Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

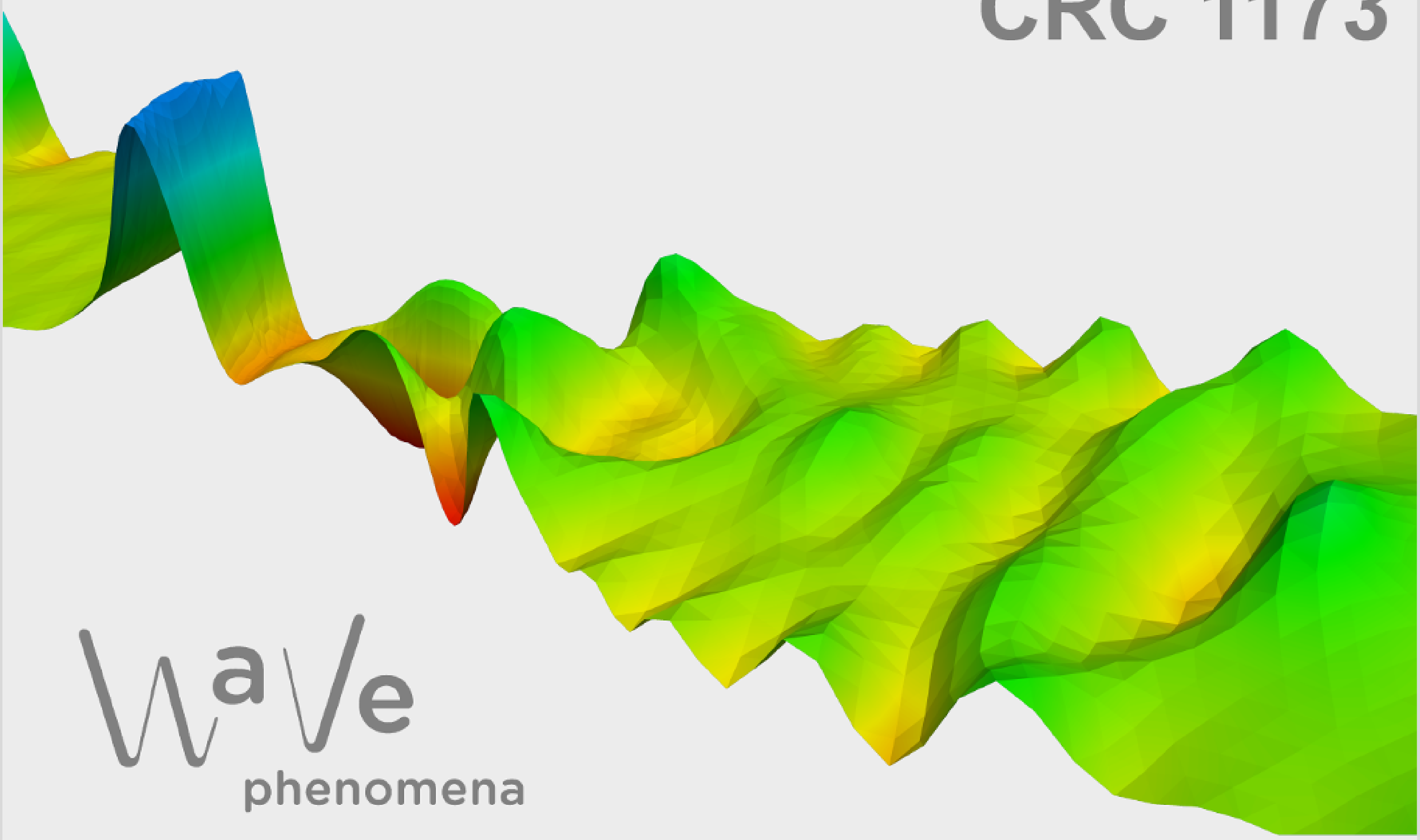
An implicit-explicit time discretization scheme for second-order semilinear wave equations with application to dynamic boundary conditions

Marlis Hochbruck, Jan Leibold

CRC Preprint 2020/20, July 2020

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Wave phenomena

Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

Numerische Mathematik manuscript No.
(will be inserted by the editor)

An implicit-explicit time discretization scheme for second-order semilinear wave equations with application to dynamic boundary conditions

Marlis Hochbruck · **Jan Leibold**

Received: date / Accepted: date

Abstract We construct and analyze a second-order implicit-explicit (IMEX) scheme for the time integration of semilinear second-order wave equations. The scheme treats the stiff linear part of the problem implicitly and the nonlinear part explicitly. This makes the scheme unconditionally stable and at the same time very efficient, since it only requires the solution of one linear system of equations per time step.

For the combination of the IMEX scheme with a general, abstract, nonconforming space discretization we prove a full discretization error bound. We then apply the method to a nonconforming finite element discretization of an acoustic wave equation with a kinetic boundary condition. This yields a fully discrete scheme and a corresponding a-priori error estimate.

Keywords implicit-explicit time integration · IMEX · dynamic boundary conditions · semilinear wave equation · nonconforming space discretization · error analysis · a-priori error bounds · semilinear evolution equations · operator semigroups

Mathematics Subject Classification (2010) Primary 65M12, 65M15; Secondary 65M60, 65J08

1 Introduction

In this paper we construct and analyze an implicit-explicit (IMEX) time integration scheme for second-order semilinear wave equations of the form

$$u''(t) + Bu'(t) + Au(t) = f(t, u(t))$$

in a suitable Hilbert space. Here, A and B are unbounded operators and f is a locally Lipschitz continuous nonlinearity. The IMEX scheme is constructed as a combination of the

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – SFB 1173

M. Hochbruck · J. Leibold
Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technologie, D-76149 Karlsruhe, Germany
E-mail: marlis.hochbruck@kit.edu, jan.leibold@kit.edu

explicit leapfrog method and the implicit Crank–Nicolson scheme. It treats the unbounded linear part of the differential equation implicitly and the nonlinear part explicitly. We show that the scheme is unconditionally stable in the sense that the time-step size is only restricted by the Lipschitz constant of f but not by the linear operators A and B.

We combine this IMEX scheme with an abstract, nonconforming space discretization within the framework of [8,9,10]. These papers provide a unified error analysis (UEA) which allows one to analyze nonconforming space discretizations of wave equations in a systematic way. Our main result is an error bound that is second order in time and contains abstract space-discretization errors. The error result can then be used to prove convergence rates for specific problems and discretizations by plugging in geometric and interpolation error results. The fully discrete scheme is very efficient. In fact, we show that one time step only requires the solution of one linear system and one application of discretizations of A, B, and f, respectively.

There is a rich literature on IMEX schemes, in particular, there is a well-developed theory for IMEX Runge–Kutta schemes [2] or IMEX multistep schemes [3]. IMEX schemes are used in many applications, e.g., in structural dynamics and fluid-structure interaction [18], hydrodynamics [12], sea-ice dynamics [16], or atmospheric dynamics, see, e.g., [6]. There exists also a so-called Crank–Nicolson-leapfrog IMEX scheme which is obtained from a combination of the Crank–Nicolson and the leapfrog scheme for first-order equations, cf. [13,14], and references therein. However, this scheme is not equivalent to the scheme we construct here, since the leapfrog schemes for first and second-order equations are not equivalent and indeed have completely different stability properties.

Due to our knowledge, the scheme we propose was not considered in literature so far. In fact, we are not aware of any IMEX scheme exploiting the structure of second-order equations. In addition, the full discretization error analysis for the scheme combined with the abstract non-conforming space discretization is new.

As an application of our abstract theory, we consider an acoustic wave equation with kinetic boundary conditions that fits into the abstract setting, cf., [10]. Kinetic boundary conditions are a special case of dynamic boundary conditions that contain tangential derivatives and are intrinsically posed on domains with (piecewise) smooth and therefore possibly curved boundaries. Hence, the spatial discretization has to be done on an approximated domain rendering the discretization nonconforming.

The paper is organized as follows: in Section 2 we present the problem setting, introduce the IMEX scheme for second-order wave equations, and state a second-order error bound for the time discretization error. In Section 3 we briefly recall the UEA and present the fully discrete scheme as a combination of the IMEX scheme with a general space discretization. Afterwards we state and prove the main result, namely the abstract full discretization error bound. Finally, in Section 4, we consider a semilinear acoustic wave equation with a kinetic boundary conditions as an example fitting into the abstract setting. We present a finite element space discretization and the full discretization error bound.We finish the paper with numerical experiments underlining the theoretical error bounds and the efficiency of the IMEX scheme.

2 The implicit-explicit (IMEX) scheme

In this section we first introduce the problem setting and then present the IMEX scheme and its properties.

2.1 Continuous problem

Let V,H be Hilbert spaces with $V \subset H$. We consider the following second-order variational equation as a prototype of second-order wave equations in weak formulation: find $u\colon [0,T] \to V$ s.t.

$$\begin{aligned} m\big(u''(t),v\big) + b\big(u'(t),v\big) + a\big(u(t),v\big) &= m\big(f(t,u(t)),v\big) \quad \text{for all } v \in V,\ t \in (0,T], \\ u(0) = u^0, \quad u'(0) &= v^0, \end{aligned} \tag{1}$$

with bilinear forms

$$\begin{aligned} m&\colon H \times H \to \mathbb{R}, \\ a&\colon V \times V \to \mathbb{R}, \\ b&\colon V \times H \to \mathbb{R}, \end{aligned}$$

and $f\colon [0,T] \times V \to H$.

For the rest of the paper we require the following assumptions without further mentioning it everywhere:

Assumption 2.1

a) The bilinear form m is a scalar product on H.
b) The bilinear form a is symmetric and there exists a constant $\alpha_{\mathrm{G}} > 0$ s.t.

$$\widetilde{a} = a + \alpha_{\mathrm{G}} m$$

is a scalar product on V. In the following we equip V with $\widetilde{a}$.
c) The space V is densely embedded in H, i.e., there exists an embedding constant C_{emb}, s.t.

$$\|v\|_m \le C_{\mathrm{emb}} \|v\|_{\widetilde{a}} \qquad \text{for all } v \in V. \tag{2}$$

d) The bilinear form b is bounded and quasi-monotone, i.e., there exist constants $C_B > 0$ and β_{qm} s.t.

$$b\big(v,w\big) \le C_B \|v\|_{\widetilde{a}} \|w\|_m \qquad \text{for all } v \in V, w \in H, \tag{3a}$$

$$b\big(v,v\big) + \beta_{\mathrm{qm}} m\big(v,v\big) \ge 0 \qquad \text{for all } v \in V. \tag{3b}$$

e) The nonlinearity f satisfies $f \in C^1([0,T] \times V; H)$ and is locally Lipschitz-continuous on V with constant $L_{T,\rho}$, i.e., for all $t \le T$ and $v,w \in V$ with $\|v\|_{\widetilde{a}}, \|w\|_{\widetilde{a}} \le \rho$ we have

$$\|f(t,v) - f(t,w)\|_m \le L_{T,\rho} \|v-w\|_{\widetilde{a}}. \tag{4}$$

In the following we consider (1) as evolution equation on H:

$$u''(t) + Bu'(t) + Au(t) = f(t,u(t)), \qquad u(0) = u^0, \quad u'(0) = v^0, \tag{5}$$

with operators $A\colon D(A) \to H$ and $B\colon V \to H$ induced by a and b, i.e.,

$$\begin{aligned} m\big(Av,w\big) &= a\big(v,w\big), \qquad \text{for all } v \in D(A), w \in V, \\ m\big(Bv,w\big) &= b\big(v,w\big), \qquad \text{for all } v \in V, w \in H, \end{aligned}$$

with

$$D(A) = \big\{ v \in V \,\big|\, \exists C = C(v) > 0 \ \forall w \in V : |a\big(v,w\big)| \le C\|w\|_m \big\}.$$

The wellposedness of (5) can be obtained by classical semigroup theory applied to the first-order formulation 6 of the equation, as detailed in [10].

Theorem 2.2 *The problem* (5) *is locally wellposed, i.e., for all* $u^0 \in D(A), v^0 \in V$ *there exists a time* $t^* = t^*(u^0, v^0)$ *s.t. for all* $T < t^*$, (5) *has a unique solution*

$$u \in C^2([0,T];H) \cap C^1([0,T];V) \cap C([0,T];D(A)).$$

2.2 Construction of the IMEX scheme

We modify the Crank–Nicolson scheme such that it treats the nonlinear part of (5) explicitly and thus avoids the solution of nonlinear equations.

To derive and analyze the scheme we state (5) in a first-order formulation. Set $u' = v$ and

$$x = \begin{bmatrix} u \\ v \end{bmatrix}, \quad S = \begin{bmatrix} 0 & -\mathrm{I} \\ A & B \end{bmatrix}, \quad g(t,x) = \begin{bmatrix} 0 \\ f(t,u) \end{bmatrix}, \quad x_0 = \begin{bmatrix} u^0 \\ v^0 \end{bmatrix}.$$

Then (5) can be written as

$$\begin{aligned} x'(t) + Sx(t) &= g(t, x(t)), \qquad t \in [0,T], \\ x(0) &= x_0. \end{aligned} \tag{6}$$

We consider this equation in the Hilbert space $(X, p) := (V, \widetilde{a}) \times (H, m)$, where p is the natural inner product, and $D(S) = D(A) \times V$.

Let $\tau > 0$ be the time step and $t_n := \tau n$ for $n \in \mathbb{N}$. By

$$\begin{bmatrix} u^n \\ v^n \end{bmatrix} = x^n \approx x(t_n)$$

we denote the numerical approximation of the exact solution of (6) at time t_n and we further use the short notations

$$g^n = \begin{bmatrix} 0 \\ f^n \end{bmatrix} := \begin{bmatrix} 0 \\ f(t_n, u^n) \end{bmatrix} = g(t_n, x^n).$$

The Crank–Nicolson scheme applied to (6) has the form

$$x^{n+1} = x^n + \frac{\tau}{2}\big(-S(x^n + x^{n+1}) + g^n + g^{n+1} \big) \tag{7a}$$

and can be written as

$$R_+ x^{n+1} = R_- x^n + \frac{\tau}{2}\left(g^n + g^{n+1}\right), \qquad R_\pm := I \pm \frac{\tau}{2} S. \tag{7b}$$

By [8, Lemma 2.14] we have the following properties of $R_\pm$:

Lemma 2.3 *Let* $c_{\mathrm{qm}} = \frac{\tau}{2}\alpha_{\mathrm{G}} C_{\mathrm{emb}} + \beta_{\mathrm{qm}}$ *with* C_{emb} *defined in* (2) *and* $\alpha_{\mathrm{G}}, \beta_{\mathrm{qm}}$ *from Assumption 2.1. Then, for* $\tau c_{\mathrm{qm}} < 2$*, the following assertions hold true:*

a) R_+ *is invertible with* $\|R_+^{-1}\|_{X \leftarrow X} \le 1$ *and* $R_+^{-1} x \in D(S)$ *for all* $x \in X$.
b) $R := R_+^{-1} R_-$ *has a continuous extension on* X *satisfying* $\|R\|_{X \leftarrow X} \le \mathrm{e}^{\tau c_{\mathrm{qm}}}$.

By applying R_+^{-1} to (7b), the Crank–Nicolson scheme reads

$$x^{n+1} = R x^n + \frac{\tau}{2} R_+^{-1}\left(g^n + g^{n+1}\right). \tag{7c}$$

Lemma 2.4 *The Crank–Nicolson scheme* (7a) *can equivalently be rewritten in a half-full-half step formulation*

$$v^{n+\frac{1}{2}} = v^n - \frac{\tau}{2}Au^n - \frac{\tau^2}{4}Av^{n+\frac{1}{2}} - \frac{\tau}{2}Bv^{n+\frac{1}{2}} + \frac{\tau}{4}\big(f^n + f^{n+1}\big), \tag{8a}$$

$$u^{n+1} = u^n + \tau v^{n+\frac{1}{2}}, \tag{8b}$$

$$v^{n+1} = v^{n+\frac{1}{2}} - \frac{\tau}{2}Au^n - \frac{\tau^2}{4}Av^{n+\frac{1}{2}} - \frac{\tau}{2}Bv^{n+\frac{1}{2}} + \frac{\tau}{4}\big(f^n + f^{n+1}\big). \tag{8c}$$

Proof Using

$$v^{n+\frac{1}{2}} := \frac{1}{2}\left(v^n + v^{n+1}\right), \tag{9}$$

we write (7a) component wise:

$$\begin{aligned} u^{n+1} &= u^n + \frac{\tau}{2}v^{n+\frac{1}{2}}, \\ v^{n+1} &= v^n + \frac{\tau}{2}\Big(-A(u^n + u^{n+1}) - Bv^{n+\frac{1}{2}} + f^n + f^{n+1}\Big). \end{aligned}$$

The first equation gives (8b). By eliminating u^{n+1} in the second equation we obtain

$$v^{n+1} = v^n - \tau Au^n - \frac{\tau^2}{4}Av^{n+\frac{1}{2}} - \frac{\tau}{2}Bv^{n+\frac{1}{2}} + \frac{\tau}{2}\big(f^n + f^{n+1}\big),$$

which is equivalent to the two half steps (8a) and (8c). □

By replacing in (8) the trapezoidal rule for the nonlinear part by the left/right rectangular rule, respectively, we obtain the following IMEX scheme:

$$v^{n+\frac{1}{2}} = v^n - \frac{\tau}{2}Au^n - \frac{\tau^2}{4}Av^{n+\frac{1}{2}} - \frac{\tau}{2}Bv^{n+\frac{1}{2}} + \frac{\tau}{2}f^n, \tag{10a}$$

$$u^{n+1} = u^n + \tau v^{n+\frac{1}{2}}, \tag{10b}$$

$$v^{n+1} = v^{n+\frac{1}{2}} - \frac{\tau}{2}Au^n - \frac{\tau^2}{4}Av^{n+\frac{1}{2}} - \frac{\tau}{2}Bv^{n+\frac{1}{2}} + \frac{\tau}{2}f^{n+1}. \tag{10c}$$

It can be interpreted as a combination of the Crank–Nicolson scheme for the linear and the leapfrog scheme for the nonlinear part, respectively. Obviously, it is time reversible.

Remark 2.5 An equivalent representation of v^{n+1} is obtained by subtracting (10a) from (10c), namely

$$v^{n+1} = -v^n + 2v^{n+\frac{1}{2}} + \frac{\tau}{2}\left(f^{n+1} - f^n\right). \tag{10d}$$

It is computationally more efficient because of the elimination of the operators A and B.

The implementation is comprised by solving the linear system in (10a), and then computing (10b), and (10d). Altogether, each time step requires the solution of one linear system, one application of A and one evaluation of the nonlinearity (note that f^{n+1} can be reused in the next time step).

2.3 Wellposedness of the IMEX scheme

By (10a) we have

$$Q_+ v^{n+\frac{1}{2}} = v^n - \frac{\tau}{2} A u^n + \frac{\tau}{2} f^n, \tag{11}$$

with $Q_\pm \colon D(A) \to H$ given by

$$Q_\pm = \mathrm{I} \pm \frac{\tau}{2} B \pm \frac{\tau^2}{4} A.$$

Since these operators play an important role in the analysis of the method, we collect some of their properties.

Lemma 2.6 (Properties of $Q_\pm$) *Let*

$$\frac{\tau^2}{2}\alpha_{\mathrm{G}} + \tau\beta_{\mathrm{qm}} \le 1. \tag{12}$$

Then, the operator Q_+ is invertible and its inverse $Q_+^{-1} \colon H \to D(A)$ satisfies

$$\Big\| \Big(\frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) Q_+^{-1} \Big\|_{H \leftarrow H} \le 1, \tag{13a}$$

$$\Big\| Q_+^{-1} \Big\|_{V \leftarrow H} \le \frac{\sqrt{2}}{\tau}, \tag{13b}$$

$$\Big\| Q_- Q_+^{-1} \Big\|_{H \leftarrow H} \le \mathrm{e}^{\frac{\tau^2}{2}\alpha_{\mathrm{G}} + \tau\beta_{\mathrm{qm}}}. \tag{13c}$$

Proof (a) We first prove that Q_+ is invertible. By (12) and Assumption 2.1, the bilinear form

$$m + \frac{\tau}{2} b + \frac{\tau^2}{4} a = \underbrace{\Big(1 - \frac{\tau}{2}\beta_{\mathrm{qm}} - \frac{\tau^2}{4}\alpha_{\mathrm{G}}\Big)}_{\ge 0} m + \frac{\tau}{2}(b + \beta_{\mathrm{qm}} m) + \frac{\tau^2}{4}\widetilde{a} \colon V \times V \to \mathbb{R}$$

is V-elliptic. Hence, by the Lax–Milgram lemma, for a given $v \in H \subset V'$ there exists a unique $z \in V$ such that

$$m\big(z,w\big) + \frac{\tau}{2} b\big(z,w\big) + \frac{\tau^2}{4} a\big(z,w\big) = m\big(v,w\big) \qquad \text{for all } w \in V,$$

or equivalently

$$\frac{\tau^2}{4} a\big(z,w\big) = m\big(v - z - \frac{\tau}{2} Bz, w\big) \qquad \text{for all } w \in V.$$

This yields $z \in D(A)$ and $Q_+ z = v$, hence Q_+ is invertible.

(b) Proof of the bounds (13): Let $v \in H$ and set $z = Q_+^{-1} v \in D(A)$. Then we have

$$\begin{aligned}
\|v\|_m^2 &= \Big\| \Big(\mathrm{I} + \frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) z \Big\|_m^2 \\
&= \|z\|_m^2 + \Big\| \Big(\frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) z \Big\|_m^2 + 2m\big(\Big(\frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) z, z\big) \\
&= \Big(1 - \frac{\tau^2}{2}\alpha_{\mathrm{G}} - \tau\beta_{\mathrm{qm}}\Big) \|z\|_m^2 + \Big\| \Big(\frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) z \Big\|_m^2 \\
&\qquad + 2\frac{\tau}{2} m\big(\Big(B + \beta_{\mathrm{qm}}\mathrm{I}\Big) z, z\big) + 2\frac{\tau^2}{4} m\big(\Big(A + \alpha_{\mathrm{G}}\mathrm{I}\Big) z, z\big) \\
&\ge \Big\| \Big(\frac{\tau}{2}B + \frac{\tau^2}{4}A\Big) Q_+^{-1} v \Big\|_m^2 + \frac{\tau^2}{2} \|Q_+^{-1} v\|_{\widetilde{a}}^2
\end{aligned}$$

due to the quasi-monotonicity of b (3b) and (12). This immediately yields (13a) and (13b). The bound (13c) can be shown similar to the bound for $R_+^{-1}R_-$ in Lemma 2.3. □

Corollary 2.7 *The IMEX scheme is wellposed in $D(A)\times H$, i.e., for $u^0 \in D(A)$ and $v^0 \in H$ the numerical approximations satisfy*

$$u^n \in D(A), \qquad v^n \in H, \qquad v^{n+\frac{1}{2}} \in D(A), \qquad n \geq 0.$$

Proof We use induction. The statement holds for $n=0$ by assumption, hence we assume that $u^n \in D(A), v^n \in H$ for some $n \geq 0$. By Lemma 2.6, Q_+ is invertible and (11) implies $v^{n+\frac{1}{2}} \in D(A)$. From (10b) and (10c) we then get $u^{n+1} \in D(A), v^{n+1} \in H$. □

2.4 Error bound for the IMEX scheme

We now state a second-order error bound for the IMEX scheme.

Theorem 2.8 *Let the assumptions of Theorem* 2.2 *be satisfied, $T < t^*(u^0, v^0)$, and let the exact solution u of* (5) *satisfy $u \in C^4([0,T],H) \cap C^3([0,T],V)$ and $f(u) \in C^2([0,T],H)$. Then for all τ sufficiently small and all $t_n < T$, the approximations u^n from the linearly implicit scheme* (10) *are bounded by*

$$\|u^n\|_{\widetilde{a}} \leq \rho := 2\|u\|_{L^\infty([0,T];V)}.$$

Moreover, u^n, v^n satisfy the error bound

$$\|u^n - u(t_n)\|_{\widetilde{a}} + \|v^n - u'(t_n)\|_m \leq C\mathrm{e}^{Mt_n} E(u)\tau^2$$

with $M = c_{\mathrm{qm}} + \frac{\left(1+(3/2)^{1/2}\right)L_{T,\rho}}{1-\left(1+(3/2)^{1/2}\right)L_{T,\rho}\tau}$, $c_{\mathrm{qm}} = \frac{\tau}{2}\alpha_{\mathrm{G}} C_{\mathrm{emb}} + \beta_{\mathrm{qm}}$,

$$\begin{aligned} E = E(u) =& \|u^{(4)}\|_{L^\infty([0,T];H)} + \|u^{(3)}\|_{L^\infty([0,T];V)} \\ &+ \left\|\frac{\mathrm{d}}{\mathrm{d}t}\big(f(u)\big)\right\|_{L^\infty([0,t_n];H)} + \left\|\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(f(u)\big)\right\|_{L^\infty([0,t_n];H)}, \end{aligned}$$

and a constant C that only depends on T but is independent of τ, L, and u.

Since the proof works with the same arguments as the the more complicated proof of Theorem 3.3 for the full discretization error of the IMEX scheme, we do not present it here, cf., also Remark 3.4 a).

2.5 The IMEX scheme in first-order formulation

For the error analysis we rewrite the IMEX scheme (10) as a perturbation of the one-step formulation of the CN scheme (7b). A similar idea was used in [11] for analyzing the leapfrog scheme and locally implicit schemes for Maxwells equations.

The formulation (7c) of the Crank–Nicolson scheme can be used to prove stability. For the IMEX scheme we now derive a similar formulation.

Lemma 2.9

(a) We have the following representations of R_+^{-1} and R:

$$R_+^{-1} = \begin{bmatrix} Q_+^{-1}\left(\mathrm{I}+\frac{\tau}{2}B\right) & \frac{\tau}{2}Q_+^{-1} \\ -\frac{2}{\tau}+\frac{2}{\tau}Q_+^{-1}\left(\mathrm{I}+\frac{\tau}{2}B\right) & Q_+^{-1} \end{bmatrix}, \tag{14a}$$

$$R = \begin{bmatrix} -\mathrm{I}+Q_+^{-1}\left(2\mathrm{I}+\tau B\right) & \tau Q_+^{-1} \\ -\frac{4}{\tau}\mathrm{I}+\frac{1}{\tau}Q_+^{-1}\left(4\mathrm{I}+2\tau B\right) & Q_-Q_+^{-1} \end{bmatrix}. \tag{14b}$$

(b) For all $w \in V$ we have

$$R_+^{-1}\begin{bmatrix} w \\ -Bw \end{bmatrix} = \begin{bmatrix} Q_+^{-1}w \\ -\left(B+\frac{\tau}{2}A\right)Q_+^{-1}w \end{bmatrix}. \tag{14c}$$

(c) The IMEX scheme (10) *is equivalent to*

$$x^{n+1} = Rx^n + \frac{\tau}{2}R_+^{-1}\left(g^n+g^{n+1}\right) + \frac{\tau^2}{4}\begin{bmatrix} Q_+^{-1}\left(f^n-f^{n+1}\right) \\ -\left(B+\frac{\tau}{2}A\right)Q_+^{-1}\left(f^n-f^{n+1}\right) \end{bmatrix}. \tag{15}$$

Proof First note that the right-hand side of (14a) is a well-defined mapping from X to $D(S)$ by Lemma 2.6. The identities (14) can be verified by straightforward calculations.

(c) To make the following calculations well defined, we assume for the moment that $v^n, v^{n+1} \in V$. We eliminate $v^{n+\frac{1}{2}}$ from the scheme (10), by subtracting (10c) from (10a). This yields

$$v^{n+\frac{1}{2}} = \frac{1}{2}\left(v^n+v^{n+1}\right) + \frac{\tau}{4}\left(f^n-f^{n+1}\right), \tag{16}$$

which differs from the Crank–Nicolson scheme by the contributions of the nonlinearity f, cf. (9). Note that we have $f^n - f^{n+1} \in V$ since, by Corollary 2.7, $v^n, v^{n+1} \in V$ and $v^{n+\frac{1}{2}} \in D(A)$. Inserting (16) into (10b) gives

$$u^{n+1} = u^n + \frac{\tau}{2}(v^n+v^{n+1}) + \frac{\tau^2}{4}\left(f^n-f^{n+1}\right). \tag{17}$$

On the other hand, by adding (10a) and (10c) and inserting (10b) and (16) we get

$$v^{n+1} = v^n - \frac{\tau}{2}A(u^n+u^{n+1}) - \frac{\tau}{2}B\left(v^n+v^{n+1}\right) + \frac{\tau}{2}\left(f^n+f^{n+1}\right) - \frac{\tau^2}{4}B\left(f^n-f^{n+1}\right). \tag{18}$$

With the definition (7b) of $R_\pm$ we can express (17) and (18) in first-order formulation as

$$R_+x^{n+1} = R_-x^n + \frac{\tau}{2}\left(g^n+g^{n+1}\right) + \frac{\tau^2}{4}\begin{bmatrix} f^n-f^{n+1} \\ -B\left(f^n-f^{n+1}\right) \end{bmatrix}.$$

Multiplying by R_+^{-1} and using (14c) shows that the IMEX scheme is equivalent to (15) under the additional assumption $v^n, v^{n+1} \in V$. Since both formulations are also well defined for $v^n, v^{n+1} \in H$ and since V is dense in H, we also get their equivalence for $v^n, v^{n+1} \in H$.

Note that one can avoid the detour over $v^n, v^{n+1} \in V$ but then one has to make the calculations more carefully and they become much more complicated. □

3 Full discretization

In this section we combine the IMEX scheme with an abstract space discretization to obtain a fully discrete scheme. We use the framework introduced in [9] for linear equations and extended in [10] to the semilinear case. It is rather general and allows one to cover conforming as well es nonconforming space discretizations, the latter being relevant for the discretization of equations with dynamic boundary conditions.

3.1 Framework

Let V_h be a finite dimensional vector space for the spatial approximation related to a mesh parameter h. We consider the following discretization of (1): find $u_h\colon [0,T] \to V_h$ s.t.

$$\begin{aligned} m_h\big(u_h'', \varphi_h\big) + b_h\big(u_h', \varphi_h\big) + a_h\big(u_h, \varphi_h\big) &= m_h\big(f_h(t,u_h), \varphi_h\big) \qquad \forall t \in (0,T], \varphi_h \in V_h, \\ u_h(0) = u_h^0, \qquad u_h'(0) &= v_h^0. \end{aligned} \tag{19}$$

Here, $m_h, a_h, b_h, f_h, u_h^0$ and v_h^0 are approximations of their corresponding continuous counterparts and satisfy similar properties as in Assumption 2.1.

Assumption 3.1 In the following statements, all constants are independent of h.

a) The bilinear form m_h is a scalar product on V_h and we denote V_h equipped with this scalar product by H_h.
b) The bilinear form a_h is symmetric and there exists a constant $\widehat{\alpha}_{\mathrm{G}} \geq 0$ s.t.

$$\widetilde{a}_h = a_h + \widehat{\alpha}_{\mathrm{G}} m_h$$

is a scalar product on V_h. In the following we equip V_h with $\widetilde{a}_h$.
c) There exists a constant $\widehat{C}_{\mathrm{emb}} > 0$ s.t. $\|v_h\|_{m_h} \leq \widehat{C}_{\mathrm{emb}} \|v_h\|_{\widetilde{a}_h}$ for all $v_h \in V_h$.
d) The bilinear form $b_h\colon V_h \times H_h \to \mathbb{R}$ is continuous and there exists a $\widehat{\beta}_{\mathrm{qm}} \geq 0$ s.t.

$$b_h\big(v,v\big) + \widehat{\beta}_{\mathrm{qm}} \|v\|_{m_h}^2 \geq 0 \quad \text{for all } v_h \in V_h.$$

e) The discrete nonlinearity $f_h\colon [0,T] \times V_h \to H_h$ is locally Lipschitz continuous on V_h with constant $\widehat{L}_M$, analogously to (4).

To reformulate (19) as an evolution equation on V_h, we define $A_h, B_h \in \mathcal{L}(V_h;V_h)$ via

$$m_h\big(A_h v_h, \varphi_h\big) = a_h\big(v_h, \varphi_h\big), \quad m_h\big(B_h v_h, \varphi_h\big) = b_h\big(v_h, \varphi_h\big) \quad \text{for all } v_h, \varphi_h \in V_h.$$

Then (19) is equivalent to

$$\begin{aligned} u_h''(t) + B_h u_h'(t) + A_h u_h(t) &= f_h(t, u_h(t)), \\ u_h(0) = u_h^0, \qquad u_h'(0) &= v_h^0. \end{aligned} \tag{20}$$

Analogously to the continuous case we can rewrite this in a first-order formulation. With the Hilbert space $X_h = V_h \times H_h$ and

$$x_h = \begin{bmatrix} u_h \\ v_h \end{bmatrix}, \quad S_h = \begin{bmatrix} 0 & -\mathrm{I} \\ A_h & B_h \end{bmatrix}, \quad g_h\left(t, x_h(t)\right) = \begin{bmatrix} 0 \\ f_h(t, u_h(t)) \end{bmatrix},$$

(20) is equivalent to

$$\begin{aligned} x_h'(t) + S_h x_h(t) &= g_h(t, x_h(t)), \qquad t \in (0,T], \\ x_h(0) &= x_h^0. \end{aligned} \tag{21}$$

The IMEX scheme (10) applied to (20), reads

$$v_h^{n+\frac{1}{2}} = v_h^n - \frac{\tau}{2} A_h u_h^n - \frac{\tau^2}{4} A_h v_h^{n+\frac{1}{2}} - \frac{\tau}{2} B_h v_h^{n+\frac{1}{2}} + \frac{\tau}{2} f_h^n, \tag{22a}$$

$$u_h^{n+1} = u_h^n + \tau v_h^{n+\frac{1}{2}}, \tag{22b}$$

$$v_h^{n+1} = v_h^{n+\frac{1}{2}} - \frac{\tau}{2} A_h u_h^n - \frac{\tau^2}{4} A_h v_h^{n+\frac{1}{2}} - \frac{\tau}{2} B_h v_h^{n+\frac{1}{2}} + \frac{\tau}{2} f_h^{n+1}, \tag{22c}$$

where we used the short notation $f_h^n := f_h(t_n, u_h^n)$. As in the continuous case, (22c) can be replaced by the more efficient update

$$v_h^{n+1} = -v_h^n + 2 v_h^{n+\frac{1}{2}} + \frac{\tau}{2} \left(f_h^{n+1} - f_h^n \right). \tag{22d}$$

3.2 Error analysis

We first introduce some notation that is required for the unified error analysis presented in [9, 10].

To relate the discrete and the continuous solution we assume that there exists a lift operator $\mathcal{L}_h^V \in \mathcal{L}(V_h; V)$ which satisfies

$$\|\mathcal{L}_h^V v_h\|_m \le C_H \|v_h\|_{m_h}, \qquad \|\mathcal{L}_h^V v_h\|_{\widetilde{a}} \le C_V \|v_h\|_{\widetilde{a}_h}, \tag{23}$$

for all $v_h \in V_h$ with constants $C_H, C_V > 0$ which are independent of h.

The adjoints

$$\mathcal{L}_h^{H*} : H \to V_h \quad \text{and} \quad \mathcal{L}_h^{V*} : V \to V_h$$

of $\mathcal{L}_h^V$ play an important role in the error analysis. They are defined via

$$\begin{aligned} m_h\big(\mathcal{L}_h^{H*} v, w_h\big) &= m\big(v, \mathcal{L}_h^V w_h\big) \qquad && \text{for all } v \in H, w_h \in H_h, \\ \widetilde{a}_h\big(\mathcal{L}_h^{V*} v, w_h\big) &= \widetilde{a}\big(v, \mathcal{L}_h^V w_h\big) && \text{for all } v \in V, w_h \in V_h. \end{aligned}$$

The corresponding first-order operators $\mathcal{L}_h : X_h \to X$ and $\mathcal{L}_h^* : X \to X_h$ are defined as

$$\mathcal{L}_h \begin{bmatrix} v_h \\ w_h \end{bmatrix} = \begin{bmatrix} \mathcal{L}_h^V v_h \\ \mathcal{L}_h^V w_h \end{bmatrix}, \qquad \mathcal{L}_h^* \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} \mathcal{L}_h^{V*} v \\ \mathcal{L}_h^{H*} w \end{bmatrix}.$$

Let $Z^V \overset{d}{\hookrightarrow} V$ be a dense subspace and $I_h \in \mathcal{L}(Z^V; V_h)$ be an interpolation operator satisfying

$$\|I_h\|_{H_h \leftarrow Z^V} \le \widehat{C}_I$$

with $\widehat{C}_I > 0$ independent of h. We define $Z = V \times Z^V \overset{d}{\hookrightarrow} X$ and the first-order reference operator $J_h : Z \to X_h$ by

$$J_h \begin{bmatrix} v \\ w \end{bmatrix} := \begin{bmatrix} \mathcal{L}_h^{V*} v \\ I_h w \end{bmatrix}.$$

Furthermore, for $v_h, w_h \in V_h$, the errors in the scalar products are defined via.

$$\begin{aligned}\Delta m\big(v_h, w_h\big) &:= m\big(\mathcal{L}_h^V v_h, \mathcal{L}_h^V w_h\big) - m_h\big(v_h, w_h\big),\\ \Delta \widetilde{a}\big(v_h, w_h\big) &:= \widetilde{a}\big(\mathcal{L}_h^V v_h, \mathcal{L}_h^V w_h\big) - \widetilde{a}_h\big(v_h, w_h\big).\end{aligned}$$

For $z = \begin{bmatrix} v \\ w \end{bmatrix} \in Z$, the discretization errors in the linear operator and the nonlinearity are given by the following remainder terms:

$$R_h z = (\mathcal{L}_h^* S - S_h J_h)\, z = \begin{bmatrix} -(\mathcal{L}_h^{V*} - I_h) w \\ \mathcal{L}_h^{H*}(Av + Bw) - (A_h \mathcal{L}_h^{V*} v + B_h I_h w) \end{bmatrix}, \tag{R1}$$

$$r_h(t,z) = \mathcal{L}_h^* g(t,z) - g_h(t, J_h z) = \begin{bmatrix} 0 \\ \mathcal{L}_h^{H*} f(t,v) - f_h(t, \mathcal{L}_h^{V*} v) \end{bmatrix}. \tag{R2}$$

Analogously to the continuous case we define the following operators:

$$\widehat{Q}_\pm := \mathrm{I} \pm \frac{\tau}{2} B_h \pm \frac{\tau^2}{4} A_h \colon V_h \to V_h, \tag{24a}$$

$$\widehat{R}_\pm := \mathrm{I} \pm \frac{\tau}{2} S_h \qquad : X_h \to X_h, \tag{24b}$$

$$\widehat{R} := \widehat{R}_+^{-1} \widehat{R}_-. \tag{24c}$$

Since the setting is the same as in the continuous case, Lemmas 2.6 and 2.3 transfer directly to the discrete case with the continuous constants replaced by the discrete ones.

Our analysis relies on the following regularity assumptions.

Assumption 3.2 Let the assumptions of Theorem 2.2 be satisfied and let u be the classical solution of (5) that satisfies additionally

$$u \in C^4\left([0,T], H\right) \cap C^3\left([0,T], V\right) \cap C^2\left([0,T], Z^V\right) \quad \text{and} \quad f(u) \in C^2([0,T], H)$$

for a $T < t^*(u^0, v^0)$.

We now state our main result.

Theorem 3.3 *Let Assumptions* 3.1 *and* 3.2 *be satisfied. For T given in Assumption* 3.2 *we set*

$$\rho = \max\left\{C_V \|u\|_{L^\infty([0,T];V)},\ \max_{t_n \le T} \|u_h^n\|_{\widetilde{a}_h}\right\}$$

and define $\widehat{c}_{\mathrm{qm}} = \widehat{\alpha}_{\mathrm{G}} \widehat{C}_{\mathrm{emb}}/2 + \widehat{\beta}_{\mathrm{qm}}$. If τ satisfies the CFL condition

$$\max\{\tau\big(1 + (3/2)^{1/2}\big)\widehat{L}_{T,\rho},\ \tau \frac{\widehat{c}_{\mathrm{qm}}}{2},\ \frac{\tau^2}{2}\widehat{\alpha}_{\mathrm{G}} + \tau \widehat{\beta}_{\mathrm{qm}}\} < 1,$$

then for all $n > 0$ with $t_n < T$, the fully discrete approximations u_h^n, v_h^n given by the scheme (22) *satisfies the error bound*

$$\|\mathcal{L}_h^V u_h^n - u(t_n)\|_{\widetilde{a}} + \|\mathcal{L}_h^V v_h^n - u'(t_n)\|_m \le C \mathrm{e}^{\widehat{M} t_n} \left(\sum_{i=0}^4 E_{h,i} + \tau^2 E\right) \tag{25}$$

with

$$\widehat{M} = \widehat{c}_{\mathrm{qm}} + \frac{\big(1 + (3/2)^{1/2}\big)\widehat{L}_{T,\rho}}{1 - \big(1 + (3/2)^{1/2}\big)\widehat{L}_{T,\rho}\,\tau}$$

and a constant C that only depends on T, but which is independent of τ, h, $\widehat{L}$, and u. The constants $E_{h,i} = E_{h,i}(u)$ contain the abstract space discretization errors and are given by

$$\begin{aligned}
E_{h,0} =& \|u_h^0 - \mathcal{L}_h^{V*} u^0\|_{\widetilde{a}_h} + \|v_h^0 - I_h v^0\|_{m_h}, \\
E_{h,1} =& \|\mathcal{L}_h^{H*} f(\cdot,u) - f_h(\cdot,\mathcal{L}_h^{V*} u)\|_{L^\infty([0,T];H_h)}, \\
E_{h,2} =& \|(\mathrm{I} - \mathcal{L}_h^V I_h)u\|_{L^\infty([0,T];V)} + \|(\mathrm{I} - \mathcal{L}_h^V I_h)u'\|_{L^\infty([0,T];V)} + \|(\mathrm{I} - \mathcal{L}_h^V I_h)u''\|_{L^\infty([0,T];H)} \\
&+ \|(\mathrm{I} - \mathcal{L}_h^V I_h) f(\cdot,u)\|_{L^\infty([0,T];V)}, \\
E_{h,3} =& \Big\| \max_{\|\varphi_h\|_{\widetilde{a}_h}=1} \Delta\widetilde{a}\big(I_h u, \varphi_h\big) \Big\|_{L^\infty(0,t)} + \Big\| \max_{\|\psi_h\|_{m_h}=1} \Delta m\big(I_h u, \psi_h\big) \Big\|_{L^\infty(0,t)} \\
&+ \Big\| \max_{\|\varphi_h\|_{\widetilde{a}_h}=1} \Delta\widetilde{a}\big(I_h u', \varphi_h\big) \Big\|_{L^\infty(0,t)} + \Big\| \max_{\|\psi_h\|_{m_h}=1} \Delta m\big(I_h u'', \psi_h\big) \Big\|_{L^\infty(0,t)} \\
E_{h,4} =& \Big\| \max_{\|\psi_h\|_{m_h}=1} \big| b\big(u', \mathcal{L}_h^V \psi_h\big) - b_h\big(I_h u', \psi_h\big) \big| \Big\|_{L^\infty(0,t)},
\end{aligned}$$

and $E = E(u)$ is given in Theorem 2.8.

Proof All error terms arising from the space discretization can be expressed within the unified error analysis and were bounded against $E_{h,i}, i = 0, \dots, 4$ in [9, Theorem 4.8], and [10, Theorem 3.9], respectively.

For the proof of the error bound (25), we use the first-order formulation of the IMEX scheme. We use the notation

$$x_h^n = \begin{bmatrix} u_h^n \\ v_h^n \end{bmatrix}, \qquad \widetilde{x}^n = \begin{bmatrix} \widetilde{u}^n \\ \widetilde{v}^n \end{bmatrix} = \begin{bmatrix} u(t_n) \\ u'(t_n) \end{bmatrix}, \qquad \widetilde{g}_h^n = g_h(t_n, J_h \widetilde{x}^n) = \begin{bmatrix} 0 \\ f_h(t_n, \mathcal{L}_h^{V*} \widetilde{u}^n) \end{bmatrix} = \begin{bmatrix} 0 \\ \widetilde{f}_h^n \end{bmatrix}.$$

The proof consists of four main steps.

(a) Splitting of the error. The error can be split via

$$\mathcal{L}_h x_h^n - \widetilde{x}^n = \mathcal{L}_h e_h^n + (\mathcal{L}_h J_h - \mathrm{I})\widetilde{x}^n, \quad \text{where} \quad e_h^n = x_h^n - J_h \widetilde{x}^n.$$

Due to the continuity of the lift operator and [9, Theorem 4.8] we have

$$\|\mathcal{L}_h x_h^n - \widetilde{x}^n\|_X \le C \|e_h^n\|_{X_h} + \|(\mathcal{L}_h J_h - \mathrm{I})\widetilde{x}^n\|_X \le C \left(\|e_h^n\|_{X_h} + E_{h,3} + E_{h,2} \right). \tag{26}$$

Hence, it remains to bound the discrete error $\|e_h^n\|_{X_h}$.

(b) Derivation of an error recursion for e_h^n. Since the discrete operators share the properties of their continuous counterparts, we can rewrite the fully discrete scheme 22 analogously to Lemma 2.9 as

$$x_h^{n+1} = \widehat{R} x_h^n + \frac{\tau}{2} \widehat{R}_+^{-1} \big(g_h^n + g_h^{n+1} \big) + \frac{\tau^2}{4} \begin{bmatrix} \widehat{Q}_+^{-1} \big(f_h^n - f_h^{n+1} \big) \\ -\big(B_h + \frac{\tau}{2} A_h \big) \widehat{Q}_+^{-1} \big(f_h^n - f_h^{n+1} \big) \end{bmatrix}. \tag{27}$$

To derive a recursion for the discrete error, we insert $J_h x$ into the fully discrete scheme (27) and obtain

$$J_h \widetilde{x}^{n+1} = \widehat{R} J_h \widetilde{x}^n + \frac{\tau}{2} \widehat{R}_+^{-1} \left(\widetilde{g}_h^{n+1} + \widetilde{g}_h^n \right) + \frac{\tau^2}{4} \begin{bmatrix} \widehat{Q}_+^{-1} \big(\widetilde{f}_h^n - \widetilde{f}_h^{n+1} \big) \\ -\big(B_h + \frac{\tau}{2} A_h \big) \widehat{Q}_+^{-1} \big(\widetilde{f}_h^n - \widetilde{f}_h^{n+1} \big) \end{bmatrix} - \Delta_h^{n+1} \tag{28}$$

with a defect Δ_h^{n+1} which is yet to be determined. We can interpret Δ_h^{n+1} as a perturbation of the defect $\Delta_{h,\mathrm{CN}}^{n+1}$ of the fully discrete Crank–Nicolson scheme given by

$$J_h\widetilde{x}^{n+1} = \widehat{R}J_h\widetilde{x}^n + \frac{\tau}{2}\widehat{R}_+^{-1}\left(\widetilde{g}_h^{n+1}+\widetilde{g}_h^n\right) - \Delta_{h,\mathrm{CN}}^{n+1}. \tag{29}$$

In fact we have

$$\Delta_{h,\mathrm{CN}}^{n+1} = \Delta_h^{n+1} - \widetilde{\delta}_h^{n+1}, \qquad \widetilde{\delta}_h^{n+1} = \frac{\tau^2}{4}\begin{bmatrix} \widehat{Q}_+^{-1}\big(\widetilde{f}_h^n - \widetilde{f}_h^{n+1}\big) \\ -\big(B_h+\frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(\widetilde{f}_h^n-\widetilde{f}_h^{n+1}\big)\end{bmatrix}. \tag{30}$$

A simple calculation shows that $J_h x$ inserted in the Crank–Nicolson scheme satisfies

$$J_h\widetilde{x}^{n+1} = J_h\widetilde{x}^n + \frac{\tau}{2}\left(-S_hJ_h(\widetilde{x}^n+\widetilde{x}^{n+1}) + \widetilde{g}_h^n + \widetilde{g}_h^{n+1}\right) - \delta_h^{n+1} - \mathcal{L}_h^*\delta_{\mathrm{CN}}^{n+1}, \tag{31}$$

where

$$\delta_h^{n+1} = -\left(J_h-\mathcal{L}_h^*\right)\left(\widetilde{x}^{n+1}-\widetilde{x}^n\right) + \frac{\tau}{2}R_h\big(\widetilde{x}^{n+1}+\widetilde{x}^n\big) - \frac{\tau}{2}\big(r_h(t_{n+1},\widetilde{x}^{n+1}) + r_h(t_n,\widetilde{x}^n)\big) \tag{32}$$

contains the abstract space discretization errors and

$$\delta_{\mathrm{CN}}^{n+1} = \frac{\tau}{2}\left(x'(t_{n+1})+x'(t_n)\right) - \left(\widetilde{x}^{n+1}-\widetilde{x}^n\right)$$

is the defect of the Crank–Nicolson scheme applied to the continuous equation (6). By applying $\widehat{R}_+^{-1}$ to (31) we obtain with (29) and (30)

$$\Delta_h^{n+1} = \widehat{R}_+^{-1}\delta_h^{n+1} + \widehat{R}_+^{-1}\mathcal{L}_h^*\delta_{\mathrm{CN}}^{n+1} + \widetilde{\delta}_h^{n+1}. \tag{33}$$

Subtracting (28) from (27) yields the error recursion

$$\begin{aligned} e_h^{n+1} &= \widehat{R}e_h^n + \frac{\tau}{2}\widehat{R}_+^{-1}\left(g_h^{n+1}-\widetilde{g}_h^{n+1}+g_h^n-\widetilde{g}_h^n\right) \\ &\quad + \frac{\tau^2}{4}\begin{bmatrix}\widehat{Q}_+^{-1}\big(f_h^n-\widetilde{f}_h^n-f_h^{n+1}+\widetilde{f}_h^{n+1}\big) \\ -\big(B_h+\frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(f_h^n-\widetilde{f}_h^n-f_h^{n+1}+\widetilde{f}_h^{n+1}\big)\end{bmatrix} + \Delta_h^{n+1}. \end{aligned} \tag{34}$$

(c) Stability. Solving (34) gives

$$\begin{aligned} e_h^n = \widehat{R}^n e_h^0 + \sum_{m=1}^n \widehat{R}^{n-m}\Bigg(&\frac{\tau}{2}\widehat{R}_+^{-1}\left(g_h^m-\widetilde{g}_h^m+g_h^{m-1}-\widetilde{g}_h^{m-1}\right) \\ &+ \frac{\tau^2}{4}\begin{bmatrix}\widehat{Q}_+^{-1}\big(f_h^{m-1}-\widetilde{f}_h^{m-1}-f_h^m+\widetilde{f}_h^m\big) \\ -\big(B_h+\frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(f_h^{m-1}-\widetilde{f}_h^{m-1}-f_h^m+\widetilde{f}_h^m\big)\end{bmatrix} + \Delta_h^{n+1}\Bigg). \end{aligned}$$

Taking the norm, using the triangle inequality, and $\|\widehat{R}\|_{X\leftarrow X} \le e^{\tau\widehat{c}_{\mathrm{qm}}}$ yields

$$\begin{aligned} \|e_h^n\|_{X_h} = \tau\sum_{m=1}^n \mathrm{e}^{(n-m)\tau\widehat{c}_{\mathrm{qm}}}\Bigg(&\frac{1}{2}\big\|\widehat{R}_+^{-1}\big(g_h^m-\widetilde{g}_h^m\big)\big\|_{X_h} + \frac{1}{2}\|\widehat{R}_+^{-1}\big(g_h^{m-1}-\widetilde{g}_h^{m-1}\big)\|_{X_h} \\ &+ \frac{\tau}{4}\left\|\begin{bmatrix}\widehat{Q}_+^{-1}\big(f_h^{m-1}-\widetilde{f}_h^{m-1}\big) \\ -\big(B_h+\frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(f_h^{m-1}-\widetilde{f}_h^{m-1}\big)\end{bmatrix}\right\|_{X_h} \\ &+ \frac{\tau}{4}\left\|\begin{bmatrix}\widehat{Q}_+^{-1}\big(f_h^m-\widetilde{f}_h^m\big) \\ -\big(B_h+\frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(f_h^m-\widetilde{f}_h^m\big)\end{bmatrix}\right\|_{X_h}\Bigg) \\ &+ \left\|\widehat{R}^n e_h^0 + \sum_{m=1}^n \widehat{R}^{n-m}\Delta_h^{n+1}\right\|_{X_h}. \end{aligned}$$

We investigate the different terms in the first sum separately. For this we use the Lipschitz-continuity of the discrete nonlinearity and the bounds from Lemmas 2.6 and 2.3 for the discrete case. Note that by (36) we have $\|\mathcal{L}_h^{V*}u(t)\|_{\widetilde{a}_h} \le \rho$ for all $t \le T$. With $\tau\widehat{c}_{\mathrm{qm}} < 2$ and $\|\widehat{R}_+^{-1}\|_{X_h\leftarrow X_h} \le 1$, we obtain:

$$\big\|\widehat{R}_+^{-1}\big(g_h^m - \widetilde{g}_h^m\big)\big\|_{X_h} \le \widehat{L}_{T,\rho}\|e_h^m\|_{X_h}.$$

Using

$$\|\widehat{Q}_+^{-1}\|_{V_h\leftarrow H_h} \le \frac{\sqrt{2}}{\tau} \quad \text{and} \quad \Big\|\Big(B_h + \frac{\tau}{2}A_h\Big)\widehat{Q}_+^{-1}\Big\|_{H_h\leftarrow H_h} \le \frac{2}{\tau},$$

we have for $\frac{\tau^2}{2}\widehat{\alpha}_{\mathrm{G}} + \tau\widehat{\beta}_{\mathrm{qm}} \le 1$

$$\left\|\begin{bmatrix} \widehat{Q}_+^{-1}\big(f_h^m - \widetilde{f}_h^m\big) \\ -\big(B_h + \frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(f_h^m - \widetilde{f}_h^m\big)\end{bmatrix}\right\|_{X_h} \le \frac{\widehat{L}_{T,\rho}}{\tau}\sqrt{6}\|e_h^m\|_{X_h}.$$

With $C_{3/2} = 1 + (3/2)^{1/2}$, this yields

$$\mathrm{e}^{-n\tau\widehat{c}_{\mathrm{qm}}}\|e_h^n\|_{X_h} \le C_{3/2}\widehat{L}_{T,\rho}\,\tau\sum_{m=1}^{n}\mathrm{e}^{-m\tau\widehat{c}_{\mathrm{qm}}}\|e_h^m\|_{X_h} + \mathrm{e}^{-n\tau\widehat{c}_{\mathrm{qm}}}\left\|\widehat{R}^n e_h^0 + \sum_{m=1}^{n}\widehat{R}^{n-m}\Delta_h^{n+1}\right\|_{X_h}.$$

By applying a discrete Gronwall Lemma, multiplying by $\mathrm{e}^{n\tau\widehat{c}_{\mathrm{qm}}}$, and inserting (33), we obtain for

$$\tau \le \frac{1}{C_{3/2}L_{T,\rho}}$$

$$\begin{aligned}
\|e_h^n\|_{X_h} &\le \mathrm{e}^{\frac{C_{3/2}\widehat{L}_{T,\rho}n\tau}{1-C_{3/2}\widehat{L}_{T,\rho}\tau}}\left(\left\|\widehat{R}^n e_h^0 + \sum_{m=1}^{n}\widehat{R}^{n-m}\left(\widehat{R}_+^{-1}\delta_h^m + \widehat{R}_+^{-1}\mathcal{L}_h^*\delta_{\mathrm{CN}}^m + \widetilde{\delta}_h^m\right)\right\|_{X_h}\right) \\
&\le \mathrm{e}^{\frac{C_{3/2}\widehat{L}_{T,\rho}t_n}{1-C_{3/2}\widehat{L}_{T,\rho}\tau}}\left(\mathrm{e}^{n\tau\widehat{c}_{\mathrm{qm}}}\left(\|e_h^0\|_{X_h} + \sum_{m=1}^{n}\left(\|\delta_h^m\|_{X_h} + \|\mathcal{L}_h^*\delta_{\mathrm{CN}}^m\|_{X_h}\right)\right)\right. \\
&\qquad \left. + \left\|\sum_{m=1}^{n}\widehat{R}^{n-m}\widetilde{\delta}_h^m\right\|_{X_h}\right).
\end{aligned} \tag{35}$$

(d) Defects. The initial error e_h^0 is given by the discretization errors of the initial values and bounded by

$$\|e_h^0\|_{X_h} \le CE_{h,0}.$$

From (32) we obtain for the defect containing the space discretization errors

$$\left\|\delta_h^m\right\|_{X_h} = \tau\left\|\frac{1}{\tau}\int_{t_{m-1}}^{t_m}(J_h - \mathcal{L}_h^*)\,x'(s)\,\mathrm{d}s + \tfrac{1}{2}R_h(\widetilde{x}^{n+1} + \widetilde{x}^n) - \frac{1}{2}\Big(r_h(t_{n+1},\widetilde{x}^{n+1}) + r_h(t_n,\widetilde{x}^n)\Big)\right\|_{X_h}.$$

By [9, Theorem 4.8] and [10, Theorem 3.9] we have

$$\left\|\delta_h^m\right\|_{X_h} \le C\tau\sum_{i=1}^{4}E_{h,i}.$$

The Crank–Nicolson defect was bounded in [8, Lemma 2.15] by

$$
\begin{aligned}
\|\mathcal{L}_h^*\delta_{\mathrm{CN}}^m\|_{X_h} \le C\|\delta_{\mathrm{CN}}^m\|_X &\le C\tau^3\|x^{(3)}\|_{L^\infty([t_m,t_{m-1}];X)} \\
&\le C\tau^3\left(\|u^{(3)}\|_{L^\infty([t_m,t_{m-1}];V)} + \|u^{(4)}\|_{L^\infty([t_m,t_{m-1}];H)}\right) \\
&\le C\tau^3 E.
\end{aligned}
$$

To bound the additional IMEX defect we split him into

$$
\widetilde{\delta}_h^m = \frac{\tau^2}{4}\begin{bmatrix} \widehat{Q}_+^{-1}\big(\widetilde{f}_h^{m-1} - \widetilde{f}_h^m\big) \\ -\big(B_h + \frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(\widetilde{f}_h^{m-1} - \widetilde{f}_h^m\big)\end{bmatrix} = \widetilde{\delta}_{h,1}^m + \widetilde{\delta}_{h,2}^m + \widetilde{\delta}_{h,3}^m
$$

with

$$
\begin{aligned}
\widetilde{\delta}_{h,1}^m &= \frac{\tau^2}{4}\begin{bmatrix} \widehat{Q}_+^{-1}\big(\widetilde{f}_h^{m-1} - \mathcal{L}_h^{H*}\widetilde{f}^{m-1} - \widetilde{f}_h^m + \mathcal{L}_h^{H*}\widetilde{f}^m\big) \\ -\big(B_h + \frac{\tau}{2}A_h\big)\widehat{Q}_+^{-1}\big(\widetilde{f}_h^{m-1} - \mathcal{L}_h^{H*}\widetilde{f}^{m-1} - \widetilde{f}_h^m + \mathcal{L}_h^{H*}\widetilde{f}^m\big)\end{bmatrix}, \\
\widetilde{\delta}_{h,2}^m &= \frac{\tau}{4}\begin{bmatrix} \tau\widehat{Q}_+^{-1}\mathcal{L}_h^{H*}\big(\widetilde{f}^{m-1} - \widetilde{f}^m\big) \\ \widehat{Q}_-\widehat{Q}_+^{-1}\mathcal{L}_h^{H*}\big(\widetilde{f}^{m-1} - \widetilde{f}^m\big)\end{bmatrix}, \\
\widetilde{\delta}_{h,3}^m &= \frac{\tau}{4}\begin{bmatrix} 0 \\ -\mathcal{L}_h^{H*}\big(\widetilde{f}^{m-1} - \widetilde{f}^m\big)\end{bmatrix},
\end{aligned}
$$

where we used the additional notation $\widetilde{f}^m = f(t_m, \widetilde{u}^m)$. The first term is bounded by $\|\widetilde{\delta}_{h,1}^m\|_{X_h} \le E_{h,1}$. The terms $\widetilde{\delta}_{h,2}^m$ and $\widetilde{\delta}_{h,3}^m$ are only of order τ^2, which is not sufficient to obtain a global error of order two. Moreover, a combination of both terms from two successive time steps allows to gain an additional factor of τ. With the explicit representation of $\widehat{R}$ analogous to that of R in (14b), we obtain

$$
\widetilde{\delta}_{h,2}^m + \widehat{R}\widetilde{\delta}_{h,3}^{m-1} = \frac{\tau}{2}\begin{bmatrix} \frac{\tau}{2}\widehat{Q}_+^{-1}\big(-\widetilde{f}^{m-2} + 2\widetilde{f}^{m-1} - \widetilde{f}^m\big) \\ \frac{1}{2}\widehat{Q}_-\widehat{Q}_+^{-1}\big(-\widetilde{f}^{m-2} + 2\widetilde{f}^{m-1} - \widetilde{f}^m\big)\end{bmatrix}.
$$

Using this together with the bound bounds for $\widehat{Q}_+^{-1}$ and $\widehat{Q}_-\widehat{Q}_+^{-1}$ from Lemma 2.6 and the continuity of the adjoint lift operator, leads to the bound

$$
\begin{aligned}
\|\widetilde{\delta}_{h,2}^m + \widehat{R}\widetilde{\delta}_{h,3}^{m-1}\|_{X_h} &\le C\tau\big\|\mathcal{L}_h^{H*}\big(-\widetilde{f}^{m-2} + 2\widetilde{f}^{m-1} - \widetilde{f}^m\big)\big\|_{m_h} \\
&\le C\tau^3\left\|\frac{\mathrm{d}^2}{\mathrm{d}t^2}\big(f(t,u(t))\big)\right\|_{L^\infty([t_{m-2},t_m];H)} \\
&\le C\tau^3 E,
\end{aligned}
$$

and hence,

$$
\begin{aligned}
\Big\|\sum_{m=1}^n \widehat{R}^{n-m}\widetilde{\delta}_h^m\Big\|_{X_h} &\le \left\|\widehat{R}^{n-1}\widetilde{\delta}_{h,1}^1 + \widetilde{\delta}_{h,2}^n + \sum_{m=2}^n \widehat{R}^{n-m}\big(\widetilde{\delta}_{h,2}^m + \widehat{R}\widetilde{\delta}_{h,3}^{m-1}\big)\right\|_{X_h} \\
&\le \mathrm{e}^{n\tau\widehat{c}_{\mathrm{qm}}}\left(\|\widetilde{\delta}_{h,1}^1\|_{X_h} + \|\widetilde{\delta}_{h,2}^n\|_{X_h} + \sum_{m=2}^n \|\widetilde{\delta}_{h,2}^m + \widehat{R}\widetilde{\delta}_{h,3}^{m-1}\|_{X_h}\right) \\
&\le C\mathrm{e}^{n\tau\widehat{c}_{\mathrm{qm}}}\tau^2 E.
\end{aligned}
$$

Inserting all bounds into (35) yields

$$\|e_h^n\|_X \leq C\mathrm{e}^{\widehat{M}t_n}\left(E_{h,0}\sum_{i=1}^{4}E_{h,i}+\tau^2 E\right).$$

Finally, the error bound (25) follows from

$$\|\mathcal{L}_h^V u_h^n - u(t_n)\|_{\widetilde{a}} + \|\mathcal{L}_h^V v_h^n - u'(t_n)\|_m \leq \sqrt{2}\|\mathcal{L}_h x_h^n - x(t_n)\|_X$$

and (26). □

Remark 3.4

a) Theorem 2.8 can be proven by replacing all discrete quantities by their continuous counterparts in the proof of Theorem 3.3. Since all assumptions in the discrete setting remain valid in the continuous case, the proof applies to the latter as well and all space discretization errors vanish.
b) Theorem 3.3 and Corollary 3.5 are also valid for the Crank–Nicolson scheme. In this case, the error recursion (34) simplifies to

$$e_h^{n+1} = \widehat{R}e_h^n + \frac{\tau}{2}\widehat{R}_+^{-1}\left(g_h^{n+1} - \widetilde{g}_h^{n+1} + g_h^n - \widetilde{g}_h^n\right) + \delta_h^{n+1} + \widehat{R}_+^{-1}\mathcal{L}_h^*\delta_{\mathrm{CN}}^{n+1}$$

and the error bound (25) holds with $E_{\mathrm{CN}}(u) = \|u^{(4)}\|_{L^\infty([0,T];H)} + \|u^{(3)}\|_{L^\infty([0,T];V)}$ (instead of E) and $1+(3/2)^{1/2}$ replaced by 1 in the CFL condition and the error bound.

Under additional consistency assumptions for the space discretization, the following corollary states that for sufficiently small τ and h, the fully discrete approximations are bounded in terms of the exact solution and converge.

Corollary 3.5 *Let the assumptions of Theorem* 3.3 *be satisfied.*

a) For T given in Assumption 3.2 *define*

$$\rho = 2C_V\|u\|_{L^\infty([0,T];V)}. \tag{36}$$

If $E_{h,i} \stackrel{h\to 0}{\longrightarrow} 0$ for $i = 0,\dots,4$, then there exist $\tau^*, h^* > 0$ s.t. for all $h < h^*, \tau < \tau^*$ we have

$$\max_{t_n\leq T}\|u_h^n\|_{\widetilde{a}_h} \leq \rho,$$

and the fully discrete solution converges, i.e.,

$$\max_{t_n\leq T}\{\|\mathcal{L}_h^V u_h^n - u(t_n)\|_{\widetilde{a}} + \|\mathcal{L}_h^V v_h^n - u'(t_n)\|_m\} \to 0, \qquad \tau, h\to 0.$$

b) If additionally $E_{h,i} \leq h^k$ for a $k > 0$ and $i = 0,\dots,4$, we obtain the error bound

$$\max_{t_n\leq T}\{\|\mathcal{L}_h^V u_h^n - u(t_n)\|_{\widetilde{a}} + \|\mathcal{L}_h^V v_h^n - u'(t_n)\|_m\} \leq C\mathrm{e}^{\widehat{M}t_n}(\tau^2 + h^k) \tag{37}$$

with $\widehat{M}$ defined in Theorem 3.3 *and a constant C independent of τ and h.*

Proof For the proof of Corollary 3.5 it remains to show that u_h^n is bounded. This can be done by contradiction using the boundedness of the exact solution u and the convergence of all error terms in (25). The other assertions follow then directly from the error bound (25). □

4 Application: Semilinear wave equation with kinetic boundary conditions

In this section we consider the IMEX scheme applied to a finite element discretization of a semilinear acoustic wave equation with a kinetic boundary condition. Kinetic boundary conditions serve as an effective model for the interaction of waves with a boundary covered by a thin layer. A derivation can be found in, e.g., [7], and the wellposedness was proven in [17]. The space discretization we present in this section was analyzed in [8,9,15].

We show that this example fits into the abstract theory presented in the previous sections.

4.1 Formulation of the equations

Let $\Omega \subset \mathbb{R}^d, d = 2,3$, be a bounded domain with smooth boundary $\Gamma = \partial\Omega$. With Δ_Γ we denote the Laplace-Beltrami operator on Γ and with $\mathbf{n}$ the outer normal vector.

We consider the following semilinear acoustic wave equations with kinetic boundary conditions. Seek $u\colon [0,T] \times \overline{\Omega} \to \mathbb{R}$ satisfying

$$\begin{cases} u_{tt} + (\alpha_\Omega + \beta_\Omega \cdot \nabla)\, u_t - \Delta u = f_\Omega(t,\mathbf{x},u), & \text{in } (0,T)\times\Omega, \\ u_{tt} + \partial_{\mathbf{n}} u + (\alpha_\Gamma + \beta_\Gamma \cdot \nabla_\Gamma)\, u_t - \Delta_\Gamma u = f_\Gamma(t,\mathbf{x},u), & \text{in } (0,T)\times\partial\Omega, \\ u(0,\mathbf{x}) = u^0(\mathbf{x}), \quad u_t(0,\mathbf{x}) = v^0(\mathbf{x}), & \text{in } \overline{\Omega}, \end{cases} \tag{38}$$

where the nonlinearities and the coefficients satisfy the following conditions.

Assumption 4.1

a) The nonlinearities satisfy

$$f_\Omega \in C^1([0,T]\times\overline{\Omega}\times\mathbb{R};\mathbb{R}), \qquad f_\Gamma \in C^1([0,T]\times\Gamma\times\mathbb{R};\mathbb{R}). \tag{39}$$

Moreover, they satisfy the following growth condition, that there exist

$$\zeta_\Omega \begin{cases} < \infty, & d=2, \\ \le \frac{d}{d-2}, & d\ge 3, \end{cases} \qquad \text{and} \qquad \zeta_\Gamma \begin{cases} < \infty, & d=2,3, \\ \le \frac{d-1}{d-3}, & d\ge 4, \end{cases} \tag{40}$$

such that for all $(t,\mathbf{x},u) \in [0,T]\times\Omega\times\mathbb{R}$

$$|f_\Omega(t,\mathbf{x},u)| \le C(1+|u|^{\zeta_\Omega}), \qquad |\nabla f_\Omega(t,\mathbf{x},u)| \le C(1+|u|^{\zeta_\Omega - 1}), \tag{41}$$

and for all $(t,\mathbf{x},u) \in [0,T]\times\Gamma\times\mathbb{R}$

$$|f_\Gamma(t,\mathbf{x},u)| \le C(1+|u|^{\zeta_\Gamma}), \qquad |\nabla f_\Gamma(t,\mathbf{x},u)| \le C(1+|u|^{\zeta_\Gamma - 1})$$

hold true.

b) The coefficients $\alpha_\Omega \in C(\overline{\Omega}), \beta_\Omega \in C^1(\overline{\Omega})^d, \alpha_\Gamma \in C(\Gamma)$ and $\beta_\Gamma \in C^1(\Gamma)^d$ are non-negative and satisfy

$$\alpha_\Omega - \frac{1}{2}\operatorname{div}\beta_\Omega \ge 0 \quad \text{in } \Omega, \qquad \alpha_\Gamma + \frac{1}{2}(\beta_\Omega \cdot n - \operatorname{div}_\Gamma \beta_\Gamma) \ge 0 \quad \text{on } \Gamma.$$

In [10] was shown that the weak formulation of (38) is of the form (2) with

$$
\begin{aligned}
H &= L^2(\Omega)\times L^2(\Gamma)\\
V &= H^1(\Omega;\Gamma) := \{v\in H^1(\Omega)\mid \gamma(v)\in H^1(\Gamma)\}\subset H^1(\Omega)\times H^1(\Gamma),\\
m\big(v,\varphi\big) &= \int_\Omega v\varphi\,\mathrm{d}x + \int_\Gamma v\varphi\,\mathrm{d}s,\\
b\big(v,\varphi\big) &= \int_\Omega (\alpha_\Omega v+\beta_\Omega\cdot\nabla v)\,\varphi\,\mathrm{d}x + \int_\Gamma (\alpha_\Gamma v+\beta_\Gamma\cdot\nabla_\Gamma v)\,\varphi\,\mathrm{d}s,\\
a\big(v,\varphi\big) &= \int_\Omega \nabla v\nabla\varphi\,\mathrm{d}x + \int_\Gamma \nabla_\Gamma v\nabla_\Gamma\varphi\,\mathrm{d}s,
\end{aligned}
\tag{42}
$$

and $f\colon [0,T]\times V\to H$ defined via

$$
m\big(f(t,v),\varphi\big) = \int_\Omega (f_\Omega(t,\cdot,v(\cdot)))\,\varphi\,\mathrm{d}x + \int_\Gamma (f_\Gamma(t,\cdot,v(\cdot)))\,\varphi\,\mathrm{d}s.
$$

Furthermore, Assumption 2.1 is satisfied and we have $D(A)=H^2(\Omega;\Gamma)$. Thus, Theorem 2.2 yields the existence of a solution u of (38) .

4.2 Space discretization

As in [10], we use the bulk-surface finite element method presented in [5] to discretize (38) in space. This discretization was also considered in [8,9] for linear problems.

We start by giving a short summary of this method and refer to [5,8] for more details.

Bulk-surface finite element method

Let $\mathcal{T}_h$ be a consistent quasi-uniform mesh of isoparametric elements K of degree p with maximal mesh width h. The discretized domain and its boundary are denoted by

$$
\Omega_h := \bigcup_{K\in\mathcal{T}_h} K \approx \Omega \qquad \text{and} \qquad \Gamma_h := \partial\Omega_h.
$$

We define the bulk and the surface finite element space of order $p\geq 1$ via

$$
\begin{aligned}
V_{h,p}^{\Omega} &:= \big\{v_h\in C(\Omega_h)\mid v_h\big|_K = \widehat{v}_h\circ(F_K)^{-1}\text{ with }\widehat{v}_h\in\mathbb{P}_p(\widehat{K})\text{ for all }K\in\mathcal{T}_h\big\},\\
V_{h,p}^{\Gamma} &:= \big\{\vartheta_h\in C(\Gamma_h)\mid \vartheta_h = v_h\big|_{\Gamma_h}\text{ with }v_h\in V_{h,p}^{\Omega}\big\}.
\end{aligned}
$$

Here $\mathbb{P}_p(\widehat{K})$ denotes the space of polynomials of degree p on a reference triangle $\widehat{K}$ and F_K is a transformation from $\widehat{K}$ to K. This discretization is nonconforming because $\Omega_h\neq\Omega$. In [5], an elementwise smooth homeomorphism $G_h\colon\Omega_h\to\Omega$ with

$$
G_h\big|_K\in C^{p+1}(K), \qquad \text{for all } p\leq k \text{ and } K\in\mathcal{T}_h
$$

is constructed. This allows us to define lifted versions of $v_h\in V_{h,p}^{\Omega}$ and $\vartheta_h\in V_{h,p}^{\Gamma}$ as

$$
v_h^\ell := v_h\circ G_h^{-1} \qquad \text{and} \qquad \vartheta_h^\ell := \vartheta_h\circ G_h^{-1}. \tag{43}
$$

The mapping G_h is constructed in such a way, that $G_h(a_i)=a_i$, $i=1,\dots,N=\dim V_h$, where $a_1,\dots,a_N\in\Omega_h$ are the nodes corresponding to the finite element discretization. This implies $v_h^\ell(a_i)=v_h(a_i)$ for $i=1,\dots,N$ and for all $v_h\in V_{h,p}^{\Omega}$.

By $I_{h,\Omega}\colon C(\overline{\Omega}) \to V_{h,p}^{\Omega}$ and $I_{h,\Gamma}\colon C(\Gamma) \to V_{h,p}^{\Gamma}$ we denote the nodal interpolation operator in Ω and on Γ, respectively. By construction, the nodes on the surface coincide with the bulk nodes and therefore we have

$$\gamma(I_{h,\Omega} v) = I_{h,\Gamma}\gamma(v) \qquad \text{for all } v \in C(\overline{\Omega}).$$

The semidiscrete equation

We now present the space discretization in the framework of Section 3. As finite element space we choose $V_h = V_{h,p}^{\Omega}$. Furthermore we set $I_h := I_{h,\Omega}\big|_{Z^V}\colon Z^V \to V_h$, where

$$Z^V := D(A) = H^2(\Omega;\Gamma) \overset{d}{\hookrightarrow} V = H^1(\Omega;\Gamma), \tag{44}$$

and define the lift operator via

$$\mathcal{L}_h^V v := v^\ell$$

with v^ℓ given in (43). The spatial discretization of (38) can then be written as (19) where the discretizations $m_h, b_h, a_h\colon V_h \times V_h \to \mathbb{R}$ of m, b, and a are defined via

$$\begin{aligned}
m_h\big(v_h,\varphi_h\big) &:= \int_{\Omega_h} v_h\varphi_h \,\mathrm{d}x + \int_{\Gamma_h} v_h\varphi_h \,\mathrm{d}s,\\
b_h\big(v_h,\varphi_h\big) &:= \int_{\Omega_h} \big(I_{h,\Omega}\alpha_\Omega v_h + I_{h,\Omega}\beta_\Omega\cdot\nabla v_h\big)\,\varphi_h \,\mathrm{d}x + \int_{\Gamma_h} \big(I_{h,\Gamma}\alpha_\Gamma v_h + I_{h,\Gamma}\beta_\Gamma\cdot\nabla_\Gamma v_h\big)\,\varphi_h \,\mathrm{d}s,\\
a_h\big(v_h,\varphi_h\big) &:= \int_{\Omega_h} \nabla v_h \nabla\varphi_h \,\mathrm{d}x + \int_{\Gamma_h} \nabla_{\Gamma_h} u_h \nabla_{\Gamma_h}\varphi_h \,\mathrm{d}s.
\end{aligned}$$

The discretized nonlinearity $f_h\colon [0,T]\times V_h \to H_h$ is given by

$$m_h\big(f_h(t,v_h),\varphi_h\big) := \int_{\Omega_h} I_{h,\Omega} f_\Omega\big(t,\cdot,v_h^\ell(\cdot)\big)(\mathbf{x})\varphi_h(\mathbf{x})\,\mathrm{d}x + \int_{\Gamma_h} I_{h,\Gamma} f_\Gamma\big(t,\cdot,v_h^\ell(\cdot)\big)(\mathbf{x})\varphi_h(\mathbf{x})\,\mathrm{d}s \tag{45}$$

for all $\varphi_h \in V_h$.

In [10] it was shown, that Assumption 3.1 is satisfied.

Remark 4.2 The nodal interpolation only requires function evaluations in the nodes $a_1,\dots,a_N$. Since they are invariant under the lift operator, the computation of v_h^ℓ is not necessary. It is only needed for the definition of f_h since the interpolation operator acts on functions defined on Ω.

4.3 Full discretization error bound

We now state an error bound for the full discretization of (38) with the bulk-surface finite element method and the IMEX scheme (22).

Corollary 4.3 *Let* $1 \le p \le k$ *and* $\Gamma \in C^{k+1}$. *Furthermore let Assumption* 4.1 *be satisfied and let* u *be a solution of* (38) *on* $[0,T]$ *satisfying*

$$\begin{aligned}
u &\in C^4([0,T];L^2(\Omega)\times L^2(\Gamma)) \cap C^3([0,T];H^1(\Omega;\Gamma)) \cap C^2([0,T];H^2(\Omega;\Gamma)),\\
u,u' &\in L^\infty\big([0,T];H^{p+1}(\Omega;\Gamma)\big),\\
u'' &\in L^\infty\big([0,T];H^{p}(\Omega;\Gamma)\big),\\
f_\Omega(t,\cdot,u(t,\cdot)) &\in L^\infty\big([0,T];H^{\max\{2,p\}}(\Omega)\big),\\
f_\Gamma(t,\cdot,u(t,\cdot)) &\in L^\infty\big([0,T];H^{\max\{2,p\}}(\Gamma)\big).
\end{aligned}$$

Then there exist $\tau^*, h^*, \rho > 0$ s.t. for all $0 < h < h^*, 0 < \tau < \tau^*$, and $t_n \le T$, the approximations u_h^n and v_h^n given by (22) *with bulk-surface elements of order p, satisfy*

$$\|(u_h^n)^\ell - u(t_n)\|_{H^1(\Omega;\Gamma)} + \|(v_h^n)^\ell - u'(t_n)\|_{L^2(\Omega)\times L^2(\Gamma)} \le C\mathrm{e}^{\left(\frac{1}{2}+\frac{\left(1+(3/2)^{1/2}\right)\widehat{L}_{T,\rho}}{1-\left(1+(3/2)^{1/2}\right)\widehat{L}_{T,\rho}\tau}\right)t_n}(h^p+\tau^2) \tag{46}$$

with a constant C independent of τ and h. The Lipschitz constant of the discretized nonlinearity is given by

$$\widehat{L}_{T,\rho} = C\left(\sigma(\Omega)^{\frac{\zeta_\Omega-1}{2\zeta_\Omega}} + \sigma(\Gamma)^{\frac{\zeta_\Gamma-1}{2\zeta_\Gamma}} + 2\rho^{\zeta_\Omega-1} + 2\rho^{\zeta_\Gamma-1}\right), \tag{47}$$

where $\sigma(\Omega)$ and $\sigma(\Gamma)$ denote the measure of Ω and Γ, respectively, and ζ_Γ and ζ_Ω are given in Assumption 4.1.

Proof In [10] it was shown, that Assumptions 2.1 and 3.1 are satisfied with $\widehat{C}_{\mathrm{emb}} = \widehat{\alpha}_{\mathrm{G}} = 1$, $\widehat{\beta}_{\mathrm{qm}} = 0$, and $\widehat{L}_{T,\rho}$ given in (47). The regularity assumptions on u are such that $u \in C^4([0,T],H) \cap C^3([0,T],V) \cap C^2([0,T],Z^V)$, cf. (42) and (44). Since additionally $Z^V = D(A) = H^2(\Omega;\Gamma)$, we have that $Au \in C^2([0,T];H)$ und hence $f(u) = u'' + Bu' + Au \in C^2([0,T];H)$. Thus, also Assumption 3.2 is satisfied, and we can apply Corollary 3.5.

Under the above assumptions, in [9,10] it was shown that the space discretization error terms are bounded by $E_{h,i} \le Ch^p$. So the the bound (46) follows then directly by (37). □

4.4 Implementation

In the following numerical experiments we compare the IMEX scheme with the Crank–Nicolson and the explicit classical Runge–Kutta scheme of order 4 applied to the space discretized wave equation with kinetic boundary conditions. For the implementation we used the `C++` finite element-library `deal.II` [1,4]. The codes used for the experiments are available from the authors on request.

To comment on the implementation we first introduce some additional notation. For a finite element function $u_h \in V_h$ we denote by $\mathbf{u} \in \mathbb{R}^N$, the corresponding coefficient vector in the finite element basis. Furthermore, $\mathbf{M} \in \mathbb{R}^{N\times N}$ is the mass matrix, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N\times N}$ are the stiffness matrices related to A_h and B_h, respectively, and $\mathbf{f}^n \in \mathbb{R}^N$ denotes the load vector corresponding to $f_h^n = f_h(t_n, u^n)$, $n \in \mathbb{N}$.

IMEX scheme

The fully discrete IMEX scheme (22) reads

$$\mathbf{M}\mathbf{v}^{n+\frac{1}{2}} = \mathbf{M}\mathbf{v}^n - \frac{\tau}{2}\mathbf{A}\mathbf{u}^n - \frac{\tau^2}{4}\mathbf{A}\mathbf{v}^{n+\frac{1}{2}} - \frac{\tau}{2}\mathbf{B}\mathbf{v}^{n+\frac{1}{2}} + \frac{\tau}{2}\mathbf{f}^n, \tag{48a}$$

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \tau\mathbf{v}^{n+\frac{1}{2}}, \tag{48b}$$

$$\mathbf{M}\mathbf{v}^{n+1} = -\mathbf{M}\mathbf{v}^n + 2\mathbf{M}\mathbf{v}^{n+\frac{1}{2}} + \frac{\tau}{2}\left(\mathbf{f}^{n+1} - \mathbf{f}^n\right). \tag{48c}$$

The linear system in (48a) has the form

$$\mathbf{Q}_+\mathbf{v}^{n+\frac{1}{2}} = \mathbf{M}\mathbf{v}^n - \frac{\tau}{2}\mathbf{A}\mathbf{u}^n + \frac{\tau}{2}\mathbf{f}^n, \qquad \mathbf{Q}_+ = \mathbf{M} + \frac{\tau}{2}\mathbf{B} + \frac{\tau^2}{4}\mathbf{A}.$$

We solve this system with the GMRES solver provided by `deal.II` and either a sparse incomplete LU or a geometric multigrid preconditioner. For the measure of the error in the GMRES iterations, the residual r with corresponding coefficient vector $\mathbf{r}$ is used. A suitable stopping criteria would be

$$\|r\|_{\widetilde{a}_h} \leq \tau^2 \,\mathrm{tol},$$

where tol is a given tolerance, since then in (48b) the error in $\mathbf{u}^{n+1}$ caused by the solution of the linear system in the $\|\cdot\|_{\widetilde{a}_h}$ norm is of order τ^3, which corresponds to the local error of the IMEX scheme. In practice, the computation of $\|r\|_{\widetilde{a}_h}$ is quite expensive and we thus used the stopping criterion

$$\|r\|_{h,2} = \|\mathbf{r}\|_{h,2} \leq \tau^2 \,\mathrm{tol},$$

in a grid dependent scaled Euclidean norm $\|\cdot\|_{h,2} = h^{d/2}\|\cdot\|_2$. This is much more efficient, since this norm is available within the GMRES code at no additional cost. The criterion worked well in our numerical experiments as we will show in Section 4.5. We always use $\mathrm{tol} = 0.01$ in our numerical examples, which is chosen s.t. the errors in solving the linear systems do not destroy the overall order of convergence.

Note that in the IMEX scheme (22), only $\mathbf{Mv}^{n+1}$ is required so that we neither compute nor store $\mathbf{v}^{n+1}$ itself.

Crank–Nicolson scheme

The fully discrete Crank–Nicolson scheme can be written in the form

$$\mathbf{Q}_+\mathbf{u}^{n+1} - \frac{\tau^2}{4}\mathbf{f}^{n+1} = (\mathbf{M} + \frac{\tau}{2}\mathbf{B} - \frac{\tau^2}{4}\mathbf{A})\mathbf{u}^n + \tau\mathbf{M}\mathbf{v}^n + \frac{\tau^2}{4}\mathbf{f}^n, \tag{49a}$$

$$\mathbf{M}\mathbf{v}^{n+1} = \mathbf{M}\mathbf{v}^n - \frac{\tau}{2}\mathbf{A}(\mathbf{u}^n + \mathbf{u}^{n+1}) + \mathbf{B}(\mathbf{u}^n - \mathbf{u}^{n+1}) + \frac{\tau}{2}\big(\mathbf{f}^n + \mathbf{f}^{n+1}\big). \tag{49b}$$

We solve the nonlinear equation (49a) with a simplified Newton method where we use $\mathbf{Q}_+$ as an approximation to the Jacobian. The linear equations are solved as in the IMEX scheme. We stop the Newton scheme when the update Δu satisfies $\|\Delta u\|_{h,2} \leq \tau^3\widetilde{\mathrm{tol}}$ with a given tolerance $\widetilde{\mathrm{tol}}$. In the numerical examples we use $\widetilde{\mathrm{tol}} = 0.1$, which is again chosen s.t. the Newton errors do not destroy the overall order of convergence. All matrix vector products appearing in (49a) and (49b) are computed only ones and saved in temporary vectors, as well as all terms that can be reused in the next time step. As in the IMEX scheme we do not compute $\mathbf{v}^{n+1}$ but only $\mathbf{Mv}^{n+1}$.

Classical Runge–Kutta scheme

The classical Runge–Kutta scheme is an explicit scheme of order four that is suited for hyperbolic problems because its stability region contains an interval on the imaginary axis. This is in contrast to the second-order schemes by Heun and Runge, which intersect with the imaginary axis in the origin only. We implemented it using mass lumping to obtain a fully explicit scheme. Note that the space discretization with mass lumping also fits into the setting of Section 3, as it was shown in [9] for a linear acoustic wave equation.

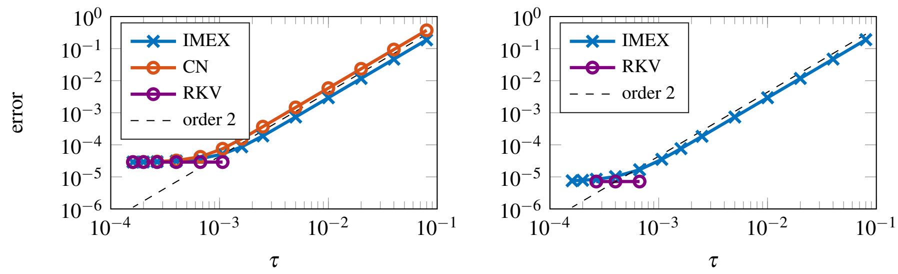


Fig. 1 Error $E_h(0.8)$ of the IMEX scheme (solved with GMRES and ILU preconditioner) and the Crank–Nicolson scheme (solved with simplified Newton method) plotted against step size τ for coarse space discretization (328 193 degrees of freedom, left) and fine space discretization (1 311 745 degrees of freedom, right)

4.5 Numerical examples

We consider the semilinear wave equation with kinetic boundary conditions (38) with $\alpha_\Omega = \beta_\Omega = 1$ and $\alpha_\Gamma = \beta_\Gamma = 0$ on the unit disc $\Omega = B(0,1) \subset \mathbb{R}^2$. As nonlinearities, we choose

$$f_\Omega(t,\mathbf{x},u) = |u|u + \eta_\Omega(t,\mathbf{x}), \qquad f_\Gamma(t,\mathbf{x},u) = |u|^2 u + \eta_\Gamma(t,\mathbf{x})$$

with

$$\eta_\Omega(t,\mathbf{x}) = -\left(4\pi^2 + |\sin(2\pi t)\mathbf{x}_1\mathbf{x}_2|\right)\sin(2\pi t)\mathbf{x}_1\mathbf{x}_2,$$
$$\eta_\Gamma(t,\mathbf{x}) = -4\pi^2\sin(2\pi t)\mathbf{x}_1\mathbf{x}_2 + 6\sin(2\pi t)\mathbf{x}_1\mathbf{x}_2 - (\sin(2\pi t)\mathbf{x}_1\mathbf{x}_2)^3,$$

and as initial values

$$u(0,\mathbf{x}) = 0, \quad u_t(0,\mathbf{x}) = 2\pi\mathbf{x}_1\mathbf{x}_2.$$

In this case the exact solution is given by

$$u(t,\mathbf{x}) = \sin(2\pi t)\mathbf{x}_1\mathbf{x}_2.$$

For the space discretization we use isoparametric elements of order $p = 2$, and choose $u_h^0 = I_{h,\Omega}u^0$ and $v_h^0 = I_{h,\Omega}v^0$ as discrete initial values.

As we cannot compute the lift of a finite element function exactly, we consider the error

$$E_h(t) := \|u_h(t) - u(t)|_{\Omega_h}\|_{H^1(\Omega_h;\Gamma_h)} + \|u_h'(t) - u'(t)|_{\Omega_h}\|_{L^2(\Omega_h)\times L^2(\Gamma_h)}.$$

We evaluate the integrals with a quadrature rule of order 4, such that the quadrature error is negligible. The restriction of u to Ω_h is possible since for convex domains we have $\Omega_h \subset \Omega$.

In Figure 1 the errors of the IMEX, the Crank–Nicolson, and the classical Runge–Kutta scheme are plotted against the time-step size τ for a coarse ($h \approx 0.014$) and a fine ($h \approx 0.007$) space discretization, respectively. As predicted by Corollary 1, the IMEX and the Crank–Nicolson scheme converge with order two until the error of the space discretization is reached. The explicit Runge–Kutta scheme is only stable under a strong CFL condition and then the error reaches immediately the space discretization error plateau.

Figure 2 shows the errors of the different schemes plotted against the runtime for the same coarse and fine space discretization as in Figure 1. It can be observed, that the IMEX scheme is significant faster than the Crank–Nicolson scheme. For errors of the magnitude of the space discretization error plateau, the classical Runge Kutta scheme is more efficient

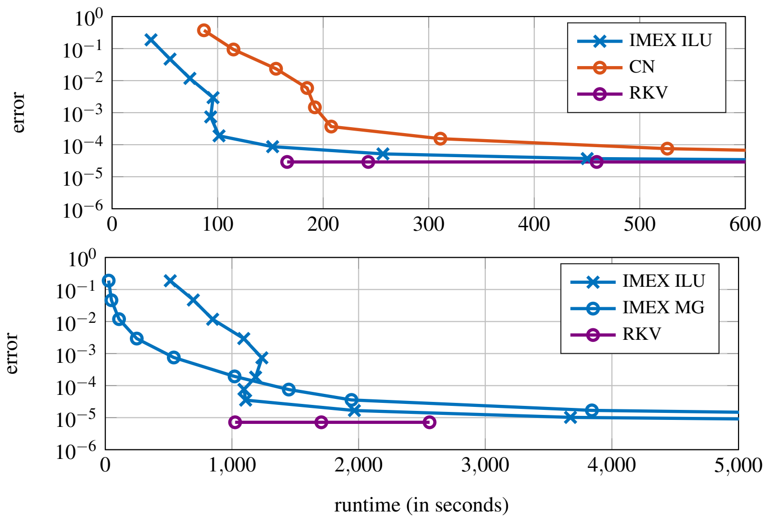


Fig. 2 Error $E_h(0.8)$ of the IMEX scheme, solved with GMRES and ILU/Multigrid(MG, F-cycle with 8 levels) preconditioner, and the Crank–Nicolson scheme (solved with simplified Newton method) plotted against runtime for coarse space discretization (328 193 degrees of freedom, top) and fine space discretization (1 311 745 degrees of freedom, bottom)

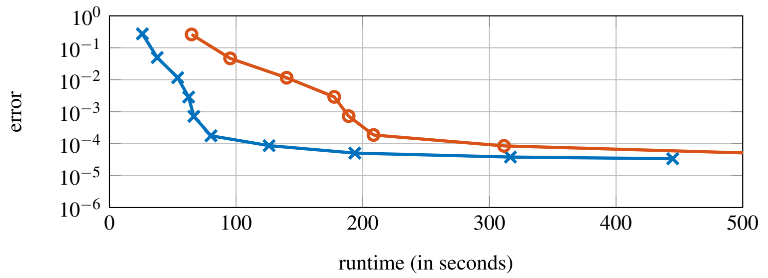


Fig. 3 Error $E_h(0.8)$ of the IMEX scheme plotted against the runtime when using the two different error estimates as stopping criteria for the GMRES scheme as discussed in Section 4.4, namely $\|r\|_{h,2}$ (blue, crosses) and $\|r\|_{\widetilde{a}_h}$ (red, circles) for a coarse space discretization (328 193 degrees of freedom)

than the IMEX scheme, but the IMEX scheme outperforms the Runge–Kutta scheme if less accuracy is sufficient. For the large system obtained by the fine space discretization and large time-step sizes, the use of the multigrid preconditioner is quite efficient. The IMEX method is significantly faster than the Crank–Nicolson method, and almost as fast as the explicit Runge–Kutta method in the non-stiff regime. The Runge Kutta method has the disadvantage that the stability limit in applications is not exactly known, and therefore there is a risk that it will not be stable if a too large time-step size is chosen, or the effort is unnecessarily high if the time-step size is too small.

Finally, Figure 3 shows a comparison of the runtimes of the IMEX scheme when using the different stopping criteria for the GMRES solver discussed in Section 4.4, namely using $\|r\|_{\widetilde{a}_h}$ or $\|r\|_{h,2}$ as estimate for the error, respectively. It can be seen that the afford of computing the (better suited) $\|r\|_{\widetilde{a}_h}$ is too high and does not pay off.

References

1. Arndt, D., Bangerth, W., Clevenger, T.C., et al.: The deal.II library, Version 9.1. J. Numer. Math. **27**(4), 203–213 (2019). DOI 10.1515/jnma-2019-0064
2. Ascher, U.M., Ruuth, S.J., Spiteri, R.J.: Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. Appl. Numer. Math. **25**(2-3), 151–167 (1997). DOI 10.1016/S0168-9274(97)00056-1. Special issue on time integration (Amsterdam, 1996)
3. Ascher, U.M., Ruuth, S.J., Wetton, B.T.R.: Implicit-explicit methods for time-dependent partial differential equations. SIAM J. Numer. Anal. **32**(3), 797–823 (1995). DOI 10.1137/0732037
4. Bangerth, W., Hartmann, R., Kanschat, G.: deal.II—a general-purpose object-oriented finite element library. ACM Trans. Math. Software **33**(4), Art. 24, 27 (2007). DOI 10.1145/1268776.1268779
5. Elliott, C.M., Ranner, T.: Finite element analysis for a coupled bulk-surface partial differential equation. IMA J. Numer. Anal. **33**(2), 377–402 (2013). DOI 10.1093/imanum/drs022
6. Gardner, D.J., Guerra, J.E., Hamon, F.P., Reynolds, D.R., Ullrich, P.A., Woodward, C.S.: Implicit–explicit (IMEX) Runge–Kutta methods for non-hydrostatic atmospheric models. Geosci. Model Dev. **11**(4), 1497 (2018)
7. Goldstein, G.R.: Derivation and physical interpretation of general boundary conditions. Adv. Differential Equations **11**(4), 457–480 (2006)
8. Hipp, D.: A unified error analysis for spatial discretizations of wave-type equations with applications to dynamic boundary conditions. Ph.D. thesis, Karlsruhe Institute of Technology (2017). URL `https://publikationen.bibliothek.kit.edu/1000070952`
9. Hipp, D., Hochbruck, M., Stohrer, C.: Unified error analysis for nonconforming space discretizations of wave-type equations. IMA J. Numer. Anal. **39**(3), 1206–1245 (2019). DOI 10.1093/imanum/dry036
10. Hochbruck, M., Leibold, J.: Finite element discretization of semilinear acoustic wave equations with kinetic boundary conditions. CRC 1173 Preprint 2019/26, Karlsruhe Institute of Technology (2019). DOI 10.5445/IR/1000105549. URL `https://www.waves.kit.edu/downloads/CRC1173_Preprint_2019-26.pdf`. To appear in Electron. Trans. Numer. Anal.
11. Hochbruck, M., Sturm, A.: Error analysis of a second-order locally implicit method for linear Maxwell's equations. SIAM J. Numer. Anal. **54**(5), 3167–3191 (2016). DOI 10.1137/15M1038037
12. Kadioglu, S.Y., Knoll, D.A., Lowrie, R.B., Rauenzahn, R.M.: A second order self-consistent IMEX method for radiation hydrodynamics. J. Comput. Phys. **229**(22), 8313 – 8332 (2010). DOI https://doi.org/10.1016/j.jcp.2010.07.019
13. Layton, W., Li, Y., Trenchea, C.: Recent developments in IMEX methods with time filters for systems of evolution equations. J. Comput. Appl. Math. **299**, 50–67 (2016). DOI 10.1016/j.cam.2015.09.038
14. Layton, W., Trenchea, C.: Stability of two IMEX methods, CNLF and BDF2-AB2, for uncoupling systems of evolution equations. Appl. Numer. Math. **62**(2), 112–120 (2012). DOI 10.1016/j.apnum.2011.10.006
15. Leibold, J.: Semilineare Wellengleichungen mit dynamischen Randbedingungen. Master's thesis, Karlsruhe Institue of Technology (2017). URL `http://na.math.kit.edu/download/thesis/2017-Leibold.pdf`
16. Lemieux, J.F., Knoll, D.A., Losch, M., Girard, C.: A second-order accurate in time IMplicit–EXplicit (IMEX) integration scheme for sea ice dynamics. J. Comput. Phys. **263**, 375–392 (2014). DOI 10.1016/j.jcp.2014.01.010
17. Vitillaro, E.: Strong solutions for the wave equation with a kinetic boundary condition. In: Recent trends in nonlinear partial differential equations. I. Evolution problems, *Contemp. Math.*, vol. 594, pp. 295–307. Amer. Math. Soc., Providence, RI (2013). DOI 10.1090/conm/594/11793
18. van Zuijlen, A.H., Bijl, H.: Implicit and explicit higher order time integration schemes for structural dynamics and fluid-structure interaction computations. Comput. Struct. **83**(2-3), 93–105 (2005)