$

Further, forasmuch as the conditions of the theorem and (35), we apply maximum principle and for $0 < m_{0*} \le \phi \le M_0^* < 1$, $0 < m_{1*} \le \rho \le M_1^* < 1$, we estimate

$$\left\|w^{(m+1)}\right\| \le \left(\left\|\frac{\frac{da}{d\phi}(\phi)\rho}{a(\phi)}\right\| + \frac{\tau_j}{h^2}\left\|4\frac{dK}{d\phi}(\phi)\frac{\mathcal{B}(\rho)\rho}{a(\phi)}\right\|\right)\left\|z^{(m+1)}\right\| + \frac{\tau_j}{h^2}\left\|4\frac{db}{d\rho}(\rho)\frac{\mathcal{K}(\phi)\rho}{a(\phi)}\right\|\left\|w^{(m)}\right\|$$
$$= \left(C_{21} + \frac{\tau_j}{h^2}C_{22}\right)\left\|z^{(m+1)}\right\| + \frac{\tau_j}{h^2}C_{23}\left\|w^{(m)}\right\|, \tag{58}$$

where

$$\left\|\frac{\frac{da}{d\phi}(\phi)\rho}{a(\phi)}\right\|_\infty \le C_{21}, \quad \left\|4\frac{dK}{d\phi}(\phi)\frac{\mathcal{B}(\rho)\rho}{a(\phi)}\right\|_\infty \le C_{22}, \quad \left\|4\frac{db}{d\rho}(\rho)\frac{\mathcal{K}(\phi)\rho}{a(\phi)}\right\|_\infty \le C_{23}.$$

Substituting (57) in (58), we derive

$$\left\|w^{(m+1)}\right\| \le \tau_j C_1\left(C_{21} + \frac{\tau_j}{h^2}C_{22}\right)\left\|z^{(m)}\right\| + \left(\tau_j C_2\left(C_{21} + \frac{\tau_j}{h^2}C_{22}\right) + C_{23}\frac{\tau_j}{h^2}\right)\left\|w^{(m)}\right\|. \tag{59}$$

The sum of (57) and (59) leads to the inequality

$$\left\|z^{(m+1)}\right\| + \left\|w^{(m+1)}\right\| \le \tau_j C_1\left(C_{21} + \frac{\tau_j}{h^2}C_{22} + 1\right)\left\|z^{(m)}\right\| + \left(\tau_j C_2\left(C_{21} + \frac{\tau_j}{h^2}C_{22} + 1\right) + C_{23}\frac{\tau_j}{h^2}\right)\left\|w^{(m)}\right\|. \tag{60}$$

Let $C_{23} > 0$. Then, if

$$\tau_j < \min\left\{\frac{h^2}{C_{23}\psi_1}, \frac{C_{23}\psi_1}{C_1\widetilde{C}}, \frac{C_{23}\psi_1}{\psi_2 C_2\widetilde{C}}\right\}, \quad \widetilde{C} = C_{23}\psi_1(C_{21}+1) + C_{22}, \quad \frac{1}{\psi_1} + \frac{1}{\psi_2} < 1, \quad \psi_1,\ \psi_2 > 1, \tag{61}$$

from (60) we get

$$\left\|z^{(m+1)}\right\| + \left\|w^{(m+1)}\right\| \le \mathcal{I}\left(\left\|z^{(m)}\right\| + \left\|w^{(m)}\right\|\right), \quad \text{where} \quad 0 < \mathcal{I} < 1. \tag{62}$$

Thus,

$$\left\|z^{(m+1)}\right\| + \left\|w^{(m+1)}\right\| \le \mathcal{I}^m\left(\left\|z^{(1)}\right\| + \left\|w^{(0)}\right\|\right). \tag{63}$$

Consider the case $C_{23} = 0$. If the time step satisfies the restriction

$$\tau_j < \min\left\{\frac{h}{\psi_1 C_{22}\max\{C_1, C_2\}}, \frac{1}{\psi_2\max\{C_1, C_2\}(C_{21}+1)}\right\}, \tag{64}$$

we consequently obtain (62), (63), which leads to the statement of the theorem. □

Studding the functions in (A1) and their derivatives, from their particular representations (8), (9), (10), (11) (14), (15), (22) and (23), we find

$$C_1 = \mathcal{D}_1\overline{p}\begin{cases} 3, & \text{if } r \in \{0\}\cup[1,2], \\ m_{0*}^{r-1}(3r+(1-3r)m_{0*}), & \text{if } r \in (0,1), \end{cases} \quad \overline{p} = \begin{cases} \ln\dfrac{M_1^*}{m_{1*}}, & \text{for (22), } p^0 = 0, \\ \dfrac{M_1^*}{m_{1*}}, & \text{for (23), } p^0 = 0, \end{cases}$$

$$C_{21} = \mathcal{D}_2\frac{M_1^*}{m_{0*}(1-m_{0*})}, \quad C_{22} = 8\mathcal{D}_2\begin{cases} M_1^*, & \text{for (8)}, \\ \dfrac{(M_1^*)^2}{m_{1*}}, & \text{for (9)}, \end{cases} \quad C_{23} = 4\mathcal{D}_2\begin{cases} 0, & \text{for (8)}, \\ \dfrac{M_1^*}{m_{1*}}, & \text{for (9)}, \end{cases}$$

$$C_2 = 2\frac{\mathcal{D}_1}{M_1^*}, \quad \mathcal{D}_2 = \frac{\bar{k}}{\mu v_1 L\beta_f}.$$

So, the time step restrictions (61), (64) become:
– for p, given by (8), (22), $p^0 = 0$:

$$\tau_j < \frac{\min\left\{\frac{h}{8\psi_1\mathcal{D}_2M_1^*}, \frac{m_{0*}(1-m_{0*})}{\psi_2(m_{0*}(1-m_{0*})+\mathcal{D}_2M_1^*)}\right\}}{\mathcal{D}_1\max\left\{3\ln\frac{M_1^*}{m_{0*}}, \frac{2}{M_1^*}\right\}}, \quad r \in \{0\}\cup[1,2];$$

$$\tau_j < \frac{\min\left\{\frac{h}{8\psi_1\mathcal{D}_2M_1^*}, \frac{m_{0*}(1-m_{0*})}{\psi_2(m_{0*}(1-m_{0*})+\mathcal{D}_2M_1^*)}\right\}}{\mathcal{D}_1\max\left\{m_{0*}^{r-1}(3r+(1-3r)m_{0*})\ln\frac{M_1^*}{m_{0*}}, \frac{2}{M_1^*}\right\}}, \quad r \in (0,1);$$

– for p, given by (9), (22), $p^0 = 0$ and $\mathcal{K} = \mathcal{D}_1(\psi_1\mathcal{D}_2M_1^* + (\psi_1+2M_1^*)m_{0*}(1-m_{0*}))$:

$$\tau_j < \left\{\frac{m_{1*}h^2}{4\mathcal{D}_2M_1^*\psi_1}, \frac{m_{1*}m_{0*}(1-m_{0*})\psi_1}{3M_1^*\mathcal{K}}\right\}, \quad r \in \{0\}\cup[1,2]; \tag{65}$$

$$\tau_j < \left\{\frac{m_{1*}h^2}{4\mathcal{D}_2M_1^*\psi_1}, \frac{m_{1*}m_{0*}(1-m_{0*})\psi_1}{\min\left\{2\psi_2m_1^*/M_1^*, M_1^*m_{0*}^{r-1}(3r+(1-3r)m_{0*})\right\}\mathcal{K}}\right\}, \quad r \in (0,1).$$

Theorem 6 (*Convergence of the Iteration Process, ItS 2*)**.** *Let the time step τ_j is sufficiently small, and the assumptions (A1), (A2) hold. Then, for the iteration process ItS 2, we have*

$$\lim_{m\to\infty}\left(\|\rho^{(m+1)} - \rho^{(m)}\| + \|\phi^{(m+1)} - \phi^{(m)}\|\right) = 0.$$

Proof. Consider the iteration process ItS 2 at time layer t_j. Let us subtract the (24) at mth iteration from (24) at $m+1$-st iteration and apply Taylor series expansion around $(\phi_i^{(m-1)}, \rho_i^{(m-1)})$. Thus in view of the notations (53), (54), we get

$$G(\phi_i^{(m+1)}) - G(\phi_i^{(m)}) = \tau_j\left(\frac{dp(\rho)}{d\rho}(\widetilde{\rho}_i)w_i^{(m)} - P_h^*\left(\phi_i^{(m)}, \rho_i^{(m)}\right) + P_h^*\left(\phi_i^{(m-1)}, \rho_i^{(m-1)}\right)\right).$$

Regarding to [2, Lemma 2], for the particular function $a_1(\phi) = a_0(\phi)\phi^{\lambda_1}(1-\phi)^{\lambda_2}$, $\lambda_1,\ \lambda_2 > 0$, $0 < \phi < 1$, there exist a positive constant $C_\phi \geq a_0(\phi)$, such that $C_\phi|G(\phi_1) - G(\phi_1)| \geq |\phi_1 - \phi_2|$. To prove this result, authors use the definition of functions $G(\phi)$, $a_1(\phi)$ and show that

$$G(\phi_1) - G(\phi_1) = \int_{\phi_1}^{\phi_2}\frac{ds}{(1-s)a_1(s)} \geq C_\phi^{-1}(\phi_1 - \phi_2).$$

Taking into account (21), incorporated in the iteration procedure of ItS 2, we have

$$G(\phi_i^{(m+1)}) - G(\phi_i^{(m)}) = \int_{\phi_i^{(m)}}^{\phi_i^{(m+1)}}\frac{ds}{(1-s)a_1(s)}$$

and therefore, since (A1) is satisfied, we state that $|G(\phi_i^{(m+1)}) - G(\phi_i^{(m)})| \geq C_\phi^{-1}|\phi_i^{(m+1)} - \phi_i^{(m)}|$. Further, using this result, applying Taylor series expansion around point $(\phi_i^{(m-1)}, \rho_i^{(m-1)})$ and from the conditions of the theorem and (56), we obtain

$$\begin{aligned}\|z_i^{(m+1)}\| &\leq 2\tau_j C_\phi^{-1}\left\|\frac{dp(\rho)}{d\rho}\right\|\left\|w^{(m)}\right\| + 2\tau_j C_\phi^{-1}\left\|\frac{df(\phi)}{d\phi}[f(\phi)]^{-1}\right\|\|p(\rho)\|\left\|z^{(m)}\right\| \\ &= \tau_j C_1\left\|z^{(m)}\right\| + \tau_j C_2\left\|w^{(m)}\right\|,\end{aligned}$$

where

$$2C_\phi^{-1}\left\|\frac{df(\phi)}{d\phi}[f(\phi)]^{-1}\right\|\|p(\rho)\| \leq C_1, \quad 2C_\phi^{-1}\left\|\frac{dp(\rho)}{d\rho}\right\| \leq C_2.$$

Next, we proceed similarly as in the proof of Theorem 5 and reach to the inequalities (62), (63) under conditions (61), (64). □

6. Numerical simulations

In this section we verify the order of convergence and the efficiency of the proposed numerical schemes IMEX 1, IMEX 2, ItS 1 and ItS 2, for coefficient functions chosen as in (10), (11), (14), (15), $r = 1$, $n = 3$ and $p(\rho_f)$, given by (9), (23), $p^0 = 0$.

For convergence test we deal with dimensionless exact solution $\phi(x,t) = 0.5e^{-t}\cos^2(0.5\pi x) + 0.45$, $\rho_f(x,t) = 0.5e^t \sin^2(0.5\pi x) + 3$. To this aim, we add appropriate residual terms in the right-hand sides of Eqs. (12), (13) and choose the initial solution according to the exact one.

We set the following model parameters for andesite magma [6,15,21,22]:

– fluid compressibility $\beta_f = 4.10^{-10}$ Pa^{-1};

– fluid viscosity $\mu = 2.6.10^{-4}$ Pa.s;

– rock share viscosity $\nu = 100$Pa.s;

– permeability porosity proportional constant $\bar{k} = 5.10^{-7}$ m^2

and for the dimensionless procedure (11), we take $v_1 = 5.10^6$ m/s.

Eq. (21) is solved by Matlab function 'fsolve' with default stopping criteria, while the tolerance in (30) is $\varepsilon = 10^{-12}$.

We provide computational results for the following methods:

– IMEX 1 and IMEX 2, where the generated system of algebraic equations (44) is solved by Thomas method (i.e. (27), (28), written for the system (44));

– ItS 1, ItS 2, realized by improved Thomas method (27), (28);

– ItS 1(1), where the system (27), (28) is solved by the MATLAB fast direct solver 'mldivide', which "employs different solvers to handle different kinds of coefficient matrices. The various cases are diagnosed automatically by examining the coefficient matrix".[1]

Let $\phi(x_i, T)$ and $\rho_f(x_i, T)$ be the solutions of the exact test solution problem (12)–(16) for $x = x_i$ and $t = T$, while ϕ_i^J and ρ_i^J are the corresponding numerical solutions at grid node (x_i, t_J). We give errors $e_\phi^N = \phi(x_i, T) - \phi_i^J$, $e_\rho^N = \rho_f(x_i, T) - \rho_i^J$ in maximal discrete norm $(\mathcal{E}_\phi^N, \mathcal{E}_\rho^N)$ and L_2 norm (E_ϕ^N, E_ρ^N):

$$\mathcal{E}_\phi^N = \max_{0\leq i\leq N}|e_\phi^N|, \quad \mathcal{E}_\rho^N = \max_{0\leq i\leq N}|e_\rho^N|, \quad E_\phi^N = h\left(\sum_{i=0}^{i=N}\left(e_\phi^N\right)^2\right)^{1/2}, \quad E_\rho^N = h\left(\sum_{i=0}^{i=N}\left(e_\rho^N\right)^2\right)^{1/2}$$

and the order of convergence:

$$\mathcal{CR}_\phi^N = \log_2\frac{\mathcal{E}_\phi^{2N}}{\mathcal{E}_\phi^N}, \quad \mathcal{CR}_\rho^N = \log_2\frac{\mathcal{E}_\rho^{2N}}{\mathcal{E}_\rho^N}, \quad CR_\phi^N = \log_2\frac{E_\phi^{2N}}{E_\phi^N}, \quad CR_\rho^N = \log_2\frac{E_\rho^{2N}}{E_\rho^N},$$

at final time $T = 0.5$.

For the considered test example, the theoretical time step restriction (61), (65) is:

$$\tau_j \lesssim \min\{0.04h^2, 0.003\}, \quad \psi_1 = \frac{9}{8}, \quad \psi_2 = 9, \quad \text{i.e. } \tau_j \lesssim 0.04h^2 \text{ for } N \geq 20.$$

It is not a surprise that we may compute the solution successfully for more relaxed time step restriction. As the expected order of convergence (in maximal norm) of the presented discretizations is $O(|\tau_j| + h^2)$, $|\tau_j| = \max_{0\leq j\leq J}\tau_j$, for the convergence test we set fixed time step $\tau = h^\gamma$.

[1] see MATLAB documentation https://www.mathworks.com/help/matlab/math/systems-of-linear-equations.html.

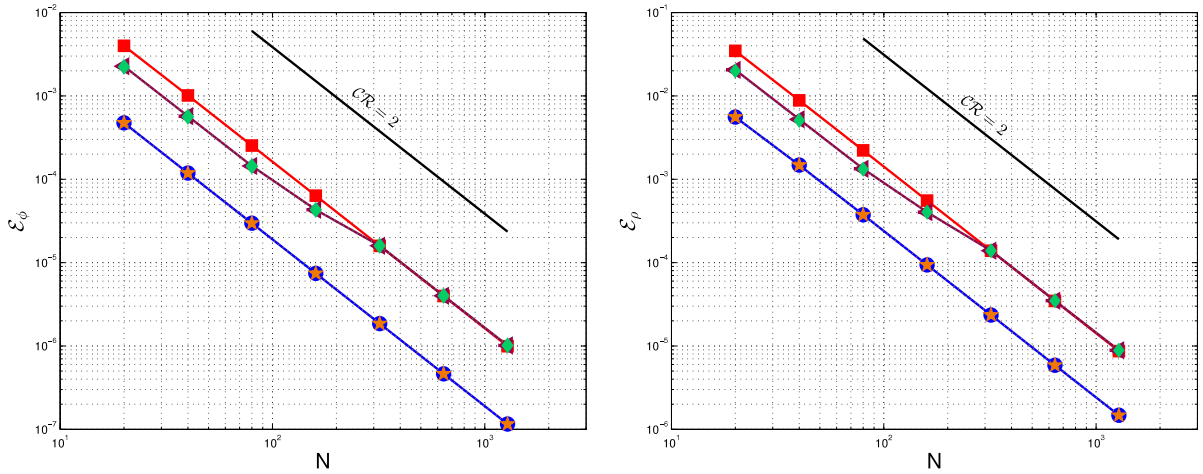

Fig. 3. Convergence rate in maximal discrete norm for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), IMEX 2 (line with triangles), ItS 1 (line with circles), ItS 2 (line with diamonds) and ItS 1(1) (line with stars), $\tau = h^2$; comparison line (solid black line), indicating the slope for the exact second order of convergence.

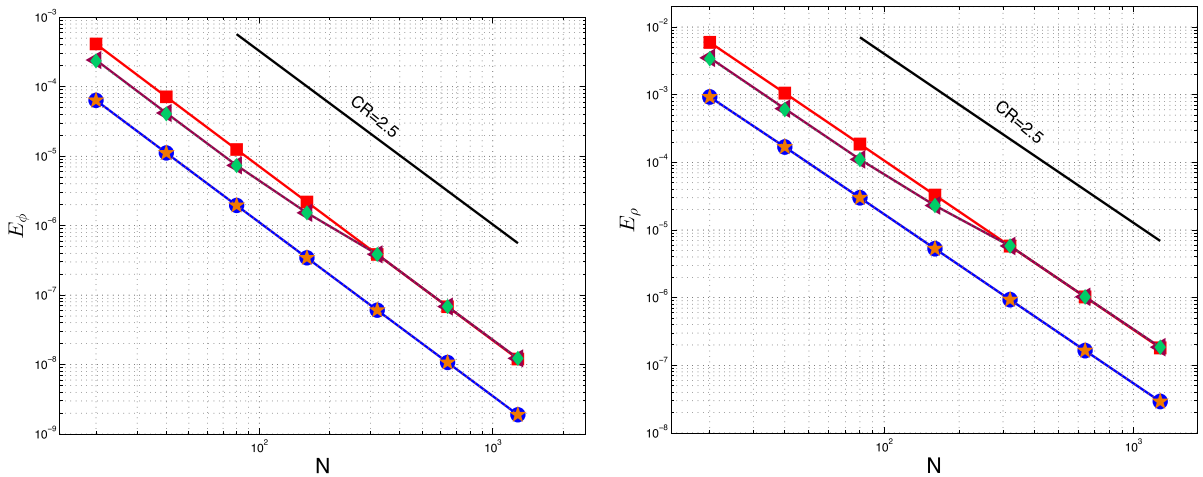

Fig. 4. Convergence rate in L_2 discrete norm for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), IMEX 2 (line with triangles), ItS 1 (line with circles), ItS 2 (line with diamonds) and ItS 1(1) (line with stars), $\tau = h^2$; comparison line (solid black line), indicating the slope for the exact order of convergence 2.5.

Let $\gamma = 2$. On Figs. 3 and 4 we plot errors in maximal and L_2 norms, respectively, versus the number of space grid nodes N in logarithmic scale. The slopes of the obtained lines correspond to the order of convergence in space — second in the maximal norm and 2.5 in L_2 norm. Because of the fixed ratio between the time and the space step size ($\tau = h^2$), we deduce that the order of convergence (in maximal norm) in time is not less than one. We detect better accuracy for ItS 1 and ItS 1(1).

On Figs. 5 and 6 we depict errors in maximal and L_2 norms, respectively, versus the CPU time (in seconds) in logarithmic scale. Obviously, the iteration scheme ItS 1 is more efficient in comparison with the corresponding non-iteration scheme (IMEX 1), while for IMEX 2 and ItS 2 we have just the opposite situation. We observe better efficiency of ItS 1 in comparison with all other considered methods.

In Table 1 we give the average number of iterations at each time level (required to reach the desired precision) for different number of space grid nodes, for ItS 1, ItS 2 and ItS 1(1). Although, the ItS 2 requires smaller number of iterations, the computational process is more time consummative in juxtaposition with ItS 1 and ItS 1(1), see Figs. 5 and 6.

Let $\gamma = 1$. With this test example, we show that despite of the theoretical time step restriction, the numerical solution, computed by ItS 1, $\tau = h$ converges to the exact one. In this case, $N = J$ and the time step is dominated over h^2. Thus, by the computations with consecutively double refined meshes, we get the order of convergence in time.

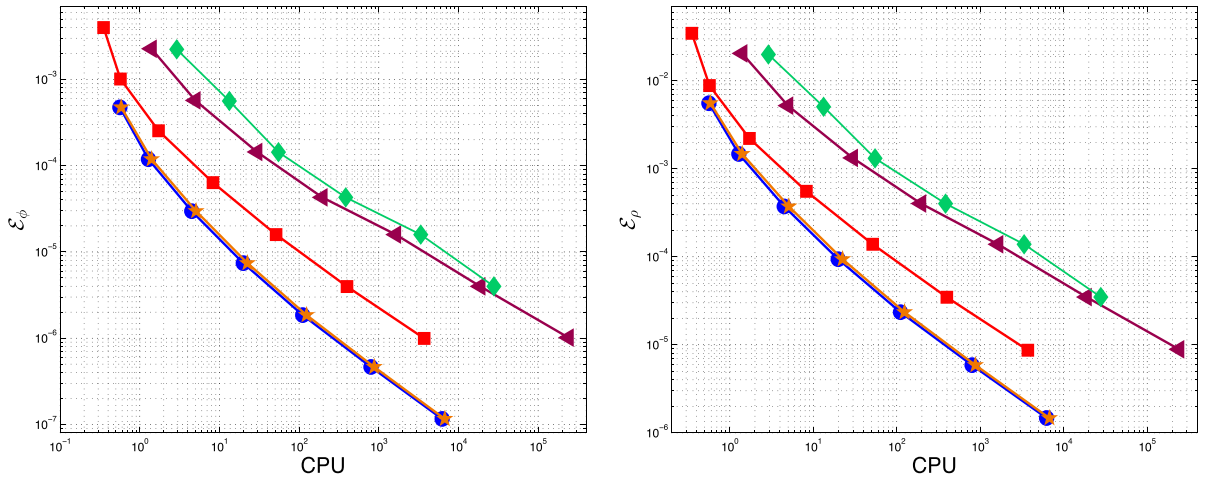

Fig. 5. Error in maximal discrete norm vs. CPU time for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), IMEX 2 (line with triangles), ItS 1 (line with circles), ItS 2 (line with diamonds) and ItS 1(1) (line with stars), $\tau = h^2$.

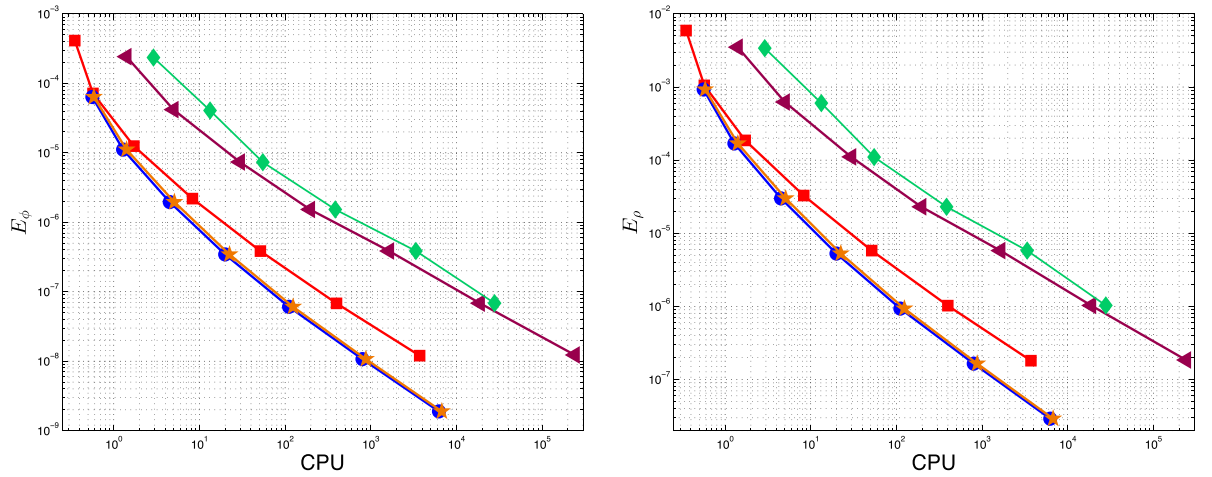

Fig. 6. Error in L_2 discrete norm vs. CPU time for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), IMEX 2 (line with triangles), ItS 1 (line with circles), ItS 2 (line with diamonds) and ItS 1(1) (line with stars), $\tau = h^2$.

Table 1
Average number of iterations at each time level for ItS 1, ItS 2 and ItS 1(1), $\tau = h^2$.

N	20	40	80	160	320	640	1280
ItS 1	5.915	4.768	4.079	3.413	3.000	3.000	2.679
ItS 2	3.980	3.995	3.000	3.000	3.000	2.000	2.000
ItS 1(1)	5.915	4.768	4.079	3.413	3.000	3.000	2.679

On Fig. 7 we plot errors in maximal norm versus the number of space grid nodes N in logarithmic scale for ItS 1, ItS 1(1) and IMEX 1. The slopes of the obtained lines correspond to first order of convergence in time. A better precision is achieved by ItS 1 and ItS 1(1).

Fig. 8 represents errors in maximal norm versus the CPU time (in seconds) in logarithmic scale for ItS 1, ItS 1(1) and IMEX 1. It is evident, that ItS 1 performs faster than ItS 1(1) and IMEX 1. The average number of iterations at each time level for ItS 1 and ItS 1(1), $\tau = h$ are given in Table 2.

7. Conclusions

We proposed accurate implicit and implicit–explicit difference schemes to simulate one-dimensional motion of magma. Specifically, we developed robust iterative algorithms for solving the non-linear systems of difference equations. Positivity,

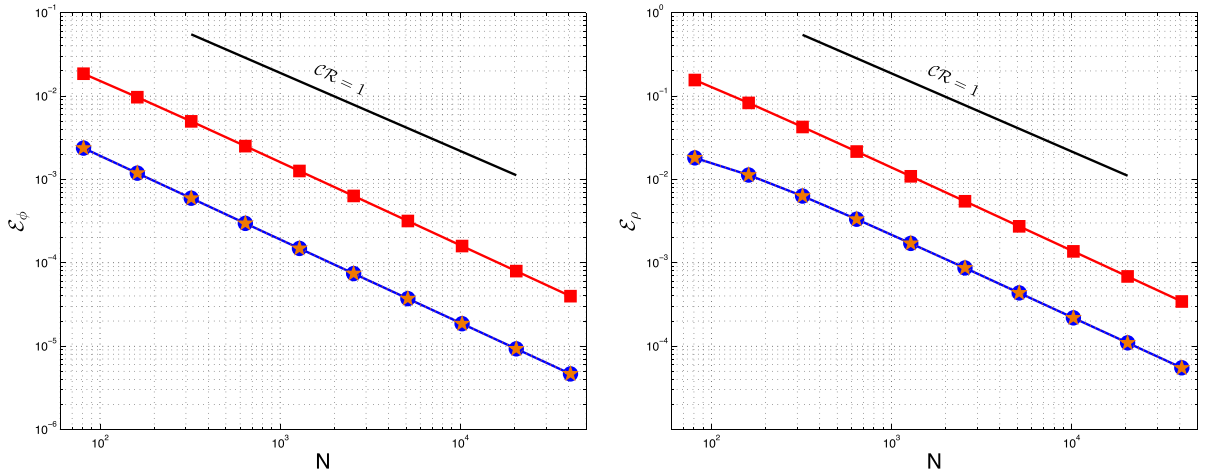

Fig. 7. Convergence rate in maximal discrete norm for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), ItS 1 (line with circles), ItS 1(1) (line with stars), $\tau = h$; comparison line (solid black line), indicating the slope for the exact first order of convergence.

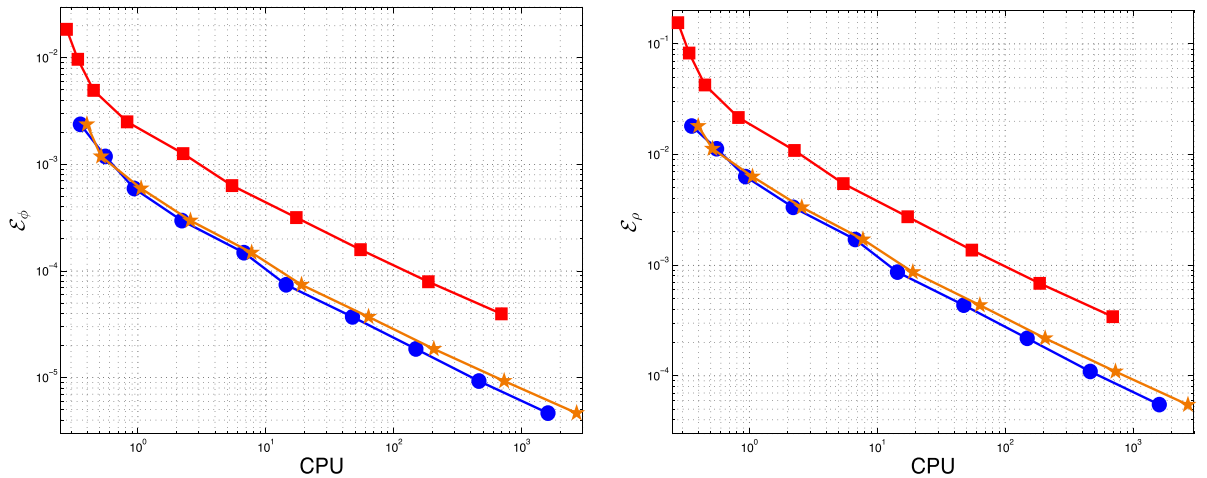

Fig. 8. Error in maximal discrete norm vs. CPU time for ϕ (*left*) and ρ (*right*) for IMEX 1 (line with squares), ItS 1 (line with circles) and ItS 1(1) (line with stars), $\tau = h$.

Table 2
Average number of iterations at each time level for ItS 1 and ItS 1(1), $\tau = h$.

N	80	160	320	640	1280	2560	5120	10241	20481	40961
ItS 1	8.341	7.086	6.230	5.470	5.003	4.407	4.140	4.000	3.538	3.217
ItS 1(1)	8.341	7.086	6.230	5.470	5.000	4.407	4.140	4.000	3.539	3.217

boundness and conservation properties of the numerical solutions are studied. The proposed schemes are tested on magma motion examples with near-real data. We observe better performance of the iteration scheme ItS 1, which is realized by improved Thomas method.

Our future work will be focused on the following directions — derivation and numerical analysis of the 2D extension of the model (12)–(16); investigation of the cases, when the porosity (volume-fraction of fluid) vanishes, see e.g. [8,10,11].

Acknowledgments

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions.

This research is supported by the Bulgarian National Science Fund under Bilateral Project DNTS/Russia 02/12 "Development and investigation of finite-difference schemes of higher order of accuracy for solving applied problems of fluid and gas mechanics, and ecology" from 2018.

References

[1] J. Bear, Dynamics of Fluids in Porous Media, Elsevier, New York, 1972.
[2] A.A. Papin, M. Tokareva, On local solvability of the system of the equations of one dimensional motion of magma, J. Sib. Fed. Univ. Math. Phys. 10 (1) (2017) 385–395.
[3] M. Simpson, M. Spiegelman, M.I. Weinstein, Degenerate dispersive equations arising in the study steady of magma dynamics, Nonlinearity 20 (2007) 21–40.
[4] Z. Chen, G. Huan, Y. Ma, Computational Methods for Multiphase Flows in Porous Media, in: Computational Science and Engineering Series, vol. 2, SIAM, Philadelphia, PA, 2006.
[5] I.G. Akhmerova, A.A. Papin, M.A. Tokareva, Mathematical Models of Heterogeneous Media, Part 1, Altai Gos. Univ. Barnaul, 2012, (in Russian).
[6] J.A.D.Connolly Y.Y.Podladchikov, Compaction-driven fluid flow in viscoelastic rock, Geodin. Acta 11 (1998) 55–84.
[7] T. Arbogast, M. Obeyeskere, M.F. Wheeler, Numerical methods for the simulation of flow in root-soil systems, SIAM J. Numer. Anal. 30 (6) (1993) 1677–1702.
[8] T. Arbogast, M.A. Hesse, A.L. Taicher, Mixed methods for two-phase Darcy-Stokes mixtures of partially melted materials with regions of zero porosity, SIAM J. Sci. Comput. 39 (2) (2017) B375–B402.
[9] F.A. Radu, K. Kumar, J.M. Nordbotten, I.S. Pop, A. robust, Mass conservative scheme for two-phase flow in porous media including Hoelder continuous nonlinearities, IMA J. Numer. Anal. 38 (2) (2018) 884–920.
[10] T. Arbogast, A.L. Taicher, A linear degenerate elliptic equation arising from two-phase mixtures, SIAM J. Numer. Anal. 54 (5) (2016) 3105–3122.
[11] T. Arbogast, A.L. Taicher, A cell-centered finite difference method for a degenerate elliptic equation arising from two-phase mixtures, Comput. Geosci. 21 (4) (2017) 701–712.
[12] K. Brenner, C. Cancès, D. Hilhorst, Finite volume approximation for an immiscible two-phase flow in porous media with discontinuous capillary pressure, Comput. Geosci. 17 (2013) 573–597.
[13] X. Cao, S.F. Nemadjieu, S. Pop, Convergence of a MPFA finite volume scheme for a two phase porous media flow model with dynamic capillarity, IMA J. Numer. Anal. (2018) (published online 27 January 2018), http://dx.doi.org/10.1093/imanum/drx078.
[14] F.A. Radu, J.M. Nordbotten, I.S. Pop, K. Kumar, A robust linearization scheme for finite volume based discretizations for simulation of two-phase flow in porous media, J. Comput. Appl. Math. 289 (2015) 134–141.
[15] C. Morency, R.S. Huismans, C. Beaumont, P. Fullsack, A numerical model for coupled fluid flow and matrix deformation with applications to disequilibrium compaction and delta stability, J. Geophys. Res. 112 (2007) B10407, http://dx.doi.org/10.1029/2006JB004701.
[16] C. Etchegaray, N. Meunier, Numerical solutions of a 2D fluid problem coupled to a nonlinear non-local reaction-advection-diffusion problem for cell crawling migration in a discoidal domain, in: R. Anguelov, M. Lachowicz (Eds.), Mathematical Methods and Models in Biosciences, pp. 122–139, http://dx.doi.org/10.11145/texts.2018.03.113.
[17] A.A. Papin, I.G. Akhmerova, Sovability of the system of equations of one-dimensional motion of a heat-conducting two-phase mixture, Math. Notes 87 (2) (2010) 230–243.
[18] A.A. Samarskii, The Theory of Difference Schemes, Marcel Dekker Inc, 2001.
[19] J.M. Peña, M-matrices whose inverses are totally positive, Linear Algebra Appl. 221 (1995) 189–193.
[20] I. Faragó, R. Horváth, Discrete maximum principle and adequate discretizations of linear parabolic problems, SIAM J. Sci. Comput. 28 (2006) 2313–2336.
[21] V.N. Nikolaevskiy, Geomechanics and Fluidodynamics with Applications to Reservoir Engineering, Springer Netherlands, 1996, 352 p., ISBN 978-90-481-4638-3 (Nedra, Moskow, 1996, 447p., ISBN 5-247-03675-1, original Russian edition).
[22] D.L. Turcotte, G. Schubert, Geodynamics: Application of Continuum Physics to Geological Problems, Wiley, 1982, 450p.