On Finite Element Method–Flux Corrected Transport Stabilization for Advection-Diffusion Problems in a Partial Differential-Algebraic Framework

Julia Niemeyer[a,*], Bernd Simeon[a]

[a] *Felix-Klein Centre for Mathematics, TU Kaiserslautern, Paul-Ehrlich Straße 31, 67663 Kaiserslautern*

Abstract

An extension of the finite element method–flux corrected transport stabilization (FEM-FCT) for hyperbolic problems in the context of partial differential-algebraic equations (PDAEs) is proposed. Given a local extremum diminishing property of the spatial discretization, the positivity preservation of the one-step $\theta-$scheme when applied to the time integration of the resulting differential-algebraic equation (DAE) is shown, under a mild restriction on the time step-size. As crucial tool in the analysis, the Drazin inverse and the corresponding Drazin ODE are explicitly derived. Numerical results are presented for non-constant and time-dependent boundary conditions in one space dimension and for a two-dimensional advection problem where the advection proceeds skew to the mesh.

Keywords: partial differential-algebraic equations, positivity preserving time integration, FEM-FCT stabilization

1. Introduction

In this paper, we are concerned with advection-dominated flow problems and study appropriate stabilization techniques that suppress unphysical oscillations and guarantee positivity preservation. Taking the recently developed Finite Element Method – Flux Corrected Transport Scheme (FEM-FCT) [1, 2] as starting point, we focus on a partial differential-algebraic framework that allows a general coupling procedure where boundary conditions are expressed as constraints and appended by means of Lagrange multipliers. The stabilization approach that we provide makes it possible to model and stabilize a flow problem as part of a more general coupled system that may stem from fluid-structure interaction, from domain decomposition or from other multiphysics models, see, e.g., [3, 4, 5, 6, 7, 8] for such applications.

It is well-known that the standard Galerkin finite element method applied to hyperbolic problems tends to produce unphysical oscillations, and stabilization is a big task in the simulation. Today, there exist many different stabilization

*Corresponding author: Phone: +49 631 205-5316

Email addresses: niemeyer@mathematik.uni-kl.de (Julia Niemeyer), simeon@mathematik.uni-kl.de (Bernd Simeon)

techniques, with the *Streamline Upwind Petrov Galerkin* (SUPG) method [9, 10] being one of earliest. In the case of time-dependent problems it has been observed for a long time that SUPG produces spurious oscillations. To deal with this issue, the class of *Spurious Oscillations at Layer Diminishing* (SOLD) schemes has been developed [11, 12]. Another stabilization technique are the so-called *Local Projection Stabilization* (LPS) methods [13, 14, 15].

The idea behind all these so far mentioned stabilization techniques is the modification of the Galerkin finite element discretization by adding additional terms to the underlying bilinear form. A different route is taken by the FEM-FCT scheme that operates directly on the system matrices derived from the standard Galerkin finite element method. This stabilization constructs first a very diffusive but non-oscillating and positivity preserving scheme and then, in a second step, corrects the diffusive low order solution by a limiting post processing based on antidiffusive terms that are computed from the standard high order solution [2, 1, 16, 17]. The FEM-FCT scheme has been shown to produce competitive results in comparison with the other techniques [18]. An important advantage of the FEM-FCT technique lies in the fact that it can be implemented in a straightforward way into an already existing finite element code.

In order to extend the FEM-FCT scheme to a partial differential-algebraic setting, we need to analyze the positivity preservation of the underlying time integration process for a given spatial discretization, cf. [19, 20], and to adjust the stabilization procedure in case of constraint terms. While the second topic turns out to be straightforward, the question of positivity preservation leads to the study of the Drazin inverse and a corresponding ordinary differential equation that we call the *Drazin ODE*. We remark that this approach has been investigated for general unstructured DAEs in [21], but here we are able to provide an explicit representation of the Drazin ODE that reflects the structure of the corresponding DAE and that enables us to prove positivity preservation for the class of one-step θ-methods, provided that a *Local Extremum Diminishing* (LED) property of the semi-discretized equations holds. As an extra benefit, our approach sheds also new light on the treatment of the boundary conditions in the standard FEM-FCT scheme since in the PDAE framework with the boundary conditions appended, the boundary nodes do not require any extra processing.

The paper is organized as follows. Taking the linear advection-diffusion equation as model problem, we discuss the positivity preservation in a partial differential-algebraic framework in Section 2. The FEM-FCT procedure is recalled and modified in Section 3. Finally, Section 4 is devoted to numerical convergence studies.

2. Positivity Preserving Schemes

Let $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$ be a bounded domain with boundary $\partial\Omega =: \Gamma$ and $\mathbb{T} = [t_0, t_m] \subset \mathbb{R}_0^+$ a given time interval. We consider the linear advection-diffusion equation, written as generic conservation law

$$\dot{u} + \nabla \cdot (\beta u) = \nabla \cdot (\kappa \nabla u) + f \quad \text{in } \Omega \times \mathbb{T} \tag{1a}$$

with Dirichlet boundary data

$$u = \bar{u} \qquad \text{on } \Gamma \times \mathbb{T} \tag{1b}$$

and initial conditions

$$u = u_0 \qquad \text{in } \Omega \times \{t_0\}. \tag{1c}$$

The scalar diffusion coefficient is denoted by κ while $\beta \in \mathbb{R}^d$ stands for the advection velocity. In the following, we restrict our discussion to constant parameters β and κ and vanishing sink / source term f.

For $|\beta| >> \kappa$, the problem is dominated by the advection and constitutes a hyperbolic partial differential equation with an additional regularization in form of the diffusion term. For such problems, the finite element method produces unphysical oscillations in the spatial discretization such that it is necessary to use a suitable stabilization technique [22]. To obtain a stable numerical solution we have to avoid two incidents:

S1 Birth and growth of local extrema
S2 Global over- and undershoots

As discussed in the Introduction, there has been a huge research effort on stabilizations of the finite element method applied to hyperbolic problems. We select here the Finite Element Method - Flux Corrected Transport (FEM-FCT) scheme [2, 1, 16, 17] as starting point of our work. In short, this stabilization operates directly on the algebraic level by modifying the system matrix and the right hand side vector, and it seems to produce less wriggles than most other techniques [18]. The idea of the *FEM-FCT* scheme is to use the knowledge about the structure of the system matrix computed in the finite element discretization to construct a local extremum diminishing (LED) scheme which fulfills requirement **S1**. Integrating the resulting differential equation system using a positivity preserving time integrator leads to the satisfaction of the maximum principle, consequently requirement **S2** is fulfilled. In the following, we will first study the time integration process in a rather general context and then, in the next section, address the details of the FEM-FCT algorithm.

2.1. Local Extremum Diminishing

The LED property is a constraint on the spatial discretization. In order to explain this notion, we discretize (1) in space by the standard Galerkin finite element method. For this purpose, let $\mathcal{T}_h$ be an admissible subdivision of Ω with mesh size h, $\Omega_h = \bigcup_{T \in \mathcal{T}_h} T$. Furthermore let φ_i be a finite element basis function. Then the semi-discrete solution u_h has the form

$$u_h(x,t) = \sum_{i=1}^{n_u} u_i(t)\varphi_i(x) \tag{2}$$

where n_u is the number of degrees of freedom and u_i is the discrete solution in the node i of the mesh. In case of a nodal basis, it holds even more $u_h(x_i,t) = u_i(t)$.

To simplify the notation, we use u_h in the following both for denoting the semi-discrete solution (2) and for the vector of unknown nodal variables $u_h = (u_1, \ldots, u_{n_u})^T$. Applying the Galerkin projection to (1), we obtain, after the usual steps, the system of ordinary differential equations

$$M\dot{u}_h + Ku_h = Q \qquad \text{in } \Omega_h \times \mathbb{T}. \tag{3}$$

Here, $M = (m)_{ij} \in \mathbb{R}^{n_u \times n_u}$, $m_{ij} := \int_{\Omega_h} \varphi_i \cdot \varphi_j \, dx$, stands for the mass matrix and $K = (k)_{ij} \in \mathbb{R}^{n_u \times n_u}$, $k_{ij} := \int_{\Omega_h} \kappa \nabla\varphi_i \cdot \nabla\varphi_j + \beta\varphi_i \cdot \nabla\varphi_j \, dx$, for the stiffness

matrix. We stress that in this case, the boundary condition $u_h = \bar{u}$ on $\Gamma_h \times \mathbb{T}$ is already built into the equation system by means of suitable modifications of M, K, and the right hand side term Q.

Without loss of generality, we assume next that the mass matrix M in (3) is diagonal. One way to obtain a diagonal mass matrix is to apply a Cholesky decomposition of M, the other would be the widespread mass lumping as described later on in (28). Setting the source / sink term to zero, we then extract from (3) at every meshnode i the semi-discrete scheme

$$\dot{u}_i = \sum_j c_{ij} u_j \tag{4}$$

where the coefficients c_{ij} depend on the spatial discretization procedure, e.g., $c_{ij} = -k_{ij}/m_{ii}$. If the coefficient matrix $(c)_{ij}$ has zero row sum, $\sum_j c_{ij} = 0$, the semi-discrete equation (4) can be decomposed into fluxes in the following way

$$\dot{u}_i = \sum_j c_{ij} u_j = \sum_{j \neq i} c_{ij}(u_j - u_i). \tag{5}$$

Suppose now that all non-diagonal entries of the coefficient matrix are non-negative, $c_{ij} \geq 0, j \neq i$. If we assume that u_i is a minimum, then $u_j - u_i \geq 0$. This leads to $\dot{u}_i \geq 0$, and therefore a minimum cannot decrease and, analogously, a maximum cannot increase. If the coefficient matrix is additionally sparse, also a local maximum cannot increase and a local minimum cannot decrease, see [1] and the references therein. A scheme which satsifies these properties is called *Local Extremum Diminishing (LED)*.

The following observation will play a role below when the FEM-FCT scheme is discussed in detail.

Remark 1. *Consider a diffusion-dominated problem* (1) *with* $\kappa \gg |\beta|$. *Then the lumped-mass Galerkin finite element method with basis functions which sum up to unity at each node fulfills the LED property for incompressible flows.*

2.2. Positivity Preserving Time Integrators

In the case of ordinary differential equations (ODEs), the positivity preservation of commonly used time stepping schemes is an intensively analyzed object [19, 20], but the transfer to DAEs is not a straightforward task. In contrast to the most general case discussed in [21], we benefit from the facts that we only want to preserve the positivity of the differential variables and that we have a detailed knowledge about the structure of the system matrices. The procedure is now the following. First we derive the corresponding Drazin ODE and show the positivity preservation for the differential variable if the one-step $\theta-$scheme is applied. Afterwards the equivalence of the discretized Drazin ODE and the original equation is briefly discussed.

Assume that the general linear advection-diffusion equation (1) is discretized in space using a finite element discretization that possesses the LED property. To enforce the discretized Dirichlet boundary conditions

$$u_h = \bar{u} \qquad \text{on } \Gamma_h \times \mathbb{T}, \tag{6}$$

we pursue a differential-algebraic approach and express (6) as constraint $Cu_h + c(t) = 0$. Observe that u_h is here, in contrast to (6), the vector of nodal variables

and that $C \in \mathbb{R}^{n_u \times n_\lambda}$ is a Boolean matrix that picks the nodes on the discrete boundary Γ_h. The linear boundary condition (6) is then coupled with the semi-discrete differential equations by means of a Lagrange multiplier $\lambda \in \mathbb{R}^{n_\lambda}$. As result, we obtain the DAE

$$\begin{aligned} M\dot{u}_h + Ku_h + C^T\lambda &= Q, && \text{(7a)} \\ 0 &= Cu_h + c(t), && \text{(7b)} \end{aligned}$$

with consistent initial data $u_h(t_0) = u_0$ and $\lambda(t_0) = \lambda_0$. While the semi-discrete system (3) includes the boundary conditions directly by corresponding modifications of the mass and stiffness matrices and extra right hand side data, the linear DAE (7) features an explicit representation of coupling information. This is of particular importance in situations where the function $\bar{u}$ stems from a second system that is solved simultaneously with the advection-diffusion problem. Such situations occur in multiphysics applications, e.g., fluid-structure interaction, and in time-dependent domain decomposition problems [23].

2.2.1. Derivation of the Drazin ODE

As discussed in [21], the key tool to analyze the linear DAE (7) with respect to positivity preservation is the Drazin inverse [24, 25]. We consider the homogenous case and write (7) as

$$A\dot{y}_h = By_h, \qquad y_h(t_0) = y_0 \tag{8}$$

with

$$y_h = (u_h, \lambda)^T,\ A = \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n\times n} \text{ and } B = \begin{pmatrix} -K & -C^T \\ C & 0 \end{pmatrix} \in \mathbb{R}^{n\times n} \tag{9}$$

where $n = n_u + n_\lambda$. We assume now that the matrix pencil $(\mu A - B), \mu \in \mathbb{C}$ is regular, i.e., $\det(\mu A - B) \neq 0$ for some $\mu \in \mathbb{C}$. Then (8) is equivalent to

$$\hat{A}\dot{y}_h = \hat{B}y_h \tag{10}$$

with

$$\hat{A} := (\mu A - B)^{-1}A, \qquad \hat{B} := A := (\mu A - B)^{-1}B = -I + \mu\hat{A}. \tag{11}$$

The reformulation of (8) with the matrix pencil in (10) is necessary to get a formulation with commutative system matrices $\hat{A}$, $\hat{B}$. Obviously, the matrix $\hat{A}$ is singular because A is singular and cannot be inverted. But we can derive an explicit representation of the solution with the help of the Drazin inverse A^D [26], which is defined by

Definition 2.1. *Let $\nu \in \mathbb{R}$ be the nilpotency index of $\hat{A} \in \mathbb{C}^{n\times n}$. Then the matrix $\hat{A}^D \in \mathbb{C}^{n\times n}$ satisfying the axioms*

$$\begin{aligned} \hat{A}\hat{A}^D &= \hat{A}^D\hat{A} && \text{(12a)} \\ \hat{A}^D\hat{A}\hat{A}^D &= \hat{A}^D && \text{(12b)} \\ \hat{A}^D\hat{A}^{\nu+1} &= \hat{A}^\nu && \text{(12c)} \end{aligned}$$

is called the Drazin inverse of $\hat{A}$.

With the help of the Drazin inverse from Definition 2.1 we can derive an analytical solution of (8), see, e.g., [27].

Lemma 2.1. *Let the matrix pencil $\mu A - B$ be regular. Then the analytical solution of the differential-algebraic equation is given by*

$$y_h(t) = \exp(\hat{A}^D \hat{B} t)\hat{A}^D \hat{A} y_0. \tag{13}$$

From the explicit representation (13) we can derive an ODE system

$$\dot{y}_h = \hat{A}^D \hat{B} y_h \tag{14}$$

with consistent initial condition

$$y_{h,0} = \hat{A}^D \hat{A} y_0 \tag{15}$$

that we call the *Drazin ODE*. Taking the same approach as in [28] to determine the Drazin ODE, we show the following result.

Theorem 2.1. *The system matrix of the Drazin ODE* (14) *is given by*

$$\hat{A}^D \hat{B} = S := \begin{pmatrix} -P_M M^{-1} K & 0 \\ H_M^T K P_M M^{-1} K & 0 \end{pmatrix} \tag{16a}$$

and the projector by

$$\hat{A}^D \hat{A} = Q := \begin{pmatrix} I & 0 \\ -H_M^T K & 0 \end{pmatrix}. \tag{16b}$$

Proof. To proof Theorem 2.1, we need to show the conditions of Definition 2.1. First, observe that it holds

$$\hat{A}^D = \mu \hat{A}^D \hat{A} - \hat{A}^D \hat{B}. \tag{17}$$

To keep the notation as short as possible we introduce the notations

$$H_M = M^{-1} C^T (C M^{-1} C^T) \tag{18}$$

and for the projection from $\mathbb{R}^{n_u}$ onto the constraint subspace $\{q \in \mathbb{R}^{n_u} \mid Cq = 0\}$

$$P_M = I - H_M C. \tag{19}$$

We start with the computation of $\hat{A}$. Assume that the matrix pencil $\mu A - B$ is regular for a fixed value $\mu \in \mathbb{C}$. With the help of blockwise inversion we compute

$$\hat{A} = \begin{pmatrix} P_Z Z^{-1} M & 0 \\ H_Z^T M & 0 \end{pmatrix}$$

where $Z := \mu M + K$, and similar to (18) - (19) we define $P_Z := I - H_Z C$, $H_Z := Z^{-1} C^T (C Z^{-1} C^T)$. Analogously to (17) one shows that

$$\hat{A}^D = \mu Q - S.$$

holds. Yet we need to verify (12a)-(12c) for (2.2.1). The satisfaction of (12a) is straightforward. We omit the calculation here and just state the result as

$$\hat{A}(\mu Q - S) = (\mu Q - S)\hat{A} = Q.$$

The condition (12b) follows from the structure of the matrix Q which implies $SQ = QS = S$ and $Q^2 = Q$. From this it follows

$$(\mu Q - S)\hat{A}(\mu Q - S) = (\mu Q - S) = \mu Q - S.$$

The last condition (12c) is satisfied for nilpotency index $\nu = 2$. Then

$$(\mu Q - S)\hat{A}^3 = \hat{A}(\mu Q - S)\hat{A}\hat{A} = \hat{A}Q\hat{A} = \hat{A}\hat{A}.$$

□

2.2.2. Positivity Preservation for DAEs

The Drazin ODE (14) is now discretized in time. We assume that (7) has been discretized in space using an LED scheme. As a positivity-preserving time stepping scheme, the one-step $\theta-$method is applied,

$$y_h^{n+1} - y_h^n = \theta \Delta t \hat{A}^D \hat{B} y_h^{n+1} + (1-\theta)\Delta t \hat{A}^D \hat{B} y_h^n. \tag{20}$$

Theorem 2.2. *Consider the linear advection-diffusion equation* (1)*. We assume a local extremum diminishing space discretization scheme in the differential part of* (14)*. Discretizing the Drazin ODE in time by the one-step $\theta-$scheme, positivity will be preserved under a CFL-like time step restriction*

$$1 + \Delta t(1-\theta) \min_i(-P_M M^{-1} K)_{ii} \geq 0. \tag{21}$$

Proof. As stated in (16a), the system matrix is given by

$$\hat{A}^D \hat{B} = S := \begin{pmatrix} -P_M M^{-1} K & 0 \\ H_M^T K P_M M^{-1} K & 0 \end{pmatrix}.$$

Following the idea of the proof in [1], we start with the special case of the implicit Euler scheme, $\theta = 1$, applied to the differential part of (14)

$$u_h^{n+1} \quad = \quad u_h^n - \Delta t P_M M^{-1} K u_h^n.$$

We expect that the solution u_h^n is positive at every node and u_h^{n+1} is negative at some nodes. Let the global minimum be located at node k. Then the discrete solution at node k satisfies

$$\dot{u}_{h,k} = \sum_{j \neq k} -[P_M M^{-1} K]_{ij} \left(u_{h,j} - u_{h,k}\right). \tag{22}$$

The coefficients $[-P_M M^{-1} K]_{ij}, i \neq j$ are positive due to the LED property. Accordingly, the assumption $u_{h,k}^{n+1} < 0$ implies $u_{h,j} - u_{h,k} < 0$, which leads to a contradiction to the global minimum property of $u_{h,k}$.

Considering next the case $\theta < 1$, we need to include the explicit terms into (22). Therefore the positivity will be preserved for

$$u_{h,i}^n - \Delta t(1-\theta) \sum_{j \neq i} [P_M M^{-1} K]_{ij} (u_{h,j}^n - u_{h,i}^n) \geq 0.$$

For positive solutions $u_{h,i}^n$ and positive coefficients $-P_M M^{-1} K$, the positivity is then preserved under the time step restriction (21). □

We remark that the time step restriction (21) is similiar to the one given in [20] where the positivity preservation of the one-step $\theta-$scheme applied to an ODE $w'(t) = Ew(t)$ is proved under a time step restriction $\delta\Delta t \leq \frac{1}{1-\theta}$ and δ depending on the coefficients of the system matrix E.

Moreover, in the proof of Theorem 2.2, we assumed an LED scheme but left it open what this precisely means in our context. Let us shortly discuss the LED property of the underlying space discretization for the FEM-FCT scheme that will be presented in more detail in the next section. We are only interested in the positivity preservation of the differential variables in the low order scheme of the FEM-FCT algorithm. Let M_L denote the lumped mass matrix and $L := K + D$ the low order stiffness matrix with K the standard Galerkin finite element stiffness matrix and D a diffusion operator designed as to improve the M-matrix property of L. The resulting ODE reads then

$$\dot{u}_h = -P_{M_L} M_L^{-1} L u_h. \tag{23}$$

Here, the matrix $M_L^{-1}L$ is an M-matrix, cf. Remark 1, and the projector P_{M_L} is defined in (19). Thanks to the structure of the constraints or boundary conditions, respectively, the structure of the system matrix $M_L^{-1}L$ is conserved under the projection P_{M_L} such that the spatial discretization scheme is LED.

2.2.3. Equivalence of Discretized Systems

Our aim was to show that the one-step $\theta-$scheme applied to the differential-algebraic equation (7) is positivity preserving with respect to the differential variable. In the last subsection, we proved this proposition for the corresponding Drazin ODE. In addition, we need to guarantee the equivalence between the discrete Drazin ODE and the discrete DAE such that we can transfer the positivity preservation property to the DAE.

For this purpose,we first adapt the one-step $\theta-$scheme to the differential-algebraic equation (7) or (8), respectively. This yields

$$\begin{aligned} u^{n+1} &= u^n - \theta\Delta t M^{-1} K u^{n+1} - (1-\theta)\Delta t M^{-1} K u^n \\ &\quad -\Delta t M^{-1} C^T \lambda^{n+1}, \end{aligned} \tag{24}$$

$$0 = C u^{n+1}, \tag{25}$$

where we enforce the constraint explicitly at the time level t_{n+1}.

Lemma 2.2. *Let the Drazin ODE* (14) *and the differential-algebraic system* (8) *be discretized in time with the one-step $\theta-$ scheme. Furthermore, we assume consistent initial conditions. Then in each time step, the discrete equations for the differential variable are equivalent.*

Proof. Discretizing (14) by the one-step $\theta-$scheme yields at a certain time instant $t = t_{n+1}$

$$\begin{aligned} u^{n+1} &= u^n - \theta\Delta t P_M M^{-1} K u^{n+1} - (1-\theta)\Delta t P_M M^{-1} K u^n, \\ \lambda^{n+1} &= \theta\Delta t H_M^T K P_M M^{-1} K u^{n+1} + (1-\theta)\Delta t H_M^T K P_M M^{-1} K u^n. \end{aligned} \tag{26}$$

We insert (24) into the constraint (25) and derive an explicit formula for the Lagrange multipliers λ^{n+1}. Afterwards, we plug this presentation of λ^{n+1} into (24). Due to the consistent initial conditions, we can then inductively proceed. Suppose that both schemes are equivalent until the time step $t_{n-1} \to t_n$, which leads to $Cu^n = 0$. With this relation in mind we can conclude the equivalence for the next step $t_n \to t_{n+1}$. □

3. Flux Correction applied to Partial Differential-Algebraic Equations

Before modifying the FEM-FCT algorithm for the handling of PDAEs we recall the standard procedure. At first, a strong artificial diffusion is introduced into the high order scheme to develop a positivity preserving scheme. Due to Godunov's theorem, see, e.g. [29], the accuracy of such a method decreases to first order. A high-resolution scheme can be reconstructed by adding locally antidiffusion. The amount of the antidiffusion is restricted at every node stencil S_i by Zalesak's limiter [30].

3.1. The Standard Algorithm

The construction of a low order scheme starts with the standard Galerkin finite element discretization using a one-step $\theta-$scheme for time integration

$$Mu_h^{n+1} + \theta\Delta tKu_h^{n+1} = Mu_h^n - (1-\theta)\Delta tKu_h^n + \Delta tQ^{n+\theta} \tag{27}$$

where M denotes the consistent mass matrix, K the system stiffness matrix collecting all advective and diffusive terms, and Q the right hand side vector. Due to Remark 1, we derive a positivity preserving scheme in two steps. First we modify the matrices by using the lumped mass matrix

$$M_L = \mathrm{diag}(m_i), \qquad m_i = \sum_j M_{ij}, \tag{28}$$

and by adding sufficient diffusion D to the stiffness matrix K,

$$\begin{aligned} L &= K + D,\ D = (d)_{ij} \qquad \text{where} && (29a)\\ d_{ij} &= -\max(0, k_{ij}, k_{ji}) \qquad \text{and} && (29b)\\ d_{ii} &= \sum_{j\neq i} d_{ij}. && (29c) \end{aligned}$$

Then the low order scheme is given by

$$M_L u_h^{n+1} + \theta\Delta tLu_h^{n+1} = M_L u^n - (1-\theta)\Delta tLu^n + \Delta tQ^{n+\theta}. \tag{30}$$

To ensure the positivity preservation we need to fulfill the time step restriction (21).

The antidiffusive fluxes are defined as the residuum between both schemes (30) and (27) via

$$\begin{aligned} \mathcal{F}(u_H^{n+1}, u^n) &= -(M - M_L)\left(u_H^{n+1} - u^n\right) \\ &\quad -\Delta tD\left(\theta u_H^{n+1} + (1-\theta)u^n\right) \end{aligned} \tag{31}$$

where u_H^{n+1} denotes the solution at time level t_{n+1} computed with the oscillatory high order scheme (27). Let $\Delta u_i := u_{i,H}^{n+1} - u_i^n$, then we can decompose (31) into fluxes

$$\begin{aligned} \mathcal{F}_{ij} &:= -(M - M_L)_{ij}(\Delta u_j - \Delta u_i) \\ &\quad -\Delta t d_{ij}\left(\theta(u_{j,H}^{n+1} - u_{i,H}^{n+1}) + (1-\theta)(u_j^n - u_i^n)\right). \end{aligned} \tag{32}$$

A high-resolution solution u^{n+1} can be computed by

$$(M_l + \theta\Delta tL)u^{n+1} = (M_l - (1-\theta)L)u^n + \Delta tQ^{n+1} + \mathcal{F}^\star(u_H^{n+1}, u^n) \tag{33}$$

where $\mathcal{F}_i^\star = \sum_{j\neq i}\alpha_{ij}\mathcal{F}_{ij}$ and $0 \leq \alpha_{ij} \leq 1$ determined via Zalesak's limiter function [1, 17].

3.1.1. One-dimensional Example

To understand how the stabilization works, we consider the one dimensional pure advection equation

$$\dot{u} + au_x = 0. \tag{34}$$

Let $S_i = \{i-1, i, i+1\}$ be the stencil of a node i. Then it holds

$$d = d_{ji} = d_{ij} = \begin{cases} -\max(0, k_{ij}), & \text{if } j \in S_i, j \neq i \\ 0, & \text{if } j \notin S_i. \end{cases}$$

Therefore,

$$d_{ii} = 2d.$$

The FEM-FCT stabilized discretization scheme at an inner node j is given by, using the Courant number C,

$$\begin{aligned} u_j^{n+1} \quad + \quad & \frac{\theta}{2} C \left(u_{j+1}^{n+1} - u_{j-1}^{n+1}\right) + d\theta\Delta t \left(u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1}\right) \\ & - \left[\alpha_{j,j-1} u_{j-1}^{n+1} - (\alpha_{j,j-1} + \alpha_{j,j+1}) u_j^{n+1} + \alpha_{j,j+1} u_{j+1}^{n+1}\right] \\ & - d\theta\Delta t \left[\alpha_{j,j-1} u_{j-1}^{n+1} - (\alpha_{j,j-1} + \alpha_{j,j+1}) u_j^{n+1} + \alpha_{j,j+1} u_{j+1}^{n+1}\right] \\ = \quad & u_j^n - \frac{1-\theta}{2} C \left(u_{j+1}^n - u_{j-1}^n\right) + d(1-\theta)\Delta t \left(u_{j-1}^n - 2u_j^n + u_{j+1}^n\right) \\ & + \left[\alpha_{j,j-1} u_{j-1}^n - (\alpha_{j,j-1} + \alpha_{j,j+1}) u_j^n + \alpha_{j,j+1} u_{j+1}^n\right] \\ & - d(1-\theta)\Delta t \left[\alpha_{j,j-1} u_{j-1}^n - (\alpha_{j,j-1} + \alpha_{j,j+1}) u_j^n + \alpha_{j,j+1} u_{j+1}^n\right] \end{aligned} \tag{35}$$

In the case of $\alpha_{ij} \equiv 1 \ \forall i, j$ we recover the high order standard Galerkin finite element scheme with the consistent mass matrix, and all additional diffusive terms cancel out. On the other hand, consider $\alpha_{ij} \equiv 0 \ \forall i, j$ where the second order central differences terms from the consistent mass matrix and the antidiffusion are canceled out and from which we derive the low order scheme (30). In all cases $0 < \alpha_{ij} < 1$, antidiffusion is added, depending on the value of α_{ij}.

3.2. The Modified Algorithm

In the PDAE case, most of the modifications of the FEM-FCT scheme concern the construction of the low order scheme. There are two major changes in the low order scheme, and in addition we need to address the limiter application.

- *Mass lumping in the low-order scheme* (28)
  The mass lumping is realized by adding all row entries to the diagonal entry. The extended mass matrix in the differential-algebraic case is given by (9), i.e. $A = \begin{pmatrix} M_{\text{FEM}} & 0 \\ 0 & 0 \end{pmatrix}$, where M_{FEM} denote the consistent mass matrix stemming from the standard Galerkin finite element discretization. This means that only additional zero blocks are added. Therefore the standard mass lumping procedure can be applied.

- *Adding additional diffusion in the low order scheme* (29)
  The extended stiffness matrix due to the differential-algebraic framework reads $B = \begin{pmatrix} -K_{\text{FEM}} & -C^T \\ C & 0 \end{pmatrix}$, and it includes the usual stiffness matrix

of the finite element discretization K_{FEM} and the additional Boolean constraint matrix C as in (9). Using the standard procedure described in (29) to modify the stiffness would mean to add diffusion to the constraint matrix C, which leads to an error in the constraints. We propose to convert the stiffness matrix blockwise into the matrix L for the lower order method where

$$L = \begin{pmatrix} -L_{\text{FCT}} & -C^T \\ C & 0 \end{pmatrix}, \qquad L_{\text{FCT}} = K_{\text{FEM}} + D_{\text{FEM}}.$$

- *Limiter application in* (33)
  The limiter controls the amount of antidiffusion added to the scheme at each node i, taking into account the solution behaviour at the local stencil S_i. Adding antidiffusion to the algebraic variables would lead to an error in the constraint. Therefore we suggest to split the limiter into $\alpha = (\alpha_u, \alpha_\lambda)^T$ with limiter $\alpha_u \in \mathbb{R}^{n_u}$ for the differential variable and limiter $\alpha_\lambda \in \mathbb{R}^{n_\lambda}$ for the algebraic variable and set $\alpha_\lambda \equiv 0$ in the algorithm.

We remark that our modified FEM-FCT scheme, in contrast to the established approach, requires no extra treatment of the boundary nodes since these are taken care of by the constraints.

4. Results

In this section we demonstrate the behaviour of the modified FEM-FCT scheme in the partial differential-algebraic case. We begin with a one-dimensional pure advection problem. The first test includes non-zero constant boundary conditions while the second test case considers time-dependent boundary conditions. We apply the FEM-FCT scheme also to a two-dimensional advection problem on a quadratic domain where the advection proceeds skew to the mesh.

Finally, the convergence behaviour of the FEM-FCT stabilization is examined. For this purpose, we consider the one-dimensional pure advection problem transporting an initial rectangular profile along the mesh and measure the error in the numerical solution with respect to the analytic solution in the L_1- and L_2-norms.

In all numerical test runs, we use linear finite elements and $\theta = 1/2$.

4.1. Non-zero Boundary Conditions

To test the robustness of the algorithm with respect to the complexity of the considered constraints we applied the FEM-FCT algorithm to two different pure advection problems.

4.1.1. Constant Boundary Condition

The first test case considers the pure advection equation on $\Omega = [0, 1]$ given by

$$\begin{aligned} \dot{u} + 0.5u_x &= 0 \quad \text{in } \Omega \times T, \\ u(x,0) &= u_0 \quad \text{in } \Omega \times \{0\}, \\ u(x,t) &= \bar{u} \quad \text{on } \partial\Omega \times T \end{aligned}$$

with boundary conditions defined as

$$\bar{u}(x,t) = \begin{cases} 1, & x = 0 \\ 0, & x = 1 \end{cases}$$

and the consistent initial profile

$$u_0(x) = \begin{cases} 1, & 0 \leq x \leq 0.4 \\ 0, & 0.4 \leq x \leq 1. \end{cases}$$

The analytic solution of this problem is given by

$$u(x,t) = u_0(x - 0.5t).$$

In Fig. 1, the numerical and the analytic solution at $t = 0.5$ are plotted. As can be seen, the sharp front of the initial profile is transported to the right and the position of the front is nicely tracked by the numerical scheme. In addition, the boundary constraints are fulfilled explicitly at both boundaries.

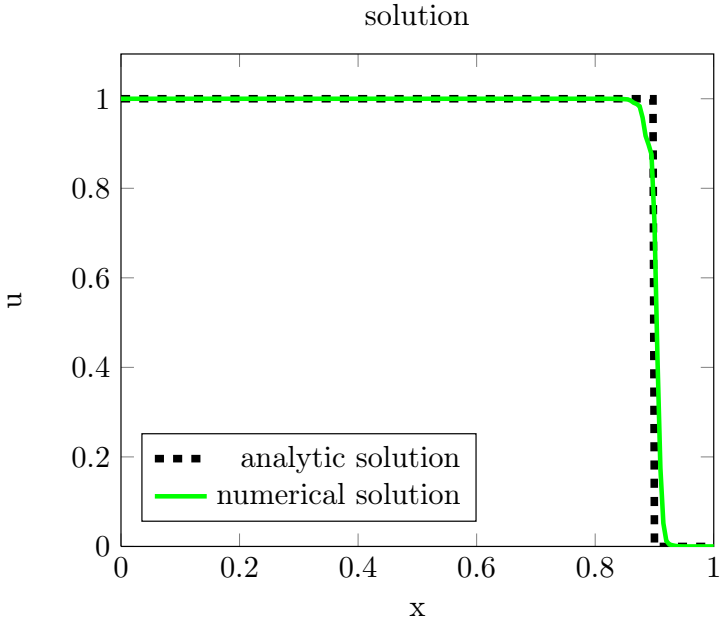

Figure 1: Constant boundary constraints, $t = 1$, $CFL = 0.1$, $\Delta t = 10^{-3}$

4.1.2. Time-Dependent Boundary Conditions

A more complicated test case are time dependent boundary conditions. Again, we consider the one-dimensional pure advection equation on $\Omega = [0, 1]$ given by

$$\begin{aligned} \dot{u} + 0.5u_x &= 0 \quad \text{in } \Omega \times T \\ u(x,0) &= u_0 \quad \text{in } \Omega \times \{0\} \\ u(x,t) &= \bar{u}(t) \quad \text{on } \partial\Omega \times T. \end{aligned}$$

Here, the boundary condition depends now on time and is defined as

$$\bar{u}(t) = \begin{cases} 0, & t < 0.1 \\ \sin^2(2\pi(t-0.1)), & 0.1 \leq t \leq 3.5 \\ 1, & t > 3.5 \end{cases}$$

and the consistent initial profile is given by

$$u_0(x) = 0 \quad \forall x \in \Omega.$$

The analytic solution is given by

$$u(x,t) = \bar{u}(t - \frac{x}{0.5}).$$

As in the test case above we plotted in fig 2 the analytic and the numerical solution at $t = 1$. In this case the moving front is smooth, and the FEM-FCT scheme is able to reproduce this front and track the front position in high quality. Once again, the constraints are fulfilled explicitly at both boundaries.

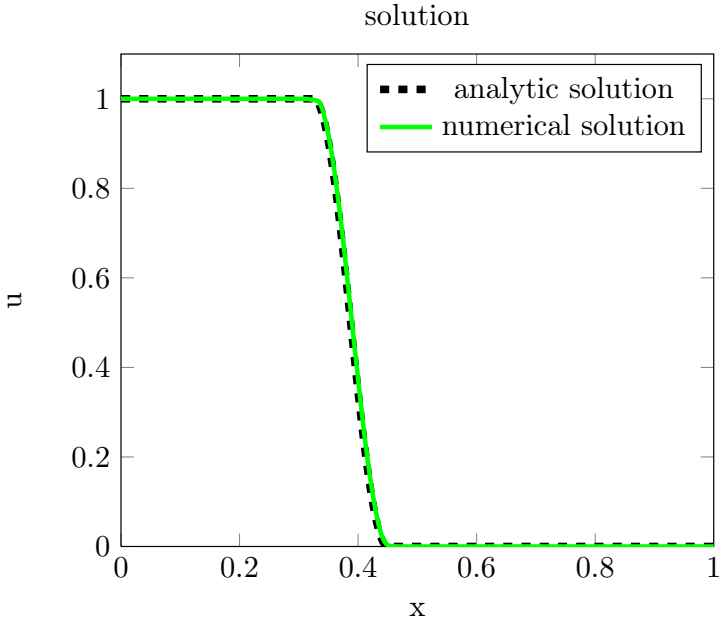

Figure 2: Time dependent boundary constraints, $t = 1$, $CFL = 0.1$, $\Delta t = 10^{-3}$

4.2. Advection Skew to Mesh

In the third test case we apply the algorithm to a two dimension pure advection equation where the advection proceeds skew to the mesh. The problem is formulated as follows

$$\dot{u} - \beta \nabla u = 0 \qquad \in \Omega$$

with $\beta = (0.5, 0.5)^T$ and zero Dirichlet boundary conditions. The initial data is a rectangular profile as plotted in Fig 3 on the left. The standard FEM solution

as well as the stabilized numerical solution are displayed in the middle and on the right in Fig. 3, respectively. As expected, the standard Galerkin finite element method produces unphysical oscillations in the numerical simulation while the FEM-FCT stabilized solution does not show any over- or undershoots. There are still a few small wriggles in the FEM-FCT solution as well but the solution stays positive everywhere, cf. [18]. This two-dimensional test case was implemented

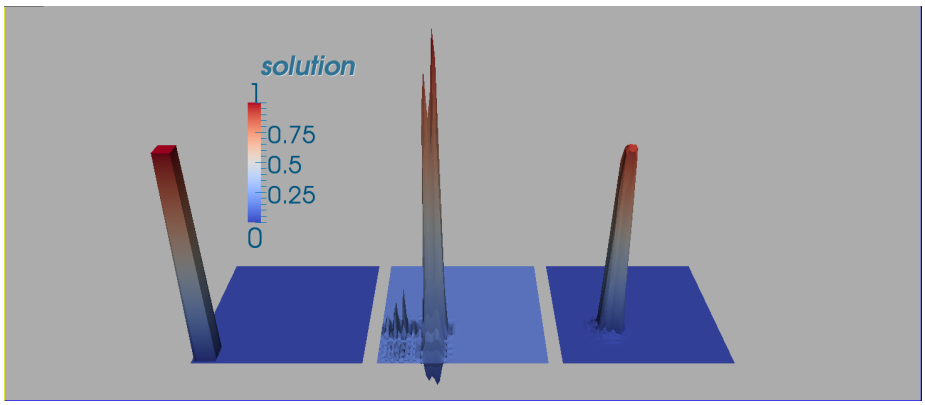

Figure 3: Left: initial profile, middle: standard Galerkin FEM solution, right: FEM-FCT solution

using the Finite Element C++ library `deal.II` [31, 32].

4.3. Convergence

To compare the convergence behaviour of the FEM-FCT algorithm with the standard Galerkin finite element method we study the one-dimensional pure advection equation

$$\dot{u} + 0.5u_x = 0 \qquad \in \Omega = [0, 1] \tag{36}$$

with discontinuous initial data

$$u_0 = \begin{cases} 1, & 0.15 < x < 0.5 \\ 0, & else. \end{cases} \tag{37}$$

The error in the numerical solution is measured by the L_1- as well as the L_2- norm of the difference between the exact solution $u(x,t) = u_0(x - 0.5t)$ of the pure advection equation (36) and the numerical solution u_h. The results are summarized in Table 1 and Table 2. Both algorithms are applied to the test problem (36) - (37) on three successively refined meshes. Here, #ref is the refinement level in the finite element simulation. The Courant number C is chosen to fulfill $C < 1$.

The FEM-FCT stabilization exhibits a monotone convergence and improves the accuracy of the numerical solution when compared with the standard Galerkin finite element method. To further increase the efficiency of the algorithm, one could think about applying a lumped mass FEM-FCT algorithm as proposed in [33], where also the pros and cons of this approach are discussed in detail.

#ref	$u_{\min}$	$u_{\max}$	$\|u_h - u\|_1$	$\|u_h - u\|_2$
1	0.0	1.0	2.24e-2	8.48e-2
2	0.0	1.0	1.34e-2	6.66e-2
3	0.0	1.0	7.1e-3	4.29e-2

Table 1: FEM-FCT

#ref	$u_{\min}$	$u_{\max}$	$\|u_h - u\|_1$	$\|u_h - u\|_2$
1	-0.2391	1.2109	4.00e-2	9.01e-2
2	-0.2169	1.1949	2.76e-2	6.68e-2
3	-0.2144	1.1925	1.97e-2	5.41e-2

Table 2: Standard Galerkin

5. Conclusion

We have extended the Finite Element Method - Flux Corrected Transport scheme to a partial differential-algebraic equation framework. While the standard FEM-FCT procedure modifies the FEM system matrices to construct a low order scheme, our extension works in a blockwise manner such that only the differential variables are affected. From the theoretical investigation based on the Drazin inverse we conclude that schemes such as the one step θ-method are positivity preserving, given a Local Extremum Diminishing spatial discretization scheme to avoid the birth and growth of local extrema. In this way, global over- and undershoots are eliminated and the resulting scheme yields stabilized solutions for hyperbolic problems. The numerical results confirm the theory.

Besides the one step θ-method, one could also study other time integration schemes, but the focus of our future work will be to include the present approach in a multiphysics simulation setting where the formulation of boundary data as constraints is a prerequisite for the modeling of general coupling conditions.

Acknowledgement. The authors would like to thank the german Federal Ministry of Education and Research (BMBF) for the support within the research project SNiMoRed (Multidisziplinäre Simulation, nichtlineare Modellreduktion und proaktive Regelung in der Fahrzeugdynamik). Moreover, we are grateful to Matthias Möller for helpful discussions and hints on the FEM-FCT scheme.

[1] D. Kuzmin and S. Turek. Flux correction tools for finite elements. *J.Comput. Phys*, 175:525 – 558, 2002.

[2] D. Kuzmin. Explicit and implicit fem-fct algorithms with flux linearization. Ergebnisberichte Angew. Math. 358, University of Dortmund, 2008.

[3] P. Crosetto, S. Deparis, G. Fourestey, and A.Quarteroni. Parallel Algorithms for Fluid-Structure Interaction Problems in Haemodynamics. *Siam Journal on Scientific Computing*, 33(4):1598–1622, 2011.

[4] U. Küttler, M.W. Gee, Ch. Förster, A. Comerford, and W.A. Wall. Coupling strategies for biomedical fluid-structure interaction problems. *Int, J. Num. Meth. Biomedical Engrg*, 26:205–321, 2010.

[5] K.-J. Bathe and H. Zhang. Finite element developments for general fluid flows with structural interactions. *International Journal for Numerical Methods in Engineering*, 60(1):213–232, 2004.

[6] U. Heck. Transient fluid structure interaction. In *Nafems Seminar "Simulation of Coupled Flow phenomena (Multified FSI)"*, 2006.

[7] Y. Bazilevs, M.-C. Hsu, and M.A. Scott. Isogeometric fluidâĂŞstructure interaction analysis with emphasis on non-matching discretizations, and with application to wind turbines. *Computer Methods in Applied Mechanics and Engineering*, Published online. DOI: http://dx.doi.org/10.1016/j.cma.2012.03.028, 2012.

[8] M.R. Doerfel and B. Simeon. Fluid-structure-interaction: Acceleration of strong coupling by preconditioning of the fixed-point iteration. In *ENUMATH Proceedings*, 2011.

[9] A.N. Brooks and T.J.R. Hughes. Streamline upwind/petrov - galerkin formulations for convection dominated flows with particular emphasis on the incomressible navier - stokes equations. *Comput. Methods Appl. Mech. Engrg.*, 32:199–259, 1982.

[10] T.J.R. Hughes and A.N. Brooks. A multidimensional upwind scheme with no crosswind diffusion,. In T.J.R. Hughes, editor, *Finite Element Methods for Convection Dominated Flows*, volume 34, chapter A multidimensional upwind scheme with no crosswind diffusion, pages 19–35. ASME, New York, 1979.

[11] V. John and P. Knobloch. A comparison of spurious oscillations at layers diminishing (sold) methods for convection-diffusion equations: part i - a review. *Comput. Methods Appl. Mech. Engrg.*, 196:2197–2215, 2007.

[12] V. John and P. Knobloch. A comparison of spurious oscillations at layers diminishing (sold) methods for convection-diffusion equations: part ii - analysis for p_1 and q_1 finite elements. *Comput. Methods Appl. Mech. Engrg.*, 197:1997–2014, 2008.

[13] S. Ganesan and L. Tobiska. Stabilization by local projection forÂăconvectionâĂŞdiffusion and incompressible flow problems. *Journal of Scientific Computing*, 43:326–342, 2010. 10.1007/s10915-008-9259-8.

[14] G. Matthies, P.Skrzypacz, and L.Tobiska. Stabilisation of local projection type applied to convection-diffusion problems with mixed boundary conditions. *Electron. Trans. Numer. Anal.*, 32:90–105, 2008.

[15] M. Braack and E. Burman. Local projection stabilization for the oseen problem and its interpretation as a variational multiscale method. *SIAM J. Numer. Anal.*, 43:2544–2566, 2006.

[16] D. Kuzmin and M. Möller. Algebraic flux correction i. scalar conservation laws. In D.Kuzmin R.Löhner and S.Turek, editors, *Flux-Corrected Transport: Principles, Algorithms and Applications*, chapter Algebraic flux correction I. Scalar conservation laws, pages 155 – 206. Springer, 2005.

[17] D. Kuzmin, M. Möller, and S. Turek. High-resolution fem-fct schmes for multidimensional conservation laws. *Comput. Methods Appl. Mech. Engrg*, 193:4915 – 4946, 2004.

[18] V. John and E. Schmeyer. Finite element methods for time-dependent convection-diffusion-reaction equations with small diffusion. *Comput. Methods Appl. Mech. Engrg.*, 198:475 – 494, 2008.

[19] C. Bolley and M. Crouzeix. Conversation de la positivite lors de la discretisation des problèmes d'evolution paraboliques. *R.A.I.R.O. Analyse numerique*, 12:81–88, 1978.

[20] W. Hundsdorfer and J.G. Verwer. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* Springer Series in Comput. Math. 33. Springer, 2003.

[21] A.-K. Baum and V. Mehrmann. Numerical integration of positive linear differential-algebraic systems. Preprint 2012-02, Institut für Mathematik, TU Berlin, Str. des 17. Juni 136, D-10623 Berlin, FRG, 2012.

[22] Ch. Großmann and H. G. Roos. *Numerische Behandlung partieller Differentialgleichungen.* Teubner, 3 edition, 2005.

[23] Z. Zheng, B. Simeon, and L.R. Petzold. A stabilized explicit lagrange multiplier based domain decomposition method for parabolic problems. *J. Comput. Phys.*, 227:5272–5285, 2008.

[24] M. P. Drazin. Pseudo-inverses in associative rings and semigroups. *The American Mathematical Monthly*, 65(7):506–514, 1958.

[25] J.H. Wilkinson. Note on practical significance of the drazin inverse. In S.L. Campbell, editor, *Recen Applications of Generalized Inverses*, pages 82–99. Pitman, London, 1982.

[26] P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations. Analysis and Numerical Solution.* EMS Publishing House, 2006.

[27] J.H. Wilkinson. Linear differential equations and kronecker's canonical form. In C. de Boor and G.H. Golub, editors, *Recent advances in numerical analysis*, pages 231–265. Proceedings Symposium University Wisconsin, Academic Press, New York, May 22-24 1978.

[28] B. Simeon, C. Führer, and P. Rentrop. The drazin inverse in multibody system dynamics. *Numer.Math.*, 64:521–539, 1993.

[29] S.K. Godunov. Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics. *Mat. Sbornik*, 47:271–306, 1959.

[30] S.T. Zalesak. Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.*, 31:335–362, 1979.

[31] W. Bangerth, R. Hartmann, and G. Kanschat. deal.ii - a general-purpose object-oriented finite element library. *ACM Trans. Math. Softw.*, 33(4), aug 2007.

[32] W. Bangerth, T. Heister, and G. Kanschat. *deal.II Differential Equations Analysis Library, Technical Reference.* `http://www.dealii.org`.

[33] D. Kuzmin and D. Kourounis. A semi-implicit fem-fct algorithm for efficient treatment of time-dependent problems. Technical Report 302, University of Dortmund, 2005.