Computers and Mathematics with Applications 102 (2021) 15–44

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

www.elsevier.com/locate/camwa

Spectral discretizations analysis with time strong stability preserving properties for pseudo-parabolic models

Eduardo Abreu [a,*], Angel Durán [b]

[a] University of Campinas, Department of Applied Mathematics, 13.083-970, Campinas, SP, Brazil
[b] University of Valladolid, Department of Applied Mathematics and Institute of Mathematics (IMUVA), 47011 Valladolid, Spain

ARTICLE INFO

Keywords:
Pseudo-parabolic equations
Spectral methods
Error estimates
Strong stability preserving methods
Non-regular data

ABSTRACT

In this work, we study the numerical approximation of the initial-boundary-value problem of nonlinear pseudo-parabolic equations with Dirichlet boundary conditions. We propose a discretization in space with spectral schemes based on Jacobi polynomials and in time with robust schemes attending to qualitative features such as stiffness and preservation of strong stability for a more correct simulation of non-regular data. Error estimates for the corresponding semidiscrete Galerkin and collocation schemes are derived. The performance of the fully discrete methods is analyzed in a computational study.

1. Introduction

1.1. Motivation

In recent years, pseudo-parabolic differential equations in a bounded domain, or in the whole space, have been studied extensively covering many several foundation aspects, e.g., modelling, numerics and theory (see the list of references in Section 1.2). We are particularly motivated in the modelling of complex and challenging multiphase flow problem in porous media [16,32,56,57,60,88,96]. Specifically, we consider the nonlinear two-dimensional system of the two-phase flow model

$$\frac{\partial}{\partial t}(\phi S_w) + \nabla \cdot \mathbf{F}(\mathbf{v}, S_w) = -\nabla \cdot [H_c(S_w)\nabla p_c], \tag{1.1a}$$

$$p_c = p_e(S_w) - \tau \frac{\partial}{\partial t}(\phi S_w). \tag{1.1b}$$

The system (1.1a)-(1.1b) involves capillary forces, general expressions for the relative permeability functions, variable porosity and permeability fields and the gravity effect. It consists of two equations: the first one (1.1a) is a nonlinear convection-reaction-diffusion equation for the wetting phase saturation S_w, [15,60], where ϕ is the rock porosity, $\mathbf{v}$ denotes the total velocity (of wetting and non-wetting phases), p_c is the capillary pressure (or pressure difference between the non-wetting and wetting phases), $H_c = H_c(S_w)$ is the capillary induced diffusion function and $\mathbf{F} = \mathbf{F}(\mathbf{v}, S_w)$ is the convection flux or water fractional flow function, with p_e denoting the equilibrium capillary pressure. The pseudo-parabolic character of the transport problem comes from the time-dependent relation, proposed in [57], and given by (1.1b), an elliptic-type equation for the dynamic capillary pressure, where τ is the so-called dynamic effect coefficient. (For a more detailed description, see [5,24] and references therein.) In addition, there is a close interplay among discretization methods, numerical analysis, and simulations discussed in this work for multiphase flow in porous media to other interesting related problems, as such Sobolev type equations, Benjamin Bona Mahony problem, the improved Boussinesq equation and the Körteweg-de Vries problem (see, e.g., [76,82,29,65,68,89,16]).

The complex dynamics of (1.1a)-(1.1b) motivates recent challenging approaches, like new conceptual notions of weak entropy solutions with specific modeling assumptions for scalar and multi-dimensional problems with discontinuous coefficients and nonlinearities in space-time, discussed in [1–4,28,35]. However, most of the work developed for (1.1a)-(1.1b) concerns the one-dimensional version

* Corresponding author.
E-mail addresses: eabreu@ime.unicamp.br (E. Abreu), angeldm@uva.es (A. Durán).

https://doi.org/10.1016/j.camwa.2021.10.001
Received 20 November 2020; Received in revised form 1 August 2021; Accepted 3 October 2021

$$\frac{\partial}{\partial t}(\phi S)+\frac{\partial F(S)}{\partial x}=-\frac{\partial}{\partial x}\left(H(S)\frac{\partial}{\partial x}\left(J(S)-\tau\frac{\partial}{\partial t}(\phi S)\right)\right). \tag{1.2}$$

The suitability of (1.2) for modelling the two-phase flow in porous media is discussed in the literature from several points of view. They are related, of course, to well-posedness (existence and uniqueness of solution), [25,34], and to the use of (1.2) as a pseudo-parabolic regularization to derive existence results for the Buckley-Leverett (BL) model, [15]. This is proposed in [33] and applied in [36] for controlling the dissipation when approximating the BL equation. In addition, the numerical simulation with discontinuous data for (1.2) and extensions incorporating capillary pressure hysteresis is considered in [41,92,113].

1.2. Scope

From the above discussion and inspired by the 1-D pseudo-parabolic equation (1.2), the present paper is concerned with the initial-boundary-value problem (ibvp) of pseudo-parabolic type in $\Omega=(-1,1)^d, d=1,2$

$$cv_t-\nabla\cdot(a\nabla v_t)=-\nabla\cdot(\alpha\nabla v)+\beta\cdot\nabla v+\gamma,\ \underline{x}\in\Omega,\ t>0, \tag{1.3}$$

$$v(\underline{x},0)=v_0(\underline{x}),\ \underline{x}\in\Omega, \tag{1.4}$$

$$v=0,\ \text{on}\ \partial\Omega,\ t>0. \tag{1.5}$$

The notation in (1.3)-(1.5) attempts to cover both the one- and two-dimensional problems, for $v=v(\underline{x},t)$, with the following conventions and assumptions:

(C1) In 2D, $\underline{x}=(x,y)$, Ω is the square $(-1,1)^2:=(-1,1)\times(-1,1)$, with $\partial\Omega$ denoting the four edges conforming its boundary, ∇ denotes the gradient operator, and the dot $\cdot$ stands for the Euclidean inner product in $\mathbb{R}^2$, either between vector fields or in operational terms. In the one-dimensional case, $\underline{x}=x, \Omega=(-1,1), \partial\Omega=\{-1,1\}, \nabla=\partial_x$, and the dot is the usual product in $\mathbb{R}$.

(C2) We assume that $a,c:\Omega\to\mathbb{R}$ are continuously differentiable functions in Ω and bounded above and below by positive constants. The right hand side of (1.3) involves linear and nonlinear terms which are assumed to be continuously differentiable functions of their arguments $\underline{x},t$, and v. While $\alpha=\alpha(\underline{x},t,v)$, and $\gamma=\gamma(\underline{x},t,v)$ are always scalar functions, β is a vector field with two components $\beta_i, i=1,2$ when $d=2$, and a scalar function when $d=1$.

As cited in [43], the origin of pseudo-parabolic (or Sobolev-type) problems of the form (1.3) can be found in [103,107]. From a mathematical point of view, their study was initiated, as far as theoretical aspects are concerned, in the paper by Showalter and Ting, [99], and in [46], while the first results on numerical analysis are obtained, to our knowledge, in [44,45], for discretizations with finite differences of the one-dimensional case. First references of (1.3) as physical model can be found in heat conduction, [27], and in fluid-flow, [106].

As an extension of the references cited above, we can make a brief review of the literature for (1.3)-(1.5) and related problems. As far as modelling is concerned, linear and nonlinear versions (with local and nonlocal terms) of pseudo-parabolic-type equations appear as mathematical models of fundamental phenomena, [11,13,14,25,27,29,31,33,34,40,41,47,57,63–65,68,70,72,74,75,79–81,83,86,90,113]. More specifically, they model imprisoned radiation through a gas, [63,64,81], fluid flow in fissured rock, [14], heat conduction in heterogeneous media, [27,93], out-of-equilibrium viscoelastic relaxation effect, [83], and porous media applications, [57,66,104] – see also [5,29,61,62,65,113] for a good survey. Several combined numerical-analytical studies about pseudo-parabolic equations linked to fluid flow problems might be found in [14,33,34,40,41]. Local pseudo-parabolic equations also appear in the study of two-phase flow models with dynamic capillary pressure and hysteresis [113,87]. On the other hand, nonlocal pseudo-parabolic equations, [13,27,72], describe a variety of physical phenomena, such as the seepage of homogeneous fluids through a fissured rock, heat conduction problems with thermodynamic temperature and conduction temperature, and the analysis of nonstationary processes in semiconductors in the presence of sources.

The mathematical theory of pseudo-parabolic equations can be covered by [19,25,30,40,95,99,100]. Existence and uniqueness of weak solutions to nonlinear pseudo-parabolic equations are proved in [90], whereas the existence of weak solutions for degenerate cases is studied in [79,80]. A homogenization of a closely related pseudo-parabolic system is considered in [86]. Travelling wave solutions and their relation to non-standard shock solutions to hyperbolic conservation laws are investigated in [31,33] for linear higher order terms. Uniqueness of weak solutions for a pseudo-parabolic equation modelling flow in porous media can be found in [25,74]. In [47], the authors study existence and uniqueness of weak solutions of the initial and boundary value problem for a fourth-order pseudo-parabolic equation with variable exponents of non-linearity, along with a long-time behaviour of weak solutions. Finally, existence of weak solutions for a nonlocal pseudo-parabolic model for the Brinkman two-phase flow model in porous media has been recently established [68].

Concerning the numerical approximation of equations of the form (1.3), the literature contains many references involving finite differences, [8–10,41,105], as well as finite elements and finite volumes [5,6,76,82,111]. We also mention some convergence results. First, stability and convergence of difference approximations to pseudo-parabolic partial differential equations is discussed in [44,45] and the time stepping Crank-Nicolson Galerkin method to approximate several nonlinear Sobolev-type problems is analyzed in [38,39]. Of particular relevance for the present study is the finite element approach for the nonlinear periodic-initial-boundary-value problem, [12], where Arnold et al. obtain optimal error estimates, in L^2 and H^1 norm, of a standard Galerkin method with continuous piecewise polynomials, and a nodal superconvergence. Moreover, Fourier spectral methods of Galerkin and collocation type for quasilinear pseudo-parabolic equations are analysed in [91]. On the other hand, in [114] an error analysis of a Chebyshev-Legendre pseudospectral discretization for the Dirichlet problem of a class of one-dimensional, nonclassical parabolic equations is performed. The equations are of the form of the 1D version of (1.3) with $c=1$, $a\in[0,1]$ constant, $\alpha=-1, \beta=0$. The estimates, in the L^2 norm and the H^1 seminorm, depend on the parameter a and the regularity of the data. In particular, they show spectral accuracy. A full discretization with the Crank-Nicolson/leapfrog method as time integrator is also introduced and illustrated with numerical experiments. More convergence results can be found in [70] (see also [69]), where an analysis of a linearization scheme for an interior penalty discontinuous Galerkin for a pseudo-parabolic model in porous media applications is considered. High-order finite differences are employed in [11] and B-spline quasi-interpolation methods in [75]. In addition, an adaptive mesh approach for pseudo-parabolic-type problems is introduced in [29] and a meshless method, based on radial basis functions (RBFs), is considered in [65]. The use of wavelets (Haar basis) in the spatial discretization of two-dimensional Sobolev equations is proposed in [53], where the temporal discretization is carried out with finite differences. Finally, unconditionally stable vector splitting schemes for

pseudo-parabolic equations are constructed and analyzed in [108]. It is worthwhile to mention that standard operator splitting may fail to capture the correct behaviour of the solutions for pseudo-parabolic type differential models. In [6], the authors presented a non-splitting numerical method which is based on a fully coupled space-time mixed hybrid finite element/volume discretization approach to account for the delicate nonlinear balance between the hyperbolic flux and the pseudo-parabolic term linked to the full pseudo-parabolic differential model.

1.3. Aims and contributions of the present work

In the search for new approaches to the numerical approximation of pseudo-parabolic problems of the form (1.3)-(1.5), several points conform the aim of the present paper. Concerning the spatial discretization, we focus on the performance of spectral methods. As in the periodic case, studied by Quarteroni, [91], and the 1D problem approximated by the Chebyshev and Legendre pseudospectral schemes in [114], our first aim is analyzing the convergence of spectral methods to approximate the Dirichlet problem (1.3)-(1.5) in space. More specifically, for Galerkin and collocation methods based on a family of Jacobi polynomials, existence of solution of the semidiscrete systems and error estimates in suitable Sobolev norms are derived. The results proved here establish the rate of convergence of the spectral approximation in terms of the regularity of the coefficients and initial condition of the problem: for data in $C^m(\Omega)$ and if N is the degree of the polynomial approximation, then spectral Galerkin error is shown to decrease as $O(N^{-m})$ or $O(N^{1-m})$, while spectral collocation error behaves like $O(N^{2-m})$.

On the other hand, temporal discretization also contributes to this search for alternatives of approximation to (1.3)-(1.5) with the introduction of additional properties in the requirements for the time integrator to improve the quality of the simulation. In addition to the classical quantitative features concerning convergence, our attention is focused on two aspects of the problem. The first one is the possible mildly stiff character (which depends on the relative magnitude of the higher-derivative terms); this point may recommend the use of fully or diagonally implicit methods. A second aspect to be taken into account concerns the use of strong stability preserving (SSP) methods, [52,50,51], as time integrators. Construction and analysis of SSP methods for hyperbolic partial differential equations have the aim at preserving the nonlinear stability (in some norm or, more generally, convex functional) of spatial discretizations with respect to the forward Euler method. This SSP property makes influence in a better simulation of discontinuous solutions, avoiding the presence of spurious oscillations and reducing the computational cost. Here we are interested in studying the performance of these methods in problems (1.3)-(1.5) with non-regular data. Our study must also take into account the so-called dispersive order of the temporal integration (cf. [67] and references therein) and, in order to reduce the phase error, choosing schemes with dispersive stability functions.

Finally, a numerical study with representative experiments, involving linear and nonlinear equations, is made to analyze computationally the performance of the proposed approach and its potential usefulness for related problems. In particular, the experiments will serve us to analyze the order of convergence of the spectral discretizations as well as to address the behaviour of the schemes with respect to the regularity of the data. For illustrative purposes and since these Legendre and Chebyshev families are mostly used in other applications, the description and computational study will be focused on two semidiscretizations: a Legendre Galerkin method and a Chebyshev collocation scheme. We believe they cover most of the numerical aspects of our whole proposal for this spectral approach but, indeed, other semidiscretizations can be chosen. As far as the temporal integration is concerned, according to the pseudo-parabolic character of the equation and possible stiffness, with oscillations, of the semidiscretizations, two SSP methods of a family of singly diagonally implicit Runge-Kutta (SDIRK) schemes are taken. Among other properties, the methods have order of convergence two and three, respectively, they are A-stable (hence L-stable), have dispersive stability functions with phase order two and four (respectively), and a good computational performance for solving iteratively the intermediate stages.

The paper is divided into two parts as follows. Section 2 concerns the analysis of convergence of the spectral semidiscretization in space. Sections 2.1 and 2.2 are devoted to some theoretical aspects of (1.3)-(1.5) as the weak formulation and some assumptions on well-posedness. These preliminaries also include a summary on inverse inequalities, as well as projection and interpolation error estimates for the family of Jacobi polynomials under consideration. All this will be used to the numerical analysis of the spectral Galerkin approximation in Section 2.3, and the collocation approximation in Section 2.4. Both contain, under suitable hypotheses on the data of the problem, results on the existence of numerical solution and convergence to the solution of (1.3)-(1.5). The second part of the paper is in Section 3 and concerns the time discretization and a computational study of the performance of the resulting fully discrete schemes. Sections 3.1 and 3.2 consist of a description of the semidiscrete systems corresponding to the Legendre Galerkin and Chebyshev collocation spectral methods. The description includes details on formulation and practical implementation. Section 3.3 is devoted to the full discretization. Dispersion and strong stability preserving properties of the selected SDIRK schemes as time integrators are discussed, as well as several implementation details. The full discretizations will be then ready for performing the computational study in Section 3.4. Section 4 summarizes the results and outlines possible directions for future research.

1.4. Preliminaries

We now describe the main notation used throughout the paper. As in the presentation of the problem (1.3)-(1.5), the distinction between the cases $d = 1$ and $d = 2$ will be made when necessary.

For positive integer p, $L^p(\Omega)$ denotes the normed space of L^p-functions on Ω with $||\cdot||_p$ as associated norm, while for nonnegative integer m, $C^m(\overline{\Omega})$ is the space of m-th order continuously differentiable functions on $\overline{\Omega} := \Omega \cup \partial\Omega$. Let $-1 < \mu < 1$ and $d = 1$. We consider the Jacobi weight function

$$w(x) = w_\mu(x) = (1 - x^2)^\mu,\ x \in \Omega. \tag{1.6}$$

($\mu = 0$ corresponds to the Legendre case and $\mu = -1/2$ to the Chebyshev case.) When $d = 2$, (1.6) becomes

$$w(x) = w_\mu(\underline{x}) = (1 - x^2)^\mu (1 - y^2)^\mu,\ \underline{x} = (x, y) \in \Omega. \tag{1.7}$$

The function (1.6) or (1.7) is used to introduce the following weighted Sobolev spaces. By $L^2_w = L^2_w(\Omega)$ we denote the space of squared integrable functions with respect to the weighted inner product

$$(\phi, \psi)_w = \int_\Omega \phi(\underline{x})\psi(\underline{x})w(\underline{x})d\underline{x},\ \phi, \psi \in L^2_w, \tag{1.8}$$

and associated norm $||\phi||_{0,w} = (\phi,\phi)_w^{1/2}$. For the Sobolev spaces $H_w^k = H_w^k(\Omega), k \geq 0$ integer (where $H_w^0 = L_w^2$) the corresponding norm will be denoted by

$$||\phi||_{k,w}^2 = \sum_{j=0}^{k} ||\frac{d^j}{dx^j}\phi||_{0,w}^2, \tag{1.9}$$

when $d = 1$ and

$$||\phi||_{k,w}^2 = \int_{\Omega} \sum_{p+q\leq k} \left(\frac{\partial^{p+q}\phi}{\partial x^p \partial y^q}\right)^2 d\underline{x}, \tag{1.10}$$

when $d = 2$, with the definition of $H_w^s = H_w^s(\Omega)$ for real $s \geq 0$ as the corresponding interpolate space, [18,7]. On the other hand, for real $s \geq 0$, $-1 < \mu < 1$, $H_{w,0}^s = H_{w,0}^s(\Omega)$ is defined as the closure of the space $D(\Omega)$ of infinitely differentiable functions with compact support in Ω. The trace theorem, [18], characterizes $H_{w,0}^s$ as the space of functions $\phi \in H_w^s$ such that

$$\phi\Big|_{\Gamma_j} = \frac{\partial\phi}{\partial n_j} = \cdots = \frac{\partial^{m-1}\phi}{\partial n_j} = 0,\ 1 \leq j \leq 4, \tag{1.11}$$

where in (1.11) n_j is the unit outward normal to Ω on the corresponding edge Γ_j of $\partial\Omega$, $1 \leq j \leq 4$ and, assuming that $s + (1-\mu)/2$ is not integer, m is its integral part. In the case $d = 1$, $H_{w,0}^k$ is then the space of functions $\phi \in H_w^k$ such that $\phi(-1) = \phi(1) = 0$. Note that in the case of the Legendre approximation ($w(x) = 1$) the spaces $H_w^s, H_{w,0}^s$ are the standard Sobolev spaces H^s, H_0^s.

For an integer $N \geq 2$, $\mathbb{P}_N = \mathbb{P}_N(\Omega)$ will stand for the space of polynomials on Ω of degree at most N with respect to each variable and let

$$\mathbb{P}_N^0 = \mathbb{P}_N^0(\Omega) := \mathbb{P}_N(\Omega) \cap H_0^1(\Omega),$$

be the subspace of polynomials in $\mathbb{P}_N(\Omega)$ vanishing on the boundary $\partial\Omega$.

If $T > 0$ and $1 \leq p \leq \infty$, $L^p(0,T)$ stands for the space of L^p functions on $[0,T]$ with norm $|\cdot|_p$. For an integer $k \geq 0$, the space of m-th order continuously differentiable functions $u : [0,T] \to X$, where $X = H_w^s$ or $H_{w,0}^s, s \geq 0$, will be denoted by $C^k(0,T,X)$. Additionally, if $0 < k < \infty$, $L^k(0,T,X)$ will stand for the normed space of measurable functions $u : [0,T] \to X$ with associated norm

$$||u||_{L^k(0,T,X)} = \left(\int_0^T ||u(t)||_{s,w}^k dt\right)^{1/k}.$$

We also denote by $L^\infty(0,T,X)$ the space of functions $u : [0,T] \to X$ with finite norm

$$||u||_{L^\infty(0,T,X)} = \text{esssup}_{t\in[0,T]}||u(t)||_{s,w},$$

where essesup stands for the essential spectrum. Furthermore, $C^1(\Omega\times[0,T]\times\mathbb{R})$ (resp. $C_b^1 = C_b^1(\Omega\times[0,T]\times\mathbb{R})$) will stand for the space of continuously differentiable (resp. uniformly bounded, continuously differentiable) functions $f = f(\underline{x},t,v)$ in $(\underline{x},t,v) \in \Omega\times[0,T]\times\mathbb{R}$.

The analysis of the collocation methods requires the introduction of discrete norms. Let $\{x_j, w_j\}_{j=0}^N$ be the nodes and weights of the Gauss-Lobatto quadrature related to $w(x)$, [18,23,78]. For ϕ,ψ continuous on $\overline{\Omega}$, the discrete inner product based on the Gauss-Lobatto data is denoted by

$$(\phi,\psi)_{N,w} = \sum_{j=0}^{N} \phi(x_j)\psi(x_j)w_j, \tag{1.12}$$

with associated norm $||\phi||_{N,w} = (\phi,\phi)_{N,w}^{1/2}$. We recall that, [23]

$$(\phi,\psi)_{N,w} = (\phi,\psi)_w, \tag{1.13}$$

if $\phi\psi \in \mathbb{P}_{2N-1}$. The equivalence of the norms $||\phi||_{N,w}$ and $||\phi||_{0,w}$ when $\phi \in \mathbb{P}_N$, established in the following lemma, was proved in [22] for the case of Legendre and Chebyshev weights and in [17] for (1.6) with $\mu > -1$.

Lemma 1.1. *Let $N \geq 2$ be an integer. Then there exist positive constants C_1, C_2, independent of N, such that for any $\phi \in \mathbb{P}_N$*

$$C_1||\phi||_{0,w} \leq ||\phi||_{N,w} \leq C_2||\phi||_{0,w}.$$

As far as the case $d = 2$ is concerned, the corresponding inner product, cf. (1.12)

$$(\phi,\psi)_{N,w} = \sum_{j=0}^{N}\sum_{k=0}^{N} \phi(\underline{x}_{jk})\psi(\underline{x}_{jk})w_{jk}, \tag{1.14}$$

for ϕ,ψ continuous on $\overline{\Omega}$, makes use of the Gauss-Lobatto grid, [17]

$$\underline{x}_{jk} = (x_j, x_k), \quad w_{jk} = (w_j, w_k), \quad 0 \leq j,k \leq N. \tag{1.15}$$

For the sake of simplicity, the corresponding norm is also denoted by $||\cdot||_{N,w}$. Observe that, since the quadrature formula

$$\int_{\Omega} \psi(\underline{x})w(\underline{x})d\underline{x} \approx \sum_{j,k=0}^{N} \psi(\underline{x}_{jk})w_{jk}, \tag{1.16}$$

is exact on $\mathbb{P}_{2N-1}$, [17], then (1.13) also holds in this case.

Remark 1.1. In some cases the previous inner products and norms will have to be extended to vector fields of two components in the expected way. Thus, if $\phi=(\phi_1,\phi_2),\psi=(\psi_1,\psi_2)\in L^2_w\times L^2_w$ we define

$$(\phi,\psi)_w := (\phi_1,\psi_1)_w+(\phi_2,\psi_2)_w, \tag{1.17}$$

where on the right-hand side (1.8) is used for the components. Similarly, for continuous vector fields $\phi=(\phi_1,\phi_2),\psi=(\psi_1,\psi_2)$

$$(\phi,\psi)_{N,w} := (\phi_1,\psi_1)_{N,w}+(\phi_2,\psi_2)_{N,w}, \tag{1.18}$$

and (1.17), (1.18) lead to similar extensions for the corresponding norms $||\cdot||_{0,w},||\cdot||_{N,w}$. Since no confusion is possible and for the sake of simplicity, we use the same notation as in the scalar case.

Throughout the paper C will be used to denote a generic, positive constant, independent of N and u, but that may depend on t (this will be specified by $C(t)$).

2. Spectral semidiscretizations and error estimates

2.1. Weak formulation

The analysis of the spectral discretizations that will be made below requires some hypotheses, properties and technical results concerning (1.3)-(1.5) and the approximation in weighted norms. From now on we will fix $\mu\in(-1,1)$ and consider the weight (1.6) or (1.7). The first property to be mentioned is the weak formulation of (1.3)-(1.5)

$$A(v_t,\psi)=B(v,\psi),\ \psi\in H^1_{w,0} \tag{2.1}$$

with $v(0)=v_0$ and

$$\begin{aligned} &A(\phi,\psi)=(c\phi,\psi)_w+L_a(\phi,\psi), \\ &B(\phi,\psi)=L_\alpha(\phi,\psi)+(\beta(\phi)\cdot\nabla\phi,\psi)_w+(\gamma(\phi),\psi)_w,\ \phi,\psi\in H^1_{w,0}, \end{aligned} \tag{2.2}$$

where, for $d=d(\underline{x},t,v)$

$$L_d(\phi,\psi)=\int_\Omega d\nabla\phi\cdot\nabla(\psi w)d\underline{x}.$$

Since a is bounded above and below by positive constants, L_a is equivalent to

$$L(\phi,\psi)=\int_\Omega \nabla\phi\cdot\nabla(\psi w)d\underline{x}, \tag{2.3}$$

and therefore, [17,18,23], the bilinear form A in (2.2) is continuous on $H^1_w\times H^1_{w,0}$ and elliptic on $H^1_{w,0}\times H^1_{w,0}$, that is, there are positive constants C_1,C_2 such that for all $\phi,\psi\in H^1_{w,0}$

$$\begin{aligned} |A(\phi,\psi)| &\le C_1\left(||\phi||_{0,w}||\psi||_{0,w}+||\phi_x||_{0,w}||\psi_x||_{0,w}\right) \\ &\le C_1||\phi||_{1,w}||\psi||_{1,w},\ \phi\in H^1_w,\psi\in H^1_{w,0}, \\ A(\phi,\phi) &\ge C_2||\phi||^2_{1,w},\ \phi\in H^1_{w,0}. \end{aligned}$$

Using the weak formulation (2.1), we will assume that (1.3)-(1.5) is well-posed in the sense given by the following theorem. Its proof can be made by adapting the steps of the proof of the corresponding results for the periodic problem given in [12] for the 1D case (see also references therein).

Theorem 2.1. *Let $T>0$ and assume that $a,c\in C^1(\Omega)$, $\alpha,\beta,\gamma\in C^1_b(\Omega\times(0,T)\times\mathbb{R})$. Given $v_0\in H^1_{w,0}$, then there is a unique solution $v\in C^1(0,T,H^1_{w,0})$ of (2.1) with $||v||_{L^\infty(0,T,H^1_w)}$ bounded by a constant depending only on $||v_0||_{1,w}$ and the data of the problem. Furthermore, if $v_0\in H^k_{w,0}$ with $k>1$ integer, $a,c\in C^{k-1}(\Omega),\alpha,\beta,\gamma,\delta\in C^{k-1}(\Omega\times(0,T)\times\mathbb{R})$, then $v(t)\in H^k_{w,0}$ for all $t\in(0,T)$ and*

$$||v||_{L^\infty(0,T,H^k_w)}+||v_t||_{L^\infty(0,T,H^k_w)}\le C, \tag{2.4}$$

where C is a constant depending only on $||v_0||_{k,w}$ and the data of the problem.

2.2. Projection and interpolation errors with Jacobi polynomials

Here we collect several results concerning projection and interpolation errors with respect to the weighted inner product (1.8) that will be used below. We refer to, e.g., [17,18,22,23,49,78,77,98] for details and additional properties. As in previous sections, the one- and two-dimensional cases will be discussed separately when necessary.

Let $\mu\in(-1,1)$ and consider the weight (1.6) or (1.7). The spectral discretizations whose error will be estimated in the sequel are based on the family of Jacobi polynomials, which form an orthogonal basis of L^2_w. They are denoted by $\{J^\mu_n\}_{n=0}^\infty$ in the 1D case. Particular cases such as Legendre and Chebyshev families correspond to $\mu=0$ and $\mu=-1/2$, respectively. Most properties of this Jacobi family (which is contained in the more general family of Jacobi polynomials $\{J^{\mu,\nu}_n\}_{n=0}^\infty$, orthogonal in $L^2_{w_{\mu,\nu}}$ with $w_{\mu,\nu}(x)=(1-x)^\mu(1+x)^\nu$) are extensions of the corresponding properties of the Legendre family, [17,18]. For the 2D case, the orthogonal basis is formed by the tensor product of the Jacobi family $\{J^\mu_n\}_{n=0}^\infty$,

$J_{\underline{n}}(\underline{x}) = J^{\mu}_{n_1}(x)J^{\mu}_{n_2}(y)$, $\underline{n} = (n_1, n_2) \in \mathbb{N}^2$.

We first consider projection errors. Let $N \geq 2$ be an integer, $v \in H^s_w, s \geq 0$, and let $P_N v \in \mathbb{P}_N$ be the orthogonal projection of v with respect to the inner product (1.8), and $P^{10}_N v \in \mathbb{P}^0_N$ be the orthogonal projection of v with respect to the inner product in $H^1_{w,0}$

$$[\phi, \psi]_w = \int_{\Omega} \nabla\phi' \cdot \nabla\psi w(\underline{x}) dx.$$

Then we have, [17,21,18]

$$||v - P_N v||_{0,w} \leq CN^{-s}||v||_{s,w},\ v \in H^s_w,\ s \geq 0, \tag{2.5}$$
$$||v - P_N v||_{r,w} \leq CN^{r-s}||v||_{s,w},\ v \in H^s_w,\ 1 \leq r \leq s,\ r \text{ integer}$$

and for $v \in H^s_w \cap H^1_{w,0}$

$$||v - P^{10}_N v||_{1,w} \leq CN^{1-s}||v||_{s,w}, s \geq 1. \tag{2.6}$$

In the Legendre and Chebyshev cases, sharper estimates hold, see [22,23].

A third projection operator used below concerns the bilinear form A given by (2.2). If $v \in H^1_{w,0}$, then the orthogonal projection $\overline{v} \in \mathbb{P}^0_N$ of v with respect to A is defined as $\overline{v} = R_N v \in \mathbb{P}^0_N$ such that

$$A(\overline{v} - v, \psi) = 0,\ \psi \in \mathbb{P}^0_N. \tag{2.7}$$

For this projection, we have, [17]

$$||v - \overline{v}||_{1,w} + N||v - \overline{v}||_{0,w} \leq CN^{1-m}||v||_{m,w}, \tag{2.8}$$

for $v \in H^m_w \cap H^1_{w,0}, m \geq 1$. Furthermore, a generalized estimate can be obtained as follows. If $v \in H^m_w, m \geq 2$, let $u^N \in \mathbb{P}_N$ be a polynomial such that, [23]

$$||u^N - v||_{k,w} \leq CN^{k-m}||v||_{m,w},\ 0 \leq k \leq 2. \tag{2.9}$$

By using (2.8), (2.9) and the inverse inequalities, [17]

$$||\psi||_{s,w} \leq CN^{2(s-r)}||\psi||_{r,w},\ \psi \in \mathbb{P}_N,\ 0 \leq r \leq s,$$

we have

$$\begin{aligned} ||v - \overline{v}||_{2,w} &\leq ||u^N - v||_{2,w} + ||u^N - \overline{v}||_{2,w} \leq CN^{2-m}||v||_{m,w} + CN^2||u^N - \overline{v}||_{1,w} \\ &\leq CN^{2-m}||v||_{m,w} + CN^2\left(||u^N - v||_{1,w} + ||v - \overline{v}||_{1,w}\right) \leq CN^{3-m}||v||_{m,w}. \end{aligned} \tag{2.10}$$

Let $N \geq 2$ be an integer, $s \geq 0, v \in H^s_w$ and let $I_N v$ denote the interpolant of v on $\mathbb{P}_N$ based on the Gauss-Lobatto-Jacobi nodes. The following estimates for the interpolation errors can be found in [17,18]: for $v \in H^s_w$

$$||v - I_N v||_{r,w} \leq CN^{r-s}||v||_{s,w}\ v \in H^s_w, \qquad 0 \leq r \leq 1, \qquad s > \sup\{\frac{d+r}{2}, \frac{r + d(1+\mu)}{2}\}, \tag{2.11}$$

where $d = 1$ or 2. An estimate comparing the continuous and discrete inner products will be necessary: if $f \in H^1_w$ and $\phi \in \mathbb{P}_N$, then

$$|(f, \phi)_w - (f, \phi)_{N,w}| \leq C\left(||f - P_{N-1}f||_{0,w} + ||f - I_N f||_{0,w}\right)||\phi||_{1,w}, \tag{2.12}$$

where $(\cdot,\cdot)_{N,w}$ is given by (1.12), [23,17].

2.3. Spectral Galerkin approximation

Let $N \geq 2$ be an integer, $T > 0$, and $v_0 \in H^1_{w,0}$. The semidiscrete Galerkin approximation is defined as the function $v^N : [0,T] \to \mathbb{P}^0_N$ satisfying

$$A(v^N_t, \psi) = B(v^N, \psi),\ \psi \in \mathbb{P}^0_N, \tag{2.13}$$
$$A(v^N(0), \psi) = A(v_0, \psi),\ \psi \in \mathbb{P}^0_N. \tag{2.14}$$

The condition (2.14), that will be used below (see proof of Theorem 2.2) states that $v^N(0)$ is the orthogonal projection of v_0 with respect to the operator A.

Remark 2.1. In what follows, we will make use of the two identities below (see, e.g., [12,91]). Let $F = F(\underline{x}, t, u)$ be a C^1 function of x, t, u. Then

$$F(v) - F(z) = (v - z)F^*(v, z), \tag{2.15}$$
$$F(v)\nabla v - F(z)\nabla z = F(v)\nabla(v - z) + (v - z)F^*(v, z)\nabla z, \tag{2.16}$$
$$F^*(v, z) = \int_0^1 F_u(v + \tau(z - v))d\tau.$$

Local existence and uniqueness of solutions of (2.13), (2.14) are ensured by standard theory of ordinary differential equations (ode) when (2.13) is considered as a finite system for the coefficients of v^N in some basis of $\mathbb{P}^0_N$, by using the property of ellipticity of A and continuity of B. Concerning this last point, some estimates on B will be required in order to prove a global existence result. This is discussed in the following remark.

Remark 2.2. We write B in (2.2) in the form

$$B(\phi,\psi)=B_1(\phi,\psi)+B_2(\phi,\psi)+B_3(\phi,\psi) \qquad B_1(\phi,\psi)=\int_\Omega \alpha(\phi)\nabla\phi\cdot\nabla(\psi w)d\underline{x},$$

$$B_2(\phi,\psi)=\int_\Omega \beta(\phi)\cdot\nabla\phi(\psi w)d\underline{x}, \qquad B_3(\phi,\psi)=\int_\Omega \gamma(\phi)(\psi w)d\underline{x}.$$

Assuming that $\alpha,\beta,\gamma\in C^1_b(\Omega\times(0,T)\times\mathbb{R})$ (for the sake of simplicity, we suppress the dependence of α,β and γ on x and t in the notation), then, using the continuity of (2.3), there are constants $\alpha_1,\beta_1>0$ such that

$$|B_1(\phi,\psi)|\le\alpha_1||\phi||_{1,w}||\psi||_{1,w}, \qquad |B_2(\phi,\psi)|\le\beta_1||\phi||_{1,w}||\psi||_{1,w}.$$

As far as B_3 is concerned, from (2.15), (2.16) with $F=\gamma$, we can write

$$\gamma(\phi)=\gamma(0)+\phi\gamma^*(\phi,0),$$

with $\gamma(0):=\gamma(\underline{x},t,0)$. Then

$$|B_3(\phi,\psi)|\le(\delta+\gamma_1||\phi||_{1,w})||\psi||_{1,w},$$

for some constant γ_1, and where $\delta=\delta(t)=||\gamma(0)||_{0,w}$ is bounded on $(0,T)$ by some constant C_δ.

Global existence and uniqueness for (2.13), (2.14) and the convergence to the solution of (2.1), (2.2) are proved in the following result.

Theorem 2.2. *For all $t\in[0,T]$, there is a unique solution $v^N(t)$ of (2.13), (2.14) satisfying*

$$||v^N||_{L^\infty(0,T,H^1_w)}\le C, \tag{2.17}$$

for some constant depending on $||v_0||_{1,w}$. Furthermore, let $m\ge1$, and assume that $v_0\in H^m_{w,0}$, $a,c\in C^m(\Omega)\cap H^m_w$, $\alpha,\beta,\gamma\in C^m(\Omega\times(0,T)\times\mathbb{R})$ with $\alpha(\cdot,t),\beta(\cdot,t),\gamma(\cdot,t)\in H^m_w, t\in[0,T]$. If v is the solution of (2.1), (2.2) given by Theorem 2.1, then

$$||v^N-v||_{L^\infty(0,T,L^2_w)}\le CN^{-m}, \tag{2.18}$$

$$||v^N-v||_{L^\infty(0,T,H^1_w)}\le CN^{1-m}, \tag{2.19}$$

for some constant C which depends on $||v_0||_{H^m_w},\alpha,\beta,\gamma,T$ but not on N.

Proof. Let v^N be the solution of (2.13), (2.14), defined locally in t. Following previous approaches, [12,91], we first assume that $\alpha,\beta,\gamma\in C^1_b(\Omega\times(0,T)\times\mathbb{R})$. First we prove that v^N exists on all $[0,T]$. By using the ellipticity of A and Remark 2.2, we set $\psi=v^N_t$ in (2.13) and have

$$||v^N_t||_{1,w}\le C||v^N||_{1,w}+\delta(t),$$

for some constant C. Then, integrating over $(0,t)\subset[0,T]$ yields

$$||v^N||_{1,w}=||v^N(0)+\int_0^t v^N_t(s)ds||_{1,w}\le||v^N(0)||_{1,w}+C\int_0^t||v^N(s)||_{1,w}ds+||\delta||_{L^1(0,T)}. \tag{2.20}$$

From (2.14) with $\psi=v^N(0)$ and continuity and coercivity of A we have $||v^N(0)||_{1,w}\le C||v_0||_{1,w}$. This and Gronwall's lemma applied to (2.20) imply the existence of $v^N(t)$ for all $t\in(0,T)$ and (2.17).

As far as the error estimates are concerned, let $\overline{v}$ be the projection given in (2.7) and define

$$\eta:=\overline{v}-v, e^N:=v^N-v, \xi^N:=\overline{v}-v^N=\eta-e^N\in\mathbb{P}^0_N. \tag{2.21}$$

Note that, due to (2.7), $A(\eta_t,\psi)=0,\forall\psi\in\mathbb{P}^0_N$, holds. Thus, (1.3) and (2.13) imply, for $\psi\in\mathbb{P}^0_N$

$$A(\xi^N_t,\psi)=-A(e^N_t,\psi)=-(B(v^N,\psi)-B(v,\psi)). \tag{2.22}$$

The right-hand side of (2.22) is written as

$$B(v^N,\psi)-B(v,\psi)=\widetilde{B}_1+\widetilde{B}_2+\widetilde{B}_3, \tag{2.23}$$

$$\widetilde{B}_1=\int_\Omega\left(\alpha(v+e^N)\nabla(v+e^N)-\alpha(v)\nabla v\right)\cdot\nabla(\psi w)d\underline{x},$$

$$\widetilde{B}_2=\int_\Omega\left(\beta(v+e^N)\cdot\nabla(v+e^N)-\beta(v)\cdot\nabla v\right)(\psi w)d\underline{x},$$

$$\widetilde{B}_3=\int_\Omega\left(\gamma(v+e^N)-\gamma(v)\right)(\psi w)d\underline{x}.$$

We use (2.15), (2.16), the hypothesis $\alpha, \beta, \gamma \in C_b^1$ and Theorem 2.1 to have

$$|\widetilde{B}_1| \leq C||e^N||_{1,w}||\psi||_{1,w}, \quad |\widetilde{B}_2| \leq C||e^N||_{1,w}||\psi||_{1,w}, \quad |\widetilde{B}_3| \leq C||e^N||_{0,w}||\psi||_{0,w}.$$

Therefore

$$|B(v^N, \psi) - B(v, \psi)| \leq C||e^N||_{1,w}||\psi||_{1,w}, \tag{2.24}$$

for $\psi \in \mathbb{P}_N^0$ and with C depending on $||v_0||_{1,w}$. Then, taking $\psi = \xi_t^N$ and using the coercivity of A, (2.22) and (2.24), we obtain

$$||\xi_t^N||_{1,w} \leq C||e^N||_{1,w}. \tag{2.25}$$

Since (2.14) implies that $v^N(0) = \overline{v}(0)$ and therefore $\xi^N(0) = 0$, then writing $e^N = \eta - \xi^N$ yields

$$||\xi^N(t)||_{1,w} = ||\int_0^t \xi_t^N(s)ds||_{1,w} \leq C\int_0^t (||\xi^N(s)||_{1,w} + ||\eta(s)||_{1,w})ds.$$

Therefore, (2.19) holds from Gronwall's lemma, the property $e^N = \eta - \xi^N$ and Theorem 2.1.

We now prove the estimate (2.18). Lax-Milgram theorem, [37], ensures the existence of $\varphi \in H_{w,0}^1$ such that, [17,77]

$$A(\psi, \varphi) = (\xi_t^N, \psi)_{0,w}, \ \psi \in H_{w,0}^1. \tag{2.26}$$

Actually (see [23]) $\varphi \in H_w^2$ and

$$||\varphi||_{2,w} \leq C||\xi_t^N||_{0,w}. \tag{2.27}$$

We take $\psi = \xi_t^N$ in (2.26) and use (2.22) to estimate

$$\begin{aligned} ||\xi_t^N||_{0,w}^2 \leq A(\xi_t^N, \varphi) &= A(\xi_t^N, \varphi - P_N^{10}\varphi) + A(\xi_t^N, P_N^{10}\varphi) \\ &= A(\xi_t^N, \varphi - P_N^{10}\varphi) - B(e^N, P_N^{10}\varphi) \\ &= A(\xi_t^N, \varphi - P_N^{10}\varphi) + B(e^N, \varphi - P_N^{10}\varphi) - B(e^N, \varphi). \end{aligned} \tag{2.28}$$

Now, the continuity of A, (2.6), and (2.27) imply

$$A(\xi_t^N, \varphi - P_N^{10}\varphi) \leq C||\xi_t^N||_{1,w}||\varphi - P_N^{10}\varphi||_{1,w} \leq CN^{-1}||\xi_t^N||_{1,w}||\varphi||_{2,w} \leq CN^{-1}||\xi_t^N||_{1,w}||\xi_t^N||_{0,w}. \tag{2.29}$$

On the other hand, Remark 2.2, (2.6), and (2.27) lead to

$$|B(e^N, \varphi - P_N^{10}\varphi)| \leq C\left(||e^N||_{1,w}|| + \delta\right)\varphi - P_N^{10}\varphi||_{1,w} \leq CN^{-1}\left(||e^N||_{1,w}|| + \delta\right)||\varphi||_{2,w} \leq CN^{-1}\left(||e^N||_{1,w}|| + \delta\right)||\xi_t^N||_{0,w}. \tag{2.30}$$

We now consider $\widetilde{B}_1$, defined in (2.23). In the one-dimensional case, integrating by parts, we can write, [12,91]

$$\widetilde{B}_1(e^N, \varphi) = \int_{-1}^{1} e^N \left(-\alpha_x - \alpha_u(v^N)v_x^N + v_x\alpha^*)(\varphi w)_x \ -\alpha(v^N)(\varphi w)_{xx}\right) dx,$$

and several applications of Hardy's inequality, see, e.g., [23] (this is not necessary of course in the Legendre case $\mu = 0$), hypothesis $\alpha \in C_b^1$, (2.17), and Theorem 2.1 imply

$$|\widetilde{B}_1(e^N, \varphi)| \leq C||e^N||_{0,w}||\varphi||_{2,w}. \tag{2.31}$$

When $d = 2$ we write $\widetilde{B}_1(e^N, \varphi) = I + II$, with

$$\begin{aligned} I &= \int_\Omega \left[(\alpha(v + e^N)(v + e^N)_x - \alpha(v)v_x\right)(\varphi w)_x d\underline{x} \\ &= \int_\Omega e^N \left(\left(\alpha^* v_x - \left[\alpha_x(v^N) + \alpha_v(v^N\right] v_x^N\right)(\varphi w)_x - \alpha(v^N)(\varphi w)_{xx}\right) d\underline{x}, \\ II &= \int_\Omega \left[(\alpha(v + e^N)(v + e^N)_y - \alpha(v)v_y\right)(\varphi w)_y d\underline{x} \\ &= \int_\Omega e^N \left(\left(\alpha^* v_y - \left[\alpha_y(v^N) + \alpha_v(v^N\right] v_y^N\right)(\varphi w)_y - \alpha(v^N)(\varphi w)_{yy}\right) d\underline{x}, \end{aligned}$$

and $\alpha \in C_b^1$, (2.17), Theorem 2.1, and the application of Hardy's inequality with respect to the corresponding variable in the integrals above lead also to (2.31).

Similarly, we write

$$\widetilde{B}_2(e^N, \varphi) = \int_{-1}^{1} e^N \left((-\beta_x - \beta_u(v^N)v_x^N + v_x\beta^*)(\varphi w)\right) dx,$$

in the 1D case and

$$\widetilde{B}_2(e^N,\varphi)=\int_\Omega e^N\left((\beta_1^* v_x^N+\beta_2^* v_y^N)(\varphi w)-\left[(\beta_{1x}+\beta_{1v}v_x^N)(\varphi w)_x+(\beta_{2y}+\beta_{2v}v_y^N)(\varphi w)_y\right](\varphi w)\right)d\underline{x},$$

and hypothesis $\beta\in C_b^1$, (2.17), Theorem 2.1 and the continuity of L in (2.3) lead to

$$|\widetilde{B}_2(e^N,\varphi)|\le C||e^N||_{0,w}||\varphi||_{1,w}. \tag{2.32}$$

As for $\widetilde{B}_3$ in (2.23), hypothesis $\gamma\in C_b^1$ implies

$$|\widetilde{B}_3(e^N,\varphi)|\le C||e^N||_{0,w}||\varphi||_{0,w}. \tag{2.33}$$

Thus (2.31)-(2.33) along with (2.27) yield

$$|B(e^N,\varphi)|\le C||e^N||_{0,w}||\varphi||_{2,w}\le C||e^N||_{0,w}||\xi_t^N||_{0,w}. \tag{2.34}$$

If we apply (2.29), (2.30) and (2.34) to (2.28) we finally obtain

$$||\xi_t^N||_{0,w}\le CN^{-1}\left(||\xi_t^N||_{1,w}+||e^N||_{1,w}+\delta\right)+C||e^N||_{0,w}.$$

If we use (2.25), (2.19), $e^N=\eta-\xi^N$, Gronwall's lemma and Theorem 2.1, then we see that (2.18) holds. Finally, the proof is completed by observing that the hypothesis $\alpha,\beta,\gamma\in C_b^1$ can be removed from the same argument as in [12,91]. □

2.4. Spectral collocation approximation

Let $N\ge 2$ be an integer, $T>0$. We define the semidiscrete collocation approximation as a mapping $v^N:[0,T]\to\mathbb{P}_N^0$ such that

$$cv_t^N-\nabla\cdot\widetilde{I}_N(a\nabla v_t^N)=-\nabla\cdot\widetilde{I}_N(\alpha\nabla v^N)+\beta\cdot\nabla v^N+\gamma(v^N), \tag{2.35}$$

at the nodes corresponding to the Gauss-Lobatto quadrature associated to (1.6) or (1.7), and introduced above, with

$$v^N(0)=I_N(v_0), \tag{2.36}$$

and where if $u_1,u_2\in H_{w,0}^1$

$$\widetilde{I}_N(u_1,u_2)=(I_N(u_1),I_N(u_2))^T.$$

(For simplicity, in this section v^N will stand for the collocation approximation and this should not be confused with the Galerkin approximation defined in Section 2.3.) A first task here is to derive a weak formulation equivalent to (2.35), (2.36) and involving the inner product (1.12). Note that if $\psi\in\mathbb{P}_N^0$ then $w^{-1}\nabla(\psi w)\in(\mathbb{P}_{N-1})^2$, [18]. Therefore, using (1.13) we have, for $\phi,\psi\in\mathbb{P}_N^0$

$$-\sum_{j=0}^N(I_N(a\phi_x)_x(x_j))\psi(x_j)w_j=-\int_{-1}^1(I_N(a\phi_x))_x\psi w dx=\int_{-1}^1 I_N(a\phi_x)(\psi w)_x dx=(a\phi_x,w^{-1}(\psi w)_x)_{N,w},$$

in 1D, and similarly in 2D, [18]

$$\begin{aligned}-\sum_{j,k=0}^N\nabla\cdot\widetilde{I}_N(a\nabla\phi)(x_{jk}))\psi(x_{jk})w_{jk}&=-\int_\Omega\nabla\cdot\widetilde{I}_N(a\nabla\phi))\psi w d\underline{x}\\&=\int_\Omega\left(I_N(a\phi_x)(\psi w)_x+I_N(a\phi_y)(\psi w)_y\right)d\underline{x}.\end{aligned}$$

Then Remark 1.1 and the comments above enable (2.35), (2.36) to admit the weak form

$$A_N(v_t^N,\psi)=B_N(v^N,\psi),\ \psi\in\mathbb{P}_N^0 \tag{2.37}$$

$$v^N(0)=I_N v_0, \tag{2.38}$$

where, for $\phi,\psi\in\mathbb{P}_N^0$

$$A_N(\phi,\psi)=(c\phi,\psi)_{N,w}+(a\nabla\phi,w^{-1}\nabla(\psi w))_{N,w}, \tag{2.39}$$

$$B_N(\phi,\psi)=(\alpha(\phi)\nabla\phi,w^{-1}\nabla(\psi w))_{N,w}+(\beta(\phi)\cdot\nabla\phi,\psi)_{N,w}+(\gamma(\phi),\psi)_{N,w}. \tag{2.40}$$

From the equivalence with the bilinear form, [17,18]

$$a_N(\phi,\psi)=(\phi,\psi)_{N,w}+(\nabla\phi,w^{-1}\nabla(\psi w))_{N,w},$$

which is continuous in $\mathbb{P}_N\times\mathbb{P}_N^0$ and coercive in $\mathbb{P}_N^0$ for all weights $w_\mu,-1<\mu<1$, we have that A_N in (2.39) satisfies the properties of continuity and ellipticity

$$|A_N(\phi,\psi)| \le C||\phi||_{1,N}||\psi||_{1,N},\ \phi \in P_N, \psi \in P_N^0,$$
$$A_N(\psi,\psi) \ge C||\psi||_{1,N}^2,\ \psi \in P_N^0, \tag{2.41}$$

where

$$||\phi||_{1.N}^2 = ||\phi||_{N,w}^2 + ||\nabla\phi||_{N,w}^2. \tag{2.42}$$

Remark 2.3. Note also that if $\alpha \in C_b^1$, then the property of continuity of the bilinear form

$$(\phi,\psi) \mapsto (\nabla\phi, w^{-1}\nabla(\psi w))_{N,w},\ \phi \in \mathbb{P}_N, \psi \in \mathbb{P}_N^0,$$

proved in [18], implies the continuity of the first term of B_N in (2.40). The other two terms can be estimated, when $\beta,\gamma \in C_b^1$, in a similar way, using the arguments of Remark 2.2, the equivalence of the norms $||\cdot||_{1,w}$, and (2.42) in $\mathbb{P}_N$ given from Lemma 1.1.

Lemma 2.1. *There is a unique solution $v^N(t)$ of (2.37), (2.38) for all $t \in [0,T]$ with*

$$||v^N||_{L^\infty(0,T,H_w^1)} \le C,$$

where C depends on $||v_0||_{H_w^1}$.

Proof. As before, let us first assume that $\alpha,\beta,\gamma \in C_b^1$. When (2.38) is viewed as a finite ode system for the coefficients of v^N with respect to some basis of P_N^0, standard theory proves local existence and uniqueness. In order to prove continuation for $t \in (0,T)$ of $v^N(t)$, previous comments on B_N in Remark 2.3 and Remark 2.2 show that for $\psi \in P_N^0$

$$|B_N(v^N,\psi)| \le (C||v^N||_{1,N} + ||\gamma(0)||_{N,w})||\psi||_{1,N}, \tag{2.43}$$

for some constant C and where $\gamma(0) = \gamma(0,\cdot,t)$. Then, from (2.41) and (2.43), taking $\psi = v_t^N$ in (2.37) leads to

$$||v_t^N||_{1,N} \le C\left(||v^N||_{1,N} + ||\gamma(0)||_{1,N}\right).$$

The equivalence of the norms from Lemma 1.1 in $\mathbb{P}_N$ yields a similar inequality

$$||v_t^N||_{1,w} \le C\left(||v^N(t)||_{1,w} + ||\gamma(0)||_{1,w}\right),$$

from which

$$||v^N(t)||_{1,w} \le ||v^N(0)||_{1,N} + C\int_0^t C\left(||v^N(s)||_{1,w} + ||\gamma(0,\cdot,s)||_{1,w}\right)ds.$$

We conclude the proof applying Gronwall's lemma and the stability of Gauss-Lobatto interpolation in the H_w^1 norm, that is, [17,23]

$$||I_N v_0||_{1,w} \le ||v_0||_{1,w}.$$

The hypothesis $\alpha,\beta,\gamma \in C_b^1$ can be finally removed as in [12,91]. □

The corresponding convergence result is as follows.

Theorem 2.3. *Let $m \ge 2$ and assume the hypotheses of Theorem 2.2. Let v be the solution of (2.1), (2.2) given by Theorem 2.1, and v^N be the solution of (2.37), (2.38), defined for all $t \in [0,T]$. Then*

$$||v^N - v||_{L^\infty(0,T,H_w^1)} \le CN^{2-m}, \tag{2.44}$$

for some constant C which depends on $||v_0||_{H_w^m}, \alpha,\beta,\gamma,T$ but not on N.

Proof. Let η, e^N, ξ^N be as given in (2.21). For $\psi \in \mathbb{P}_N^0$, since $A(\eta_t,\psi) = 0$, we can write

$$A_N(\xi_t^N,\psi) = A_N(\overline{v}_t,\psi) - A(\overline{v}_t,\psi) + B(v,\psi) - B_N(v^N,\psi). \tag{2.45}$$

Note first that

$$|A_N(\overline{v}_t,\psi) - A(\overline{v}_t,\psi)| \le |(c\overline{v}_t,\psi)_{N,w} - (c\overline{v}_t,\psi)_{0,w}| + \left|\int_\Omega (E - \widetilde{I}_N)(a\nabla\overline{v}_t)(\psi w)_x d\underline{x}\right|, \tag{2.46}$$

where E denotes the identity operator. Since $\overline{v}_t = I_N\overline{v}_t$, then (2.12), applied to the first term on the right-hand side of (2.46), and the hypothesis on c implies that

$$|(c\overline{v}_t,\psi)_{N,w} - (c\overline{v}_t,\psi)_{0,w}| \le C||(E - P_{N-1})\overline{v}_t||_{0,w}||\psi||_{0,w}. \tag{2.47}$$

Now, (2.5) and (2.8) lead to

$$||(E-P_{N-1})\overline{v}_t||_{0,w} \le ||(E-P_{N-1})v_t||_{0,w} + ||(E-P_{N-1})(v_t-\overline{v}_t)||_{0,w}$$
$$\le C(N-1)^{-m}||v_t||_{m,w} + C(N-1)^{-1}||v_t-\overline{v}_t||_{1,w}$$
$$\le C(N-1)^{-m}||v_t||_{m,w}. \tag{2.48}$$

As far as the second term on the right-hand side of (2.46) is concerned, from the continuity of L in (2.3) we have, for $\psi \in \mathbb{P}_N^0$

$$\Big|\int_{-1}^{1}(E-\widetilde{I}_N)(a\nabla\overline{v}_t)\cdot\nabla(\psi w)d\underline{x}\Big| \le ||(E-\widetilde{I}_N)(a\nabla\overline{v}_t)||_{0,w}||\psi||_{1,w}. \tag{2.49}$$

Now, from (2.11), the hypothesis on a and (2.11) we can write

$$||(E-\widetilde{I}_N)(a\nabla\overline{v}_t)||_{0,w} \le ||(E-\widetilde{I}_N)(a\nabla v_t)||_{0,w}$$
$$+||(E-\widetilde{I}_N)(a(\nabla v_t-\nabla\overline{v}_t))||_{0,w}$$
$$\le CN^{1-m}||a\nabla v_t||_{m-1,w} + CN^{-1}||a(\nabla v_t-\nabla\overline{v}_t)||_{1,w}$$
$$\le CN^{1-m}||v_t||_{m,w} + CN^{2-m}||v_t||_{m,w}, \tag{2.50}$$

where C depends on $||a||_{m-1,w}$. In the last inequality the property, [17,77]

$$||uv||_{s,w} \le C||u||_{s,w}||v||_{s,w},\ u,v \in H^s_w, s \ge 1,$$

was used.

We now consider the last two terms in (2.45), written as

$$B(v,\psi)-B_N(v^N,\psi) = B(v,\psi)-B(\overline{v},\psi)+B(\overline{v},\psi)-B_N(\overline{v},\psi)$$
$$+B_N(\overline{v},\psi)-B_N(v^N,\psi), \tag{2.51}$$

and estimate each couple of (2.51). Assume first that $\alpha,\beta,\gamma \in C_b^1$. Similar arguments to those of (2.24) along with (2.8) imply

$$|B(v,\psi)-B(\overline{v},\psi)| \le C||v-\overline{v}||_{1,w}||\psi||_{1,w} \le CN^{1-m}||v||_{m,w}||\psi||_{1,w}. \tag{2.52}$$

On the other hand

$$B(\overline{v},\psi)-B_N(\overline{v},\psi) = \int_\Omega (E-\widetilde{I}_N)(\alpha\nabla\overline{v})(\psi w)_x dx + (\beta\cdot\nabla\overline{v},\psi)_{0,w} - (\beta\cdot\nabla\overline{v},\psi)_{N,w} + (\gamma(\overline{v}),\psi)_{0,w} - (\gamma(\overline{v}),\psi)_{N,w}.$$

As in (2.49), (2.50)

$$\Big|\int_\Omega (E-\widetilde{I}_N)(\alpha\nabla\overline{v})(\psi w)_x d\underline{x}\Big| \le CN^{2-m}||v_t||_{m,w}||\psi||_{1,w}.$$

Next, using (2.12)

$$|(\beta\cdot\nabla\overline{v},\psi)_{0,w}-(\beta\cdot\nabla\overline{v},\psi)_{N,w}| \le C\left(||(E-P_{N-1})\beta(\overline{v})\cdot\nabla\overline{v}||_{0,w} + ||(E-\widetilde{I}_N)\beta(\overline{v})\cdot\nabla\overline{v}||_{0,w}\right)||\psi||_{0,w},$$

and $\beta \in C_b^1$ along with (2.5) and (2.11) lead to

$$||(E-P_{N-1})\beta(\overline{v})\overline{v}_x||_{0,w} \le C(N-1)^{1-m}||v_x||_{m-1,w} + C(N-1)^{-1}||\overline{v}_x-v_x||_{1,w} \le C(N-1)^{2-m}||v||_{m,w},$$

while, similarly to (2.50)

$$||(E-I_N)\beta(\overline{v})\overline{v}_x||_{0,w} \le CN^{2-m}||v||_{m,w}.$$

Finally, (2.12) also implies

$$|(\gamma(\overline{v}),\psi)_{0,w}-(\gamma(\overline{v}),\psi)_{N,w}| \le C\left(||(E-P_{N-1})\gamma(\overline{v})||_{0,w} + ||(E-I_N)\gamma(\overline{v})||_{0,w}\right)||\psi||_{0,w}.$$

Now

$$||(E-P_{N-1})\gamma(\overline{v})||_{0,w} \le ||(E-P_{N-1})\gamma(v)||_{0,w} + ||(E-P_{N-1})(\gamma(v)-\gamma(\overline{v}))||_{0,w}.$$

Using (2.15) with $F=\gamma$ we have, cf. Remark 2.2

$$\gamma(v) = \gamma(0) + v\gamma^*(v,0),$$
$$\gamma(\overline{v}) = \gamma(v) + (\overline{v}-v)\gamma^*(\overline{v},v).$$

Then from (2.5) and (2.8) we obtain

$$||(E-P_{N-1})\gamma(v)||_{0,w} \le CN^{1-m}(||\gamma(0,\cdot,t)||_{m-1,w} + ||v||_{m,w}),$$
$$||(E-P_{N-1})(\gamma(v)-\gamma(\overline{v}))||_{0,w} \le CN^{1-m}||v||_{m,w}.$$

All this leads to

$$|B(\overline{v},\psi)-B_N(\overline{v},\psi)|\leq \left(CN^{2-m}||v||_{m,w}+N^{1-m}||\gamma(0,\cdot,t)||_{m-1,w}\right)||\psi||_{0,w}. \tag{2.53}$$

Finally, Remark 2.3 on the continuity of B_N and Lemma 1.1 imply

$$|B_N(\overline{v},\psi)-B_N(v^N,\psi)|\leq C||e^N||_{1,w}||\psi||_{1,w}. \tag{2.54}$$

We may now take $\psi=\xi_t^N$ in (2.45), use the coercivity of A_N and (2.46)-(2.54) to obtain an estimate of the form

$$||\xi_t^N||_{1,w}\leq CN^{2-m}(||\gamma(0,\cdot,t)||_{m-1,w}+||v(t)||_{m,w}+||v_t(t)||_{m,w}+||e^N||_{1,w}).$$

Then Gronwall's lemma, the property $e^N=\eta-\xi^N$ and Theorem 2.1 conclude the proof of (2.44). The condition $\alpha,\beta,\gamma\in C_b^1$ can be removed as in previous results. □

Remark 2.4. From the previous proof we can observe that the difference with respect to that of Theorem 2.2 is the use of the generalized estimate (2.11). This results in the term N^{2-m} in (2.44), instead of N^{1-m} as in (2.19).

3. Full discretization and a computational study

In this section we complete our proposal for the numerical approximation of the ibvp (1.3)-(1.5) with the introduction of the time integrator. The resulting fully discrete schemes will be used in a series of numerical experiments with the aim at illustrating the error estimates obtained in Sections 2.3 and 2.4, as well as investigating the general performance of the method with non-smooth data.

As mentioned in the Introduction, the spectral discretization, analyzed in Section 2 for the general Jacobi polynomials family and for the Galerkin and collocation versions, will be represented in the computational study below by the Legendre Galerkin method and the Chebyshev collocation scheme. These two subfamilies are widely used when approximating with spectral methods and their implementation is well known and explained in many references (see below). Our choice might have involved, in any of the two approaches (Galerkin or collocation), any other family of Jacobi polynomials, although the implementation is, to our knowledge, not so developed in the literature.

The spectral collocation approach and the practical formulation of the spectral Galerkin method (based on numerical integration, the so-called Galerkin with numerical integration formulation, [23]) involve some properties of the discrete norm associated to the Gauss-Lobatto quadrature that are now discussed. In the one-dimensional case, and for the given weight function w, the Gauss-Lobatto quadrature formula is obtained as follows, [22,23,49,78]. Let $N>0$ be an integer and $\mathbb{P}_N$ be the space of polynomials of degree at most N on $\overline{\Omega}$, with p_N the N-th degree polynomial of the orthogonal Legendre or Chebyshev family. Let

$$q(x)=p_{N+1}(x)+\widetilde{a}p_N(x)+\widetilde{b}p_{N-1}(x), \tag{3.1}$$

with $\widetilde{a},\widetilde{b}$ chosen such that $q(-1)=q(1)=0$. If $-1=x_0<x_1<\cdots<x_N=1$ are the roots of (3.1), then there are weights $w_0,\ldots,w_N$ such that

$$\int_{-1}^{1}p(x)w(x)dx=\sum_{j=0}^{N}p(x_j)w_j,\ p\in\mathbb{P}_{2N-1}.$$

In the case of Legendre polynomials (where p_N is denoted by L_N), the $x_j, j=1,\ldots,N-1$, are shown to be the zeros of L_N' and

$$w_j=\frac{2}{N(N+1)}\frac{1}{L_N(x_j)^2},\ j=0,\ldots,N,$$

while for the Chebyshev case (where p_N is denoted by T_N)

$$x_j=\cos\frac{j\pi}{N},\ \ w_j=\begin{cases}\frac{\pi}{2N} & j=0,N,\\ \frac{\pi}{N} & j=1,\ldots,N.\end{cases}$$

Recall that the Gauss-Lobatto quadrature is related to a discrete inner product (1.12) and satisfies (1.13). In the two-dimensional case, the Gauss-Lobatto quadrature formula is given by (1.16) for $\psi\in\mathbb{P}_{2N-1}$ with nodes and weights obtained from the 1D case and (1.15), [23,98].

For the computational study developed in this section, some cases of (1.3), relevant in the applications, will be used in the numerical experiments. We make here a brief description of them:

- The following linear pseudo-parabolic problem will be considered as a first model example

$$v_t-a\nabla\cdot\nabla v_t=b\nabla\cdot\nabla v,\ x\in\Omega=(-1,1)^d,\ t>0,\ d=1,2, \tag{3.2}$$

$$v(\underline{x},0)=v_0(\underline{x}),\ x\in\Omega, \tag{3.3}$$

$$v(t)\Big|_{\partial\Omega}=0,\ t>0, \tag{3.4}$$

  where a and b are positive constants and $v_0:(-1,1)^d\to\mathbb{R}$. Equation (3.2) is a linearized version of the BBM-Burgers equation. It can be solved, for general enough initial conditions v_0, by using the technique of separation of variables. In the one-dimensional case, from the basis of the corresponding eigenvalue problem

$$X_n(x)=\sin\frac{n\pi}{2}(x+1),\ n=1,2,\ldots,$$

  with eigenvalues $\lambda_n=-(n\pi/2)^2, n=1,2,\ldots$, the solution of (3.2)-(3.4) can be formally written in the form

$$v(x,t)=\sum_{n=1}^{\infty} C_n e^{\alpha_n t} X_n(x), \quad \text{where} \quad \alpha_n=\frac{b\lambda_n}{1-a\lambda_n}, \tag{3.5}$$

and C_n is the n-th coefficient of v_0 in the corresponding sine Fourier expansion

$$v_0(x)=\sum_{n=1}^{\infty} C_n X_n(x), \tag{3.6}$$

assuming that this does exist. In order to check the convergence when dealing with problems (3.2)-(3.4), the representation (3.5)-(3.6) will be used in Section 4 as follows. For different initial data v_0, the corresponding numerical approximation will be compared with the associated solution of (3.2)-(3.4), computed exactly or in an accurate enough, approximate way, via the sine Fourier expansion (3.5). In this last case, acceleration techniques, [101,102], will be used when necessary. For an alternative way to estimate the numerical order of convergence, see e.g. [20]. In the two-dimensional case, the corresponding representation of v, analogous to (3.5), is

$$v(x,y,t)=\sum_{n,m=1}^{\infty} C_{n,m} e^{\alpha_{n,m} t} X_{n,m}(x,y), \quad \text{where} \quad \alpha_{n,m}=\frac{b\lambda_{n,m}}{1-a\lambda_{n,m}}, \quad \lambda_{n,m}=-(n\pi/2)^2-(m\pi/2)^2, \tag{3.7}$$

with $X_{n,m}(x,y)=\sin\frac{n\pi}{2}(x+1)\sin\frac{n\pi}{2}(y+1)$, $n,m=1,2,\ldots$, and $C_{n,m}$ is the (n,m)th coefficient of the sine Fourier expansion

$$v_0(x,y)=\sum_{n,m=1}^{\infty} C_{n,m} X_{n,m}(x,y). \tag{3.8}$$

- A second case study will be the Dirichlet problem of equations of the form

$$v_t-a\nabla\cdot\nabla v_t+\alpha\cdot\nabla v+\beta\nabla\cdot\nabla v+\gamma\cdot\nabla f(v)=F, \tag{3.9}$$

with $a>0, \beta\in\mathbb{R}$, $f=f(v)$ some nonlinear function of v, $F=F(\underline{x},t)$ a source term, and $\alpha,\gamma\in\mathbb{R}^d$. Two particular important examples of f will be used in the numerical experiments:

– The case of the BBM-Burgers equation, for which

$$f(v)=v^2. \tag{3.10}$$

– The function

$$f(v)=\begin{cases} 0 & \text{if } v<0 \\ \frac{v^2}{v^2+2(1-v)^2} & \text{if } 0\le v\le 1 \\ 1 & \text{if } v>1 \end{cases}. \tag{3.11}$$

The nonlinear term (3.11) appears in modelling two-phase flow porous media, see e.g. [5,6,24,96].

Some details on the implementation, for both Galerkin and collocation approximations, will be given below. They will be firstly described for the case of one dimension, and then we will explain how the procedures may be extended to the two-dimensional case.

3.1. Legendre spectral Galerkin approximation

Let $N\ge 2$ be an integer, $T>0$. In this section $v^N:[0,T]\to\mathbb{P}_N^0$ will denote the semidiscrete Galerkin approximation, solution of (2.13), (2.14). We are interested in the representation of v^N to implement (2.13), (2.14) in the Legendre case, and the discussion of several formulations. We consider the 1D case first. Due to the presence of nonlinear terms, we write v^N in terms of the nodal basis functions, [23]

$$\psi_j(x)=\frac{1}{N(N+1)}\frac{(1-x^2)}{(x_j-x)}\frac{L_N'(x)}{L_N(x_j)},\ j=0,\ldots,N, \tag{3.12}$$

where $x_j, j=0,\ldots,N$, denotes the nodes associated to the Legendre-Gauss-Lobatto quadrature, L_N is the N-th Legendre polynomial. The basis (3.12) satisfies

$$\psi_j(x_k)=\delta_{jk},\ j,k=0,\ldots,N. \tag{3.13}$$

A Galerkin with numerical integration (G-NI) formulation will be also adopted. This means that, [23], from the expansion of the numerical approximation

$$v^N(x,t)=\sum_{k=0}^{N} V_k(t)\psi_k(x),\ V_k(t)=v^N(x_k,t), \tag{3.14}$$

the integrals in the weak formulation are approximated by the Legendre-Gauss-Lobatto quadrature. The resulting system for $V(t)=(V_0(t),\ldots,V_N(t))^T$ will have the form

$$\left(K_N^{(0)}(c)+K_N^{(2)}(a)\right)\frac{d}{dt}V=K_N^{(2)}(\alpha)(V)+K_N^{(1)}(\beta)(V)+\Gamma_N(V), \tag{3.15}$$

where, for $0\le j,k\le N$

$$\left(K_N^{(0)}(c)\right)_{jk} = c(x_j)w_j\delta_{jk},$$

$$\left(K_N^{(2)}(a)\right)_{jk} = \sum_{h=0}^{N} a(x_h)\frac{d\psi_j}{dx}(x_h)\frac{d\psi_k}{dx}(x_h)w_h, \tag{3.16}$$

$$\left(K_N^{(2)}(\alpha)(V)\right)_{jk} = \sum_{h=0}^{N} \alpha(V_h)\frac{d\psi_j}{dx}(x_h)\frac{d\psi_k}{dx}(x_h)w_h, \tag{3.17}$$

$$\left(K_N^{(1)}(\beta)(V)\right)_{jk} = \sum_{h=0}^{N} \beta(V_h)\frac{d\psi_j}{dx}(x_h)\psi_k(x_h)w_h = \beta(V_k)\frac{d\psi_j}{dx}(x_k)w_k, \tag{3.18}$$

$$(\Gamma_N(V))_j = \sum_{h=0}^{N} \gamma(V_h)\psi_j(x_h)w_h = \gamma(V_j)w_j. \tag{3.19}$$

In general, matrices (3.16)-(3.18) are full and require $O(N^3)$ operations, with the grid values of the derivatives computed from the Legendre differentiation matrix, [23]. The coefficients in (3.17)-(3.19) are obtained from the use of the nodal basis and (3.13). Thus if $\mathbb{F} = \alpha, \beta$ or γ, then the computation of $\mathbb{F}(V)(x_h), h = 0, \ldots, N$, is understood as $\mathbb{F}(v^N(x_h, t))$, that is $\mathbb{F}(V_h(t))$. The general formulation (3.15) can be simplified in particular cases of (1.3). For the pseudo-parabolic problem (3.9), the description is made with $F = 0$ on $\Omega = (-1, 1)$ and homogeneous boundary conditions. (In the numerical experiments, though, problems on other intervals, with inhomogeneous terms in (3.9) and/or nonhomogeneous boundary data may be considered. This means that the implementation is adapted from the homogeneous problem in Ω to the corresponding case at hand via suitable change of variables to homogenize the boundary data, cf. Remark 3.1 below.) The formulation simplifies to

$$(M_N + aK_N^{(2)})\frac{d}{dt}V(t) + \alpha C_N V(t) - \beta K_N^{(2)}V(t) + \gamma C_N f(V(t)) = 0, \tag{3.20}$$

where now

$$M_N = \mathrm{diag}(w_0, \ldots, w_N), \tag{3.21}$$

$$C_N = -K_N^{(1)},\ (K_N^{(1)})_{ij} = (\psi_i, \frac{d}{dx}\psi_j)_{N,w}, \tag{3.22}$$

$$(K_N^{(2)})_{ij} = (\frac{d}{dx}\psi_i, \frac{d}{dx}\psi_j)_{N,w}, \tag{3.23}$$

and the computation of $f(V)$ must be understood component wise. (For example, if $f(v) = v^2$, then $f(V) = V \cdot V$, where the dot denotes the Hadamard product of the vectors.) Note that this formulation makes the Galerkin method be essentially equivalent to the collocation approach. The reason is that in this case, [23]

$$K_N^{(1)} = -M_N D_N,\ K_N^{(2)} = -M_N D_N^2,$$

where D_N and D_N^2 denote here the first- and second-derivative matrix at the Legendre-Gauss-Lobatto nodes respectively. This is used, along with the boundary conditions, to write (3.20) in the form

$$(I_{N-1} - a\widetilde{D}_N^{(2)})\frac{d}{dt}\widetilde{V}(t) + \alpha\widetilde{D}_N\widetilde{V}(t) + \beta\widetilde{D}_N^{(2)}\widetilde{V}(t) + \gamma\widetilde{D}_N\widetilde{f}(V(t)) = 0,$$

where I_{N-1} is the $(N-1)\times(N-1)$ identity matrix and the tilde means that the first and last rows and columns (for matrices) and the first and last components (in column vectors) are removed from (3.20).

For the 2D case, we write

$$v^N(x, y, t) = \sum_{h,i=0}^{N} V_{h,i}(t)\psi_h(x)\psi_i(y), \quad V_{h,i}(t) = v^N(x_i, x_j, t), \tag{3.24}$$

and the G-NI formulation for $V(t) = (V_{ij}(t))_{i,j=0}^N$, based on the grid (1.15), will have the matrix form

$$\begin{aligned} M_N(Q_N(c)\cdot V')M_N + K_N^{(2)}(L_N(a)\cdot V')M_N + M_N(L_N(a)\cdot V')(K_N^{(2)})^T = K_N^{(2)}\alpha_N(V))M_N + M_N(\alpha_N(V))(K_N^{(2)})^T \\ + K_N^{(1)}\beta_N(V))M_N + M_N(\beta_N(V))(K_N^{(1)})^T \\ + M_N\gamma_N(V)M_N, \end{aligned} \tag{3.25}$$

where the dot stands for the Hadamard product of matrices,

$$Q_N(d) := (d(x_i, x_j))_{i,j=0}^N, \quad d = a, c,$$
$$F_N(V) := (F_N(V_{ij}))_{i,j=0}^N, \quad F = \alpha, \beta, \gamma,$$

and $M_N, K_N^{(1)}$ and $K_N^{(2)}$ are given by (3.21)-(3.23). The formulation (3.25) can be written in terms of tensor products when $V(t)$ is vectorized in the form

$$V = (V_{00}, \ldots, V_{N0}, \ldots, V_{0N}, \ldots, V_{NN})^T. \tag{3.26}$$

Thus, for example, in the case of (3.9) with homogeneous conditions, making use of some properties of the Kronecker product $\otimes$, (3.25) can be simplified to (cf. [98])

$$\left(I_{N-1}\otimes I_{N-1}-a\mathcal{A}_N\right)\frac{d}{dt}\widetilde{V}+\mathcal{B}_N\widetilde{V}(t)+\mathcal{G}_N\widetilde{f}(V(t))=0, \tag{3.27}$$

with

$$\begin{aligned}\mathcal{A}_N &:= I_{N-1}\otimes\widetilde{D}_N^{(2)}+\widetilde{D}_N^{(2)}\otimes I_{N-1},\\ \mathcal{B}_N &:= \alpha_1(\widetilde{D}_N\otimes I_{N-1})-\alpha_2(I_{N-1}\otimes\widetilde{D}_N)\\ &\quad+\beta\left((\widetilde{D}_N^{(2)}\otimes I_{N-1})+(I_{N-1}\otimes\widetilde{D}_N^{(2)})\right)\\ \mathcal{G}_N &:= \gamma_1(\widetilde{D}_N\otimes I_{N-1})-\gamma_2(I_{N-1}\otimes\widetilde{D}_N),\end{aligned} \tag{3.28}$$

where the tilde conserves the same meaning as before (adapted to the two-dimensional case), and where $f(V)$ is also understood component wise.

Remark 3.1. It may be worth mentioning some implementation details concerning the treatment of the nonhomogeneous case. The semidiscretization of any source term (independent of v) is similar to that of the coefficients a and c in (1.3), with t as parameter. On the other hand, the presence of nonhomogeneous Dirichlet boundary conditions requires a previous step to homogenize the problem in the usual way. Searching for some function w satisfying the boundary conditions and defining $\widetilde{v}=v-w$, then $\widetilde{v}$ satisfies a nonhomogeneous version of (1.3) but with homogeneous boundary conditions and initial data $\widetilde{v}(\underline{x},0)=v(\underline{x},0)-w(\underline{x},0)$. By way of illustration, for the case of (3.9), the equation for $\widetilde{v}$ would be of the form

$$\widetilde{v}_t-a\nabla\cdot\nabla\widetilde{v}_t+\alpha\cdot\nabla\widetilde{v}+\beta\nabla\cdot\nabla\widetilde{v}+\gamma\cdot\nabla f(\widetilde{v}+w)=\widetilde{F},$$

where $\widetilde{f}=F-w_t-a\nabla\cdot\nabla w_t+\alpha\cdot\nabla w+\beta\nabla\cdot\nabla w$.

In one dimension, the auxiliary function w is determined in the usual way: If $v(\pm1,t)=c_\pm(t)$, then we can take

$$w(x,t)=\frac{c_+(t)-c_-(t)}{2}x+\frac{c_+(t)+c_-(t)}{2}.$$

In the two-dimensional case, an extension of this procedure can be derived from the technique described in e.g. [97,98] (see also references therein). If

$$v(\pm1,y,t)=a_\pm(y,t),\quad v(x,\pm1,t)=b_\pm(x,t), \tag{3.29}$$

we consider

$$u^{(1)}(x,y,t)=\frac{b_+(x,t)-b_-(x,t)}{2}y+\frac{b_+(x,t)+b_-(x,t)}{2}.$$

Defining

$$\widetilde{a}_\pm(y,t)=a_\pm(y,t)-u^{(1)}(\pm1,y,t),$$

then $\widetilde{a}(\pm1,t)=0$. Let

$$u^{(2)}(x,y,t)=\frac{\widetilde{a}_+(y,t)-\widetilde{a}_-(y,t)}{2}x+\frac{\widetilde{a}_+(y,t)+\widetilde{a}_-(y,t)}{2}.$$

Then, by construction, $w(x,y,t)=u^{(1)}(x,y,t)+u^{(2)}(x,y,t)$ satisfies (3.29).

For the linear problem (3.2)-(3.4), the general formulation (3.15), (3.25) can be also simplified. To this end, the implementation of the Legendre Galerkin method for (3.2)-(3.4) will follow the compact representation described in [97] for linear elliptic problems (see also [23,98]). The main idea is choosing a suitable basis for $\mathbb{P}_N^0$ such that the linear system obtained from (2.13) is as simple as possible. (In the experiments and for simplicity, $v^N(0)$ will be taken as $v_0(x)$, so that the second equation (2.14) is satisfied.) In Lemma 2.1 of [97] this is given by $\phi_0,\dots,\phi_{N-2}$ with

$$\phi_k(x)=c_k(L_k(x)-L_{k+2}(x)),\ c_k=\frac{1}{\sqrt{4k+6}},\ k=0,\dots,N-2,$$

where L_k denotes the Legendre polynomial of degree k. By using the representation

$$v^N(x,t)=\sum_{k=0}^{N-2}v_k^N(t)\phi_k(x),$$

and evaluating (2.13) for $\psi=\phi_j, j=0,\dots,N-2$, we obtain the system for $V(t)=(v_0^N(t),\dots,v_{N-2}^N(t))^T$

$$K_NV'(t)+S_NV(t)=0, \tag{3.30}$$

with K_N, S_N matrices with entries

$$\begin{aligned}(K_N)_{jk} &= \underbrace{(\phi_k,\phi_j)_w}_{b_{jk}}+a\underbrace{(\phi_k',\phi_j')_w}_{a_{jk}},\\ (S_N)_{jk} &= b(\phi_k',\phi_j')_w=ba_{jk},\end{aligned} \tag{3.31}$$

where, [97]

$$a_{jk}=\begin{cases}1 & k=j\\ 0 & k\neq j\end{cases}, \tag{3.32}$$

and $b_{kj} = b_{jk}$, defined, for $k \geq j$, as

$$b_{jk} = \begin{cases} c_k c_j \left(\frac{2}{2j+1} + \frac{2}{2j+5} \right) & k = j \\ -c_k c_j \frac{2}{2k+1} & k = j+2 \\ 0 & \text{otherwise} \end{cases} \tag{3.33}$$

Then (3.31) is of the form

$$K_N = aI_{N-1} + B_N,\ S_N = bI_{N-1}, \tag{3.34}$$

where $B_N = (b_{jk})_{j,k=0}^{N-2}$. Note that this matrix is pentadiagonal with only three nonzero diagonals.

This approach may also be used in the two-dimensional case from the representation

$$v^N(x,y,t) = \sum_{k,j=0}^{N-2} v_{kj}^N(t)\phi_k(x)\phi_j(y), \tag{3.35}$$

and the analysis made in [97,98], leading to an alternative to the Kronecker formulation. A description of the procedure for the problem (3.9) follows by way of illustration. In the linear case ($\gamma = 0$), the semidiscrete Legendre spectral Galerkin approximation $v^N(t) = (v_{kj}^N(t))_{k,j=0}^{N-2}$ will satisfy

$$B\frac{dv^N}{dt}B + a\left(A\frac{dv^N}{dt}B + B\frac{dv^N}{dt}A^T \right) + \alpha_1 C v^N B + \alpha_2 B v^N C^T - \beta\left(Av^N B + Bv^N A^T \right) = \overline{F}, \tag{3.36}$$

where $A = (a_{jk})$ and $B = (b_{jk})$ are given by (3.32) and (3.33), respectively, $C = (c_{jk})$ is the matrix, [97]

$$c_{jk} := (\phi_k', \phi_j)_w = -c_{kj} = \begin{cases} 2c_k c_j & k = j+1 \\ 0 & k \neq j+1 \end{cases}, \tag{3.37}$$

and $\overline{F} = (F_{ij})_{i,j=0}^{N-2}$, $F_{ij} = \int_\Omega F\phi_i(x)\phi_j(y)d\underline{x}$. We may use the matrix decomposition method described in [97,98] to solve (3.36). Since in the Legendre case $A = I_{N-1}$ and B is symmetric and definite positive, we can decompose

$$BE = \Lambda E, \tag{3.38}$$

where $\Lambda := \text{diag}(\lambda_0, \ldots, \lambda_{N-2})$ is the diagonal matrix of eigenvalues of B and E is an orthogonal matrix whose columns form an orthonormal basis of eigenvectors of B. If $U := E^T V$, then $V = EU$ and applying (3.38) to (3.36) leads to

$$E\Lambda U' B + a\left(EU'B + E\Lambda U' \right) + \alpha_1 CEUB + \alpha_2 E\Lambda U C^T - \beta(EUB + E\Lambda U) = \overline{F}. \tag{3.39}$$

Multiplying (3.39) by $E^T = E^{-1}$, we have

$$\Lambda U' B + a\left(U'B + \Lambda U' \right) + \alpha_1 \widetilde{C} U B + \alpha_2 \Lambda U C^T - \beta(UB + \Lambda U) = G := E^T \overline{F}, \tag{3.40}$$

where $\widetilde{C} = E^T C E$. If we transpose (3.40) and use that $C^T = -C$, we can write

$$B(U')^T\Lambda + a\left(B(U')^T + (U')^T\Lambda \right) - \alpha_1 BU^T\widetilde{C} + \alpha_2 CU^T\Lambda - \beta\left(BU^T + U^T\Lambda \right) = G^T. \tag{3.41}$$

The matrix system (3.41) may be decoupled in some cases. For example, if we assume $\alpha = 0$, and u_j, g_j denote, respectively, the jth column of U^T and G^T, then (3.41) reads

$$(a\lambda_j I_{N-1} + (1 + a\lambda_j)B)u_j - \beta(\lambda_j I_{N-1} + B)u_j = g_j, \tag{3.42}$$

for $j = 0, \ldots, N-2$. Then the implementation of the procedure consists of the eigendecomposition (3.38), the computation of G, the resolution of (3.42) (where we recall that B is pentadiagonal with only three nonzero diagonals), and the change $V = EU$. In the nonlinear case ($\gamma \neq 0$) we can still combine the previous approach with a G-NI approximation of the nonlinearity, using the relation between the two representations (3.24) and (3.35), that is

$$V_{h,i}(t) = \sum_{k,j=0}^{N-2} v_{kj}^N(t)\phi_k(x_h)\phi_j(x_i), \quad h,i = 0, \ldots, N.$$

Remark 3.2. As it is well known, [85,48,23,110,109], spectral discretizations are generally ill conditioned because of the condition number of the matrices involved, which typically grows like $O(N^4)$. For the experiments below (cf. Section 3.4) we did not observe limitations in this sense. This was probably due to the use of the basis chosen to expand the spectral approximation. The case illustrated above for the linear problem is an example of the basis recombination technique, which expands the spectral approximation in a basis of polynomials fulfilling the boundary conditions and that may lead to well-conditioned systems, specially in the Dirichlet case, [58,98]. In addition, and due (among other reasons) to the estimates proved in section 2, the values of N considered in the experiments of section 3.4 were not needed to be too large.

The technique of basis recombination has several alternatives, useful for more general boundary conditions, like the boundary bordering considered in [84], and references therein. Furthermore, in order to reduce ill conditioning, several strategies of preconditioning and stabilization are indeed available, [23,59,112,98,84].

3.2. Spectral collocation approximation

Consider first the 1D case. For an integer $N \geq 2$, $v^N : [0,T] \to \mathbb{P}_N^0$ stands for the semidiscrete spectral collocation approximation satisfying (2.37), (2.38) and based on Chebyshev polynomials

$$v^N(x,t) = \sum_{k=0}^{N} v_k^N(t) T_k(x), \tag{3.43}$$

with $T_k(x)$ standing for the Chebyshev polynomial of degree k. The approximation v^N is usually represented by the nodal values

$$V(t) = V^N(t) = (v^N(x_0,t), \dots, v^N(x_N,t))^T, \tag{3.44}$$

at the Gauss-Lobatto nodes $x_j, j = 0, \dots, N$. The vector (3.44) is related to (3.43) by the formula, [23,89]

$$v_k(t) = \sum_{j=0}^{N} C_{kj} V_j^N(t),\ V_j^N(t) = v^N(x_j,t),$$
$$C_{kj} = \frac{2}{c_k c_j} \cos\frac{jk\pi}{N},\ c_j = \begin{cases} 2 & j = 0, N \\ 1 & j = 1, \dots, N-1 \end{cases}.$$

The general formulation of the semidiscrete system for (3.44) can be derived by using a representation of V in the nodal basis (3.12). Thus, a full-matrix system, similar to that of the G-NI approach (3.15), can be obtained (but indeed with different nodes and weights). For practical purposes, it may be more interesting to describe the simplified formulations for the special cases (3.2) and (3.9). In the first one, we have

$$Z_N \left((I_N - aD_N^2) \frac{d}{dt} V^N(t) - bD_N^2 V^N(t) \right) = 0,$$

where

- D_N is now the $N \times N$ Chebyshev interpolation differentiation matrix, [23], and $D_N^2 = D_N D_N$.
- Z_N is the $N \times N$ matrix that represents setting the first and the last components of a vector equals zero, enforcing in this way the boundary conditions (3.4) directly.

Similarly, for (3.9), the semidiscrete system in 1D is

$$(I_N - aD_N^2) \frac{d}{dt} V(t) + \alpha D_N V(t) + \beta D_N^2 V(t) + \gamma D_N f(V(t)) = \overline{F}, \tag{3.45}$$

where $\overline{F}$ stands for the vector of nodal values of F. In the 2D case, the semidiscretization is formulated using the tensor product approach, and the 2D version of (3.43), given by

$$v^N(x,y,t) = \sum_{k,m=0}^{N} v_{k,m}^N(t) T_k(x) T_m(y),$$

and represented by the nodal values

$$V(t) = (v^N(x_i, x_j, t))_{i,j=0}^N.$$

The resulting formulas are similar to those in (3.27), (3.28). The two-dimensional version of (3.45) then reads

$$\begin{aligned} &\left(I_N \otimes I_N - a\left(D_N^2 \otimes I_N + I_N \otimes D_N^2\right)\right) \frac{d}{dt} V(t) + \left(\alpha_1 D_N \otimes I_N - \alpha_2 I_N \otimes D_N\right) V(t) \\ &\quad + \beta \left(D_N^2 \otimes I_N + I_N \otimes D_N^2\right) V(t) + \left(\gamma_1 D_N \otimes I_N - \gamma_2 I_N \otimes D_N\right) f(V(t)) = \overline{F}, \end{aligned} \tag{3.46}$$

for $V(t)$, $f(V(t))$ and $\overline{F}$ in the vectorized form, cf. (3.26).

Remark 3.3. We note that the matrix $I_N - aD_N^2$ in (3.45) is always invertible for $a > 0$, since the eigenvalues of D_N^2 are real and negative, [48]. Possible ill conditioning in the resolution of (3.45) can also overcome (besides with some of the procedures already mentioned in Remark 3.2) with standard direct or iterative techniques for ill-conditioned systems, [94] and references therein. For the experiments below, see the description in section 3.4.

On the other hand, the strategy, described above, to simplify the implementation of the Legendre Galerkin method also implies an important saving in the computational effort. In the case of the Chebyshev collocation scheme, the performance in the computational cost can be improved with the combined use of a fast (FFT-based) transform and an efficient, preconditioned iterative technique for the resolution of the resulting systems, [98] (cf. Section 3.4 for details concerning the experiments of the present paper).

3.3. Full discretization

Our proposal for a numerical treatment of (1.3)-(1.5) is completed by the choice of schemes to approximate in time the spectral ode semidiscrete systems. To this end, we consider the singly diagonally implicit Runge-Kutta (SDIRK) methods of Butcher tableau

$$\begin{array}{c|cc} \mu & \mu & 0 \\ 1-\mu & 1-2\mu & \mu \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \tag{3.47}$$

with $\mu = 1/2$ (implicit midpoint rule, order two) and $\mu = \frac{3+\sqrt{3}}{6}$ (order three). Among other properties, the methods are A-stable (and therefore L-stable), cf. [54,55].

Our motivation for the choice of (3.47) mainly concerns the possible oscillatory stiff character (1.3) or the corresponding spectral semidiscrete systems. This phenomenon may happen, for example, when discretizing in space hyperbolic problems. In our case, the presence of the third order derivative $\nabla \cdot \nabla \partial_t$ typically tends to regularize the evolution (and, as mentioned before, the stiff character) but the presence of oscillations, from discontinuous data, during the numerical simulation is not discarded if the hyperbolic terms in (1.3) are dominant. For example, in the BBM-Burgers case (3.9), (3.10), the previous situation may occur if $\gamma >> a$. In this sense, SDIRK methods like (3.47) combine two favourable qualitative properties that may ensure the stabilization of the discretization.

The first property is the generation of small dispersion errors of the oscillations in the numerical approximation. The literature on this topic is numerous (cf. e.g. [67] and references therein for details). A brief description follows. For a system of ordinary differential equations

$$u'(t) = F(u), \tag{3.48}$$

obtained from a semidiscretization in space of some partial differential equations, we consider a s-stage, implicit Runge-Kutta method of Butcher tableau

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} \tag{3.49}$$

The stability properties of (3.49) are usually studied using the stability function

$$R(z) = \frac{\det(I - zA + z\widetilde{e}b^T)}{\det(I - zA)}, \tag{3.50}$$

with $\widetilde{e} = (1, 1, \dots, 1)^T$. The function R is a rational approximation to the exponential function

$$e^z - R(z) = \sum_{j=0}^{\infty} C_j z^j,$$

for some coefficients C_j. The linear order of (3.49) is defined as the integer p for which $C_j = 0, j = 0, \dots, p, C_{p+1} \neq 0$. On the other hand, dispersion errors when integrating (3.48) with (3.49) are measured by the phase error function

$$\Phi(y) := y - \arg(R(iy)), \quad y \in \mathbb{R}, \tag{3.51}$$

(where $\arg(z)$ denotes an argument of z). Analyticity at the origin enables to expand (3.51)

$$\Phi(y) = \sum_{j=0}^{\infty} C_{p,j} y^j,$$

and to define the phase order of (3.49) with stability function (3.50) as the integer q such that

$$C_{p,j} = 0, j = 0, \dots, q, C_{p,q+1} \neq 0.$$

Then (3.49) is said to be dispersive of order q.

Several papers in the literature develop the theory of construction of dispersive RK (and multi-step) methods and, in the RK case, the relation of the order q with the linear order p and the number of stages s. For the case of DIRK and SDIRK methods, this relation is characterized in e.g. [73]. In particular, for the schemes (3.47) with $\mu = 1/2$ (considered as a 1-stage method) and $\mu = \frac{3+\sqrt{3}}{6}$, the phase order is $q = 2$ and $q = 4$, respectively. This ensures small dispersion errors in the case of the appearance of oscillatory behaviour in the simulation, which is specially useful for long times.

A second aspect that we considered for the choice of (3.47) is the strong stability preserving (SSP) property and the use of the so-called SSP methods. These time integration schemes preserve the strong stability properties of spatial discretizations under the forward Euler time integration. Their formulation relies on the following SSP property (see [52] for details). For the system (3.48), assume that the forward Euler method

$$u_{FE}^{n+1} = u_{FE}^n + \Delta t F(u_{FE}^n),$$

satisfies, in some convex functional $||\cdot||_C$

$$||u_{FE}^{n+1}||_C \leq ||u_{FE}^n||_C,$$

when $\Delta t \leq \Delta t_{FE}$ for some Δt_{FE}. Given a s-stage Runge-Kutta (RK) method for (3.48), written in the form

$$\begin{aligned} y_i &= u^n + \Delta t \sum_{j=1}^{s} a_{ij} F(y_j),\ 1 \leq i \leq s+1, \\ u^{n+1} &= y_{s+1}, \end{aligned} \tag{3.52}$$

the SSP coefficient of (3.52) is defined as the largest constant $\kappa \geq 0$ such that

$$||y_i||_C \leq ||u^n||_C,\ 1 \leq i \leq s+1,$$

(which in particular implies $||u^{n+1}||_C \leq ||u^n||_C$) whenever

Table 1
Six smallest eigenvalues λ of B_N for several N.

$N = 16$	$N = 32$	$N = 64$
3.8483E-03	3.0081E-04	2.0239E-05
3.1038E-03	2.6739E-04	1.9040E-05
1.9343E-03	1.3810E-04	9.0673E-06
1.5451E-03	1.2251E-04	8.5278E-06
5.2100E-04	3.5183E-05	2.2777E-06
4.1274E-04	3.1177E-05	2.1418E-06

$$\Delta t \leq \kappa \Delta t_{FE}. \tag{3.53}$$

If $\kappa > 0$, the method (3.52) is said to be strong stability preserving under (3.53). Examples of convex functionals $||\cdot||_C$ are the classical norms and the total variation seminorm.

Our motivation for the use of SSP methods in (1.3)-(1.5) is similar to that mentioned above for the use of methods with dispersive stability functions; it can be found in the search for a way to ensure the stabilization of the discretization when dealing with discontinuous data. Note that several examples, see e.g. [52,50], reveal the advantages of SSP methods in hyperbolic problems like Burgers or Euler equations.

In our case, this application would also require a previous analysis on the behaviour of the spectral semidiscretizations with respect to the Euler method. Our confidence here is based on the stability results of the Euler method in other related approaches, [12,91]. By way of illustration, we may analyze the approximation to the Legendre semidiscrete system (3.30) by the forward Euler scheme. For $t_n = n\Delta t, n = 0, 1, \ldots$ let $V_{FE}^n \in \mathbb{R}^{N-1}$ be an approximation to $V(t_n)$ such that

$$K_N \left(\frac{V_{FE}^{n+1} - V_{EF}^n}{\Delta t} \right) + S_N V_{FE}^n = 0,\ n = 0, 1, \ldots$$

Using (3.34), this can be written as

$$V_{FE}^{n+1} = (I_{N-1} - b\Delta t K_N^{-1}) V_{FE}^n,\ n = 0, 1, \ldots \tag{3.54}$$

Note that since B_N is symmetric, all its eigenvalues are real. Furthermore, it is not hard to check that

$$b_{jj} b_{j+2,j+2} - b_{j,j+2}^2 > 0,$$

which implies that B_N is also positive definite. Therefore all the eigenvalues λ are positive. Thus, from (3.54), we have

$$||V_{FE}^{n+1}|| \leq ||V_{FE}^n||, \tag{3.55}$$

(where here $||\cdot||_C = ||\cdot||$ denotes the usual Euclidean norm in $\mathbb{R}^{N-1}$) when $\Delta t < \mu/b$, for all $\mu = a + \lambda$ eigenvalue of K_N. This implies that, taking $\Delta t_{FE} = \mu_{min}/b$, where $\mu_{\min} = \min\{\mu, \mu \text{ eigenvalue of } K_N\}$, we obtain that the semidiscretization (3.30) satisfies the monotonicity property (3.55) with respect to the Euler method. It is experimentally observed (see Table 1) that as $N \to \infty$ the smallest eigenvalue $\lambda = \lambda_N$ of B_N tends to zero. This means that asymptotically Δt_{FE} behaves like a/b and in practice the choice $\Delta t_{FE} = a/b$ would imply (3.55) for $\Delta t \leq \Delta t_{FE}$.

We finally observe that the SDIRK methods (3.47) are SSP methods and both were shown to be optimal (within the corresponding SDIRK schemes with the same stages and order) in the sense that the value κ in property (3.53) is maximal, [42,71]. They will be denoted by SSP12 ($\mu = 1/2$, 1 stage, order 2) and SSP23 ($\mu = \frac{3+\sqrt{3}}{6}$, 2 stages, order 3). It is indeed possible the use of higher-order methods and of different type (other Runge-Kutta families or multisteps methods), [50,51].

3.4. A computational study

In this section we make a computational study to check the performance of the numerical methods described above, considering (3.2) and (3.9) as model problems.

The implementation of the fully discrete schemes is performed as follows. For the experiments with nonlinear problems below, and taking advantage of the diagonally implicit structure of the temporal discretization schemes, the corresponding implicit systems for the intermediate stages are numerically solved by the classical fixed point iteration. (Note that we have one intermediate stage for SSP12 and two for SSP23.) In all the computations we did not require more than two iterations per stage. In the case of the discretization of (3.20), the matrices (3.21)-(3.23) are computed directly, and this is also used in the resolution of the systems of the iterative process. Other alternatives, based on differentiation in frequency space, [98], may be somehow adapted to the representation (3.14). (To our knowledge, the approach in [98] would be the closest idea to what might be called fast transform in this Legendre case.) On the other hand, the full discretization of (3.45) takes advantage of the computation of $D_N V$ with FFT techniques, [23,89]. For the resolution of the systems in the fixed point iterative process, we implemented two procedures: the first one was a direct resolution, based on a unique factorization of the matrix $I_N - aD_N^2, a > 0$. The second one was carried out with Krylov methods, [94], making use of direct computations with $I_N - aD_N^2$ to implement the Conjugate Gradient method. This second approach was experimentally observed more efficient (probably because of the performance of the FFT in the computations) and was taken in the numerical results below. We also recall that in the case of linear problems, iteration is not necessary. For the computations in 2D, we made use of the Kronecker formulations (3.27) and (3.46); in some linear cases we also implemented the procedure based on the eigendecomposition (3.35). The application of this approach to the nonlinear case, and the use of dynamical low-rank approximation (cf. [26] and references therein) for the matrix or tensor systems will be considered for a future research.

In the 1D case and concerning the Legendre Galerkin method, the numerical solution at a final time $T = M\Delta t$ is evaluated at a grid of Chebyshev points in $(-1, 1)$

$$x_j = \cos \frac{j\pi}{P},\ j = 0, \ldots, P,$$

Table 2
Numerical approximation of (3.9), (3.10), (3.56): L^2 and H^1 norms of the error at $T = 1$ with Legendre Galerkin method and $N = 256$.

Δt	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	L^2 Error	H^1 Error	L^2 Error	H^1 Error
0.1	2.8114E-04	1.0382E-03	2.6531E-05	9.6292E-05
0.05	7.0232E-05	2.5936E-04	3.4773E-06	1.2570E-05
0.0025	1.7555E-05	6.4830E-05	4.4547E-07	1.6069E-06
0.00125	4.3885E-06	1.6207E-05	5.6383E-08	2.0316E-07

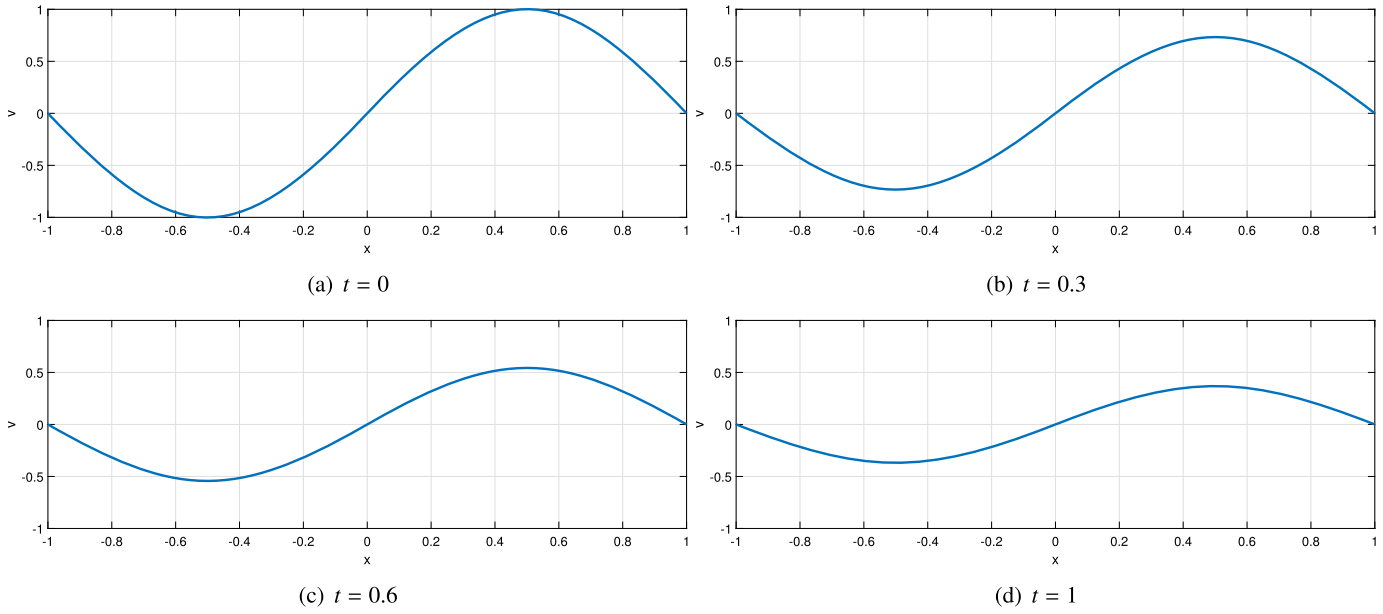


Fig. 1. Numerical solution with Legendre Galerkin and SSP23 for the problem (3.9), (3.10), (3.56) at $t = 0, 0.3, 0.6, 1$.

and compared with the solution at the grid using the L^2, H^1 and L^∞ norms

$$||E(h)||_2 = \left(h \sum_{j=1}^{P} (v^M(x_j) - v(x_j, T))^2 \right)^{1/2},$$

$$||E(h)||_{H^1} = \left(h \sum_{j=1}^{P} ((v^M)'(x_j) - v'(x_j, T))^2 + ||E(h)||_{L^2}^2 \right)^{1/2},$$

$$||E(h)||_\infty = \max_{1 \le j \le P} |v^M(x_j) - v(x_j, T)|,$$

where $h = 2/N$. (For the computations in 2D, we considered the same formulas for the errors, but with respect to the grid $\{(x_i, x_j) : i, j = 0, \ldots, P\}$ and using the vectorized form of the approximations and the solutions.) For the Chebyshev collocation scheme, the comparisons are made in the corresponding weighted, discrete norms, computing the derivative with the matrix D_N. Note that in this case, we take into account that the formulation of the scheme gives the role of representation of the numerical solution to the vector of approximation at the quadrature nodes. In most of the computations the L^2 and L^∞ norms give similar conclusions. For that reason, the L^∞ norm of the error will be shown only in those experiments for which it provides new features.

3.4.1. Problem 1. Spectral convergence for Legendre Galerkin approximation

In order to check the spectral convergence for regular data, we first consider the BBM-Burgers problem (3.9), (3.10) in $\Omega = (-1, 1)$ with homogeneous boundary conditions, $a = \alpha = 1, \beta = -1, \gamma = 1/2$, and

$$v(x, 0) = \sin(\pi x),$$
$$F(x, t) = e^{-t} \left(-\sin(\pi x) + \pi \cos(\pi x)(1 + e^{-t} \sin(\pi x)) \right). \tag{3.56}$$

The exact solution is $v(x, t) = e^{-t} \sin(\pi x)$, [82]. The problem is approximated by the Legendre G-NI method and the two SSP time integrators. L^2 and H^1 errors at $T = 1$ with $N = 256$ and several values of the time stepsize Δt are shown in Table 2. The results show the corresponding order of convergence of the time integrators. (We checked that larger values of N did not give any change in this behaviour.) The form of the numerical solution at several times is shown in Fig. 1.

Table 3
Numerical approximation of (3.9), (3.10), (3.57): L^2 norms of the error at $T = 1$ and rates of convergence with Legendre Galerkin method and $N = 64$.

Δt	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	L^2 Error	Rate	L^2 Error	Rate
0.1	4.6655E-04		4.4400E-05	
0.05	1.1660E-04	2.000	5.6934E-06	2.963
0.0025	2.9146E-05	2.000	7.2100E-07	2.981
0.00125	7.2864E-06	2.000	9.0721E-08	2.991

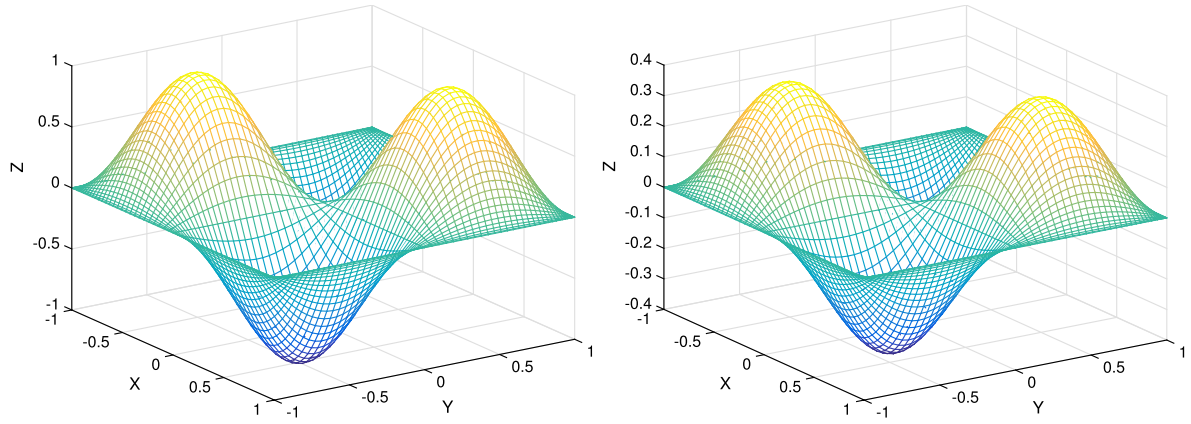


Fig. 2. Numerical solution with Legendre Galerkin and SSP23 for the problem (3.9), (3.10), (3.57) at $t = 0, 1$ with $\Delta t = 0.025$.

The spectral convergence and the order of the temporal discretization are also checked in Table 3, corresponding to the L^2 errors at $T = 1$, with $N = 64$, for the 2D problem (3.9), (3.10) in $\Omega = (-1, 1)^2$, with homogeneous boundary conditions, $a = 2, \alpha = (1, 1)^T, \beta = -1, \gamma = (1/2, 1/2)^T$, and

$$\begin{aligned} &v(x, y, 0) = \sin(\pi x)\sin(\pi y), \\ &F(x, y, t) = e^{-t}\left(-(1 + 2(a + \beta)\pi^2)\sin(\pi x)\sin(\pi y) + \pi(\alpha_1 \cos(\pi x)\sin(\pi y) + \alpha_2 \sin(\pi x)\cos(\pi y))\right) \\ &\qquad + 2\pi e^{-2t}\sin(\pi x)\sin(\pi y)(\gamma_1 \cos(\pi x)\sin(\pi y) + \gamma_2 \sin(\pi x)\cos(\pi y)). \end{aligned} \tag{3.57}$$

The exact solution is $v(x, y, t) = e^{-t}\sin(\pi x)\sin(\pi y)$. Fig. 2 shows the form of the numerical approximation at the initial and final time given by the discretization with the spectral Legendre Galerkin G-NI method and SSP23.

3.4.2. *Problem 2. Spectral convergence for Chebyshev collocation approximation*

The Chebyshev collocation scheme is now used to approximate the BBM-Burgers problem (3.9), (3.10) in $\Omega = (-20, 30)$ with homogeneous boundary conditions, $a = \alpha = \beta = 1, \gamma = -1/2$ and, [76]

$$\begin{aligned} &v(x, 0) = \operatorname{sech}(x), \\ &F(x, t) = \operatorname{sech}(x - t)\left(1 - 6\tanh^3(x - t) - 2\tanh^2(x - t)\right. \\ &\qquad \left. + \tanh(x - t)(5 + \operatorname{sech}(x - t))\right). \end{aligned} \tag{3.58}$$

The function $v(x, t) = \operatorname{sech}(x - t)$ is the solution of the corresponding initial-value problem. Strictly speaking, it does not satisfy the homogeneous boundary conditions. But its values at the boundaries $x = -20, 30$ are, for each $t > 0$, small enough to take it for comparison with the numerical solutions given by Chebyshev collocation and SSP12, SSP23 methods. The L^2 and H^1 errors at $T = 10$ are shown in Table 4, while the travelling wave form for the numerical profile is illustrated in Fig. 3. The errors show again the order of convergence in time of the fully discrete methods. Here a larger value of N is required. This is probably related with the approximation at the maximum height of the wave and the fact that the Chebyshev points are not equally distributed.

The performance of the Chebyshev collocation scheme is also illustrated by approximating the 2D problem (3.9), (3.10) in $\Omega = (-\pi, \pi) \times (-1, 1)$, with $a = 5, \alpha = (0, 0)^T, \beta = 1, \gamma = (1, 1)^T$, boundary conditions

$$v(-\pi, y, t) = -1 - y, \quad v(\pi, y, t) = 1 - y, \quad v(x, -1, t) = x + 1, \quad v(x, 1, t) = x - 1,$$

and

$$v(x, y, 0) = \sin(x)\sin(\pi y) + x - y,$$

Table 4
Numerical approximation of (3.9), (3.10), (3.58): L^2 and H^1 norms of the error at $T = 1$ with Chebyshev collocation method and $N = 1024$.

Δt	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	L^2 Error	H^1 Error	L^2 Error	H^1 Error
0.1	5.2981E-04	9.8639E-04	2.2235E-05	3.8118E-05
0.05	1.3237E-04	2.4636E-04	32.9139E-06	4.9823E-06
0.0025	3.3087E-05	6.1631E-05	3.7482E-07	6.3747E-07
0.00125	8.2715E-06	1.5638E-05	4.7773E-08	8.1216E-08

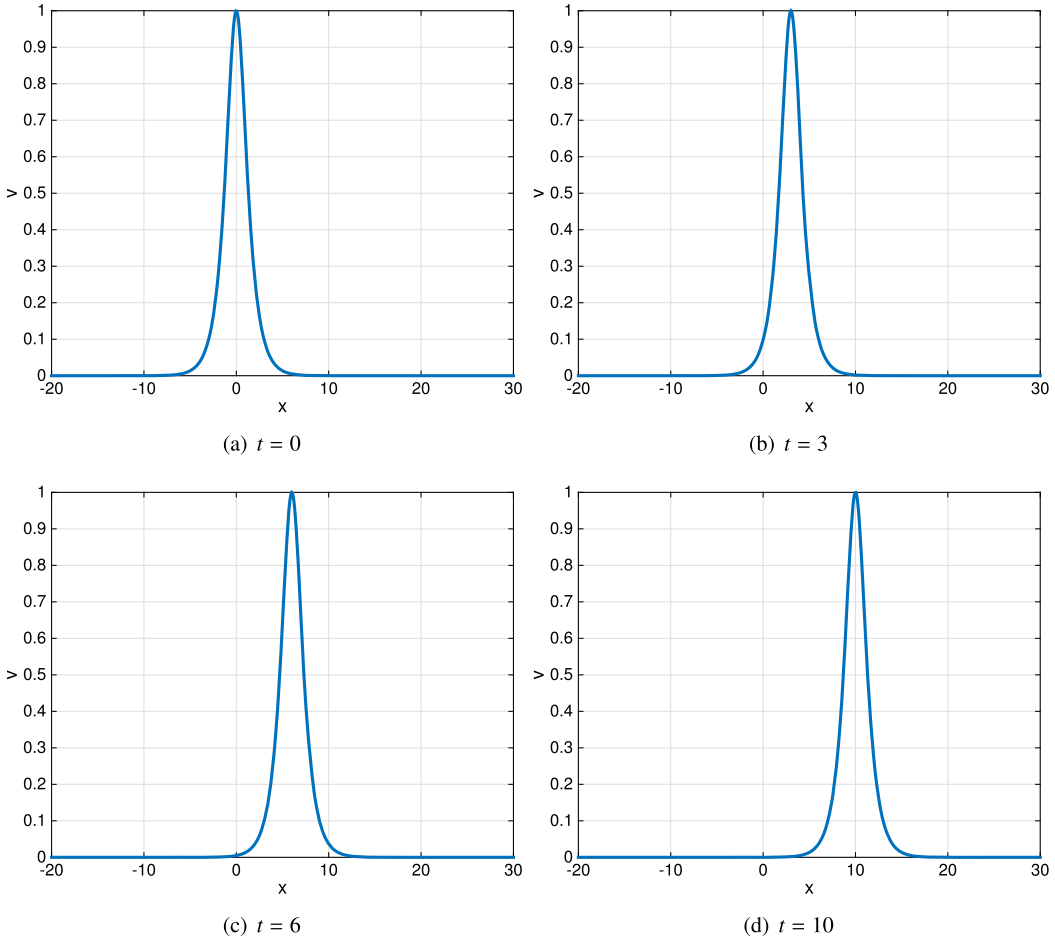


Fig. 3. Numerical solution with Chebyshev collocation and SSP23 for the problem (3.9), (3.10), (3.58) at $t = 0, 3, 6, 10$.

$$
\begin{aligned}
F(x,y,t) = e^{-t}\big(&-(1+(a+\beta)(1+\pi^2))\sin(x)\sin(\pi y) + \alpha_1\cos(x)\sin(\pi y) + \pi\alpha_2\sin(x)\cos(\pi y)\\
&+2(x-y)(\gamma_1\cos(x)\sin(\pi y) + \pi\gamma_2\sin(x)\cos(\pi y)) + 2(\gamma_1-\gamma_2)\sin(x)\sin(\pi y)\big)\\
&+2e^{-2t}\sin(x)\sin(\pi y)(\gamma_1\cos(x)\sin(\pi y) + \pi\gamma_2\sin(x)\cos(\pi y)) + 2(\gamma_1-\gamma_2)(x-y).
\end{aligned}
\tag{3.59}
$$

The exact solution is $v(x,y,t) = e^{-t}\sin(x)\sin(\pi y) + x - y$. Table 5 shows the L^2 errors at $T = 10$ given by the time integrators with $N = 64$; this is enough to confirm the spectral convergence of the spatial approximation and the order of convergence of the temporal discretization. The form of the numerical solution is shown in Fig. 4.

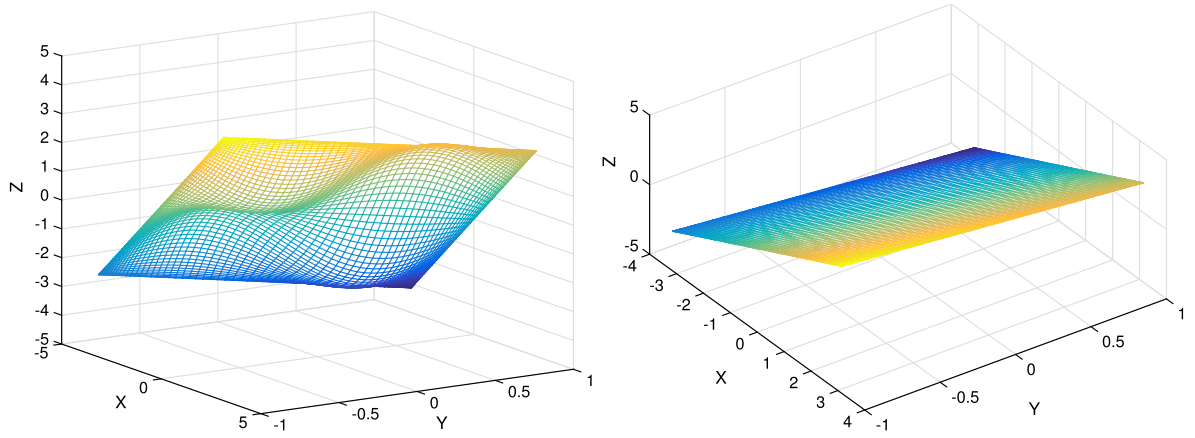
3.4.3. Problem 3. Nonsmooth data

We are now interested in studying the performance of the methods when the initial data has low regularity. To this end we perform numerical experiments to compute the numerical rates of convergence of the spatial discretization in the corresponding norms. In all cases, we checked with several ranges of time stepsize Δt that errors and orders do not change with smaller values than those that were finally taken. Unless otherwise stated, we fix $\Delta t = h/2$. We first consider (3.2)-(3.4) with $a = b = 1$ and

$$
v_0(x) = \begin{cases} 1 & |x| \le 2 \\ 0 & \text{otherwise} \end{cases}. \tag{3.60}
$$

Table 5
Numerical approximation of (3.9), (3.10), (3.59): L^2 norms of the error at $T = 10$ and rates of convergence with Legendre Galerkin method and $N = 64$.

Δt	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	L^2 Error	Rate	L^2 Error	Rate
0.1	2.9201E-02		4.2167E-04	
0.05	7.3030E-03	1.999	5.2502E-05	3.006
0.0025	1.8259E-03	1.999	6.5480E-06	3.003



(a) $t = 0$ (b) $t = 10$

Fig. 4. Numerical solution with Chebyshev collocation and SSP23 for the problem (3.9), (3.10), (3.59) at $t = 0, 10$ with $\Delta t = 0.025$.

In this case, the expansion (3.6) has coefficients $C_n = \frac{2}{n\pi}\left(\cos\frac{n\pi}{4} - \cos\frac{3n\pi}{4}\right)$. The corresponding solution (3.5) is represented by a truncated series whose accuracy is checked by using acceleration techniques, [102]. Table 6 shows the errors and convergence rates at $T = 1$ of the Legendre Galerkin approximation for the two fully discrete methods using $h = 2/N$. In both, the lack of regularity makes the H^1 norm unable to control the error, but in the case of the other two norms, the error in space seems to be dominant and, according to the rates, like $O(N^{-1})$. The form of the numerical solution at different times is illustrated in Fig. 5. As the regularity of the initial condition is increasing, an increment in the spatial order of convergence is expected. Thus, taking

$$v_0(x) = 1 - |x|, \tag{3.61}$$

(see Fig. 6) the behaviour of the errors in L^2 norm corresponding to SSP23 (which is third-order), observed in Table 7, suggests an error in space of $O(N^{-2})$, while in the case of the H^1 norm this seems to be $O(N^{-1/2})$. In the case of SSP12, since $\Delta t = O(h)$, the order of the spatial error in L^2 norm would coincide with the second order in time.

A final experiment concerns the initial condition

$$v_0(x) = \begin{cases} 1 + 2x + x^2 & -1 \le x \le 0 \\ 1 + 2x - 3x^2 & 0 \le x \le 1 \end{cases}, \tag{3.62}$$

whose second derivative has a discontinuity at $x = 0$. The corresponding results, displayed in Table 8, show that the dominant error in time is recovered. (See Fig. 7 to illustrate the form of the approximation at several times.) In order to determine the order in space, Table 9 shows the errors given by SSP23 with $\Delta t = O(h^2)$. The rates for the L^2 error norm suggest an spatial error of $O(N^{-4})$, while the H^1 norm of the error is similar to that of Table 8, and this behaves like $O(N^{-3/2})$. According to the experiments with (3.60)-(3.62) and at least for the case of the (3.2)-(3.4), the numerical results suggest that, for $v_0 \in H^m_w, m \ge 1$

$$\max_{0\le t\le T} ||v^N(t) - v(t)||_{1,w} \le CN^{1/2-m}, \qquad \max_{0\le t\le T} ||v^N(t) - v(t)||_{0,w} \le CN^{-2m}.$$

The experiments with (3.60) also suggest to conjecture that if $v_0 \in L^2_w$ then

$$\max_{0\le t\le T} ||v^N(t) - v(t)||_{0,w} \le CN^{-1}.$$

The results corresponding to the Chebyshev collocation method do not show any qualitative change with respect to those given by the Legendre Galerkin discretization. Thus previous experiments suggest a similar error behaviour for Galerkin and collocation methods. (Recall that the implementation makes the schemes have a related formulation.)

Similar effects were observed in the 2D case. By way of illustration we consider (3.2)-(3.4) with $a = b = 1$ and

$$v_0(x, y) = \begin{cases} 1 & x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.63}$$

Table 6
Numerical approximation of (3.2)-(3.4) from (3.60): L^2 and L^∞ norms at $T = 1$ of the error with Legendre Galerkin method and $\Delta t = 0.5h, h = 2/N$.

N	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_\infty$	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_\infty$
32	6.0479E-04	6.2190E-04	6.2133E-04	6.4695E-04
64	3.0184E-04	3.1317E-04	2.9785E-04	3.0665E-04
128	1.5098E-04	1.5723E-04	1.5201E-04	1.5887E-04
256	7.5451E-05	8.1659E-05	7.5194E-05	8.1246E-05

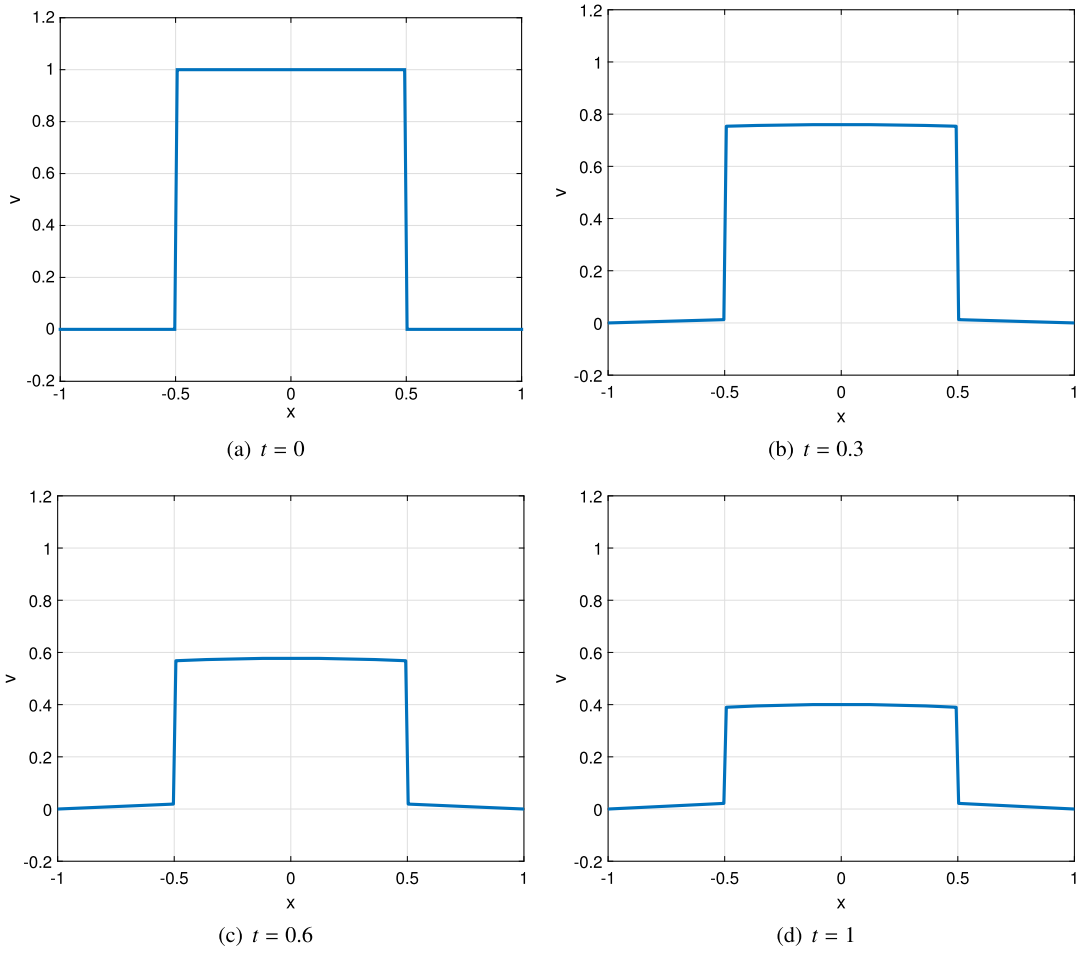


Fig. 5. Numerical approximation of (3.2)-(3.4) with Legendre Galerkin and SSP23 from the initial condition (3.60) at $t = 0, 0.3, 0.6, 1$.

(cf. Fig. 8(a)). The numerical solution given by the Legendre Galerkin method and SSP23 at $T = 1$ is shown on Fig. 8(b). This can be compared with an approximation to the solution computed by an accurate truncation of (3.7).

In the nonlinear case, the performance of the methods with less regular conditions is illustrated by the following experiments. Starting now with the 2D case, we consider the problem (3.9), (3.11) with $\gamma = (1,1), \beta = -1, a = 5, \alpha = (0,0)$ on $\Omega = (-1,1)^2$, homogeneous boundary conditions, and initial condition and source term given by

$$v(x,y,0) = 1 - |x| - |y|, \quad F = F(x,y,t) = 1/(1 + x^2 + y^2), \quad (x,y) \in \Omega. \tag{3.64}$$

The initial profile and its projection onto $y = 0$ are shown in Figs. 9(a) and (b) respectively. The problem was integrated numerically with the Legendre G-NI scheme and the SSP23 method up to a final time $T = 10$. The resulting numerical solution and corresponding projection onto $y = 0$ can be seen in Figs. 9(c) and (d) respectively. The discretization captures the smoothing effect of the evolution and does not seem to develop spurious oscillations.

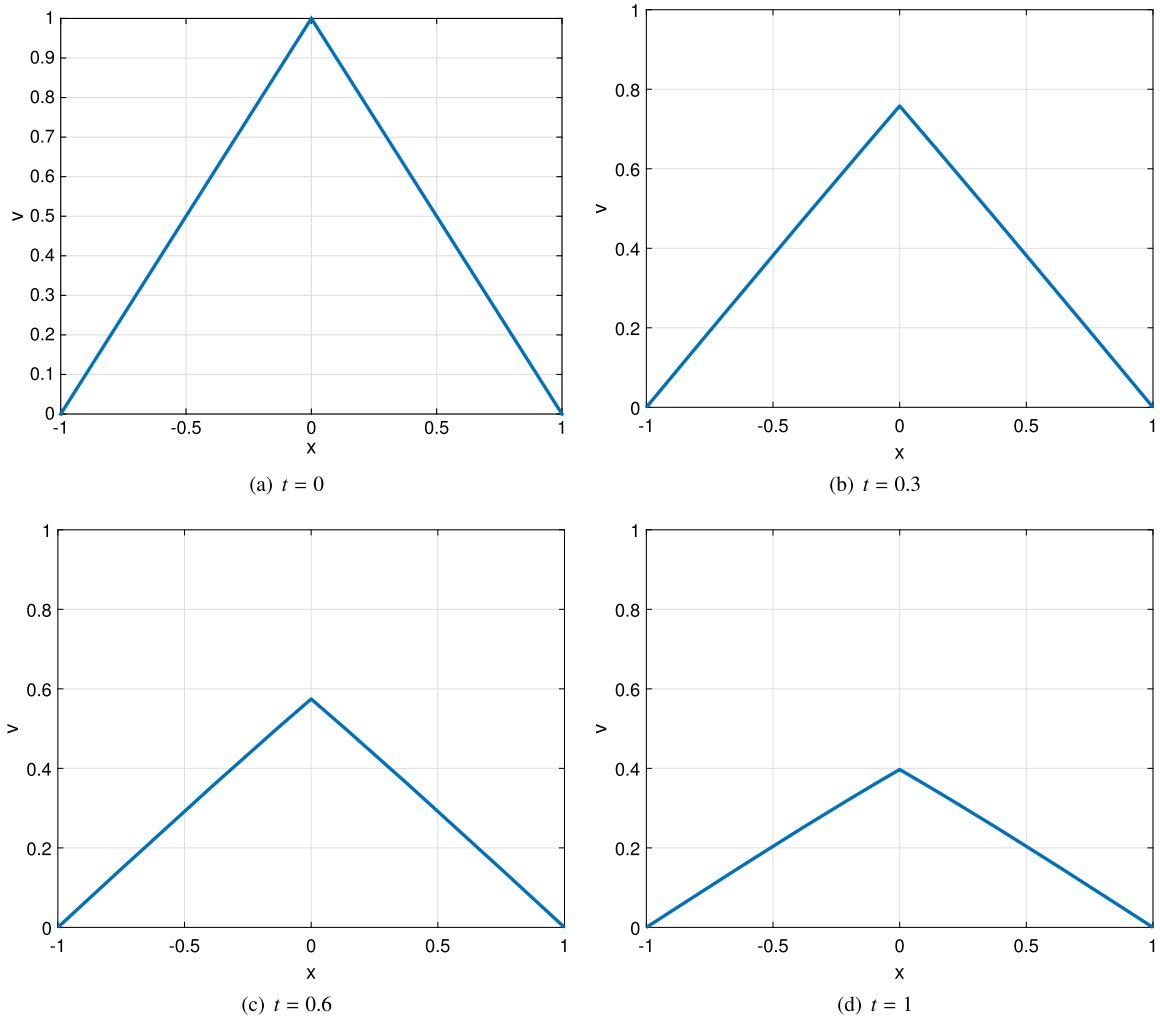
A similar behaviour, in the one-dimensional case, can be observed when approximating (3.9), (3.11) on $\Omega = (-60, 210)$ with $\alpha = 0, a = 5, \beta = -1, \gamma = 1$, $F = 0$, nonhomogeneous boundary conditions

$$v(-60,t) = S_L, \ \ v(210,t) = S_R,$$

and Riemann type initial data

Table 7
Numerical approximation of (3.2)-(3.4) from (3.61): L^2 and H^1 norms at $T = 1$ of the error with Legendre Galerkin method and $\Delta t = 0.5h, h = 2/N$.

N	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_{H^1}$	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_{H^1}$
32	3.0806E-05	7.1528E-02	4.6914E-05	7.1537E-02
64	7.6988E-06	5.0138E-02	1.1779E-05	5.0140E-02
128	1.9260E-06	3.5374E-02	2.9535E-06	3.5374E-02
256	4.8245E-07	2.5000E-02	7.4021E-07	2.5001E-02



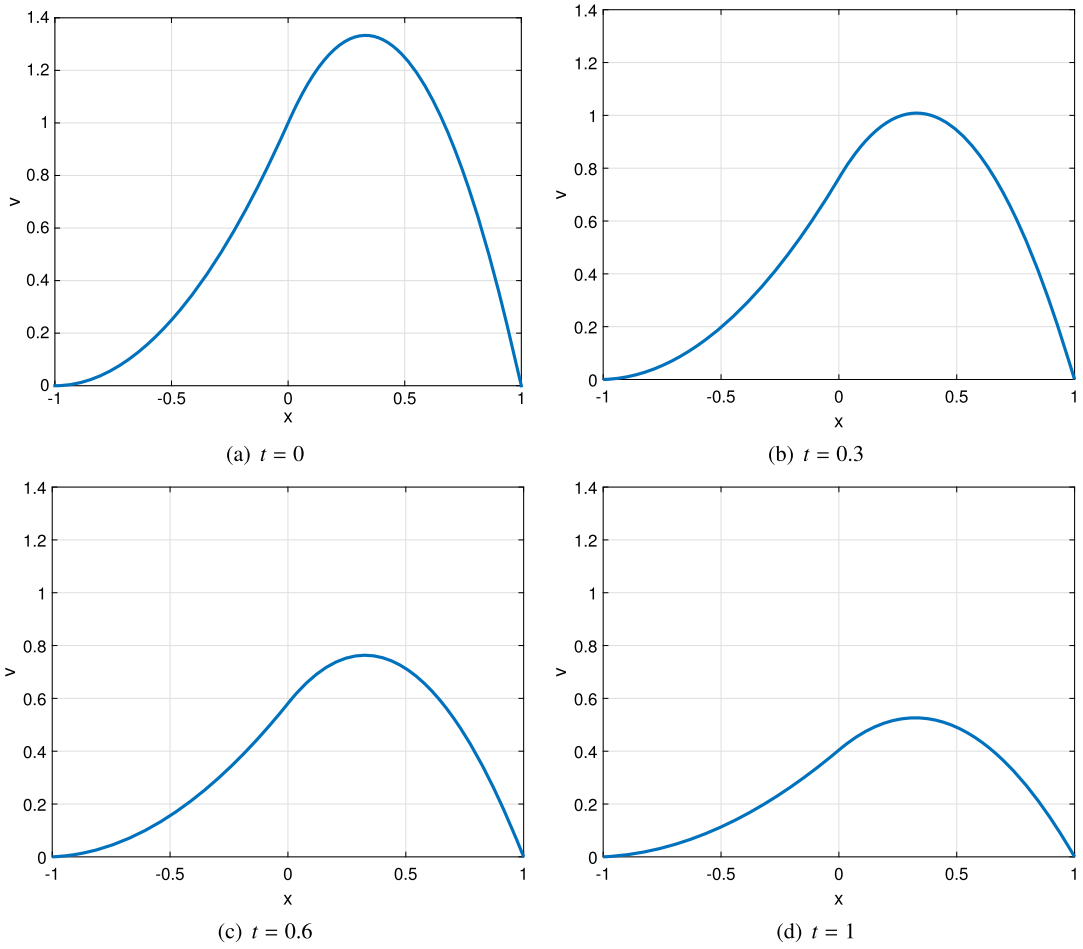
(a) $t = 0$ (b) $t = 0.3$

(c) $t = 0.6$ (d) $t = 1$

Fig. 6. Numerical approximation of (3.2)-(3.4) with Legendre Galerkin and SSP23 from the initial condition (3.61) at $t = 0, 0.3, 0.6, 1$.

Table 8
Numerical approximation of (3.2)-(3.4) from (3.62): L^2 and H^1 norms at $T = 1$ of the error with Legendre Galerkin method and $\Delta t = 0.5h, h = 2/N$.

N	$\mu = 1/2$		$\mu = \frac{3+\sqrt{3}}{6}$	
	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_{H^1}$	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_{H^1}$
32	2.5019E-05	9.6161E-03	8.0169E-07	9.6074E-03
64	6.2468E-06	3.3927E-03	9.8674E-08	3.3911E-03
128	1.5612E-06	1.1987E-03	1.2254E-08	1.1984E-03
256	3.9028E-07	4.2369E-04	1.5272E-09	4.2364E-04



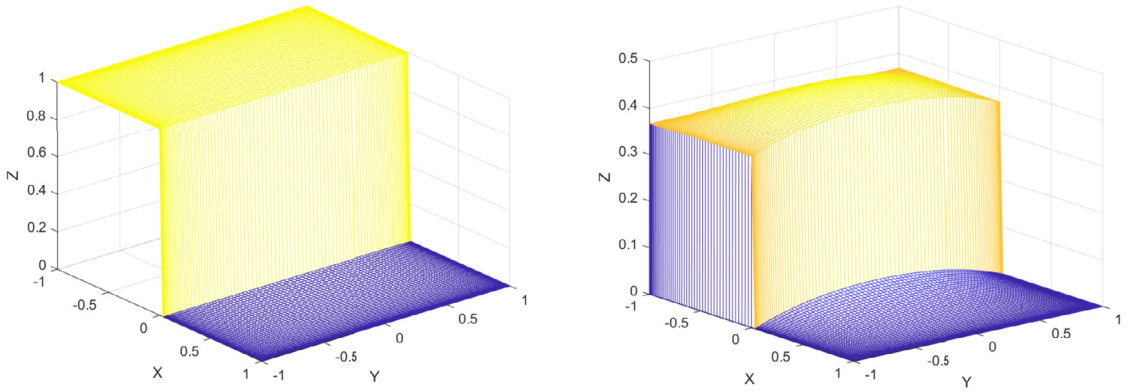
(a) $t = 0$ (b) $t = 0.3$

(c) $t = 0.6$ (d) $t = 1$

Fig. 7. Numerical approximation of (3.2)-(3.4) with Legendre Galerkin and SSP23 from the initial condition (3.62) at $t = 0, 0.3, 0.6, 1$.

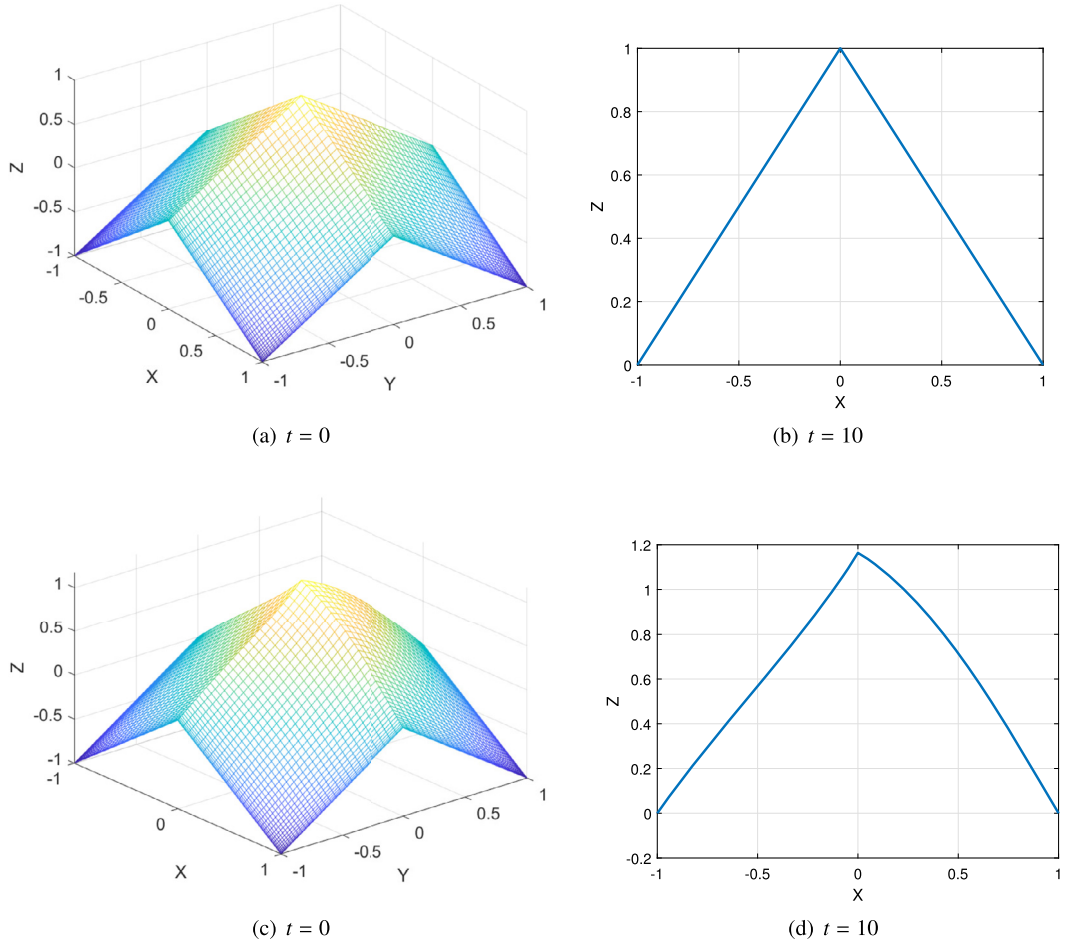
Table 9
Numerical approximation from (3.62): L^2 and H^1 norms at $T = 1$ of the error with Legendre Galerkin method, $\gamma = \frac{3+\sqrt{3}}{6}$ and $\Delta t = 0.25h^2, h = 2/N$.

N	$\|\|E(h)\|\|_2$	$\|\|E(h)\|\|_{H^1}$
32	9.3545E-08	9.6074E-03
64	5.8890E-09	3.3911E-03
128	6.5649E-10	1.1984E-03
256	4.1507E-11	4.2364E-04



(a) $t = 0$ (b) $t = 1$

Fig. 8. Numerical solution with Legendre Galerkin G-NI and SSP23 for the problem (3.2)-(3.4), (3.63) at $t = 0, 1$ with $\Delta t = 10^{-2}$.



(a) $t = 0$ (b) $t = 10$

(c) $t = 0$ (d) $t = 10$

Fig. 9. Numerical solution with Legendre G-NI and SSP23 for the problem (3.9), (3.11) with $\gamma = (1,1), \beta = -1, a = 5, \alpha = (0,0)$ on $\Omega = (-1,1)^2$ and (3.64) with $\Delta t = 0.05$. (a) Initial profile; (b) Numerical solution at $t = 10$; (c) and (d) are the corresponding projections onto $y = 0$.

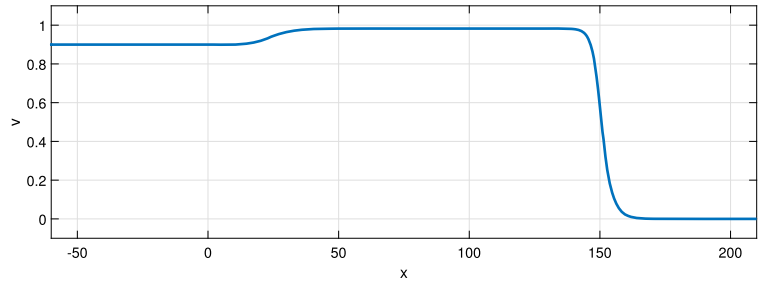
$$\eta(x) = \begin{cases} S_L & \text{if } x < 0 \\ S_R & \text{if } x \geq 0 \end{cases} \tag{3.65}$$

Two cases are considered: $S_L = 0.9, S_R = 0$ and $S_L = 0.55, S_R = 0$. (For the relevance of the models in porous media flows, see [6,5] and references therein.) For a final time of simulation $T = 150$, the numerical solution given by the Legendre G-NI method with SSP23 is shown in Fig. 10. The profiles do not seem to develop disturbances from the discontinuous initial data and the final structure of the solution is in accordance with that in the literature, [41,33].

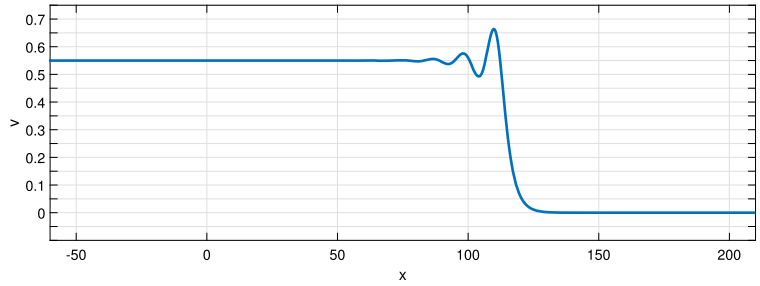
4. Concluding remarks

This paper attempts to contribute to the approximation to the initial-boundary-value problem, with Dirichlet boundary conditions, of pseudo-parabolic equations in several ways. First, a rigorous numerical analysis of the spectral Galerkin and collocation discretizations based on Jacobi polynomials, is made. Existence of numerical solution and convergence results are proved. The error estimates indeed depend on the regularity of the data. Specifically, if m denotes the order of regularity of the data of the problem and N is the degree of the polynomial approximation, then the error analysis for the spectral Galerkin method behaves like $O(N^{1-m})$ while the spectral collocation method behaves like $O(N^{2-m})$. On the other hand, the different exponent in the decay of the error (which depends on this regularity) between the estimates in the Galerkin and collocation schemes is due to the corresponding inverse inequalities associated to the Jacobi polynomials, cf. the Fourier spectral case studied in [91].

In the second part of the paper, a computational study of spectral discretizations in space, combined with SDIRK-SSP schemes for the time integration and without operator splitting strategies, is performed. The study is focused on the families of Legendre and Chebyshev polynomials, as widely used in the applications. We first make a detailed description of the Legendre Galerkin and Chebyshev collocation schemes. Then our choice for the full discretization is determined (beyond classical aspects of quantitative accuracy) by two qualitative aspects: the possibility that the semidiscretizations are affected by stiffness and the behaviour with respect to the integration with nonregular data, with the subsequent generation of a spurious oscillatory behaviour in the numerical solution. In order to avoid these problems without a great computational effort, we consider SDIRK methods with dispersive stability functions. In addition, for controlling the numerical solution from nonsmooth data, we propose to study the benefits of the SSP property in this pseudo-parabolic case and the corresponding use of SSP methods. Our suggestion is based on the performance of these schemes to approximate discontinuous solutions of hyperbolic partial differential equations. All this finally takes us to consider two SDIRK-SSP methods: one with second order of convergence and phase order two (SSP12), and the second with third order of convergence and phase order four (SSP23). In both cases, the SSP coefficients, within the corresponding number of stages and order, are optimal. The alternative of using higher-order and/or explicit SSP methods (for example, if the problem is known to be nonstiff) is always possible.



(a) $S_L = 0.9, S_R = 0$



(b) $S_L = 0.55, S_R = 0$

Fig. 10. Numerical solution with Legendre G-NI and SSP23 for the problem (3.9), (3.11), (3.65) at $t = 150$. (a) $S_L = 0.9, S_R = 0$; (b) $S_L = 0.55, S_R = 0$.

We present numerical experiments, in 1D and 2D, with linear and nonlinear problems, to check the performance of the methods. For smooth initial data (Problem 1, in Section 3.4.1 and Problem 2, in Section 3.4.2), the numerical experiments show the spectral order of convergence in space (confirming the theoretical results in the first part of the paper). We also study numerically the behaviour of the approximation from initial conditions with low regularity. The computations suggest some estimates on the behaviour of the error with respect to the smoothness of the initial condition. The use of SSP methods with dispersive stability functions seems to avoid, as in the hyperbolic case, spurious oscillations in the numerical approximation when simulating not regular solutions.

Keeping in mind the challenging task to preserve the numerical efficiency and robustness to the overall spectral approach being analyzed, several potential continuations of this work can be mentioned. The first one is the theoretical confirmation of some conjectures, experimentally suggested in the present paper and not covered by the convergence results, on the error behaviour for nonsmooth data. On the other hand, the good performance observed by the inclusion of the SSP property and the order of dispersion in the time discretization may deserve a deeper study, in order to analyze their extent and influence on the time behaviour of the simulation. In addition, some alternatives for the implementation in two-dimensional problems, suggested in the present study, as well as those based on the dynamical low-rank approximation, [26], can be explored in order to improve the computational effort. Finally, the introduction of different boundary conditions, always subject to modelling requirements, may have the challenge of considering a different family of polynomials for the spectral approach.

Acknowledgements

The authors would like to thank Prof. V. A. Dougalis and Prof. W. Lambert for discussions to improve the paper. E. Abreu was supported by FAPESP 2019/20991-8, CNPq 306385/2019-8 and Petrobras 2015/00398-0. A. Durán was supported by the Spanish Ministerio de Ciencia e Innovación under Research Grant PID2020-113554GB-100, and by the Junta de Castilla y León and FEDER funds (EU) under Research Grant VA193P20.

References

[1] E. Abreu, A. Bustos, P. Ferraz, W. Lambert, A relaxation projection analytical numerical approach in hysteretic two-phase flows in porous media, J. Sci. Comput. 79 (3) (2019) 1936–1980.
[2] E. Abreu, M. Colombeau, E.Y. Panov, Weak asymptotic methods for scalar equations and systems, J. Math. Anal. Appl. 444 (2016) 1203–1232.
[3] E. Abreu, M. Colombeau, E.Y. Panov, Approximation of entropy solutions to degenerate nonlinear parabolic equations, Z. Angew. Math. Phys. 68 (2017) 133.
[4] E. Abreu, C. Diaz, J. Galvis, A convergence analysis of generalized multiscale finite element methods, J. Comput. Phys. 396 (2019) 303–324.
[5] E. Abreu, P. Ferraz, J. Vieira, Numerical resolution of a pseudo-parabolic Buckley-Leverett model with gravity and dynamic capillary pressure in heterogeneous porous media, J. Comput. Phys. 411 (2020) 109395.
[6] E. Abreu, J. Vieira, Computing numerical solutions of pseudo-parabolic Buckley-Leverett equation with dynamic capillary pressure, Math. Comput. Simul. 137 (2017) 29–48.
[7] R.A. Adams, J.F. Fournier, Sobolev Spaces, 2nd ed., Academic Press, New York, 2003.
[8] G.M. Amiraliyev, Difference method for the solution of one problem of the theory of dispersive waves, USSR Differ. Equ. 26 (1990) 2146–2154.
[9] G.M. Amiraliyev, Y.D. Mamedov, Difference schemes on the uniform mesh for a singularly perturbed pseudo-parabolic equations, Turk. J. Math. 19 (1995) 207–222.
[10] G.M. Amiraliyev, I.G. Amiraliyeva, Difference schemes for the singularly perturbed Sobolev equations, in: Proceeding of the International Conference on Difference Equations, Special Functions and Applications, Elsevier, Munich, 2005.
[11] G.M. Amiraliyev, E. Cimen, I. Amirali, M. Cakir, High-order finite difference technique for delay pseudo-parabolic equations, J. Comput. Appl. Math. 321 (2017) 1–7.
[12] D.N. Arnold, J. Douglas, V. Thomée, Superconvergence of a finite element approximation to the solution of a Sobolev equation in a single space variable, Math. Comput. 36 (1981) 53–63.
[13] G. Barenblat, I. Zheltov, I. Kochiva, Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks, J. Appl. Math. Mech. 24 (1960) 1286–1303.
[14] G. Barenblatt, I. Zheltov, I. Kochina, Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks [strata], J. Appl. Math. Mech. 24 (1960) 1286–1303.
[15] J. Bear, Dynamics of Fluids in Porous Media, American Elsevier, New York-London-Amsterdam, 1972.

[16] J. Bear, Modeling Phenomena of Flow and Transport in Porous Media, Springer, Berlin, 2018.
[17] C. Bernardi, Y. Maday, Properties of some weighted Sobolev spaces and application to spectral approximations, SIAM J. Numer. Anal. 26 (1989) 769–829.
[18] C. Bernardi, Y. Maday, Spectral methods, in: P.J. Ciarlet, J.L. Lions (Eds.), Handbook of Numerical Analysis, Vol. 5: Techniques of Scientific Computing, North Holland, Amsterdam, 1997, pp. 209–486.
[19] M. Böhm, R.E. Showalter, A nonlinear pseudo-parabolic diffusion equation, SIAM J. Math. Anal. 16 (1985) 980–999.
[20] G. Bretti, R. Natalini, B. Piccoli, Numerical algorithms for simulation of a traffic model on road networks, J. Comput. Appl. Math. 210 (2007) 71–77.
[21] C. Canuto, A. Quarteroni, Spectral and pseudo-spectral methods for parabolic problems with nonperiodic boundary conditions, Calcolo 18 (1981) 197–218.
[22] C. Canuto, A. Quarteroni, Approximation results for orthogonal polynomials in Sobolev spaces, Math. Comput. 38 (1982) 67–86.
[23] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang, Spectral Methods in Fluid Dynamics, Springer-Verlag, New York-Heidelberg-Berlin, 1988.
[24] X. Cao, S.F. Nemadjieu, I.S. Pop, Convergence of an MPFA finite volume scheme for a two-phase porous media flow model with dynamic capillarity, IMA J. Numer. Anal. 39 (1) (2019) 512–544.
[25] X. Cao, I.S. Pop, Uniqueness of weak solutions for a pseudo-parabolic equation modeling two phase flow in porous media, Appl. Math. Lett. 46 (2015) 25–30.
[26] G. Ceruti, Ch. Lubich, An unconventional robust integrator for dynamical low-rank approximation, BIT Numer. Math. (2021), https://doi.org/10.1007/s10543-021-00873-0.
[27] P.J. Chen, M.E. Gurtin, On a theory of heat conduction involving two temperatures, Z. Angew. Math. Phys. 19 (1968) 614–627.
[28] E.T. Chung, Y. Efendiev, W.T. Leung, An adaptive generalized multiscale discontinuous Galerkin method for high-contrast flow problems, SIAM J. Multiscale Model. Simul. 16 (3) (2018) 1227–1257.
[29] A.B. Chiyaneh, H. Duru, On adaptive mesh for the initial boundary value singularly perturbed delay Sobolev problems, Numer. Methods Partial Differ. Equ. 36 (2) (2020) 228–248.
[30] D. Colton, Pseudoparabolic equations in one space variable, J. Differ. Equ. 12 (1972) 559–565.
[31] C. Cuesta, J. Hulshof, A model problem for groundwater flow with dynamic capillary pressure: stability of travelling waves, Nonlinear Anal. 52 (2003) 1199–1218.
[32] D.B. Das, S.M. Hassanizadeh (Eds.), Upscaling Multiphase Flow in Porous Media, Springer, Berlin, 2005.
[33] C. van Duijn, L.A. Peletier, I.S. Pop, A new class of entropy solutions of the Buckley-Leverett equation, SIAM J. Math. Anal. 39 (2007) 507–536.
[34] C. van Duijn, Y. Fan, L.A. Peletier, I.S. Pop, Travelling wave solutions for degenerate pseudo-parabolic equations modelling two-phase flow in porous media, Nonlinear Anal., Real World Appl. 14 (2013) 1361–1383.
[35] Y. Efendiev, J. Galvis, T. Hou, Generalized multiscale finite element methods, J. Comput. Phys. 251 (2013) 116–135.
[36] J. Ernest, Schemes with well-controlled dissipation (wcd) for scalar conservations laws with pseudo-parabolic regularization, in: F. Ancona, A. Bressan, P. Marcati, A. Marson (Eds.), Proceeding of 14th International Conference: Theory, Numerics and Applications of Hyperbolic Problems 2012, in: AIMS Appl. Math., vol. 8, 2014, pp. 233–240.
[37] L.C. Evans, Partial Differential Equations, 2nd ed., American Mathematical Society, New York, 2010.
[38] R.E. Ewing, Numerical solution of Sobolev partial differential equations, SIAM J. Numer. Anal. 12 (1975) 345–363.
[39] R.E. Ewing, Time-stepping Galerkin methods for nonlinear Sobolev partial differential equations, SIAM J. Numer. Anal. 15 (1978) 1125–1150.
[40] Y. Fan, I.S. Pop, A class of pseudo-parabolic equations: existence, uniqueness of weak solutions and error estimates for the Euler-implicit discretization, Math. Methods Appl. Sci. 34 (2011) 2329–2339.
[41] Y. Fan, I.S. Pop, Equivalent formulations and numerical schemes for a class of pseudo-parabolic equations, J. Comput. Appl. Math. 246 (2013) 86–93.
[42] L. Ferracina, M.N. Spijker, Strong stability of singly-diagonally-implicit Runge-Kutta methods, Appl. Numer. Math. 58 (2008) 1675–1686.
[43] W.H. Ford, Galerkin approximations to non-linear pseudo-parabolic partial differential equations, Aequ. Math. 14 (1976) 271–291.
[44] W.H. Ford, T.W. Ting, Stability and convergence of difference approximations to pseudo-parabolic partial differential equations, Math. Comput. 27 (1973) 737–743.
[45] W.H. Ford, T.W. Ting, Uniform error estimates for difference approximations to nonlinear pseudo-parabolic partial differential equations, SIAM J. Numer. Anal. 11 (1974) 155–169.
[46] H. Gajewski, K. Zacharias, Zur starken Konvergenz des Galerkinverfahrens bei einer Klasse pseudoparabolischer partieller Differentialgleichungen, Math. Nachr. 47 (1970) 365–376.
[47] Y. Gao, B. Guo, W. Gao, Weak solutions for a high-order pseudo-parabolic equation with variable exponents, Appl. Anal. 93 (2) (2014) 322–338.
[48] D. Gottlieb, L. Lustman, The spectrum of the Chebyshev collocation operator for the heat equation, SIAM J. Numer. Anal. 20 (1983) 909–921.
[49] D. Gottlieb, S.A. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications, SIAM, Philadelphia, 1977.
[50] S. Gotlieb, On high order strong stability preserving Runge-Kutta and multi step time discretizations, J. Sci. Comput. 25 (2005) 105–128.
[51] S. Gotlieb, D.I. Ketcheson, C.-W. Shu, High order strong stability preserving time discretizations, J. Sci. Comput. 38 (2009) 251–289.
[52] S. Gottlieb, C.-W. Shu, E. Tadmor, Strong stability preserving high-order time discretization methods, SIAM Rev. 43 (2001) 89–112.
[53] S. Haq, A. Ghafoor, M. Hussain, S. Arifeen, Numerical solutions of the two dimensional Sobolev and generalized Benjamin-Bona-Mahony-Burgers equations via Haar wavelets, Comput. Math. Appl. 77 (2019) 565–575.
[54] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations, Springer-Verlag, New York-Heidelberg-Berlin, 2004.
[55] E. Hairer, S.P. Norsett, G. Wanner, Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, 2nd edition, Springer Series in Computational Mathematics, vol. 8, Springer, Berlin, 1996.
[56] S.B. Harza, Mathematical model of multiphase flow through porous media, in: Large-Scale PDE-Constrained Optimization in Applications, in: NACM, vol. 49, Springer-Verlag, Berlin Heidelberg, 2010, pp. 29–41.
[57] S. Hassanizadeh, W. Gray, Thermodynamic basis of capillary pressure in porous media, Water Resour. Res. 29 (1993) 3389–3405.
[58] W. Heinrichs, Improved condition number for spectral methods, Math. Comput. 53 (1989) 103–119.
[59] W. Heinrichs, Stabilization techniques for spectral methods, J. Sci. Comput. 6 (1991) 1–19.
[60] R. Helmig, Multiphase Flow and Transport Processes in the Subsurface: A Contribution on the Modelling of Hydrosystems, Springer, Berlin, 1997.
[61] R. Hilfer, F. Doster, P. Zegeling, Nonmonotone saturation profiles for hydrostatic equilibrium in homogeneous porous media, Vadose Zone J. 11 (3) (2012), vzj2012.0021.
[62] R. Hilfer, R. Steinle, Saturation overshoot and hysteresis for twophase flow in porous media, Eur. Phys. J. Spec. Top. 223 (11) (2014) 2323–2338.
[63] T. Holstein, Imprisonment of resonance radiation in gases, Phys. Rev. 72 (1947) 1212.
[64] T. Holstein, Imprisonment of resonance radiation in gases. II, Phys. Rev. 83 (1951) 1159.
[65] M. Hussain, S. Haq, A. Ghafoor, Meshless RBFs method for numerical solutions of two-dimensional high order fractional Sobolev equations, Comput. Math. Appl. 79 (3) (2020) 802–816.
[66] O. Iliev, G. Printsypar, S. Rief, A two-dimensional model of the pressing section of a paper machine including dynamic capillary effects, J. Eng. Math. 83 (1) (2013) 81–107.
[67] G. Izzo, J. Jackiewicz, Construction of SDIRK methods with dispersive stability functions, Appl. Numer. Math. 160 (2021) 265–280.
[68] A-A. Juber, C. Rohde, Existence of weak solutions for a nonlocal pseudo-parabolic model for Brinkman two-phase flow in asymptotically flat porous media, J. Math. Anal. Appl. 477 (1) (2019) 592–612.
[69] S. Karpinski, I.S. Pop, Analysis of an interior penalty discontinuous Galerkin scheme for two phase flow in porous media with dynamic capillary effects, Numer. Math. 136 (1) (2017) 249–286.
[70] S. Karpinski, I.S. Pop, F.A. Radu, Analysis of a linearization scheme for an interior penalty discontinuous Galerkin method for two-phase flow in porous media with dynamic capillarity effects, Int. J. Numer. Methods Eng. 112 (6) (2017) 553–577.
[71] D.I. Ketcheson, C.B. Macdonald, S. Gottlieb, Optimal implicit strong stability preserving Runge-Kutta methods, Appl. Numer. Math. 59 (2009) 373–392.
[72] M.O. Korpusov, A.G. Sveshnikov, Three-dimensional nonlinear evolution equations of pseudo-parabolic type in problems of mathematical physics, Zh. Vychisl. Mat. Mat. Fiz. 43 (2003) 1835–1869.
[73] T. Koto, Phase-lag analysis of diagonally implicit Runge-Kutta methods, J. Inf. Process. 13 (1990) 361–366.
[74] D. Kroener, S. Luckhaus, Flow of oil and water in a porous medium, J. Differ. Equ. 55 (1984) 276–288.
[75] R. Kumar, S. Baska, B-spline quasi-interpolation based numerical methods for some Sobolev type equations, J. Comput. Appl. Math. 292 (2016) 41–66.
[76] C. Lu, Q. Gao, C. Fu, H. Yang, Finite element method of BBM-Burgers equation with dissipative term based on adaptive moving mesh, Discrete Dyn. Nat. Soc. (2017) 3427376.
[77] Y. Maday, A. Quarteroni, Legendre and Chebyshev spectral approximations of Burgers' equation, Numer. Math. 37 (1981) 321–332.
[78] B. Mercier, An Introduction to the Numerical Analysis of Spectral Methods, Springer-Verlag, Berlin, 1989.

[79] A. Mikelić, A global existence result for the equations describing unsaturated flow in porous media with dynamic capillary pressure, J. Differ. Equ. 248 (2010) 561–1577.
[80] A. Mikelić, H. Bruining, Analysis of model equations for stress-enhanced diffusion in coal layers. Part I: existence of a weak solution, SIAM J. Math. Anal. 40 (2008) 1671–1691.
[81] E. Milne, The diffusion of imprisoned radiation through a gas, J. Lond. Math. Soc. 1 (1926) 40–51.
[82] Y. Ning, M. Sun, G. Piao, A finite element solution for the conservation form of BBM-Burgers' equation, East Asian Math. J. 33 (2017) 495–509.
[83] A. Novick-Cohen, R. Pego, Stable patterns in a viscous diffusion equation, Trans. Am. Math. Soc. 324 (1991) 331–351.
[84] S. Olver, A. Townsend, A fast and well-conditioned spectral method, SIAM Rev. 55 (2013) 462–489.
[85] S.A. Orszag, Spectral methods for problems in complex geometries, J. Comput. Phys. 37 (1980) 70–92.
[86] M. Peszyńska, R. Showalter, S. Yi, Homogenization of a pseudo-parabolic system, Appl. Anal. 88 (2009) 265–1282.
[87] M. Peszynska, S. Yi, Numerical methods for unsaturated flow with dynamic capillary pressure in heterogeneous porous media, Int. J. Numer. Anal. Model. 5 (2008) 126–149.
[88] G.F. Pinder, W.G. Gray, Essentials of Multiphase Flow and Transport in Porous Media, J. Wiley & Sons, New Jersey, 2008.
[89] G. Plonka, D. Potts, G. Steidl, M. Tasche, Numerical Fourier Analysis, Springer, Berlin, 2018.
[90] M. Ptashnyk, Nonlinear pseudo-parabolic equations as singular limit of reaction-diffusion equations, Appl. Anal. 85 (2006) 1285–1299.
[91] A. Quarteroni, Fourier spectral methods for pseudo-parabolic equations, SIAM J. Numer. Anal. 24 (1987) 323–335.
[92] A. Rätz, B. Schweizer, Hysteresis models and gravity fingering in porous media, J. Appl. Math. Mech. (Z. Angew. Math. Mech.) 94 (7–8) (2014) 645–654.
[93] L. Rubinstein, On the problem of the process of propagation of heat in heterogeneous media, Izv. Akad. Nauk SSSR, Ser. Geogr. 1 (1948) 12–45.
[94] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, Philadelphia, 2003.
[95] N. Seam, G. Vallet, Existence results for nonlinear pseudo-parabolic problems, Nonlinear Anal., Real World Appl. 12 (2011) 2625–2639.
[96] D. Seus, K. Mitra, I.S. Pop, F.A. Radud, C. Rohde, A linear domain decomposition method for partially saturated flow in porous media, Comput. Methods Appl. Mech. Eng. 333 (2018) 331–355.
[97] J. Shen, Efficient spectral-Galerkin method I. Direct solvers for the second and fourth order equations using Legendre polynomials, SIAM J. Sci. Comput. 15 (1994) 1489–1505.
[98] J. Shen, T. Tang, L. Wang, Spectral Methods, Algorithms, Analysis and Applications, Springer, Berlin, 2011.
[99] R.E. Showalter, T.W. Ting, Pseudoparabolic partial differential equations, SIAM J. Math. Anal. 1 (1970) 1–26.
[100] R.E. Showalter, Sobolev equations for nonlinear dispersive systems, Appl. Anal. 7 (1978) 297–308.
[101] A. Sidi, Acceleration of convergence of (generalized) Fourier series by the d-transformation, Ann. Numer. Math. 2 (1995) 381–406.
[102] A. Sidi, Practical Extrapolation Methods: Theory and Applications, Cambridge Monographs on Applied and Computational Mathematics, vol. 10, Cambridge University Press, 2003.
[103] S. Sobolev, Some new problems in mathematical physics, Izv. Akad. Nauk SSSR, Ser. Mat. 18 (1954) 3–50.
[104] F. Stauffer, Time dependence of the relations between capillary pressure, water content and conductivity during drainage of porous media, in: IAHR Symposium on Scale Effects in Porous Media, Thessaloniki, Greece, vol. 29, 1978, pp. 3–35.
[105] T. Sun, D. Yang, The finite difference streamline diffusion methods for Sobolev equations with convection-dominated term, Appl. Math. Comput. 125 (2002) 325–345.
[106] T.W. Ting, Certain non-steady flows of second-order fluids, Arch. Ration. Mech. Anal. 14 (1963) 1–26.
[107] Y.W. Ting, Parabolic and pseudo-parabolic partial differential equations, J. Math. Soc. Jpn. 21 (1969) 440–453.
[108] P.N. Vabishchebich, A.V. Grigoriev, Splitting schemes for pseudo-parabolic equations, Differ. Equ. 49 (2013) 807–814.
[109] H. Vandeven, On the eigenvalues of second-order spectral differentiation operators, in: C. Canuto, A. Quarteroni (Eds.), Spectral and High Order Methods for PDE's, North-Holland, Amsterdam, 1990, pp. 313–318.
[110] J.A.C. Weideman, L.N. Trefethen, The eigenvalues of second-order spectral differentiation matrices, SIAM J. Numer. Anal. 25 (1988) 1279–1298.
[111] M. Yang, Analysis of second order finite volume element methods for pseudo-parabolic equations in three spatial dimensions, Appl. Math. Comput. 196 (2008) 94–104.
[112] E. Zampieri, On the condition number of some spectral collocation operators and their finite element preconditioning, J. Sci. Comput. 9 (1994) 419–443.
[113] H. Zhang, P.A. Zegeling, A numerical study of two-phase flow models with dynamic capillary pressure and hysteresis, Transp. Porous Media 116 (2017) 825–846.
[114] T. Zhao, Y. Wu, H. Ma, Error analysis of Chebyshev-Legendre pseudo-spectral method for a class of nonclassical parabolic equations, J. Sci. Comput. 52 (2012) 588–602.