

An IMEX-RK scheme for capturing similarity solutions in the multidimensional Burgers's equation

JENS ROTTMANN-MATTHES

*Department of Mathematics, Institute for Analysis, Karlsruhe Institute of Technology, Engler Str. 2,
76131 Karlsruhe, Germany
jens.rottmann-matthes@kit.edu*

[Received on 14 December 2016; revised on 31 May 2017]

In this article, we introduce a new, simple and efficient numerical scheme for the implementation of the freezing method for capturing similarity solutions in partial differential equations (PDEs). The scheme is based on an implicit-explicit (IMEX) Runge-Kutta approach for a method of lines (semi)discretization of the freezing partial differential algebraic equation (PDAE). We prove second-order convergence for the time discretization at smooth solutions in the ordinary differential equation sense, and we present numerical experiments that show second-order convergence for the full discretization of the PDAE. The multidimensional Burgers's equations serves as an example. By considering very different values of viscosity, Burgers's equation can be considered a prototypical example of general coupled hyperbolic–parabolic PDEs. Numerical experiments show that our method works perfectly well for all values of viscosity, suggesting that the scheme is indeed suitable for capturing similarity solutions in general hyperbolic–parabolic PDEs by direct forward simulation with the freezing method.

Keywords: IMEX-Runge–Kutta; central scheme; hyperbolic–parabolic partial differential algebraic equations; similarity solutions; Burgers's equation; freezing method; scaling symmetry.

1. Introduction

Many time-dependent partial differential equations (PDEs) from applications exhibit simple patterns. When these patterns are stable, solutions with sufficiently close initial data develop these patterns as time increases. They often have important implications on the actual interpretation of the system's behavior. A very important and well-known example of a simple pattern are the traveling wave solutions that appear in the nerve–axon equations of [Hodgkin & Huxley \(1952\)](#) (see [Hastings, 1976](#) for a discussion of their existence). Here they model the transport of information along the axon of a nerve cell.

Traveling waves are one of the simplest examples of patterns called relative equilibria. Relative equilibria are solutions to evolution equations of the form $u(x, t) = [a(g(t))\underline{u}](x)$, where $g : \mathbb{R} \rightarrow G$ is some curve in a symmetry group G acting by some action a on a fixed profile \underline{u} , so that the time evolution of the solution u is completely described by the curve g .

In a more mathematical way, we consider evolution equations

$$u_t = F(u), \quad x \in \mathbb{R}^d, \quad t \geq 0, \quad u(x, t) \in \mathbb{R}^m, \quad (1.1)$$

where F is some vector field given by a partial differential operator. In $d = 1$ spatial dimension, a traveling wave is a solution $u(x, t)$ of (1.1) of the form $u(x, t) = \underline{u}(x - \underline{c}t)$, with profile \underline{u} and velocity \underline{c} . Here, the symmetry group is just the real numbers \mathbb{R} , which act on functions by left translation and g is a

linear curve in \mathbb{R} . From an application point of view, the velocity \underline{c} of the traveling wave in the Hodgkin–Huxley system is of great relevance as it quantifies how fast information is passed on in this system. This motivates that one is often interested in the relative equilibria of the underlying PDE problem and also in their specific constants of motion, like their velocity. When (1.1) is considered in a comoving frame, i.e., with the new spatial coordinate $\xi = x - \underline{c}t$, the profile \underline{u} becomes a steady state of the comoving equation

$$v_t = F(v) + \underline{c}v_\xi, \quad \xi \in \mathbb{R}, \quad t \geq 0.$$

Similarly, there are evolution equations that exhibit rotating waves or spiral waves. In $d = 2$ spatial dimensions, these are solutions $u : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^m$ of (1.1) of the form $u(x, t) = \underline{u} \left(\begin{pmatrix} \cos(\underline{c}t) & -\sin(\underline{c}t) \\ \sin(\underline{c}t) & \cos(\underline{c}t) \end{pmatrix} x \right)$. Considering (1.1) in a corotating frame, i.e., with the new spatial coordinates $\xi = \begin{pmatrix} \cos(\underline{c}t) & -\sin(\underline{c}t) \\ \sin(\underline{c}t) & \cos(\underline{c}t) \end{pmatrix} x$, the profile \underline{u} of the rotating wave is a steady state of the corotating equation

$$v_t = F(v) + \underline{c}\xi_2 v_{\xi_1} - \underline{c}\xi_1 v_{\xi_2}, \quad \xi \in \mathbb{R}^2, \quad t \geq 0.$$

Again, not only the profile \underline{u} is of interest but also its rotational velocity \underline{c} .

Typically, the velocities \underline{c} of a traveling or rotating wave (or similar numbers for other relative equilibria) are not known in advance, and thus the optimal comoving coordinate frame, in which the solution becomes a steady state, cannot be used. A method that calculates the solution to the Cauchy problem for (1.1) and in parallel a suitable reference frame is the freezing method, independently introduced in Beyn & Thümmeler (2004) and Rowley *et al.* (2003) (see also Beyn *et al.*, 2014). A huge advantage of this method is that asymptotic stability with asymptotic phase of a traveling (or rotating) wave for the original system becomes asymptotic stability in the sense of Lyapunov of the wave profile and its velocity for the freezing system (see Thümmeler, 2008; Rottmann-Matthes, 2012a,b; Beyn *et al.*, 2014). As such, the method allows us to approximate the profile and its constants of motion by a direct forward simulation. The method works not only for traveling, rotating or meandering waves but also for other relative equilibria and similarity solutions with a more complicated symmetry group (e.g., Beyn *et al.*, 2014; Rottmann-Matthes, 2016).

The idea of the method is to write the equation in new (time-dependent) coordinates and to split the evolution of the solution into an evolution of the profile and an evolution in a symmetry group that brings the profile into the correct position via the group action. For example, in the case of an $E(2)$ equivariance of the evolution equation ($E(2)$ is the Euclidean group of the plane), when we allow for rotation and translation in \mathbb{R}^2 , this leads to

$$v_t = F(v) + \mu_1 \partial_{\xi_1} v + \mu_2 \partial_{\xi_2} v + \mu_3 (\xi_2 \partial_{\xi_1} v - \xi_1 \partial_{\xi_2} v). \quad (1.2a)$$

In (1.2a) we have the new time-dependent unknowns v and μ_1, μ_2, μ_3 , where $\mu_j \in \mathbb{R}$. To cope with the additional degrees of freedom due to the μ_j , (1.2a) is supplemented by algebraic equations, the so-called phase conditions, which abstractly read

$$0 = \psi(v, \mu). \quad (1.2b)$$

The complete system (1.2) is called the freezing system and the Cauchy problem for it can be implemented on a computer. Note that (1.2) in fact is a partial differential algebraic equation (PDAE).

In many cases, the Cauchy problem for (1.2) can be solved using standard software packages like COMSOL Multiphysics (see, e.g., Beyn *et al.*, 2014; Beyn *et al.*, 2016). Nevertheless, in other cases,

these standard toolboxes may not work at all or may not give reliable results. For example, this is the case for partly parabolic reaction–diffusion equations, i.e., reaction–diffusion equations in which not all components diffuse, as is the case in the important Hodgkin–Huxley equations. In this case, the freezing method leads to a parabolic equation that is nonlinearly coupled to a hyperbolic equation with a time-varying principal part. Other examples that cannot be easily solved using standard packages include hyperbolic conservation laws or coupled hyperbolic–parabolic PDEs, which appear in many important applied problems.

In this article, we present a new, simple and robust numerical discretization of the freezing PDAE, which allows us to do long-term simulations of time-dependent PDEs and to capture similarity solutions, also for viscous and even inviscid conservation laws, by the freezing method.

We derive our fully discrete scheme in two stages. First, we do a spatial discretization of the freezing PDAE and obtain a method of lines (MOL) system. For the spatial discretization, we employ a central scheme for hyperbolic conservation laws from Kurganov & Tadmor (2000), namely, we adapt the semidiscrete scheme derived in Kurganov & Tadmor (2000) to the case when the flux may also depend on the spatial variable. This yields a second-order semidiscrete central scheme. It has the important property that it does not require information on the local wave structure besides an upper bound on the local wave speed. In particular, no solutions to Riemann problems are needed.

The resulting MOL system is a huge ordinary differential algebraic equation (ODAE) system, which has parts with very different properties. On the one hand, parabolic parts lead to fine spatial discretizations to very stiff parts in the equation, for which one should employ implicit time-marching schemes. On the other hand, a hyperbolic term leads to a medium stiffness but becomes highly nonlinear due to the spatial discretization, so an explicit time-marching scheme is preferable. To couple these conflicting requirements, we use an implicit–explicit (IMEX) Runge–Kutta (RK) scheme for the time discretization. Such schemes were considered in Ascher *et al.* (1997), but here we will apply them to DAE problems.

As an example, we consider Burgers’s equation,

$$u_t + \left(\frac{1}{2}u^2\right)_x = \nu u_{xx}, \quad x \in \mathbb{R}, \quad t \geq 0, \quad u(x, t) \in \mathbb{R}, \quad (1.3)$$

with $\nu > 0$, which was originally introduced by J.M. Burgers (see, e.g., Burgers, 1948) as a mathematical model of turbulence. We also consider the multidimensional generalizations of (1.3),

$$u_t + \frac{1}{p} \operatorname{div}(a|u|^p) = \nu \Delta u, \quad x \in \mathbb{R}^d, \quad t \geq 0, \quad u(x, t) \in \mathbb{R}, \quad (1.4)$$

where $a \in \mathbb{R}^d \setminus \{0\}$ and $p > 1$ are fixed. Equation (1.3) and its generalization to d dimensions are among the simplest truly nonlinear PDEs, and, moreover, in the inviscid ($\nu = 0$) case they develop shock solutions. As such, they are often used as test problems for shock-capturing schemes (e.g., Hu & Shu, 1999). But they are also of interest from a physical point of view as they are special cases of the ‘multidimensional Burgers’s equation’

$$\partial_t \vec{u} + (\vec{u} \cdot \nabla) \vec{u} = \nu \Delta \vec{u}, \quad x \in \mathbb{R}^d, \quad t \geq 0, \quad \vec{u}(x, t) \in \mathbb{R}^d, \quad (1.5)$$

which has applications in different areas of physics (see, e.g., Bec & Khanin, 2007 for a review).

There are three main reasons for us to choose Burgers’s equation as an example. First of all, it is a very simple (scalar) nonlinear equation. Second, despite its simplicity, it provides an example for which the hyperbolic part dominates by choosing $0 < \nu \ll 1$, and it also provides as an example for which the parabolic part dominates by choosing $\nu \gg 0$. Third, as we will see subsequently, the new terms,

introduced by the freezing method for freezing similarity solutions, have properties very similar to the terms appearing in the method of freezing for rotating waves ((1.2a) and (1.2b)).

The plan of the article is as follows. In Section 2, we derive the continuous freezing method for Burgers's equation and present the analytic background. In Section 3, we explain the spatial discretization of the freezing PDAE and obtain the MOL ODAE approximation. In the subsequent step in Section 4, we then do the time discretization of the DAE with our IMEX-RK scheme and show that it is a second-order method for the DAE (with respect to the differential variables). In the final Section 5, we present several numerical results that show that our method and its numerical discretization are suitable for freezing patterns in equations for which the parabolic part dominates and also for equations for which the hyperbolic part dominates. Because of this, we expect our method to be well suited also for capturing traveling or rotating waves in hyperbolic–parabolic coupled problems.

2. The continuous freezing system

In this section, we briefly review the results from Rottmann-Matthes (2016) and explain the freezing system for Burgers's equation (1.4). For the benefit of the reader, though, we present a simplified and direct derivation without using the abstract language and theory of Lie groups and Lie algebras. We refer to Rottmann-Matthes (2016) for the abstract approach.

2.1 The comoved system

To formally derive the freezing system, we assume that the solution u of the Cauchy problem for (1.4),

$$u_t = v \Delta u - \frac{1}{p} \operatorname{div}_x (a|u|^p) =: F(u), \quad x \in \mathbb{R}^d, t \geq 0, \quad u(x, t) \in \mathbb{R}, \quad u(0) = u_0, \quad (2.1)$$

with some fixed $a = (a_1, \dots, a_d)^\top \in \mathbb{R}^d \setminus \{0\}$, and $p > 1$ is of the form

$$u(x, t) = \frac{1}{\alpha(\tau(t))} v\left(\frac{x - b(\tau(t))}{\alpha(\tau(t))^{p-1}}, \tau(t)\right), \quad (2.2)$$

where $v : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$, $b = (b_1, \dots, b_d)^\top : \mathbb{R} \rightarrow \mathbb{R}^d$, $\alpha : \mathbb{R} \rightarrow (0, \infty)$ and $\tau : [0, \infty) \rightarrow [0, \infty)$ are smooth functions and $\dot{\tau}(t) > 0$ for all $t \in [0, \infty)$.

REMARK 2.1

- (i) One can interpret the function τ as a transformation of the time t to a new time τ ; the action of the scalar function α on the function v can be understood as a scaling of the function and the space. Finally, we interpret the meaning of b as a spatial shift.
- (ii) Note that in Rottmann-Matthes (2016) we also allow for a spatial rotation. We actually do not use the rotational symmetry here, because it does not appear in one and two spatial dimensions. Moreover, the symmetries we consider here are also present in the multidimensional Burgers's system (1.5), whereas the rotational symmetry is not.

A simple calculation shows

$$F(u)(x, t) = \frac{1}{\alpha^{2p-1}(\tau)} F(v)(\xi, \tau), \quad (2.3)$$

where $\xi = \frac{x-b}{\alpha^{p-1}}$ denotes new spatial coordinates. Moreover, from (2.2), we obtain with the chain rule

$$\begin{aligned} \frac{\partial}{\partial t} u(x, t) &= \frac{d}{dt} \left(\frac{1}{\alpha(\tau(t))} v \left(\frac{x-b(\tau(t))}{\alpha(\tau(t))^{p-1}}, \tau(t) \right) \right) \\ &= -\frac{\alpha'(\tau)}{\alpha^2(\tau)} \dot{\tau} v(\xi, \tau) - (p-1) \frac{\alpha'(\tau)}{\alpha^2} \xi^\top \nabla_\xi v(\xi, \tau) \dot{\tau} - \frac{b'(\tau)^\top}{\alpha(\tau)^p} \nabla_\xi v(\xi, \tau) \dot{\tau} + \frac{1}{\alpha(\tau)} v_\tau(\xi, \tau) \dot{\tau}. \end{aligned} \quad (2.4)$$

By setting

$$\mu_1(\tau) = \frac{\alpha'(\tau)}{\alpha(\tau)} \in \mathbb{R} \quad \text{and} \quad \mu_{i+1}(\tau) = \frac{b'_i(\tau)}{\alpha(\tau)^{p-1}} \in \mathbb{R}, \quad i = 1, \dots, d,$$

$$\phi_1(\xi, v) = -(p-1) \operatorname{div}_\xi(\xi v) - (1-d(p-1))v \quad \text{and} \quad \phi_2(\xi, v) = -\nabla_\xi v = -\left(\frac{\partial}{\partial \xi_j} v \right)_{j=1}^d, \quad (2.5)$$

we can write (2.4) as

$$\frac{\partial}{\partial t} u(x, t) = \frac{1}{\alpha(\tau)} (\mu_1(\tau) \phi_1(\xi, v) + \mu_{(2:d+1)}(\tau) \phi_2(\xi, v) + v_\tau) \dot{\tau}, \quad (2.6)$$

where $\mu_{(2:d+1)} = (\mu_2, \dots, \mu_{d+1})$. Because u solves (2.1), we obtain from (2.3) and (2.6) under the assumption that τ satisfies $\dot{\tau} = \alpha(\tau)^{2-2p}$ in the new ξ, τ, v coordinates for v the equation,

$$v_\tau = F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v). \quad (2.7)$$

The key observation for the freezing method is (Rottmann-Matthes, 2016, Theorem 4.2), which relates the solution of the original Cauchy problem (2.1) to the solution of the Cauchy problem for (2.7) in the new (ξ, τ) coordinates. Using the spaces $X = L^2(\mathbb{R}^d)$ and $Y_1 = \{v \in H^2(\mathbb{R}^d) : \xi^\top \nabla_\xi v \in L^2(\mathbb{R}^d)\}$, the result from Rottmann-Matthes (2016) can be rephrased as follows.

THEOREM 2.2 A function $u \in \mathcal{C}([0, T]; Y_1) \cap \mathcal{C}^1([0, T]; X)$ solves the Cauchy problem (2.1) if and only if $v \in \mathcal{C}([0, \hat{T}]; Y_1) \cap \mathcal{C}^1([0, \hat{T}]; X)$, $\mu_i \in \mathcal{C}([0, \hat{T}]; \mathbb{R})$, $i = 1, \dots, d+1$, $\alpha \in \mathcal{C}^1([0, \hat{T}]; (0, \infty))$, $b \in \mathcal{C}^1([0, \hat{T}]; \mathbb{R}^d)$ and $\tau \in \mathcal{C}^1([0, T]; [0, \hat{T}])$ solve the system

$$\begin{aligned} v_\tau &= F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v), & v(0) &= u_0, \\ \alpha_\tau &= \mu_1 \alpha, & \alpha(0) &= 1, \\ b_\tau &= \alpha^{p-1} \mu_{(2:d+1)}^\top, & b(0) &= 0, \\ \frac{\partial}{\partial t} \tau &= \alpha(\tau)^{2-2p}, & \tau(0) &= 0, \end{aligned} \quad (2.8)$$

and u and v, α, b, τ are related by (2.2).

2.2 The freezing system

To cope with the $d + 1$ additional degrees of freedom, due to μ_1, \dots, μ_{d+1} , we complement (2.8) with $d + 1$ algebraic equations, the so-called phase conditions (see Beyn & Thümmler, 2004).

Type 1: Orthogonal phase condition. We require that v and μ_1, \dots, μ_{d+1} , which solve (2.7), are chosen such that at each time instance, $\|v_\tau\|_{L^2}^2$ is minimized with respect to μ_1, \dots, μ_{d+1} . Therefore, we have

$$0 = \frac{1}{2} \frac{d}{d\mu} \|F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v)\|_{L^2}^2,$$

which is equivalent to

$$\begin{cases} 0 = \langle \phi_1(\xi, v), F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v) \rangle_{L^2}, \\ 0 = \left\langle \frac{\partial}{\partial \xi_j} v, F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v) \right\rangle_{L^2}, j = 1, \dots, d, \end{cases} \quad (2.9)$$

where ϕ_1 and ϕ_2 are given in (2.5). We abbreviate (2.9) as $0 = \psi^{\text{orth}}(v, \mu)$.

Type 2: Fixed phase condition. The idea of the fixed phase condition is to require that the v -component of the solution always lies in a fixed, $(d + 1)$ -codimensional hyperplane, which is given as the level set of a fixed, linear mapping. Here, we assume that a ‘suitable’ reference function \hat{u} is given and the v -component of the solution always satisfies

$$\begin{cases} 0 = \langle \phi_1(\xi, \hat{u}), v - \hat{u} \rangle_{L^2}, \\ 0 = \left\langle \frac{\partial}{\partial \xi_j} \hat{u}, v - \hat{u} \right\rangle_{L^2}, j = 1, \dots, d, \end{cases} \quad (2.10)$$

where ϕ_1 and ϕ_2 are given in (2.5). We abbreviate (2.10) as $0 = \psi^{\text{fix}}(v)$.

We augment system (2.8) with one of the phase conditions (2.9) or (2.10) and obtain

$$v_\tau = F(v) - \mu_1 \phi_1(\xi, v) - \mu_{(2:d+1)} \phi_2(\xi, v), \quad v(0) = u_0, \quad (2.11a)$$

$$0 = \psi(v, \mu), \quad (2.11b)$$

$$\alpha_\tau = \mu_1 \alpha, \quad b_\tau = \alpha^{p-1} \mu_{(2:d+1)}^\top, \quad \alpha(0) = 1 \in \mathbb{R}, \quad b(0) = 0 \in \mathbb{R}^d, \quad (2.11c)$$

$$\frac{d}{d\tau} t = \alpha(\tau)^{2p-2}, \quad t(0) = 0, \quad (2.11d)$$

where $\psi(v, \mu)$ is either ψ^{orth} from (2.9) or ψ^{fix} from (2.10).

REMARK 2.3

- (i) Observe that (2.11) consists of the PDE (2.11a), coupled to a system of ODEs (2.11c) and (2.11d) and coupled to a system of algebraic equations (2.11b). Moreover, in (2.11a) the hyperbolic part dominates for $0 < v \ll 1$ and the parabolic part dominates for $v \gg 0$. Finally, note that the ODEs (2.11c) and (2.11d) decouple from (2.11a), and (2.11b) can be solved in a postprocessing step.

- (ii) Note that the ϕ_1 term in (2.5) resembles the generator of rotation, cf. (1.2a).
- (iii) Under suitable assumptions on the solution and the reference function, the system (2.11a), (2.11b) is a PDAE of ‘time index’ 1 in the case $\psi = \psi^{\text{orth}}$ and of ‘time index’ 2 in the case $\psi = \psi^{\text{fix}}$, where the index is understood as a differentiation index (see Martinson & Barton, 2000).

3. Implementation of the numerical freezing method I: spatial discretization

In this section, we derive a spatial semidiscretization (MOL system) of the freezing partial differential algebraic evolution equation system (2.11). First, we separately consider the PDE part (2.11a) of the equation on the full domain. Afterwards, we consider the case of a bounded domain with artificial no-flux boundary conditions. Finally, we consider the spatial discretization of the (low-dimensional) remaining equations (2.11b)–(2.11d).

3.1 Spatial semidiscretization of the PDE part

Recall from (2.5) that $\mu_1 \phi_1(\xi, v)$ is of the form

$$\mu_1 \phi_1(\xi, v) = \mu_1 \sum_{j=1}^d \frac{d}{d\xi_j} f_{1,j}(\xi, v) - \mu_1 f_{1,0}(v),$$

where $f_{1,0}(v) = (1 - d(p - 1))v$ and $f_{1,j}(\xi, v) = -(p - 1)\xi_j v$. Note that $\frac{d}{d\xi_j} f(\xi, v) = \frac{\partial}{\partial \xi_j} f + \frac{\partial}{\partial v} f \frac{\partial}{\partial \xi_j} v$. Similarly, the term $\mu_{(2:d+1)} \phi_2(\xi, v)$ can be written as

$$\mu_{(2:d+1)} \phi_2(\xi, v) = - \sum_{i=1}^d \mu_{1+i} \frac{\partial}{\partial \xi_i} v = \sum_{i=1}^d \mu_{1+i} \sum_{j=1}^d \frac{d}{d\xi_j} f_{1+i,j}(\xi, v),$$

where $f_{1+i,j}(\xi, v) = -v$ if $i = j$ and $f_{1+i,j}(\xi, v) = 0$ if $i \neq j$, $i, j = 1, \dots, d$. Furthermore, by also setting $f_{0,j}(v) = \frac{1}{p} a_j |v|^p$ and $Q_j(\frac{\partial}{\partial \xi_j} v) = \frac{\partial}{\partial \xi_j} v$, equation (2.11a) can be recast as

$$v_\tau + \sum_{i=1}^{d+1} \mu_i \left[\sum_{j=1}^d \frac{d}{d\xi_j} f_{i,j}(\xi, v) \right] + \sum_{j=1}^d \frac{d}{d\xi_j} f_{0,j}(v) = \sum_{j=1}^d \frac{d}{d\xi_j} Q_j\left(\frac{\partial}{\partial \xi_j} v\right) + \mu_1 f_{1,0}(v). \quad (3.1)$$

For the spatial semidiscretization, we adapt the second-order semidiscrete central scheme for hyperbolic conservation laws from Kurganov & Tadmor (2000) to our situation. The details of the adaptation to space-dependent hyperbolic conservation laws of the form

$$v_\tau + \sum_{j=1}^d \frac{d}{d\xi_j} f_j(\xi, v) = 0 \quad (3.2)$$

is presented in the appendix. As is noted in Kurganov & Tadmor (2000, Section 4), diffusive flux terms and zero-order terms, which appear in (3.1), can easily be appended to the semidiscretization of (3.2) by simple second-order finite difference approximations.

For the actual discretization, we choose a uniform spatial grid in each coordinate direction,

$$\xi_{k-\frac{1}{2}}^j = \left(k - \frac{1}{2}\right) \Delta \xi_j, \quad k \in \mathbb{Z}, j = 1, \dots, d$$

and obtain the rectangular cells (finite volumes) $C_{\mathbf{k}} := C_{k_1, \dots, k_d} := \times_{j=1}^d (\xi_{k_j-\frac{1}{2}}^j, \xi_{k_j+\frac{1}{2}}^j) \subset \mathbb{R}^d$, with centers $\xi_{\mathbf{k}} := (\xi_{k_1}^1, \dots, \xi_{k_d}^d) \in \mathbb{R}^d$, where $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{K}$ with index set $\mathbb{K} = \mathbb{Z}^d$. In the following, we interpret for each $\mathbf{k} \in \mathbb{K}$, $v_{\mathbf{k}}(\tau)$ as an approximation of the cell average of v in the cell $C_{\mathbf{k}}$ at time τ ,

$$v_{\mathbf{k}}(\tau) \approx \frac{1}{\text{vol}(C_{\mathbf{k}})} \int_{C_{\mathbf{k}}} v(\xi, \tau) \, d\xi, \quad \text{vol}(C_{\mathbf{k}}) = \prod_{j=1}^d \Delta \xi_j.$$

Then the MOL system for (3.1) is given by

$$\begin{aligned} v'_{\mathbf{k}} = & - \sum_{j=1}^d \frac{H_{\mathbf{k}^j+\frac{1}{2}}^{0j} - H_{\mathbf{k}^j-\frac{1}{2}}^{0j}}{\Delta \xi_j} - \sum_{i=1}^{d+1} \mu_i \sum_{j=1}^d \frac{H_{\mathbf{k}^j+\frac{1}{2}}^{ij} - H_{\mathbf{k}^j-\frac{1}{2}}^{ij}}{\Delta \xi_j} + \mu_1 f_{1,0}(v_{\mathbf{k}}) \\ & + \sum_{j=1}^d \frac{P_{\mathbf{k}^j+\frac{1}{2}}^j - P_{\mathbf{k}^j-\frac{1}{2}}^j}{\Delta \xi_j} =: F_{\mathbf{k}}(v_{\mathbb{Z}^d}, \mu), \quad \mathbf{k} \in \mathbb{K}. \end{aligned} \quad (3.3)$$

In (3.3), we use the following notation and abbreviations:

For $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{K}$, we define $\mathbf{k}^j \pm \frac{1}{2} := (k_1, \dots, k_{j-1}, k_j \pm \frac{1}{2}, k_{j+1}, \dots, k_d)$. Then, $H_{\mathbf{k}^j \pm \frac{1}{2}}^{ij}(t)$ is an approximation of the ‘hyperbolic flux’ through the boundary face

$$(\xi_{k_1-\frac{1}{2}}^1, \xi_{k_1+\frac{1}{2}}^1) \times \dots \times \{\xi_{k_j \pm \frac{1}{2}}^j\} \times \dots \times (\xi_{k_d-\frac{1}{2}}^d, \xi_{k_d+\frac{1}{2}}^d) =: \partial C_{\mathbf{k}^j \pm \frac{1}{2}}$$

of $C_{\mathbf{k}}$ due to the flux function $f_{i,j}$, $i = 0, \dots, d+1$, $j = 1, \dots, d$. Moreover, we add the regularizing part of the Kurganov and Tadmor scheme completely to the H^{0j} -terms, namely the term $\frac{\mathbf{a}}{2\Delta \xi} (u_{k-\frac{1}{2}}^- - u_{k-\frac{1}{2}}^+ - u_{k+\frac{1}{2}}^- + u_{k+\frac{1}{2}}^+)$ from (A.10) is added to the discretization of $\frac{d}{d\xi} f_{0,1}(v)$ in the one-dimensional case. In the two-dimensional case, the term $\frac{\mathbf{a}}{2\Delta \xi} (u_{k_1-\frac{1}{2}, k_2}^- - u_{k_1-\frac{1}{2}, k_2}^+ - u_{k_1+\frac{1}{2}, k_2}^- + u_{k_1+\frac{1}{2}, k_2}^+)$ from (A.12) is added to the discretization of $\frac{d}{d\xi_1} f_{0,1}(v)$ and the term $\frac{\mathbf{a}}{2\Delta \eta} (u_{k_1, k_2-\frac{1}{2}}^- - u_{k_1, k_2-\frac{1}{2}}^+ - u_{k_1, k_2+\frac{1}{2}}^- + u_{k_1, k_2+\frac{1}{2}}^+)$ from (A.12) is added to the discretization of $\frac{d}{d\xi_2} f_{0,2}(v)$. The reason for adding this term to the hyperbolic part is that it is a highly nonlinear term, due to the nonlinear reconstruction procedure and, moreover, that the value of \mathbf{a} may change during computation. If it is instead added to the parabolic part of the equation, this term would cause severe difficulties in the efficient solution of our IMEX-RK scheme introduced in Section 4; see also Section 4.3. Hence, the H^{ij} , $i = 0, \dots, d+1$, $j = 1, \dots, d$ take the form

$$H_{\mathbf{k}^j \pm \frac{1}{2}}^{ij}(t) = \frac{f_{ij}(\xi_{\mathbf{k}^j \pm \frac{1}{2}}, v_{\mathbf{k}^j \pm \frac{1}{2}}^+) + f_{ij}(\xi_{\mathbf{k}^j \pm \frac{1}{2}}, v_{\mathbf{k}^j \pm \frac{1}{2}}^-)}{2} - \delta_{i,0} \mathbf{a} \left(v_{\mathbf{k}^j \pm \frac{1}{2}}^+ - v_{\mathbf{k}^j \pm \frac{1}{2}}^- \right). \quad (3.4)$$

In (3.4), the point of evaluation is

$$\xi_{\mathbf{k}j \pm \frac{1}{2}} = (\xi_{k_1}, \dots, \xi_{k_d}) \pm \frac{1}{2} \mathbf{e}_j \Delta \xi_j \in \partial C_{\mathbf{k}}, \quad (3.5)$$

and $v_{\mathbf{k}j \pm \frac{1}{2}}^+$, $v_{\mathbf{k}j \pm \frac{1}{2}}^-$ denote the limits ‘from above’ and ‘from below’ at the boundary face $\partial C_{\mathbf{k}j \pm \frac{1}{2}}$ of the piecewise linear minmod reconstruction \hat{v} of $(v_{\mathbf{k}})_{\mathbf{k} \in \mathbb{Z}^d}$, explained in the appendix, i.e.,

$$v_{\mathbf{k}j \pm \frac{1}{2}}^+ = \lim_{h \searrow \frac{1}{2}} \hat{v}(x_{\mathbf{k}} + h \Delta x_j \mathbf{e}_j), \quad v_{\mathbf{k}j \pm \frac{1}{2}}^- = \lim_{h \nearrow \frac{1}{2}} \hat{v}(x_{\mathbf{k}} + h \Delta x_j \mathbf{e}_j). \quad (3.6)$$

For explicit formulas of $v_{\mathbf{k}j \pm \frac{1}{2}}^\pm$ in the one-dimensional and two-dimensional cases, see (A.11) and (A.13), respectively.

Moreover, the value \mathbf{a} , appearing in $H_{\mathbf{k}j \pm \frac{1}{2}}^{0j}$, denotes an upper bound for the maximal wave speeds of the hyperbolic terms. The choice of the value \mathbf{a} is only formal at the moment, because such a number does not exist in the case of a nontrivial scaling. But note that for the actual numerical computations performed in Section 5, we have to focus on a bounded domain and on bounded domains the number \mathbf{a} is always finite.

The term $f_{1,0}(v_{\mathbf{k}})$ in (3.3) is an approximation of the average source term $f_{1,0}(v)$ on the cell $C_{\mathbf{k}}$.

Finally, $P_{\mathbf{k}j \pm \frac{1}{2}}^j$, $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{K}$, denotes an approximation of the ‘diffusion flux’ through the boundary face $\partial C_{\mathbf{k}j \pm \frac{1}{2}}$, due to the flux function $Q_j(\frac{\partial}{\partial \xi_j} v)$. This is simply approximated by

$$P_{\mathbf{k}j \pm \frac{1}{2}}^j = Q_j\left(\frac{v_{\mathbf{k}+\mathbf{e}_j} - v_{\mathbf{k}}}{\Delta \xi_j}\right). \quad (3.7)$$

There is no difficulty in generalizing to more general diffusion fluxes of the form $Q_j(v, \frac{\partial}{\partial \xi_j} v)$ (see Kurganov & Tadmor, 2000, Section 4).

REMARK 3.1

- (i) The upper index j in (3.3)–(3.7) corresponds to the j th coordinate direction, i.e.,

$$-\frac{H_{\mathbf{k}j \pm \frac{1}{2}}^{0j} - H_{\mathbf{k}j - \frac{1}{2}}^{0j}}{\Delta \xi_j} - \sum_{i=1}^{d+1} \mu_i \frac{H_{\mathbf{k}j \pm \frac{1}{2}}^{ij} - H_{\mathbf{k}j - \frac{1}{2}}^{ij}}{\Delta \xi_j} + \frac{P_{\mathbf{k}j \pm \frac{1}{2}}^j - P_{\mathbf{k}j - \frac{1}{2}}^j}{\Delta \xi_j}$$

is the numerical flux in the j th direction and H stands for the hyperbolic flux and P for the dissipative flux.

- (ii) In the derivation of (3.3), we make essential use of the the property that the method (A.10) and its multidimensional variants (e.g., (A.12)) depend linearly on the flux function; see Remark A1.

3.2 Artificial no-flux boundary conditions

Equation (3.3) is an infinite-dimensional system of ODEs and hence not implementable on a computer. Therefore, we restrict (2.11) to a bounded domain. For simplicity, we consider only rectangular domains

of the form

$$\Omega = \{\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d : R_j^- \leq \xi_j \leq R_j^+, j = 1, \dots, d\},$$

with $R_j^- < R_j^+ \in \mathbb{R}$. For the PDE part (2.11a), we then impose no-flux boundary conditions on $\partial\Omega$. We perform the MOL approach as presented in the previous Section 3.1. For this purpose, we assume that the grid in the j th coordinate direction, $j = 1, \dots, d$, is given by

$$\xi_{k_j - \frac{1}{2}}^j = R_j^- + k_j \Delta \xi_j, \quad k_j = 0, \dots, N_j + 1, \quad \text{with } \xi_{N_j + \frac{1}{2}}^j = R_j^+.$$

Again $\mathbb{K} := \{\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d : 0 \leq k_j \leq N_j + 1, j = 1, \dots, d\}$ denotes the index set and the cells (finite volumes) are

$$C_{\mathbf{k}} := C_{k_1, \dots, k_d} := \bigtimes_{j=1}^d (\xi_{k_j - \frac{1}{2}}^j, \xi_{k_j + \frac{1}{2}}^j), \quad \mathbf{k} \in \mathbb{K}.$$

The cell $C_{\mathbf{k}}$ has center $\xi_{\mathbf{k}} = (\xi_{k_1}^1, \dots, \xi_{k_d}^d)$.

For the resulting discretization of (2.11a) on Ω , which is of the form (3.3), it is easy to implement no-flux boundary conditions by imposing for $i = 0, \dots, d + 1, j = 1, \dots, d$,

$$H_{\mathbf{k} \pm \frac{1}{2}}^{ij} = 0 \quad \text{and} \quad P_{\mathbf{k} \pm \frac{1}{2}}^j = 0 \quad \text{if} \quad \xi_{\mathbf{k} \pm \frac{1}{2}} \in \partial\Omega. \quad (3.8)$$

Moreover, an explicit upper bound \mathbf{a} of the local wave speeds in (3.4) is

$$\mathbf{a} = \max_{\xi \in \Omega} \max_j \left[\sum_{i=1}^{d+1} \mu_i \frac{\partial}{\partial v} f_{i,j}(\xi, v) + \frac{\partial}{\partial v} f_{0,j}(v) \right]. \quad (3.9)$$

Therefore, an MOL approximation of (2.11a) on Ω subject to (artificial) no-flux boundary conditions on $\partial\Omega$ is given by formula (3.3) for $\mathbf{k} \in \mathbb{K}$ with (3.4), (3.5), (3.6), (3.7), (3.8) and (3.9).

3.3 Spatial semidiscretization of the ODE and algebraic parts

Because (2.11c) and (2.11d) are already (low-dimensional) ODEs and do not explicitly depend on the values of the function v , nothing has to be done for their spatial discretizations. Thus it remains to discretize the phase conditions (2.11b). In this article we assume that they are given by one of the formulas (2.9) or (2.10). These integral conditions can easily be discretized using average values of the function on a cell, namely, we approximate (2.9) by the system

$$0 = \sum_{\mathbf{k} \in \mathbb{K}} \text{vol}(C_{\mathbf{k}}) \left[\left(- \sum_{j=1}^d \frac{H_{\mathbf{k} + \frac{1}{2}}^{ij} - H_{\mathbf{k} - \frac{1}{2}}^{ij}}{\Delta \xi_j} + \delta_{i,1} f_{1,0}(v_{\mathbf{k}}) \right) \cdot F_{\mathbf{k}}(v_{\mathbf{k}}, \mu) \right], \quad i = 1, \dots, d + 1 \quad (3.10)$$

of $d + 1$ algebraic equations. Similarly, we approximate (2.10) by

$$0 = \sum_{\mathbf{k} \in \mathbb{K}} \text{vol}(C_{\mathbf{k}}) \left[\left(- \sum_{j=1}^d \frac{H_{\mathbf{k} + \frac{1}{2}}^{ij}(\widehat{u}_{\mathbf{k}}) - H_{\mathbf{k} - \frac{1}{2}}^{ij}(\widehat{u}_{\mathbf{k}})}{\Delta \xi_j} + \delta_{i,1} f_{1,0}(\widehat{u}_{\mathbf{k}}) \right) \cdot (v_{\mathbf{k}} - \widehat{u}_{\mathbf{k}}) \right] \quad (3.11)$$

for $i = 1, \dots, d+1$. Here $\widehat{u}_{\mathbf{k}}$ is an approximation of the average value of the reference function \widehat{u} in the cell $C_{\mathbf{k}}$, $\widehat{u}_{\mathbb{K}} = (\widehat{u}_{\mathbf{k}})_{\mathbf{k} \in \mathbb{K}}$, and $H_{\mathbf{k}' \pm \frac{1}{2}}^{ij}(\widehat{u}_{\mathbb{K}})$ is given by (3.4) with v replaced by $(\widehat{u}_{\mathbf{k}})_{\mathbf{k} \in \mathbb{K}}$.

REMARK 3.2 The evaluation of the phase conditions (3.10) and (3.11) is cheap and easily obtained because the terms $H_{\mathbf{k}' \pm \frac{1}{2}}^{ij}$ and $f_{1,0}(v_{\mathbf{k}})$ in (3.10) are needed anyway for the evaluation of $F_{\mathbf{k}}$.

Similarly, for the discretization of (3.11), one needs to calculate the $H_{\mathbf{k}' \pm \frac{1}{2}}^{ij}(\widehat{u})$ only once and one has to remember the result of this calculation. In the special case, when one chooses some previous value $v_{\mathbb{K}}$ as reference solution, the value of $H_{\mathbf{k}' \pm \frac{1}{2}}^{ij}(\widehat{u})$ is, in fact, already known.

3.4 The final MOL system

Recollecting the above discussion, the MOL approximation of the freezing PDAE (2.11) takes the final form

$$v'_{\mathbf{k}} = -H_{\mathbf{k}}^0(v_{\mathbb{K}}) - H_{\mathbf{k}}^1(v_{\mathbb{K}})\mu + P_{\mathbf{k}}(v_{\mathbb{K}}) =: F_{\mathbf{k}}(v_{\mathbb{K}}, \mu), \quad \mathbf{k} \in \mathbb{K}, \quad (3.12a)$$

$$0 = \begin{cases} H^1(v_{\mathbb{K}})^{\top} F(v_{\mathbb{K}}, \mu) =: \Psi^{\text{orth}}(v_{\mathbb{K}}, \mu), & \text{orthogonal phase,} \\ H^1(\widehat{u}_{\mathbb{K}})^{\top} (v_{\mathbb{K}} - \widehat{u}_{\mathbb{K}}) =: \Psi^{\text{fix}}(v_{\mathbb{K}}), & \text{fixed phase,} \end{cases} \quad (3.12b)$$

$$g' = r_{\text{alg}}(g, \mu), \quad (3.12c)$$

$$t' = r_{\text{time}}(g), \quad (3.12d)$$

with initial data $v_{\mathbf{k}} \approx \frac{1}{\text{vol}(C_{\mathbf{k}})} \int_{C_{\mathbf{k}}} u_0(\xi) \, d\xi$ for $\mathbf{k} \in \mathbb{K}$, $g(0) = (\alpha(0), b(0)^{\top})^{\top} = (1, 0, \dots, 0)^{\top}$ and $t(0) = 0$. In (3.12), we use the abbreviations $\mu = (\mu_1, \dots, \mu_{d+1})^{\top} \in \mathbb{R}^{d+1}$ and

$$H_{\mathbf{k}}^0(v_{\mathbb{K}}) = \sum_{j=1}^d \frac{H_{\mathbf{k}' + \frac{1}{2}}^{0j} - H_{\mathbf{k}' - \frac{1}{2}}^{0j}}{\Delta \xi_j}, \quad H^0(v_{\mathbb{K}}) = (H_{\mathbf{k}}^0(v_{\mathbb{K}}))_{\mathbf{k} \in \mathbb{K}},$$

$$H_{\mathbf{k}}^1(v_{\mathbb{K}}) = \left(-f_{1,0}(v_{\mathbf{k}}) + \sum_{j=1}^d \frac{H_{\mathbf{k}' + \frac{1}{2}}^{1j} - H_{\mathbf{k}' - \frac{1}{2}}^{1j}}{\Delta \xi_j}, \dots, \sum_{j=1}^d \frac{H_{\mathbf{k}' + \frac{1}{2}}^{d+1,j} - H_{\mathbf{k}' - \frac{1}{2}}^{d+1,j}}{\Delta \xi_j} \right), \quad H^1(v_{\mathbb{K}}) = (H_{\mathbf{k}}^1(v_{\mathbb{K}}))_{\mathbf{k} \in \mathbb{K}},$$

$$P_{\mathbf{k}}(v_{\mathbb{K}}) = \sum_{j=1}^d \frac{P_{\mathbf{k}' + \frac{1}{2}}^j - P_{\mathbf{k}' - \frac{1}{2}}^j}{\Delta \xi_j}, \quad P(v_{\mathbb{K}}) = (P_{\mathbf{k}}(v_{\mathbb{K}}))_{\mathbf{k} \in \mathbb{K}},$$

$$F(v_{\mathbb{K}}, \mu) = (F_{\mathbf{k}}(v_{\mathbb{K}}, \mu))_{\mathbf{k} \in \mathbb{K}},$$

$$g = \begin{pmatrix} \alpha \\ b \end{pmatrix}, \quad r_{\text{alg}}(g, \mu) = \begin{pmatrix} g_1 \mu_1 \\ g_1^{p-1} \mu_{(2:d+1)}^{\top} \end{pmatrix} = \begin{pmatrix} \alpha \mu_1 \\ \alpha^{p-1} \mu_{(2:d+1)}^{\top} \end{pmatrix},$$

$$r_{\text{time}}(g) = g_1^{2p-2} = \alpha^{2p-2},$$

with all terms explained in Section 3.1. Note, that we included the $f_{1,0}$ -term in the first column of $H_{\mathbf{k}}^1$. We also impose (3.8) to implement no-flux boundary conditions. As before $v_{\mathbb{K}} = (v_{\mathbf{k}})_{\mathbf{k} \in \mathbb{K}}$ is replaced by

$\widehat{u}_{\mathbb{K}} = (\widehat{u}_{\mathbf{k}})_{\mathbf{k} \in \mathbb{K}}$ in $H_{\mathbf{k}j \pm \frac{1}{2}}^{ij}$ and $f_{1,0}$ for the definition of $H^1(\widehat{u}_{\mathbb{K}})$. Equation (3.12) is written in matrix times vector form, where we understand $H^0(v_{\mathbb{K}})$, $f^1(v_{\mathbb{K}})$ and $P(v_{\mathbb{K}})$ as vectors in $\mathbb{R}^{\#(\mathbb{K})}$ and $H^1(v_{\mathbb{K}})$ as a matrix in $\mathbb{R}^{\#(\mathbb{K}), d+1}$, with $\#(\mathbb{K})$ being the number of elements in the index set \mathbb{K} .

4. Implementation of the numerical freezing method II: time discretization

In this section, we introduce a new time discretization of the DAE (3.12). First, we motivate and explain our scheme and then we analyse the local truncation error introduced by it. Note that for a full convergence analysis, the PDE approximation error has to be analysed and one has to cope with the unbounded operators. We finish this section with a description of how the implicit equations can be solved efficiently.

4.1 An IMEX-RK scheme

Originating from the structure of the original problem (2.11a) (see Remark 2.3), the ODE part (3.12a) of (3.12) has very different components:

The $P(v_{\mathbb{K}})$ -part is linear in $v_{\mathbb{K}}$, but as the discretization of a (negative) elliptic operator has a spectrum that extends far into the left complex half-plane. Therefore, the ‘subproblem’ $v'_{\mathbb{K}} = P(v_{\mathbb{K}})$ becomes very stiff for fine spatial grids and an efficient numerical scheme requires an implicit time discretization.

On the other hand, the terms $H^0(v_{\mathbb{K}})$ and $H^1(v_{\mathbb{K}})\mu$ are highly nonlinear in the argument $v_{\mathbb{K}}$ due to the nonlinearity f and the reconstruction of a piecewise linear function from cell averages; see (3.4), (3.6) and the appendix. But these terms, which originate from the spatial discretization of a hyperbolic problem, have a moderate CFL number, and the maximal temporal step size that satisfies the CFL restriction scales linearly with the spatial step size. Therefore, the ‘subproblem’ $v'_{\mathbb{K}} = -H^0(v_{\mathbb{K}}) - H^1(v_{\mathbb{K}})\mu$ is most efficiently implemented by an explicit time-marching scheme.

To couple these contradicting requirements, we introduce a half-explicit IMEX-RK time discretization for coupled DAEs of the form (3.12). For half-explicit RK schemes for DAE problems, see Hairer *et al.* (1989).

To simplify the notation, we write V for $v_{\mathbb{K}}$ and restrict the discussion mainly to (3.12a) and (3.12b), because (3.12c) and (3.12d) decouple. Hence, consider an ODAE of the form

$$\begin{aligned} V' &= P(V) - H^1(V)\mu - H^0(V) + G(V), \\ 0 &= H^1(V)^{\top} (P(V) - H^1(V)\mu - H^0(V) + G(V)) \end{aligned} \quad (4.1)$$

in the case of the orthogonal phase condition (2.9) (respectively (3.10)) and

$$\begin{aligned} V' &= P(V) - H^1(V)\mu - H^0(V) + G(V), \\ 0 &= H^1(\widehat{U})^{\top} V - H^1(\widehat{U})^{\top} \widehat{U} \end{aligned} \quad (4.2)$$

in the case of the fixed phase condition (2.10) (respectively (3.11)).

From now on, consider general DAEs of the form (4.1) or (4.2), i.e., P , H^1 , H^0 and G are given as nonlinear functions of V , sufficiently smooth in their arguments.

Let two Butcher tableaux be given,

$$\text{tableau 1: } \begin{array}{c|c} c & A \\ \hline & b^{\top} \end{array}$$

$$\text{tableau 2: } \begin{array}{c|c} c & \widehat{A} \\ \hline & \widehat{b}^{\top} \end{array}$$

with $c = (c_0, \dots, c_s)^\top$, $A = (a_{ij})_{i,j=0,\dots,s}$, $\hat{A} = (\hat{a}_{ij})_{i,j=0,\dots,s}$ and $b = (a_{s0}, \dots, a_{ss})^\top$, $\hat{b} = (\hat{a}_{s0}, \dots, \hat{a}_{ss})^\top$. We assume that the RK method with tableau 1 is explicit and the RK method with tableau 2 is diagonally implicit, i.e., $a_{ij} = 0$ for all $j \geq i$ and $\hat{a}_{ij} = 0$ for all $j > i$, $i, j = 0, \dots, s$.

Now assume that at some time instance τ_n , a consistent approximation V^n and μ^n of $V(\tau_n)$ and $\mu(\tau_n)$ is given. As usual in the theory of DAEs, by consistent approximation, we imply that V^n and μ^n satisfy the algebraic constraints as well as the hidden constraints, which are obtained by differentiating the algebraic equation with respect to time and using the PDE for the resulting time derivatives (see Hairer & Wanner, 1996, Chapter VII.1). A step for (4.1) from τ_n to $\tau_{n+1} = \tau_n + h$ with step size h is then performed as follows:

$$\text{Set } V_0 = V^n,$$

and for $i = 1, \dots, s$, the internal values V_i and μ_{i-1} of the scheme are given as solutions to

$$\begin{cases} V_i = V_0 - h \sum_{v=0}^{i-1} a_{iv} (H^1(V_v)\mu_v + H^0(V_v) - G(V_v)) + h \sum_{v=0}^i \hat{a}_{iv} P(V_v), \\ 0 = H^1(V_{i-1})^\top (P(V_{i-1}) - H^1(V_{i-1})\mu_{i-1} - H^0(V_{i-1}) + G(V_{i-1})), \end{cases} \quad i = 1, \dots, s. \quad (4.3)$$

As an approximation of V at the new time instance τ_{n+1} ,

$$\text{set } V^{n+1} := V_s.$$

A suitable approximation of μ^{n+1} at the new time instance is actually the internal value μ_0 of the next time instance, i.e., $\mu = \mu^{n+1}$ solves $0 = H^1(V^{n+1})^\top (P(V^{n+1}) - H^1(V^{n+1})\mu - H^0(V^{n+1}) + G(V^{n+1}))$.

Similarly, a step for (4.2) is performed as follows:

$$\text{Set } V_0 = V^n,$$

and for $i = 1, \dots, s$, let the internal values V_i and μ_{i-1} be given as solutions to

$$\begin{cases} V_i = V_0 - h \sum_{v=0}^{i-1} a_{iv} (H^1(V_v)\mu_v + H^0(V_v) - G(V_v)) + h \sum_{v=0}^i \hat{a}_{iv} P(V_v), \\ 0 = H^1(\hat{U})^\top V_i - H^1(\hat{U})^\top \hat{U}, \end{cases} \quad i = 1, \dots, s. \quad (4.4)$$

In this case we set

$$V^{n+1} := V_s \quad \text{and} \quad \mu^{n+1} := \mu_{s-1}$$

as approximations of V and μ at the new time instance $\tau_{n+1} = \tau_n + h$.

REMARK 4.1

- (i) A suitable time discretization for the ‘parabolic subproblem’ $v'_\mathbb{K} = P(v_\mathbb{K})$ is the Crank–Nicolson method, for which there is no CFL restriction and the resulting full discretization of $v' = \Delta v$ is of second order.

- (ii) Concerning the semidiscrete ‘hyperbolic subproblem’ $v'_{\mathbb{K}} = -H^0(v_{\mathbb{K}}) - H^1(v_{\mathbb{K}})\mu$, it is shown in [Kurganov & Tadmor \(2000, Theorem 5.1\)](#) that an explicit Euler discretization with a suitable CFL condition satisfies a stability property in the form of a maximum principle (cf. nonoscillatory). It is possible to retain this property for higher-order methods, if they can be written as convex combinations of explicit Euler steps. A simple second-order method for which this is possible is Heun’s method (see [Kurganov & Tadmor, 2000, Corollary 5.1](#)). It has already been observed, in [Shu & Osher \(1988\)](#), that Heun’s method is the optimal (concerning the CFL restrictions) second-order explicit RK-type scheme.

For concreteness and motivated by Remark 4.1, we choose

$$c = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, b^{\top} = (\frac{1}{2}, \frac{1}{2}, 0), \hat{A} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \hat{b}^{\top} = (\frac{1}{2}, 0, \frac{1}{2}), \quad (4.5)$$

i.e., we couple Heun’s method with the Crank–Nicolson method.

As noted above, the ODEs (3.12c) and (3.12d) can be solved in a postprocessing step, but it is more convenient to do the calculation in parallel and use the same explicit scheme with tableau 1, because the intermediate stages of V and μ are already calculated. Therefore, given the states g^n and t^n at τ_n , we perform a time step for (3.12c) and (3.12d) from τ_n to $\tau_{n+1} = \tau_n + h$ by

$$\text{let } g_0 = g^n, \quad t_0 = t^n,$$

compute for $i = 1, \dots, s$,

$$g_i = g_0 + h \sum_{v=0}^{i-1} a_{iv} r_{\text{alg}}(g_v, \mu_v), \quad t_i = t_0 + h \sum_{v=0}^{i-1} a_{iv} r_{\text{time}}(g_v)$$

and set

$$g^{n+1} := g_s, \quad t^{n+1} := t_s.$$

4.2 Order of the time discretization

Now we consider the local truncation error of our IMEX-RK scheme for DAEs (4.3), respectively (4.4), with tableaux (4.5). Hence, consider a system of ODEs

$$V' = P(V) + H(V, \mu), \quad g' = r_{\text{alg}}(\mu, g), \quad t' = r_{\text{time}}(g), \quad (4.6)$$

coupled to a system of algebraic equations of the form

$$0 = \Psi(V, \mu), \quad (4.7a)$$

or of the form

$$0 = \Psi(V). \quad (4.7b)$$

We assume that for consistent initial data $V(0) = V^0$, $g(0) = g^0$, $t(0) = t^0$, $\mu(0) = \mu^0$, a smooth solution $V \in \mathcal{C}^3([0, T]; \mathbb{R}^m)$, $\mu \in \mathcal{C}^3([0, T]; \mathbb{R}^p)$ of (4.6), (4.7a), respectively, (4.6), (4.7b) exists. Furthermore, we assume the following hypothesis.

HYPOTHESIS 4.2

- (i) In the case (4.6), (4.7a) the matrix $\partial_\mu \Psi(V(\tau), \mu(\tau))$ is invertible for any $\tau \in [0, T)$.
- (ii) In the case (4.6), (4.7b) the matrix $\partial_V \Psi(V(\tau)) \partial_\mu H(V(\tau), \mu(\tau))$ is invertible for any $\tau \in [0, T)$.

It is easy to check that under Hypothesis 4.2 the system (4.6), (4.7a) is a DAE of (differentiation) index 1 and the system (4.6), (4.7b) is a DAE of (differentiation) index 2 (see, e.g., Hairer & Wanner, 1996, Chapter VII).

Given consistent data V^n, g^n, t^n of the DAE (4.6), (4.7a) or (4.6), (4.7b) at some time instance τ_n , a step of size h of the method with these data takes the following explicit form: one first solves the following system for $(V_1, \mu_0, g_1, t_1) \in \mathbb{R}^{m+p+p+1}$ and $(V_2, \mu_1, g_2, t_2) \in \mathbb{R}^{m+p+p+1}$ in two consecutive steps:

$$V_0 = V^n, \quad g_0 = g^n, \quad t_0 = t^n, \quad (4.8a)$$

$$\left[\begin{array}{l} V_1 = V_0 + \frac{h}{2}P(V_0) + \frac{h}{2}P(V_1) + hH(V_0, \mu_0), \\ 0 = \begin{cases} \Psi(V_0, \mu_0), & \text{case (4.7a), or} \\ \Psi(V_1), & \text{case (4.7b),} \end{cases} \\ g_1 = g_0 + hr_{\text{alg}}(\mu_0, g_0), \\ t_1 = t_0 + hr_{\text{time}}(g_0), \end{array} \right. \quad (4.8b)$$

$$\left[\begin{array}{l} V_2 = V_0 + \frac{h}{2}P(V_0) + \frac{h}{2}P(V_2) + \frac{h}{2}H(V_0, \mu_0) + \frac{h}{2}H(V_1, \mu_1), \\ 0 = \begin{cases} \Psi(V_1, \mu_1), & \text{case (4.7a), or} \\ \Psi(V_2), & \text{case (4.7b),} \end{cases} \\ g_2 = g_0 + \frac{h}{2}r_{\text{alg}}(\mu_0, g_0) + \frac{h}{2}r_{\text{alg}}(\mu_1, g_1), \\ t_2 = t_0 + \frac{h}{2}r_{\text{time}}(g_0) + \frac{h}{2}r_{\text{time}}(g_1). \end{array} \right. \quad (4.8c)$$

And then the approximations of V , μ , g and t at the new time instance $\tau^{n+1} = \tau^n + h$ are given by

$$V^{n+1} := V_2, \quad g^{n+1} := g_2, \quad t^{n+1} := t_2 \quad \text{and} \quad \mu^{n+1} \begin{cases} \text{by solving } \Psi(V^{n+1}, \mu^{n+1}) = 0, & \text{case (4.7a),} \\ \mu_1, & \text{case (4.7b).} \end{cases} \quad (4.9)$$

Since the g - and t -equations decouple from the system, we first consider the V and μ variables separately. To analyse the local error, we assume that (V_\star, μ_\star) is a given consistent value at $\tau = 0$, so that the DAE satisfies all assumptions from above with (V^0, μ^0) replaced by (V_\star, μ_\star) . For brevity, a subindex \star denotes the evaluation of a function at $\tau = 0$, e.g., $(V_\star, \mu_\star) = (V(0), \mu(0))$, $\partial_V P_\star = \frac{\partial}{\partial V} P(V(0))$.

Taylor expansion of the exact solution. First, we consider the differential variables V and obtain from the ODE (4.6) and anticipating $\mu(0) = \mu_*$ (see (4.11a), respectively, (4.12b), below),

$$V(0) = V_*, \quad (4.10a)$$

$$V'(0) = P_* + H_* =: V'_*, \quad (4.10b)$$

$$V''(0) = \partial_V P_* V'_* + \partial_V H_* V'_* + \partial_\mu H_* \mu'_*(0) =: V''_*. \quad (4.10c)$$

Similarly, we can use the differential equation (4.6) together with the algebraic constraint (4.7a) to obtain in the index-1 case for the algebraic variable,

$$0 = \Psi(V_*, \mu_*) \quad (\text{locally unique solvable for } \mu_* \text{ by Hypothesis 4.2 (i)}), \quad (4.11a)$$

$$\mu'_*(0) = -(\partial_\mu \Psi_*)^{-1} \partial_V \Psi_*(P_* + H_*) =: \mu'_*, \quad (4.11b)$$

$$\mu''_*(0) = -(\partial_\mu \Psi_*)^{-1} \left\{ \partial_V^2 \Psi_* V_*'^2 + \partial_V \Psi_* V_*'' + 2\partial_V \partial_\mu \Psi_* V'_* \mu'_* + \partial_\mu^2 \Psi_* \mu_*'^2 \right\} =: \mu''_*. \quad (4.11c)$$

For the index-2 case (i.e., (4.7b)), the first coefficients of the Taylor expansion of the algebraic variables μ are obtained by differentiating the algebraic condition (4.7b) and using the ODE (4.6) to find

$$0 = \Psi(V_*), \quad (4.12a)$$

$$0 = \partial_V \Psi_*(P_* + H_*) \quad (\text{locally unique solvable for } \mu_* \text{ by Hypothesis 4.2 (ii)}), \quad (4.12b)$$

$$\mu'_*(0) = -(\partial_V \Psi_* \partial_\mu H_*)^{-1} \left\{ \partial_V^2 \Psi_* V_*'^2 + \partial_V \Psi_* \partial_V P_* V'_* + \partial_V \Psi_* \partial_V H_* V'_* \right\} =: \mu'_*. \quad (4.12c)$$

Taylor expansion of the numerical solution. We assume that the numerical solution and all intermediate stages depend smoothly on the step size h . To emphasize this dependence, we explicitly include the dependence on h in the notation.

First consider the differential variable V , anticipating that the algebraic variables $\mu_0(h)$ and $\mu_1(h)$ are known and satisfy $\mu_0(0) = \mu_1(0) = \mu_*$, which will be justified in (4.17) and (4.18), respectively (4.22) and (4.24). By (4.8) the numerical values $V_1(h)$ and $V_2(h)$ satisfy

$$\begin{aligned} V_1(h) &= V_0 + \frac{h}{2} P(V_0) + \frac{h}{2} P(V_1(h)) + hH(V_0, \mu_0(h)) \\ &= V_* + \frac{h}{2} P_* + \frac{h}{2} P(V_1(h)) + hH(V_*, \mu_0(h)), \\ V_2(h) &= V_0 + \frac{h}{2} P(V_0) + \frac{h}{2} P(V_2(h)) + \frac{h}{2} H(V_0, \mu_0(h)) + \frac{h}{2} H(V_1(h), \mu_1(h)) \\ &= V_* + \frac{h}{2} P_* + \frac{h}{2} P(V_2(h)) + \frac{h}{2} H(V_*, \mu_0(h)) + \frac{h}{2} H(V_1(h), \mu_1(h)). \end{aligned} \quad (4.13)$$

For the differential variables, we then obtain for the first intermediate stage,

$$V_1(0) = V_*, \quad (4.14a)$$

$$V'_1(h) = \frac{1}{2} P_* + \frac{1}{2} P(V_1(h)) + \frac{h}{2} \partial_V P(V_1(h)) V'_1(h) + H(V_*, \mu_0(h)) + h \partial_\mu H(V_*, \mu_0(h)) \mu'_0(h), \quad (4.14b)$$

$$V'_1(0) = P_* + H_* = V'_*, \quad (4.14c)$$

$$V''_1(0) = \partial_V P_*(P_* + H_*) + 2\partial_\mu H_* \mu'_0(0). \quad (4.14d)$$

Similarly, for the final (second) stage we obtain

$$V_2(0) = V_*, \quad (4.15a)$$

$$V_2'(h) = \frac{1}{2}P_* + \frac{1}{2}P(V_2(h)) + \frac{h}{2}\partial_V P(V_2(h))V_2'(h) + \frac{1}{2}H(V_*, \mu_0(h)) + \frac{h}{2}\partial_\mu H(V_*, \mu_0(h))\mu_0'(h) \\ + \frac{1}{2}H(V_1(h), \mu_1(h)) + \frac{h}{2}\partial_V H(V_1(h), \mu_1(h))V_1'(h) + \frac{h}{2}\partial_\mu H(V_1(h), \mu_1(h))\mu_1'(h), \quad (4.15b)$$

$$V_2'(0) = P_* + H_* = V_*', \quad (4.15c)$$

$$V_2''(0) = \partial_V P_*(P_* + H_*) + \partial_V H_*(P_* + H_*) + \partial_\mu H_*(\mu_0'(0) + \mu_1'(0)). \quad (4.15d)$$

Now we consider the Taylor expansion of the numerical solution of the algebraic variables. We begin with the *index-1 case*, i.e., (4.13) is closed by imposing the algebraic constraints

$$0 = \Psi(V_0(h), \mu_0(h)), \quad 0 = \Psi(V_1(h), \mu_1(h)). \quad (4.16)$$

Since $V_0(h) = V_*$ for all $h \geq 0$, (4.16) implies, because of Hypothesis 4.2 (i),

$$\mu_0(h) = \mu_* \text{ and } \mu_0'(h) = 0 \quad \forall h \geq 0. \quad (4.17)$$

This justifies (4.14) and shows $V_1''(0) = \partial_V P_*(P_* + H_*)$. Then we can calculate the Taylor expansion of $\mu_1(h)$ around $h = 0$ from (4.16) to find

$$\begin{aligned} 0 = \Psi(V_1(h), \mu_1(h)) &\Rightarrow \mu_1(0) = \mu_* \text{ since } V_1(0) = V_* \text{ by (4.14a),} \\ 0 = \partial_V \Psi(V_1, \mu_1)V_1' + \partial_\mu \Psi(V_1, \mu_1)\mu_1' &\Rightarrow \mu_1'(0) = \mu_*' \text{ by (4.11b) and (4.14c).} \end{aligned} \quad (4.18)$$

Thus, inserting the findings for $\mu_0'(0)$ and $\mu_1'(0)$ into the Taylor expansion of V_2 , (4.15) shows $V_2(0) = V_*$, $V_2'(0) = V_*'$, and $V_2''(0) = V_*''$, which proves for the differential variable V_2 second-order consistency,

$$V_2(h) = V(h) + \mathcal{O}(h^3) \quad \text{as } h \searrow 0. \quad (4.19)$$

Setting $\mu_2(h)$ as a solution of $\Psi(V_2(h), \mu_2(h)) = 0$ also implies second order for the algebraic variable

$$\mu_2(h) = \mu(h) + \mathcal{O}(h^3) \quad \text{as } h \searrow 0. \quad (4.20)$$

In the *index-2 case* (4.13) is closed by appending the algebraic constraints

$$0 = \Psi(V_1(h)), \quad 0 = \Psi(V_2(h)). \quad (4.21)$$

Differentiating the first of these two equations yields $0 = \partial_V \Psi(V_1(h))V_1'(h)$ and at $h = 0$,

$$0 = \partial_V \Psi_*(P_* + H(V_*, \mu_0(0))), \quad (4.22)$$

which by Hypothesis 4.2 (ii) has the locally unique solution $\mu_0(0) = \mu_*$. Considering the second derivative at $h = 0$ leads to

$$0 = \partial_V^2 \Psi_*(P_* + H_*)^2 + \partial_V \Psi_* (\partial_V P_*(P_* + H_*) + 2\partial_\mu H_* \mu'_0(0)),$$

so that

$$\mu'_0(0) = -(\partial_V \Psi_* \partial_\mu H_*)^{-1} \left\{ \frac{1}{2} \partial_V^2 \Psi_*(P_* + H_*)^2 + \frac{1}{2} \partial_V \Psi_* (\partial_V P_*(P_* + H_*)) \right\}. \quad (4.23)$$

Similarly, differentiating the second equation in (4.21) once and evaluating at $h = 0$ yields

$$0 = \partial_V \Psi_* (P_* + \frac{1}{2} H_* + \frac{1}{2} H(V_*, \mu_1(0))),$$

which has the locally unique solution

$$\mu_1(0) = \mu_*. \quad (4.24)$$

Considering the second derivative at $h = 0$, inserting (4.15d) and comparing with (4.12c) shows

$$\mu'_0(0) + \mu'_1(0) = \mu'_*. \quad (4.25)$$

Therefore, we obtain again (4.19) for V_2 . But for the algebraic variables we find only the estimate

$$\mu_1(h) + \mathcal{O}(h) = \mu_2(h) + \mathcal{O}(h) = \mu(h) \quad \text{as } h \searrow 0, \quad (4.26)$$

which is a severe order reduction.

Nevertheless, when the group variables g and the transformed time t are calculated in parallel with V and μ , these variables are again second-order accurate, i.e.,

$$g_2(h) = g(h) + \mathcal{O}(h^3), \quad t_2(h) = t(h) + \mathcal{O}(h^3) \quad \text{as } h \searrow 0. \quad (4.27)$$

This can be seen by either including the equations $g' = r_{\text{alg}}(\mu, g)$ and $t' = r_{\text{time}}(g)$ into the V equation, or by performing a similar analysis to above.

Summarizing the above analysis we obtain the following theorem.

THEOREM 4.3 Assume that the DAE consisting of the ODE system (4.6) and either (4.7a) or (4.7b) with consistent initial data (V_*, μ_*, g_*, t_*) at $\tau = \tau_n$ has a smooth solution in some nonempty interval $[\tau_n, T)$ and satisfies Hypothesis 4.2. Then, the method (4.8), (4.9) is consistent of second order at the differential variables, i.e., (4.19) and (4.27) hold. Moreover, in the case of (4.7a) also the algebraic variables μ are second order consistently approximated, i.e., (4.20) holds.

REMARK 4.4 We observe here the well-known difficulty with higher index problems that one faces a loss in the order of the approximation of the algebraic variables. But note, that one is often mainly interested in the behavior of the original solution, so that the approximation of μ is not important, but the approximations of V , g and t are crucial.

Moreover, when the main focus is on the asymptotic behavior and the final rest state, one is interested in the limit $\lim_{\tau \rightarrow \infty} \mu(\tau)$ but not on its actual evolution. The approximation of this is actually not dependent on the time discretization but only on the spatial discretization, because we are using an RK-type method. In fact, in this case, it might even be better to use a scheme that has a less restrictive CFL condition but maybe has a lower order in the time approximation. See [Shu \(1988\)](#) for such ideas in the case of hyperbolic conservation laws.

4.3 Efficiently solving the RK equations

We remark that in the Burgers's case and similar cases with a linear operator P , the actual equations (4.3) and (4.4) from the IMEX time discretization can be solved very efficiently. To see this, we write (4.3) in block matrix form

$$\begin{pmatrix} I - h\widehat{a}_{i,i}P & ha_{i,i-1}H^1(V_{i-1}) \\ 0 & H^1(V_{i-1})^\top H^1(V_{i-1}) \end{pmatrix} \begin{pmatrix} V_i \\ \mu_{i-1} \end{pmatrix} = \begin{pmatrix} R_i^1 \\ R_i^2 \end{pmatrix}, \quad i = 1, \dots, s, \quad (4.28)$$

where $R_i^1 = R_i^1(V_0, \dots, V_{i-1}, \mu_0, \dots, \mu_{i-2})$ and $R_i^2 = R_i^2(V_{i-1})$ are given by

$$\begin{aligned} R_i^1 &= V_0 - h \left(\sum_{v=0}^{i-2} a_{iv} H^1(V_v) \mu_v + \sum_{v=0}^{i-1} a_{iv} (H^0(V_v) - G(V_v)) - \sum_{v=0}^{i-1} \widehat{a}_{iv} P(V_v) \right), \\ R_i^2 &= H^1(V_{i-1})^\top (P(V_{i-1}) - H^0(V_{i-1}) + G(V_{i-1})). \end{aligned}$$

A solution to (4.28) can be calculated by solving for each $i = 1, \dots, s$ one small linear system with the $(d+1) \times (d+1)$ -matrix $H^1(V^{(i-1)})^\top H^1(V^{(i-1)})$ and one large linear system with the matrix $I - h\widehat{a}_{i,i}P$.

Similarly, (4.4) can be written in the block matrix form

$$\begin{pmatrix} I - h\widehat{a}_{i,i}P & ha_{i,i-1}H^1(V_{i-1}) \\ H^1(\widehat{U})^\top & 0 \end{pmatrix} \begin{pmatrix} V_i \\ \mu_{i-1} \end{pmatrix} = \begin{pmatrix} R_i^1 \\ R^2 \end{pmatrix}, \quad i = 1, \dots, s, \quad (4.29)$$

where R_i^1 is given as above and $R^2 = H^1(\widehat{U})^\top \widehat{U}$ does not depend on the stage i . It is possible to obtain a solution to (4.29) by solving $d+2$ large linear systems with $I - h\widehat{a}_{i,i}P$ and one small linear $(d+1) \times (d+1)$ system:

$$\begin{aligned} A_\star &= (I - h\widehat{a}_{i,i}P)^{-1} ha_{i,i-1}H^1(V_{i-1}) && ((d+1) \times \text{large}), \\ V_\star &= (I - h\widehat{a}_{i,i}P)^{-1} R_i^1 && (1 \times \text{large}), \\ \mu_{i-1} &= (H^1(\widehat{U})^\top A_\star)^{-1} (H^1(\widehat{U})^\top V_\star - R^2) && (1 \times \text{small}), \\ V_i &= V_\star - A_\star \mu_{i-1}. \end{aligned}$$

Therefore, if h does not vary during computation, factorizations of $I - h\widehat{a}_{i,i}P$, $i = 1, \dots, s$ can be calculated in a preprocessing step and then the subsequent time steps are rather cheap. Note that for our specific scheme with Butcher tableaux (4.5) it holds that $s = 2$ and $\widehat{a}_{1,1} = \widehat{a}_{2,2} = \frac{1}{2}$, so that only one matrix factorization is needed, namely of $I - h\frac{1}{2}P$.

5. Numerical experiments

In this article, we are primarily interested in the actual discretization of the freezing PDAE (2.11) and the error introduced by the discretization. We present the results of several numerical experiments that support second-order convergence of the method for the full discretization of the PDAE for our scheme.

For further numerical findings, we refer to Rottmann-Matthes (2016), where we use the same scheme and see that our method enables us to do long-time simulations for Burgers's equation on fixed (bounded) computational domains also for different parameter values of $p > 1$ in (1.4). We also note that our method is able to capture the metastable solution behavior present in (1.4) for the case $0 < \nu < |a|$, where a is the parameter in (1.4). We again refer to Rottmann-Matthes (2016). Note that in that article we completely ignored the errors introduced by the discretization.

In the following, we focus on the special choice $p = \frac{d+1}{d}$ in (1.4). For convenience, we explicitly state the systems for freezing similarity solutions of the 'conservative' Burgers's equation (i.e., $p = \frac{d+1}{d}$) in one and two spatial dimensions. These are the systems we numerically solve with the scheme proposed in Sections 3 and 4. In the *one-dimensional case*, we solve the PDAE system

$$\begin{cases} v_\tau = \nu v_{\xi\xi} - \frac{1}{2} (v^2)_\xi + \mu_1 (\xi v)_\xi + \mu_2 v_\xi, & v(0) = u_0, \\ 0 = \psi(v, \mu) \in \mathbb{R}^2, \\ \alpha_\tau = \alpha \mu_1, \quad b_\tau = \alpha \mu_2, \quad t_\tau = \alpha^2, & \alpha(0) = 1, \quad b(0) = 0, \quad t(0) = 0. \end{cases} \quad (5.1)$$

And the functional ψ is given by either

$$\psi^{\text{orth}}(v, \mu) = \left(\int_{\mathbb{R}} (\xi v)_\xi \cdot \left(\nu v_{\xi\xi} - \frac{1}{2} (v^2)_\xi + \mu_1 (\xi v)_\xi + \mu_2 v_\xi \right) d\xi \right) \quad (\text{orthogonal phase condition})$$

or

$$\psi^{\text{fix}}(v) = \left(\int_{\mathbb{R}} (\xi \hat{u})_\xi \cdot (v - \hat{u}) d\xi \right) \quad (\text{fixed phase condition}).$$

Similarly, in the *two-dimensional case*, we consider (1.4) with $a = (1, 1)^\top$ and $p = \frac{3}{2}$ and numerically solve

$$\begin{cases} v_\tau = \underbrace{\nu \Delta v - \frac{2}{3} (\partial_{\xi_1} + \partial_{\xi_2}) |v|^{\frac{3}{2}} + \frac{1}{2} \mu_1 ((\xi_1 v)_{\xi_1} + (\xi_2 v)_{\xi_2}) + \mu_2 v_{\xi_1} + \mu_3 v_{\xi_2}}_{=: \text{RHS}(v, \mu)}, & v(0) = u_0, \\ 0 = \psi(v, \mu) \in \mathbb{R}^3, \\ \alpha_\tau = \alpha \mu_1, \quad b_{1,\tau} = \alpha^{\frac{1}{2}} \mu_2, \quad b_{2,\tau} = \alpha^{\frac{1}{2}} \mu_3, \quad t_\tau = \alpha, & \alpha(0) = 1, \quad b_1(0) = 0, \quad b_2(0) = 0, \quad t(0) = 0. \end{cases} \quad (5.2)$$

The functional ψ is given by either

$$\psi^{\text{orth}}(v, \mu) = \left(\begin{array}{l} \int_{\mathbb{R}^2} ((\xi_1 v)_{\xi_1} + (\xi_2 v)_{\xi_2}) \cdot \text{RHS}(v, \mu) d\xi \\ \int_{\mathbb{R}^2} v_{\xi_1} \cdot \text{RHS}(v, \mu) d\xi \\ \int_{\mathbb{R}^2} v_{\xi_2} \cdot \text{RHS}(v, \mu) d\xi \end{array} \right) \quad (\text{orthogonal phase condition})$$

or

$$\psi^{\text{fix}}(v) = \left(\begin{array}{c} \int_{\mathbb{R}^2} ((\xi_1 \widehat{u})_{\xi_1} + (\xi_2 \widehat{u})_{\xi_2}) \cdot (v - \widehat{u}) \, d\xi \\ \int_{\mathbb{R}^2} \widehat{u}_{\xi_1} \cdot (v - \widehat{u}) \, d\xi \\ \int_{\mathbb{R}^2} \widehat{u}_{\xi_2} \cdot (v - \widehat{u}) \, d\xi \end{array} \right) \quad (\text{fixed phase condition}).$$

In the following, we write $\Delta\tau$ for the time-step size h , and $\Delta\xi$, respectively, $\Delta\xi_1$ and $\Delta\xi_2$, for the spatial step sizes.

5.1 One-dimensional experiments

We choose a fixed spatial domain Ω , as explained in Section 3.2. Due to the nonlinearity, and because the algebraic variables μ depend implicitly on the solution, we choose a very rough upper estimate for the value of \mathbf{a} in (3.4) for simplicity. We take (cf. (3.9))

$$\mathbf{a} = \sup_{\xi \in \Omega} |v(\xi)| + |\mu_1| \max\{|\xi| : \xi \in \Omega\} + |\mu_2|, \quad (5.3)$$

where μ_1, μ_2 are the current approximations obtained in the last time step, so that the number \mathbf{a} is updated after each time step.

As time-step size we choose $\Delta\tau$, which satisfies

$$\Delta\tau \leq \frac{\Delta\xi}{\mathbf{a}} \cdot \lambda_{\text{CFL}}, \quad (5.4)$$

so the time-step size may increase or decrease during the calculation, depending on the evolution of \mathbf{a} from (5.3). Experimentally, we found that $\lambda_{\text{CFL}} = \frac{1}{3}$ is a suitable choice for all spatial step sizes $\Delta\xi$ and all viscosities $\nu \geq 0$. In our code, we actually do not update $\Delta\tau$ after each step, but only if (5.4) is violated or $\Delta\tau$ might be enlarged significantly according to (5.4).

In Fig. 1(a–c), we show the time evolution of the solution to the freezing PDAE (5.1) for $\nu = 0.4$, $\nu = 0.01$ and $\nu = 0$. The initial function is $u_0(x) = \sin(2x)1_{[-\frac{\pi}{2}, 0]} + \sin(x)1_{[0, \pi]}$, where 1_A is the indicator function. One observes that the solution converges to a stationary profile as time increases. We stopped the calculation when the solution virtually did not change anymore. Actually, one could choose the norm of v_τ and μ_τ as indicators of whether the solution has reached a steady state, and build a stopping criterion based on these norms.

The initial value together with the final state of these simulations is shown in Fig. 1(d–f). Note that in all plots the solution is given in the comoving coordinates of (5.1) and the solution in the original coordinates is obtained by the transformation $u(x, t) = \frac{1}{\alpha(\tau)} v(\frac{x-b(\tau)}{\alpha(\tau)}, \tau)$. At the final time, the algebraic variables α and μ have the values $\alpha(10) \approx 1.2 \cdot 10^6$ and $b(10) \approx -1.6 \cdot 10^6$ in Fig. 1(d), $\alpha(300) \approx 1.3 \cdot 10^{27}$ and $b(300) \approx -1.8 \cdot 10^{26}$ in Fig. 1(e), $\alpha(10) \approx 36$ and $b(10) \approx -11$ for Fig. 1(f).

Order of the method. Now we numerically check the order of our method. For this purpose, we calculate the solution for different step sizes $\Delta\xi$ until time $\tau = 1$ and compare the final state with a reference solution, obtained using a much smaller step size $\Delta\xi = 0.0005$. We calculate on the fixed spatial domain $[R^-, R^+]$ with no-flux boundary conditions as described in Section 3. We calculate the error of all solution components separately. That is, we calculate the L^2 -error of the differential variable ('v error') $\int_{R^-}^{R^+} |v_{\Delta\xi}(\xi, 1) - v_{\text{ref}}(\xi, 1)|^2 \, d\xi$, where for simplicity, we have approximated $v_{\Delta\xi}(\cdot, 1)$ and $v_{\text{ref}}(\cdot, 1)$ by piecewise linear interpolation of the cell averages appointed to the midpoints of the cell. We did not

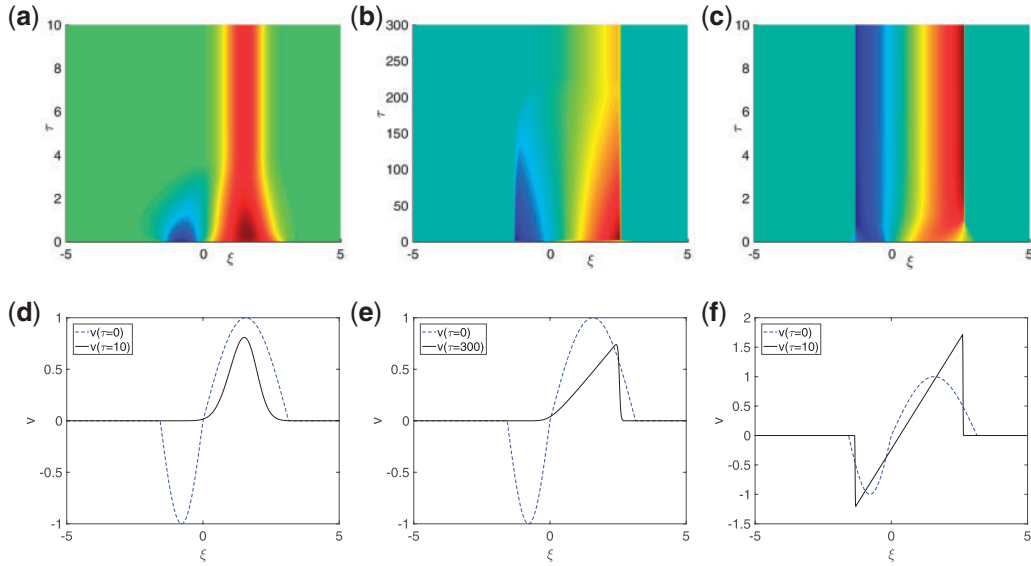


FIG. 1. Time evolution (a–c) and final states after the solution stabilized (d–f) for the freezing method for the one-dimensional Burgers's equation with different values of viscosity. The initial value is given by the dashed line and the final state by the solid line. All plots are in the scaled and translated coordinates of the freezing method.

choose a more sophisticated piecewise linear reconstruction, because already the simple interpolation yields convergence of order 2. Nevertheless, we note in passing that we observed that a piecewise constant reconstruction leads to a convergence rate of only order 1. And we also calculate the error of the algebraic variables (' μ error') $|\mu_{\Delta\xi}(1) - \mu_{\text{ref}}(1)|_\infty$, the error of the reconstructed time ('time error') $|t_{\Delta\xi}(1) - t_{\text{ref}}(1)|$ and the error of the reconstructed transformation (' α, b error') $\max(|\alpha_{\Delta\xi}(1) - \alpha_{\text{ref}}(1)|, |b_{\Delta\xi}(1) - b_{\text{ref}}(1)|)$. Here a subindex $\Delta\xi$ denotes the result of the numerical approximation with step size $\Delta\xi$ and a subindex ref refers to the reference solution. The results are shown in Fig. 2, where we consider large and small viscosities ($\nu = 1$ and $\nu = 0.01$) and in both cases orthogonal phase and fixed phase conditions. The results for other values of viscosity ν look very similar and are not shown. In all experiments, $\Delta\xi$ is prescribed, and $\Delta\tau$ is related to $\Delta\xi$, so that (5.4) holds, as described above. Moreover, we choose $R^+ = -R^- = 10$ for $\nu = 1$ and $R^+ = -R^- = 5$ for $\nu = 0.01$, so we may neglect the influence of the boundary conditions.

On the horizontal axis in Fig. 2, we plot the $\Delta\xi$ value and the vertical axis is the error of the respective solution components. In all experiments one observes second-order convergence for the orthogonal phase condition, as was proved (for ODEs) in Theorem 4.3. One also observes second-order convergence for the differential variables in the case of the fixed phase condition, which was also shown in Theorem 4.3.

Violation of the CFL condition. Our final experiment for the one-dimensional Burgers's equation concerns violation of (5.4). In Fig. 3(a), we consider the inviscid equation and choose $\lambda_{\text{CFL}} = 1.2$ in (5.4) (instead of $\lambda_{\text{CFL}} = \frac{1}{3}$). The spatial step size is $\Delta\xi = 0.1$. One observes nicely how oscillations develop. Actually, the calculation breaks down soon after the plotted solution at $\tau = 1$. For this choice of λ_{CFL} , we observed the same behavior also for all other step sizes $\Delta\xi$ we have tried and we do not present these experiments here.

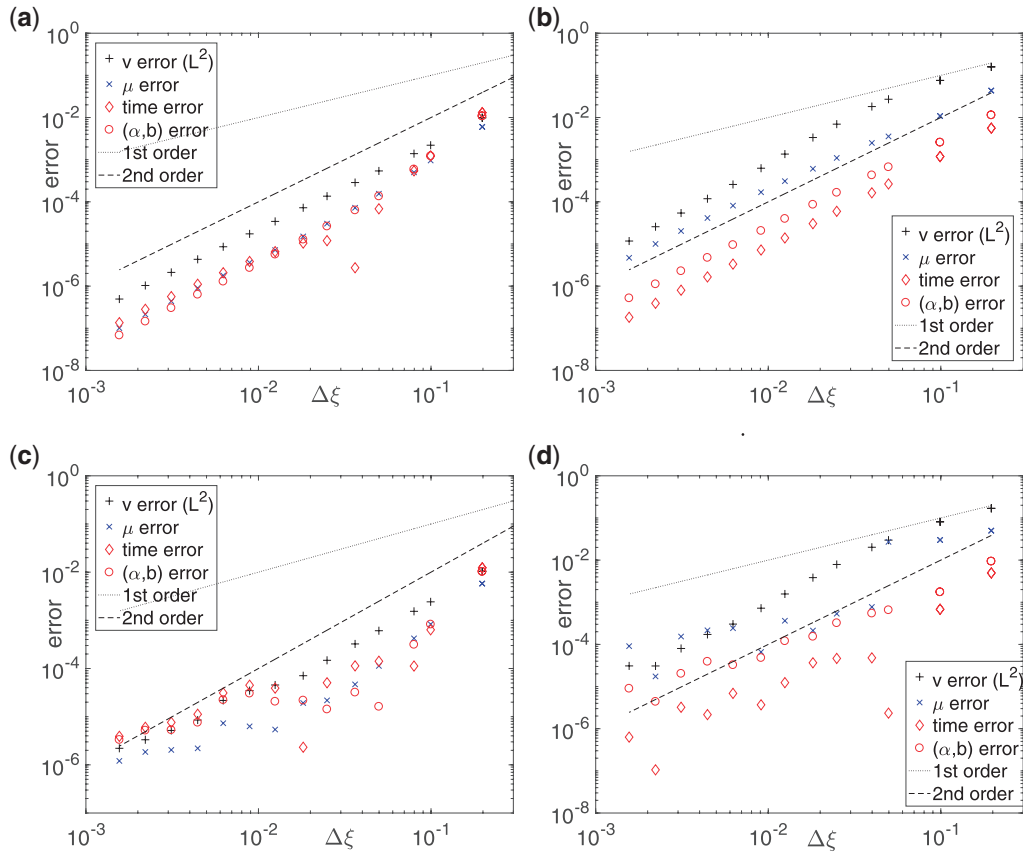


FIG. 2. Convergence plots for the freezing one-dimensional Burgers's equation. With large viscosity (left column) and very small viscosity (right column), and orthogonal phase condition (top row) and fixed phase condition (bottom row).

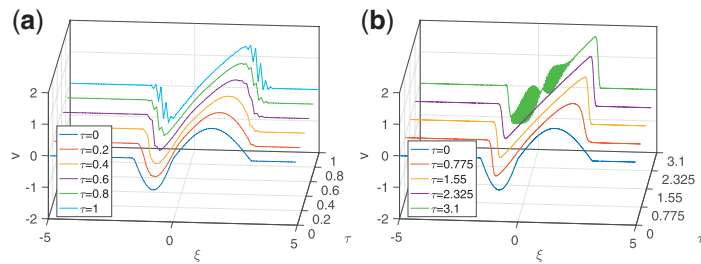


FIG. 3. Appearance of oscillations, when the ratio $\frac{\Delta\tau}{\Delta\xi}$ is too large.

In Fig. 3(b), we show the result for $\nu = 0.02$. In this case, we have chosen $\lambda_{\text{CFL}} = 5$ in (5.4) and the spatial step size was $\Delta\xi = 10/350$. Further experiments with different step sizes $\Delta\xi$ seem to suggest that the numerical discretization of the parabolic part is strongly smoothing and there is no linear barrier of $(\Delta\tau)/(\Delta\xi)$ which prevents oscillations, but rather the ratio may increase as $\Delta\xi$ decreases. Nevertheless, we expect that in the coupled hyperbolic–parabolic case the linear relation dominates.

5.2 Two-dimensional experiments

We also apply our method to the following two-dimensional Burgers's equations:

$$\partial_t u = \nu \Delta u - \frac{2}{3}(\partial_x + \partial_y) \left(|u|^{\frac{3}{2}} \right). \quad (5.5)$$

That is, we numerically solve (5.2) with the scheme introduced in Sections 3–4. First, we choose a fixed rectangular spatial domain Ω , as explained in Section 3.2. As in the one-dimensional case, we bound the maximal local wave speeds by the rough upper estimate

$$\mathbf{a} = \sup_{\xi \in \Omega} \sqrt{|\nu(\xi)|} + \max \left(|\mu_1| \max_{\xi \in \Omega} |\xi_1| + |\mu_2|, |\mu_1| \max_{\xi \in \Omega} |\xi_2| + |\mu_3| \right)$$

and choose in each time step a step size $\Delta\tau$ so that

$$\Delta\tau \leq \frac{\min(\Delta\xi_1, \Delta\xi_2)}{\mathbf{a}} \cdot \lambda_{\text{CFL}}. \quad (5.6)$$

We found that $\lambda_{\text{CFL}} = 0.2$ is a suitable choice. Note that in [Kurganov & Tadmor \(2000\)](#), there are bounds for λ_{CFL} given, which guarantee a maximum principle.

Plots, showing the time evolution of the solution to the freezing method, calculated with our scheme can be found in [Rottmann-Matthes \(2016\)](#), and we do not repeat them here, as we focus on the discretization and the error introduced by the discretization in this article. In all our experiments, we solve the Cauchy problem for (1.4) with initial data u_0 , given by

$$u_0(x, y) = \begin{cases} \cos(y) \cdot \sin(2x), & -\frac{\pi}{2} < y < \frac{\pi}{2}, -\frac{\pi}{2} < x < 0, \\ \cos(y) \cdot \sin(x), & -\frac{\pi}{2} < y < \frac{\pi}{2}, 0 < x < \pi, \\ 0 & \text{otherwise,} \end{cases}$$

depicted in Fig. 4(a), by the freezing method (5.2). We choose the fixed domain $\Omega = [-5, 5]^2$ with no-flux boundary conditions.

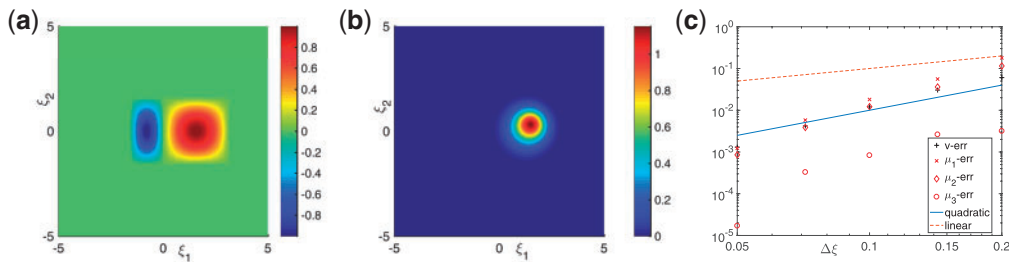


FIG. 4. The final state of the reference solution $v_{\text{ref}}(6)$ is shown in (b), it is obtained by the freezing method (5.2) with $\nu = 0.4$ and initial function shown in (a). In (c) we show the relative errors of solutions obtained on coarser grids compared with the solution on the fine grid.

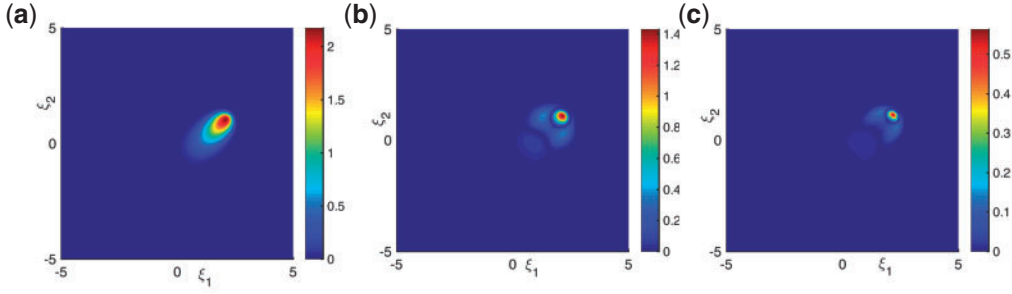


FIG. 5. Comparison of the solutions obtained with the freezing method at $\tau = 300$ for $\nu = 0.05$ and different cell sizes 0.2×0.2 vs. 0.05×0.05 in (b) and 0.1×0.1 vs. 0.05×0.05 in (c).

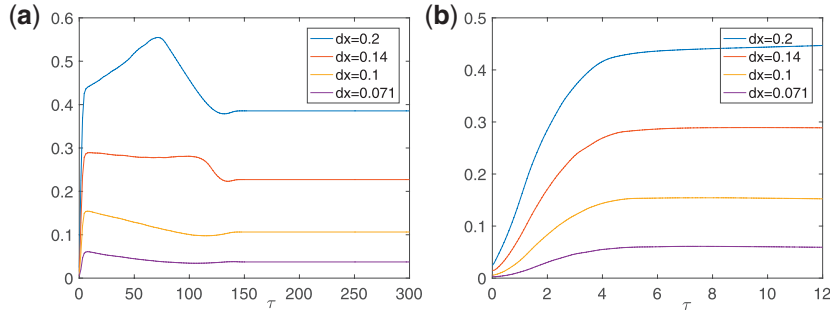


FIG. 6. The relative L^2 -error $\frac{1}{\|v_{\text{ref}}(\tau)\|} \|v_{\text{ref}}(\tau) - v(\tau)\|$ for different step sizes $\Delta\xi_1 = \Delta\xi_2 = dx$ drawn as a function of time.

In our first experiment we choose viscosity $\nu = 0.4$ and solve the freezing system until $\tau = 6$, which is a time when the solution has settled. The final state of this calculation obtained for the fine grid with $\Delta\xi = \Delta\xi_1 = \Delta\xi_2 = \frac{1}{28}$ is shown in Fig. 4(b). We choose this solution as the reference solution. In Fig. 4(c), we plot the difference of the solution components for different step sizes. More precisely we plot the relative L^2 -difference $\frac{1}{\|v_{\text{ref}}\|} \|v_{\Delta\xi} - v_{\text{ref}}\|$ ('v-err') at time $\tau = 6$, where we reconstruct $v_{\Delta\xi}(\cdot, 6)$ and $v_{\text{ref}}(\cdot, 6)$ by linear interpolation as in the one-dimensional case; and we also plot the relative differences of the algebraic variables $\frac{1}{\|\mu_{j,\text{ref}}\|} |\mu_{j,\Delta\xi} - \mu_{j,\text{ref}}|$ (' μ_j error') at the same time $\tau = 6$. The numerical findings suggest that the scheme is in fact second-order convergent (in $\Delta\xi$). We repeat the experiment with the very small viscosity $\nu = 0.05$. The solution to the two-dimensional Burgers's equation (5.5) with this viscosity and the same initial data as before shows a metastable behavior similar to the one-dimensional case. Therefore, we have to calculate for a very long time, and we show the final state of the calculation at time $\tau = 300$ for $\Delta\xi_1 = \Delta\xi_2 = 0.05$ in Fig. 5(a) and plot in Fig. 5(b), (c), the deviation of solutions obtained for coarser grids with $\Delta\xi_1 = \Delta\xi_2 = 0.2$ and $\Delta\xi_1 = \Delta\xi_2 = 0.1$ from the reference solution.

In Fig. 6 and 7, we again consider the solution to the freezing method for (5.5) with $\nu = 0.05$. We choose the solution with step size $\Delta\xi_1 = \Delta\xi_2 = 0.05$ as reference solution and plot, for different step sizes $\Delta\xi_1 = \Delta\xi_2 = dx$, how the error of the solution components depends on time. For all solution components the result is similar: one observes that the error initially rapidly increases, as expected since the error accumulates over time. But after a short time ($\tau \approx 6$) it settles and even slowly decreases later on ($\tau \approx 100$) until it finally reaches a stationary value ($\tau \approx 150$).

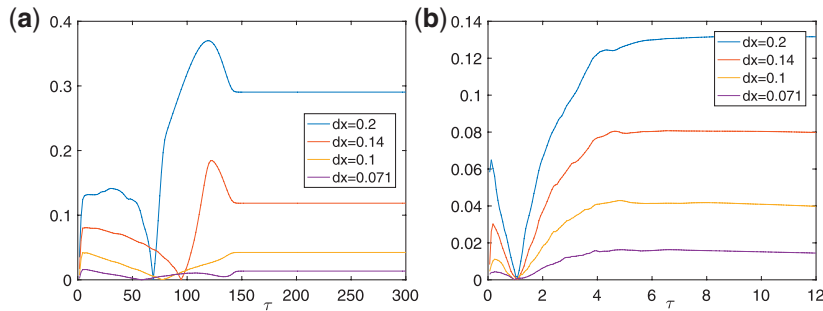


FIG. 7. The relative error $\frac{|\mu_{1,\text{ref}}(\tau) - \mu_1(\tau)|}{|\mu_{1,\text{ref}}(\tau)|}$ of the algebraic variable corresponding to the speed of the scaling μ_1 for different step sizes $\Delta\xi_1 = \Delta\xi_2 = dx$ drawn as a function of time.

Violation of the CFL condition. We also observed that oscillations appear when we violate (5.6). We found that oscillations for example appear for the same problem with $\nu = 0.02$ and $\lambda_{\text{CFL}} = 0.8$. But we do not present the results here.

Acknowledgements

We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) through CRC 1173. The author would like to thank Wolf-Jürgen Beyn for many helpful discussions and encouraging him to see this project through to the finish.

REFERENCES

- ASCHER, U. M., RUUTH, S. J. & SPITERI, R. J. (1997) Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Appl. Numer. Math.*, **25**, 151–167. Special issue on time integration (Amsterdam, 1996).
- BEC, J. & KHANIN, K. (2007) Burgers turbulence. *Phys. Rep.*, **447**, 1–66.
- BEYN, W.-J., OTTEN, D. & ROTTMANN-MATTHES, J. (2014) Stability and computation of dynamic patterns in PDEs. *Current Challenges in Stability Issues for Numerical Differential Equations* (L. Dieci & N. Guglielmi eds). Lecture Notes in Mathematics, vol. 2082. Cham: Springer, pp. 89–172.
- BEYN, W.-J., OTTEN, D. & ROTTMANN-MATTHES, J. (2016) *Computation and Stability of Traveling Waves in Second Order Evolution Equations*. Preprint 2016/15. Karlsruhe: Karlsruhe Institute of Technology, CRC 1173.
- BEYN, W.-J. & THÜMMLER, V. (2004) Freezing solutions of equivariant evolution equations. *SIAM J. Appl. Dyn. Syst.*, **3**, 85–116.
- BURGERS, J. M. (1948) A mathematical model illustrating the theory of turbulence. *Adv. Appl. Mech.*, **1**, 171–199.
- HAIRER, E., LUBICH, C. & ROCHE, M. (1989) *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*. Lecture Notes in Mathematics, vol. 1409. Berlin: Springer, pp. viii+139.
- HAIRER, E. & WANNER, G. (1996) *Solving Ordinary Differential Equations. II: Stiff and Differential-Algebraic Problems*, 2nd rev. edn. Berlin: Springer, pp. xvi + 614.
- HASTINGS, S. (1976) On travelling wave solutions of the Hodgkin-Huxley equations. *Arch. Ration. Mech. Anal.*, **60**, 229–257.
- HODGKIN, A. L. & HUXLEY, A. F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, **117**, 500–544.
- HU, C. & SHU, C.-W. (1999) Weighted essentially non-oscillatory schemes on triangular meshes. *J. Comput. Phys.*, **150**, 97–127.

- KURGANOV, A. & TADMOR, E. (2000) New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations. *J. Comput. Phys.*, **160**, 241–282.
- MARTINSON, W. S. & BARTON, P. I. (2000) A differentiation index for partial differential-algebraic equations. *SIAM J. Sci. Comput.*, **21**, 2295–2315.
- ROTTMANN-MATTHES, J. (2012a) Stability and freezing of nonlinear waves in first order hyperbolic PDEs. *J. Dynam. Differential Equations*, **24**, 341–367.
- ROTTMANN-MATTHES, J. (2012b) Stability and freezing of waves in non-linear hyperbolic-parabolic systems. *IMA J. Appl. Math.*, **77**, 420–429.
- ROTTMANN-MATTHES, J. (2016) Freezing similarity solutions in multi-dimensional Burgers' equation. CRC 1173-Preprint 2016/27. Karlsruhe Institute of Technology.
- ROWLEY, C. W., KEVREKIDIS, I. G., MARSDEN, J. E. & LUST, K. (2003) Reduction and reconstruction for self-similar dynamical systems. *Nonlinearity*, **16**, 1257–1275.
- SHU, C.-W. (1988) Total-variation-diminishing time discretizations. *SIAM J. Sci. Statist. Comput.*, **9**, 1073–1084.
- SHU, C.-W. & OSHER, S. (1988) Efficient implementation of essentially nonoscillatory shock-capturing schemes. *J. Comput. Phys.*, **77**, 439–471.
- THÜMLER, V. (2008) The effect of freezing and discretization to the asymptotic stability of relative equilibria. *J. Dynam. Differential Equations*, **20**, 425–477.

Appendix: Discretization of the hyperbolic part

For the discretization of the hyperbolic part in (2.11a), we briefly review and adapt a scheme introduced in Kurganov & Tadmor (2000) (KT-scheme). In particular, we need an adaptation for nonhomogeneous flux functions, i.e., the case of hyperbolic conservation laws of the form

$$u_\tau + \frac{d}{d\xi} f(\xi, u) = 0.$$

In this appendix u may be \mathbb{R}^m valued, but in the actual application in Section 3, it is always a scalar. For the KT-scheme, a uniform grid $\xi_j = \xi_0 + j\Delta\xi$, $j \in \mathbb{Z}$ in \mathbb{R} is chosen and one assumes that at a time instance τ^n an approximation of the average mass of the function u in the spatial cells $(\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}})$, $\xi_{j\pm\frac{1}{2}} = \xi_j \pm \frac{\Delta\xi}{2}$, $j \in \mathbb{Z}$ is given as $u_j^n \in \mathbb{R}^m$, i.e., $u_j^n \approx \frac{1}{\Delta\xi} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} u(\xi, t^n) d\xi$. From this grid function, one obtains the piecewise linear reconstruction at the time instance τ^n as

$$\tilde{u}(\xi, t^n) = \sum_j \left(u_j^n \mathbf{1}_{\left[\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}}\right)}(\xi) + (u_\xi)_j^n (\xi - \xi_j) \mathbf{1}_{\left[\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}}\right)}(\xi) \right),$$

where $(u_\xi)_j^n$ is a suitable choice for the slope of this piecewise linear function in the cell $(\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}})$. These 'suitable' slopes are given by the 'minmod' reconstruction, i.e., for a fixed $\theta \in [1, 2]$ the slopes are given as

$$(u_\xi)_j^n = \text{mm} \left(\theta \frac{u_j - u_{j-1}}{\Delta\xi}, \frac{u_{j+1} - u_{j-1}}{2\Delta\xi}, \theta \frac{u_{j+1} - u_j}{\Delta\xi} \right),$$

where

$$\text{mm}(w_1, \dots, w_k) = \max(\min(w_1, 0), \dots, \min(w_k, 0)) + \min(\max(w_1, 0), \dots, \max(w_k, 0)). \quad (\text{A.1})$$

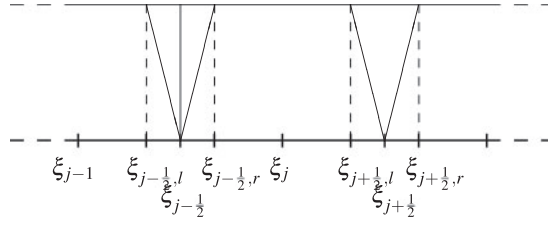


FIG. A.1. Sketch of the grid and smooth and nonsmooth parts of the solution.

In the vector-valued case, (A.1) is understood in the componentwise sense. The idea of the KT-scheme is to use an artificial finer grid, which depends on the time-step size $\Delta\tau$ and separates the smooth and nonsmooth regions of the solution. In the original article [Kurganov & Tadmor \(2000\)](#), local (maximal) wave speeds are calculated and a spatial grid based on these is chosen. Since we ultimately intend to apply the scheme to the PDAE system (2.11), for which the wave speeds nonlinearly depend on the solution, we choose a sufficiently large $\mathbf{a} > 0$, which bounds the spectral radius of $\frac{\partial}{\partial u} f(\xi, u)$. Note that this is obviously not possible for $f(\xi, u) = \xi u$, $\xi \in \mathbb{R}$, but because we will focus on compact domains in the end, we assume that the spectral radius is uniformly bounded in ξ for bounded u .

The new grid is then given by

$$\cdots < \xi_{j-1} < \xi_{j-\frac{1}{2},l} = \xi_{j-\frac{1}{2}} - \mathbf{a}\Delta\tau < \xi_{j-\frac{1}{2},r} = \xi_{j-\frac{1}{2}} + \mathbf{a}\Delta\tau < \xi_j < \cdots$$

for $\Delta\tau$ sufficiently small. We remark that for $\Delta\tau \searrow 0$, the new grid points $\xi_{j-\frac{1}{2},l}$ and $\xi_{j-\frac{1}{2},r}$ both converge to $\xi_{j-\frac{1}{2}}$, the first from the left and the second from the right. The principal idea now is to use the integral form of the conservation law in the smooth and nonsmooth regions for the calculation of the mass at the new time instance τ^{n+1} . We begin with the nonsmooth part:

$$\begin{aligned} & \frac{1}{\Delta\xi_{\frac{1}{2}}} \int_{\xi_{j-\frac{1}{2},l}}^{\xi_{j-\frac{1}{2},r}} u(\xi, \tau^{n+1}) \, d\xi \\ &= \frac{1}{2\mathbf{a}\Delta\tau} \left(\int_{\xi_{j-\frac{1}{2},l}}^{\xi_{j-\frac{1}{2},r}} \tilde{u}(\xi, \tau^n) \, d\xi - \int_{\tau^n}^{\tau^{n+1}} f\left(\xi_{j-\frac{1}{2},r}, u\left(\xi_{j-\frac{1}{2},r}, \tau\right)\right) \, d\tau \right. \\ & \quad \left. + \int_{\tau^n}^{\tau^{n+1}} f\left(\xi_{j-\frac{1}{2},l}, u\left(\xi_{j-\frac{1}{2},l}, \tau\right)\right) \, d\tau \right), \end{aligned}$$

where $\Delta\xi_{\frac{1}{2}} = 2\mathbf{a}\Delta\tau = \xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2},l}$. The time integrals are in the smooth regions of the solution (the dashed vertical lines in Fig. A.1). We approximate them by the midpoint rule, so that we need the value of the solution u at $\xi_{j-\frac{1}{2},l}$, $\xi_{j-\frac{1}{2},r}$ and time instance $\tau^n + \frac{\Delta\tau}{2}$. In a smooth region this value of u is approximately given by

$$\begin{aligned} u\left(\xi_{j-\frac{1}{2},l}, \tau^{n+\frac{1}{2}}\right) &\approx \tilde{u}\left(\xi_{j-\frac{1}{2},l}, \tau^n\right) - \frac{\Delta\tau}{2} \left(\frac{\partial}{\partial \xi} f\left(\xi_{j-\frac{1}{2},l}, u_{j-\frac{1}{2},l}^n\right) + \frac{\partial}{\partial u} f\left(\xi_{j-\frac{1}{2},l}, u_{j-\frac{1}{2},l}^n\right) (u_\xi)_j^n \right) \\ &:= u_{j-\frac{1}{2},l}^{n+\frac{1}{2}}, \end{aligned}$$

and the same for l replaced by r . For the average value of the solution u at the new time instance τ^{n+1} in the spatial interval $(\xi_{j-\frac{1}{2},l}, \xi_{j+\frac{1}{2},r})$ this yields the approximation

$$\begin{aligned} w_{j-\frac{1}{2}}^{n+1} &:= \frac{u_{j-1}^n + u_j^n}{2} + \frac{\Delta\xi - \mathbf{a}\Delta\tau}{4} ((u_\xi)_j^n - (u_\xi)_j^n) - \frac{1}{2\mathbf{a}} \left(f\left(\xi_{j-\frac{1}{2},r}, u_{j-\frac{1}{2},r}^{n+\frac{1}{2}}\right) - f\left(\xi_{j-\frac{1}{2},l}, u_{j-\frac{1}{2},l}^{n+\frac{1}{2}}\right) \right) \\ &\approx \frac{1}{\Delta\xi_{\frac{1}{2}}} \int_{\xi_{j-\frac{1}{2},l}}^{\xi_{j-\frac{1}{2},r}} u(\xi, \tau^{n+1}) d\xi. \end{aligned} \quad (\text{A.2})$$

Similarly, the average of the solution at the new time instance τ^{n+1} in the spatial interval $(\xi_{j-\frac{1}{2},r}, \xi_{j+\frac{1}{2},l})$ is approximated by

$$\begin{aligned} w_j^{n+1} &:= u_j^n - \frac{\Delta\tau}{\Delta\xi - 2\mathbf{a}\Delta\tau} \left(f\left(\xi_{j+\frac{1}{2},l}, u_{j+\frac{1}{2},l}^{n+\frac{1}{2}}\right) - f\left(\xi_{j-\frac{1}{2},r}, u_{j-\frac{1}{2},r}^{n+\frac{1}{2}}\right) \right) \\ &\approx \frac{1}{\Delta\xi - \Delta\xi_{\frac{1}{2}}} \int_{\xi_{j-\frac{1}{2},r}}^{\xi_{j+\frac{1}{2},l}} u(\xi, \tau^{n+1}) d\xi. \end{aligned} \quad (\text{A.3})$$

To come back from the new grid to the original one, a piecewise linear reconstruction \tilde{w} of the grid function w^{n+1} is calculated and then u_j^{n+1} is obtained by integration over the cells $(\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}})$. Finally, one obtains the fully discrete scheme:

$$\begin{aligned} u_j^{n+1} &= \frac{1}{\Delta\xi} \left[\left(\xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2}} \right) w_{j-\frac{1}{2}}^{n+1} + (w_\xi)_{j-\frac{1}{2}}^{n+1} \frac{\left(\xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2}} \right)^2}{2} \right. \\ &\quad \left. + \left(\xi_{j+\frac{1}{2},l} - \xi_{j-\frac{1}{2},r} \right) w_j^{n+1} \right. \\ &\quad \left. + \left(\xi_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2},l} \right) w_{j+\frac{1}{2}}^{n+1} - (w_\xi)_{j+\frac{1}{2}}^{n+1} \frac{\left(\xi_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2},l} \right)^2}{2} \right], \end{aligned} \quad (\text{A.4})$$

where $w_{j\pm\frac{1}{2}}^{n+1}$ and w_j^{n+1} are given by (A.2) and (A.3) and

$$(w_\xi)_{j-\frac{1}{2}}^{n+1} = \frac{2}{\Delta\xi} \text{mm} \left(w_j^{n+1} - w_{j-\frac{1}{2}}^{n+1}, w_{j-\frac{1}{2}}^{n+1} - w_{j-1}^{n+1} \right).$$

The fully discrete scheme (A.4) admits a semidiscrete version, i.e., an MOL system, which we will use in our discretization of (2.11). To see this, let $\lambda = \Delta\tau/\Delta\xi$, in which we consider $\Delta\xi$ as a fixed quantity and are interested in the limit $\Delta\tau \rightarrow 0$, so that $\mathcal{O}(\lambda^2) = \mathcal{O}(\Delta\tau^2)$. We note that the summands in (A.4)

can be written in the following form, where we use $\xi_{j-\frac{1}{2}} - \xi_{j-\frac{1}{2},l} = \xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2}} = \mathbf{a}\Delta\tau$:

$$\begin{aligned} \left(\xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2}}\right) w_{j-\frac{1}{2}}^{n+1} &= \frac{\mathbf{a}\Delta\tau}{2} \left(u_{j-1}^n + \frac{\Delta\xi}{2} (u_\xi)_{j-1}^n + u_j^n - \frac{\Delta\xi}{2} (u_\xi)_j^n \right) \\ &\quad - \frac{\Delta\tau}{2} \left(f\left(\xi_{j-\frac{1}{2},r}, u_{j-\frac{1}{2},r}^{n+\frac{1}{2}}\right) - f\left(\xi_{j-\frac{1}{2},l}, u_{j-\frac{1}{2},l}^{n+\frac{1}{2}}\right) \right) + \mathcal{O}(\lambda^2), \end{aligned} \quad (\text{A.5})$$

$$(w_\xi)_{j-\frac{1}{2}}^{n+1} \frac{(\xi_{j-\frac{1}{2},r} - \xi_{j-\frac{1}{2}})^2}{2} = \mathcal{O}(\lambda^2), \quad (\text{A.6})$$

$$(\xi_{j+\frac{1}{2},l} - \xi_{j-\frac{1}{2},r}) w_j^{n+1} = \Delta\xi u_j^n - 2\mathbf{a}\Delta\tau u_j^n - \Delta\tau \left(f\left(\xi_{j+\frac{1}{2},l}, u_{j+\frac{1}{2},l}^{n+\frac{1}{2}}\right) - f\left(\xi_{j-\frac{1}{2},r}, u_{j-\frac{1}{2},r}^{n+\frac{1}{2}}\right) \right), \quad (\text{A.7})$$

$$\begin{aligned} \left(\xi_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2},l}\right) w_{j+\frac{1}{2}}^{n+1} &= \frac{\mathbf{a}\Delta\tau}{2} \left(u_j^n + \frac{\Delta\xi}{2} (u_\xi)_j^n + u_{j+1}^n - \frac{\Delta\xi}{2} (u_\xi)_{j+1}^n \right) \\ &\quad - \frac{\Delta\tau}{2} \left(f\left(\xi_{j+\frac{1}{2},r}, u_{j+\frac{1}{2},r}^{n+\frac{1}{2}}\right) - f\left(\xi_{j+\frac{1}{2},l}, u_{j+\frac{1}{2},l}^{n+\frac{1}{2}}\right) \right) + \mathcal{O}(\lambda^2), \end{aligned} \quad (\text{A.8})$$

$$(w_\xi)_{j+\frac{1}{2}}^{n+1} \frac{(\xi_{j+\frac{1}{2}} - \xi_{j+\frac{1}{2},l})^2}{2} = \mathcal{O}(\lambda^2). \quad (\text{A.9})$$

The semidiscrete version is then obtained by subtracting u_j^n from both sides of (A.4), dividing by $\Delta\tau$ and considering the limit $\Delta\tau \rightarrow 0$. Using (A.5)–(A.9) and the convergences $\xi_{j\pm\frac{1}{2},l/r} \rightarrow \xi_{j\pm\frac{1}{2}}$, $u_{j-\frac{1}{2},l}^{n+\frac{1}{2}} \rightarrow u_{j-\frac{1}{2}}^{n,-}$, $u_{j-\frac{1}{2}}^{n,-} = u_{j-1}^n + \frac{\Delta\xi}{2} (u_\xi)_{j-1}^n$ and $u_{j-\frac{1}{2},r}^{n+\frac{1}{2}} \rightarrow u_{j-\frac{1}{2}}^{n,+}$, $u_{j-\frac{1}{2}}^{n,+} = u_j^n - \frac{\Delta\xi}{2} (u_\xi)_j^n$ as $\Delta\tau \rightarrow 0$, the continuity properties of f imply that the limit of the difference quotient exists. The final result is the following semidiscrete scheme (MOL system):

$$\begin{aligned} \frac{d}{d\tau} u_j &= \frac{1}{2\Delta\xi} \left(\left(f\left(\xi_{j-\frac{1}{2}}, u_{j-\frac{1}{2}}^-\right) + f\left(\xi_{j-\frac{1}{2}}, u_{j-\frac{1}{2}}^+\right) \right) - \left(f\left(\xi_{j+\frac{1}{2}}, u_{j+\frac{1}{2}}^-\right) + f\left(\xi_{j+\frac{1}{2}}, u_{j+\frac{1}{2}}^+\right) \right) \right) \\ &\quad + \frac{\mathbf{a}}{2\Delta\xi} \left(u_{j-\frac{1}{2}}^- - u_{j-\frac{1}{2}}^+ - u_{j+\frac{1}{2}}^- + u_{j+\frac{1}{2}}^+ \right), \quad j \in \mathbb{Z}, \end{aligned} \quad (\text{A.10})$$

where

$$\begin{aligned} u_{j-\frac{1}{2}}^- &= u_{j-1} + \frac{1}{2} \text{mm} \left(\theta(u_{j-1} - u_{j-2}), \frac{u_j - u_{j-2}}{2}, \theta(u_j - u_{j-1}) \right), \\ u_{j-\frac{1}{2}}^+ &= u_j - \frac{1}{2} \text{mm} \left(\theta(u_j - u_{j-1}), \frac{u_{j+1} - u_{j-1}}{2}, \theta(u_{j+1} - u_j) \right). \end{aligned} \quad (\text{A.11})$$

Before we continue, we make some simple remarks.

REMARK A1

- (i) In the case $f(\xi, u) = f(u)$, (A.10) reduces to the semidiscrete version of the KT-scheme from Kurganov & Tadmor (2000).
- (ii) In the case of a piecewise constant reconstruction of u , the last summand in (A.10) reduces to $\frac{\mathbf{a}}{2\Delta\xi}(u_{j-1} - 2u_j + u_{j+1})$ and can be interpreted as artificial viscosity that vanishes as $\Delta\xi \rightarrow 0$.
- (iii) The flux function f enters linearly into the semidiscrete scheme. This is an important observation, because it allows us to derive an MOL system for the PDE-part of the PDAE (2.11), without knowing the algebraic variables μ in advance. To the resulting DAE system, we may apply a so-called half-explicit scheme (see Hairer *et al.*, 1989).
- (iv) The MOL system (A.10) can be written in conservative form as

$$u'_j = -\frac{H_{j+\frac{1}{2}} - H_{j-\frac{1}{2}}}{\Delta\xi},$$

where

$$H_{j+\frac{1}{2}} = \frac{f\left(\xi_{j+\frac{1}{2}}, u_{j+\frac{1}{2}}^+\right) + f\left(\xi_{j+\frac{1}{2}}, u_{j+\frac{1}{2}}^-\right)}{2} - \mathbf{a}\left(u_{j+\frac{1}{2}}^+ - u_{j+\frac{1}{2}}^-\right).$$

Multidimensional version

There is no difficulty in performing the same calculation in the multidimensional setting. We present only the result for the two-dimensional case: for this purpose, we consider a system of hyperbolic conservation laws of the form

$$u_\tau + \frac{\mathrm{d}}{\mathrm{d}\xi}f(\xi, \eta, u) + \frac{\mathrm{d}}{\mathrm{d}\eta}g(\xi, \eta, u) = 0.$$

Choose a uniform spatial grid $(\xi_j, \eta_k) = (\xi_0 + j\Delta\xi, \eta_0 + k\Delta\eta)$ and assume that a grid function $(u_{j,k})$ is given on this grid, where $u_{j,k}(\tau)$ denotes the average mass in the cell $\left(\xi_{j-\frac{1}{2}}, \xi_{j+\frac{1}{2}}\right) \times \left(\eta_{k-\frac{1}{2}}, \eta_{k+\frac{1}{2}}\right)$, $\xi_{j\pm\frac{1}{2}} = \xi_j \pm \frac{\Delta\xi}{2}$ and similarly $\eta_{k\pm\frac{1}{2}} = \eta_k \pm \frac{\Delta\eta}{2}$ at time τ . Adapting the above derivation, we obtain the following two-dimensional version of the semidiscrete KT-scheme:

$$\begin{aligned} u'_{j,k} = & \frac{1}{2\Delta\xi} \left[\left(f\left(\xi_{j-\frac{1}{2}}, \eta_k, u_{j-\frac{1}{2},k}^-\right) + f\left(\xi_{j-\frac{1}{2}}, \eta_k, u_{j-\frac{1}{2},k}^+\right) \right) - \left(f\left(\xi_{j+\frac{1}{2}}, \eta_k, u_{j+\frac{1}{2},k}^-\right) + f\left(\xi_{j+\frac{1}{2}}, \eta_k, u_{j+\frac{1}{2},k}^+\right) \right) \right] \\ & + \frac{1}{2\Delta\eta} \left[\left(g\left(\xi_j, \eta_{k-\frac{1}{2}}, u_{j,k-\frac{1}{2}}^-\right) + g\left(\xi_j, \eta_{k-\frac{1}{2}}, u_{j,k-\frac{1}{2}}^+\right) \right) - \left(g\left(\xi_j, \eta_{k+\frac{1}{2}}, u_{j,k+\frac{1}{2}}^-\right) + g\left(\xi_j, \eta_{k+\frac{1}{2}}, u_{j,k+\frac{1}{2}}^+\right) \right) \right] \\ & + \frac{\mathbf{a}}{2\Delta\xi} \left(u_{j-\frac{1}{2},k}^- - u_{j-\frac{1}{2},k}^+ - u_{j+\frac{1}{2},k}^- + u_{j+\frac{1}{2},k}^+ \right) \\ & + \frac{\mathbf{a}}{2\Delta\eta} \left(u_{j,k-\frac{1}{2}}^- - u_{j,k-\frac{1}{2}}^+ - u_{j,k+\frac{1}{2}}^- + u_{j,k+\frac{1}{2}}^+ \right). \end{aligned} \tag{A.12}$$

In (A.12), the number \mathbf{a} is again assumed to be an upper bound for the maximum of the spectral radii of $\frac{\partial}{\partial u}f(\xi, \eta, u)$ and $\frac{\partial}{\partial u}g(\xi, \eta, u)$ for all relevant ξ, η, u . We assume that such an upper bound exists.

The one-sided limits appearing in (A.12) are given by

$$\begin{aligned}
 u_{j-\frac{1}{2},k}^- &= u_{j-1,k} + \frac{1}{2} \operatorname{mm} \left(\theta(u_{j-1,k} - u_{j-2,k}), \frac{u_{j,k} - u_{j-2,k}}{2}, \theta(u_{j,k} - u_{j-1,k}) \right), \\
 u_{j-\frac{1}{2},k}^+ &= u_{j,k} - \frac{1}{2} \operatorname{mm} \left(\theta(u_{j,k} - u_{j-1,k}), \frac{u_{j+1,k} - u_{j-1,k}}{2}, \theta(u_{j+1,k} - u_{j,k}) \right), \\
 u_{j,k-\frac{1}{2}}^- &= u_{j,k-1} + \frac{1}{2} \operatorname{mm} \left(\theta(u_{j,k-1} - u_{j,k-2}), \frac{u_{j,k} - u_{j,k-2}}{2}, \theta(u_{j,k} - u_{j,k-1}) \right), \\
 u_{j,k-\frac{1}{2}}^+ &= u_{j,k} - \frac{1}{2} \operatorname{mm} \left(\theta(u_{j,k} - u_{j,k-1}), \frac{u_{j,k+1} - u_{j,k-1}}{2}, \theta(u_{j,k+1} - u_{j,k}) \right).
 \end{aligned} \tag{A.13}$$