Karlsruhe Institute of Technology

wave phenomena
analysis and numerics

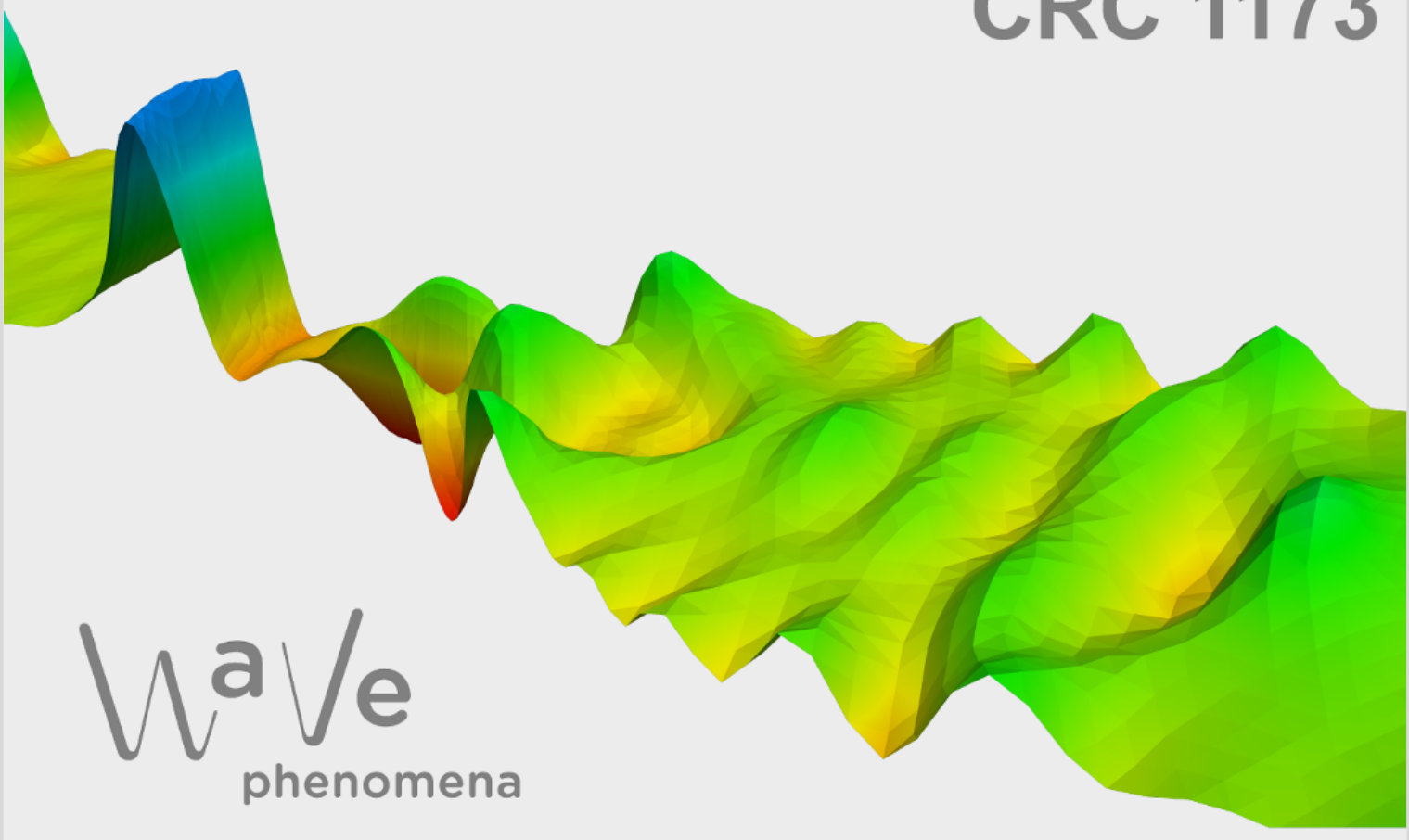
Error analysis of multirate leapfrog-type methods for second-order semilinear ODEs

Constantin Carle, Marlis Hochbruck

CRC Preprint 2021/26, June 2021

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Funded by

ISSN 2365-662X

ERROR ANALYSIS OF MULTIRATE LEAPFROG-TYPE METHODS FOR SECOND-ORDER SEMILINEAR ODES*

CONSTANTIN CARLE† AND MARLIS HOCHBRUCK†

Abstract. In this paper we consider the numerical solution of second-order semilinear differential equations, for which the stiffness is induced by only a few components of the linear part. For such problems, the leapfrog scheme suffers from severe restrictions on the step size to ensure stability. We thus propose a general class of multirate leapfrog-type methods which allows to use step sizes which are independent on the stiff part of the equation and also very efficient to implement. This class comprises local time-stepping schemes [5, 7] but also locally implicit or locally trigonometric integrators. Our main contribution is a rigorous error and stability analysis with special emphasis on explicit multirate methods, which are based on stabilized leapfrog-Chebyshev polynomials introduced in [4].

Key words. time integration, second-order ode, leapfrog method, stability analysis, error analysis, Hamiltonian systems, CFL condition, Chebyshev polynomials, wave equation

AMS subject classifications. Primary 65L04, 65L20. Secondary 65L05, 65L06, 65L70, 65P10.

1. Introduction.

In this paper we consider the numerical solution of second-order differential equations in $\mathbb{R}^d$ of the form

$$\mathbf{M}\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t) + \mathbf{M}g\big(t, \mathbf{q}(t)\big), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \mathbf{p}_0, \tag{1.1}$$

where $\mathbf{L} \in \mathbb{R}^{d\times d}$ is a symmetric, positive definite or positive semidefinite matrix, $\mathbf{M} \in \mathbb{R}^{d\times d}$ is symmetric, positive definite, and g is a sufficiently regular function. This equation serves as a model for various applications, e.g., Hamiltonian equations of motions occurring in astronomy or in molecular dynamics, as well as spatially discretized wave-type partial differential equations.

The method of choice for approximating the solution of (1.1) is the *leapfrog* (LF) scheme, also known as Störmer or Verlet scheme. It is explicit, easy to implement, very efficient, of order two and it also has nice geometric properties such as symplecticity and symmetry; cf. [9, 10] for many more details on its properties.

However, for stability the LF scheme requires step sizes which are smaller than the period of the fastest oscillations arsing in the system (1.1). To be more precise, for $g \equiv 0$, we need step sizes $\tau^2 \le 4/\|\mathbf{M}^{-1/2}\mathbf{L}\mathbf{M}^{-1/2}\|$. Hence, in the past decades, several variants of the LF scheme have been proposed to overcome this limitation. Most of them are designed and motivated for particular applications such as multiple time scales in molecular dynamics; cf., for example, the monographs [10, Chapters VIII.4, XIII.1], [14, Chapter 4], and [15, Chapter 10].

Here, we are interested in situations where only a few components of the solution, or, equivalently, a small principal submatrix of $\mathbf{L}$ is responsible for the high frequencies. Then the step-size restriction of the classical LF method causes a significant loss of efficiency. Such situations appear for instance for spatial discretizations of wave-type equations on meshes, where only a small part of the mesh consists of tiny elements whereas the majority of the mesh elements are significantly larger. For such applications, local time-stepping schemes were invented in [5] for linear, homogeneous

*Version of June 16, 2021

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

†Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany (constantin.carle@kit.edu, marlis.hochbruck@kit.edu)

wave equations. A closely related situation appears in inhomogeneous materials if there is a scale separation in the material coefficient such that the main stiffness is induced only by few mesh elements.

Recently, we proposed and analyzed leapfrog–Chebyshev methods (LFC) [4] which rely on a splitting of the right-hand side of (1.1) into the "stiff" part $\mathbf{Lq}$ and the "nonstiff" part $g(t, \mathbf{q})$. The method is applicable to splittings where $\mathbf{L}$ is symmetric and positive semidefinite.

However, if we want to split the components of the solution vector $\mathbf{q}$, it seems more appropriate to split $\mathbf{L}$ into

$$\mathbf{L} = \mathbf{LR} + \mathbf{L}(\mathbf{I}_d - \mathbf{R}), \tag{1.2}$$

where $\mathbf{I}_d$ denotes the identity matrix and $\mathbf{R}$ is a diagonal matrix with diagonal entries being either zero or one, where the ones refer to the stiff components; cf. [5]. Here we present a general class of two-step methods for semilinear problems which is based on a splitting of the form (1.2), where the second term $\mathbf{L}(\mathbf{I}_d - \mathbf{R})$ is "nonstiff" and treated together with the nonlinear part g. The general class comprises a variety of options to deal with the "stiff" part $\mathbf{LR}$ of the operator $\mathbf{L}$, ranging from local time-stepping schemes to locally implicit or locally trigonometric integrators; cf. Section 2 for details. In particular, for $g \equiv 0$, the method contains the one by [5] as a special case.

Our main contribution is to rigorously analyze the whole class of methods, although we are mostly interested in methods based on Chebyshev polynomials. For linear, homogeneous wave equations, such methods have also been analyzed in [7] using discrete semigroup techniques. By using generating functions as a main tool we are able to weaken their CFL condition and regularity assumption on the exact solution (1.1). Moreover, our analysis includes the case of positive semidefinite matrices $\mathbf{L}$, to which the technique in [7] does not apply.

Outline of the paper. In Section 2 we introduce a general class of two-step methods for (1.1) based on the LF scheme and the matrix splitting of $\mathbf{L}$ in (1.2). Sections 3 and 4 are devoted to the stability and error analysis of these schemes, respectively. In there, we prove stability under a step-size restriction and second-order convergence of the scheme. In Section 5 we show that the LFC polynomials from [4] satisfy the abstract assumptions made beforehand and give explicit formulas for all occurring constants. Moreover, we discuss the implementation of these schemes and their efficiency compared to the LF scheme. We conclude our paper with some examples in Section 6. After an analytical example illustrating the necessity of our abstract assumptions for obtaining stability, we numerically show the benefit of the schemes for two realistic examples.

2. General multirate leapfrog-type two-step schemes. We start by stating the well-known LF scheme for the semilinear problem (1.1). For a step size $\tau > 0$ we denote by $\mathbf{q}_n$ the approximation of the exact solution $\mathbf{q}(t_n)$ at time $t_n = n\tau$. Then, the LF scheme is given as two-step method by

$$\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1} = -\tau^2 \mathbf{M}^{-1}\mathbf{L}\mathbf{q}_n + \tau^2 \mathbf{g}_n, \qquad n = 1, 2, \dots, \tag{2.1a}$$

$$\mathbf{q}_1 = \mathbf{q}_0 + \tau \mathbf{p}_0 - \tfrac{1}{2}\tau^2 \mathbf{M}^{-1}\mathbf{L}\mathbf{q}_0 + \tfrac{1}{2}\tau^2 \mathbf{g}_0, \tag{2.1b}$$

where $\mathbf{g}_n = g(t_n, \mathbf{q}_n)$.

For the sake of presentation, in the following we restrict ourselves to the case of $\mathbf{M} = \mathbf{I}_d$, since the general case can be transformed to this one; cf. Remark 2.5 below.

With $(\cdot,\cdot)$ we denote the standard Euclidean inner product in $\mathbb{R}^d$ and with $\|\cdot\|$ the corresponding (matrix) norm.

ASSUMPTION 2.1. *Let* $\mathbf{L}$ *be symmetric, positive semidefinite, and (possibly after permutation) be partitioned as*

$$\mathbf{L} = \begin{pmatrix} \mathbf{S} & \mathbf{K}^T \\ \mathbf{K} & \mathbf{N} \end{pmatrix}, \tag{2.2}$$

where the norms of the "nonstiff" *and* "stiff" *submatrices* $\mathbf{N} \in \mathbb{R}^{(d-s)\times(d-s)}$ *and* $\mathbf{S} \in \mathbb{R}^{s\times s}$*, respectively, satisfy* $\|\mathbf{S}\| = r\|\mathbf{N}\|$ *with* $r \gg 1$ *and* $d \gg s$*. For the* coupling *matrix* $\mathbf{K} \in \mathbb{R}^{(d-s)\times s}$ *it holds* $\|\mathbf{K}\| = \kappa\|\mathbf{N}\|$ *with* $0 \le \kappa \ll r^{1/2}$*.*

With this assumption we guarantee that the stiffness of (1.1) is only induced by the submatrix $\mathbf{S}$. Clearly, the symmetry and positive semidefiniteness of $\mathbf{L}$ transfers to the matrices $\mathbf{S}$ and $\mathbf{N}$. The assumption $\kappa \ll r^{1/2}$ originates from the positive semidefiniteness of $\mathbf{L}$. Note that the reordering of $\mathbf{L}$ in (2.2) is only for the ease of representation and not necessary for the implementation.

Further, due to (2.2) we define the *restriction matrix* $\mathbf{R} \in \mathbb{R}^{d\times d}$ occurring in (1.2), which maps to the "stiff" part of $\mathbf{L}$, by

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_s & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \quad \widetilde{\mathbf{S}} = \mathbf{RLR} = \begin{pmatrix} \mathbf{S} & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.3}$$

In order to propose a scheme with a step-size restriction depending only on $\mathbf{N}$ and κ but not on $\mathbf{S}$, we now modify the LF scheme by multiplying the right-hand side by a suitable matrix function $\widehat{\Psi}(\tau^2\mathbf{LR})$:

$$\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1} = \tau^2\widehat{\Psi}(\tau^2\mathbf{LR})\big(-\mathbf{L}\mathbf{q}_n + \mathbf{g}_n\big), \qquad n = 1, 2, \dots, \tag{2.4a}$$

$$\mathbf{q}_1 = \mathbf{q}_0 + \tau\mathbf{p}_0 + \tfrac{1}{2}\tau^2\widehat{\Psi}(\tau^2\mathbf{LR})\big(-\mathbf{L}\mathbf{q}_0 + \mathbf{g}_0\big). \tag{2.4b}$$

Our interest is mainly in choosing $\widehat{\Psi}$ as a polynomial, a rational, or a trigonometric function, since then the scheme results in an explicit, a locally implicit, or a locally trigonometric integrator, respectively.

The consistency order two of the LF scheme is preserved if $\widehat{\Psi}$ fulfills the following assumption.

ASSUMPTION 2.2 (Consistency). $\widehat{\Psi}\colon [0,\infty) \to \mathbb{R}$ *is sufficiently smooth and satisfies* $\widehat{\Psi}(0) = 1$*.*

This assumption implies that for $\mathbf{R} = 0$ or $\widehat{\Psi} \equiv 1$ the scheme (2.4) reduces to the LF scheme (2.1).

DEFINITION 2.3. *For* $\widehat{\Psi}$ *satisfying Assumption 2.2 we define functions* Ψ*,* $\mathcal{X}$ *as*

$$\Psi(z) = z\widehat{\Psi}(z) \qquad \text{and} \qquad \mathcal{X}(z) = \frac{\widehat{\Psi}(z) - 1}{z}, \quad z > 0, \quad \mathcal{X}(0) = \widehat{\Psi}'(0). \tag{2.5}$$

Obviously, Assumption 2.2 implies

$$\Psi(0) = 0, \qquad \Psi'(0) = 1. \tag{2.6}$$

Below, we will prove that

$$\widehat{\Psi}(\tau^2\mathbf{LR}) = \begin{pmatrix} \widehat{\Psi}(\tau^2\mathbf{S}) & 0 \\ \tau^2\mathbf{K}\mathcal{X}(\tau^2\mathbf{S}) & \mathbf{I}_{d-s} \end{pmatrix}.$$

Hence, the analysis and also the implementation rely only on matrix functions acting on the small symmetric submatrix $\mathbf{S}$. The actual implementation of this scheme depends on the choice of $\widehat{\Psi}$; cf. Section 5.2 for a specific case. For the analysis, working with Ψ turns out to be more convenient.

The modification (2.4) of the LF scheme is motivated by LF-based local time-stepping schemes [5, 7] and locally implicit schemes, where the LF and the Crank–Nicolson (CN) scheme are combined [11, 16]. However, it is worth mentioning that the application of these schemes is not restricted to problems arising from spatially discretized partial differential equations with local mesh refinements; cf. Section 6.2.

Remark 2.4. A variant to (2.4) is given by

$$\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1} = -\tau^2 \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\mathbf{L}\mathbf{q}_n + \tau^2 \mathbf{g}_n, \qquad n = 1, 2, \dots, \tag{2.7a}$$

$$\mathbf{q}_1 = \mathbf{q}_0 + \tau \mathbf{p}_0 - \tfrac{1}{2}\tau^2 \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\mathbf{L}\mathbf{q}_0 + \tfrac{1}{2}\tau^2 \mathbf{g}_0, \tag{2.7b}$$

where $\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})$ is only applied to the linear part but not to g. The analysis of this scheme can be done analogously and leads to similar results under similar assumptions.

Remark 2.5. If the matrix $\mathbf{M}$ in (1.1) is not the identity but some arbitrary symmetric and positive definite matrix, then its Cholesky decomposition $\mathbf{M} = \mathbf{C}_\mathbf{M}\mathbf{C}_\mathbf{M}^T$ exists. In this case, we rewrite (1.1) in terms of the transformed variable $\mathbf{y} = \mathbf{C}_\mathbf{M}^T\mathbf{q}$ equivalently as

$$\ddot{\mathbf{y}}(t) = -\widetilde{\mathbf{L}}\mathbf{y}(t) + \widetilde{g}\big(t, \mathbf{y}(t)\big), \qquad \mathbf{y}(0) = \mathbf{C}_\mathbf{M}^T\mathbf{q}_0, \qquad \dot{\mathbf{y}}(0) = \mathbf{C}_\mathbf{M}^T\mathbf{p}_0,$$

where $\widetilde{\mathbf{L}} = \mathbf{C}_\mathbf{M}^{-1}\mathbf{L}\mathbf{C}_\mathbf{M}^{-T}$ is again symmetric and positive semidefinite. Hence, one can apply scheme (2.4) to this equation and transform the approximations back to the original variables. This yields

$$\mathbf{M}\big(\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1}\big) = \tau^2 \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{C}_\mathbf{M}^{-T}\mathbf{R}\mathbf{C}_\mathbf{M}^{-1})\big(-\mathbf{L}\mathbf{q}_n + \mathbf{M}\mathbf{g}_n\big) \tag{2.8}$$

for $n \geq 1$, and analogously for the starting value $\mathbf{q}_1$. Note that the structure of the Cholesky factor yields

$$\mathbf{C}_\mathbf{M}^{-T}\mathbf{R}\mathbf{C}_\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{M}_S^{-1} & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{for} \qquad \mathbf{M} = \begin{pmatrix} \mathbf{M}_S & \mathbf{M}_K^T \\ \mathbf{M}_K & \mathbf{M}_N \end{pmatrix},$$

which makes the implementation of (2.8) very efficient. In particular, the evaluation of $\widehat{\Psi}$ only requires the small dimensional matrix $\mathbf{M}_S^{-1}$ and not the full Cholesky factor of $\mathbf{M}$.

We would like to emphasize that our analysis also applies to this more general situation, if the standard norm $\|\cdot\|$ is replaced by $\|\cdot\|_\mathbf{M}^2 = \big(\,\cdot\,, \mathbf{M}\,\cdot\,\big)$.

In this paper we pay special attention to Ψ being the LFC polynomials from [4] given by

$$\Psi(z) = \Psi_p(z) = 2 - \frac{2}{T_p(\nu)} T_p\Big(\nu - \frac{z}{\alpha_p}\Big), \qquad \alpha_p = 2\frac{T_p'(\nu)}{T_p(\nu)}, \tag{2.9}$$

where T_p denotes the pth Chebyshev polynomial of first kind ($p \in \mathbb{N}$) and $\nu \geq 1$. Obviously, Ψ_p is a polynomial of degree $p \geq 1$ and satisfies Assumption 2.2 for arbitrary $\nu \geq 1$. We abbreviate the combination of the general scheme (2.4) and the polynomials (2.9) with *sLFC schemes*. We note that for $g \equiv 0$ the scheme (2.4a) equipped

with (2.9) coincides with the stabilized local time-stepping scheme proposed in [7] (with $\eta^2/2$ replaced by η). Further, for $\mathbf{R} = \mathbf{I}_d$ the variant (2.7a) coincides with the multirate LFC scheme in [4].

Other choices for Ψ fitting into the setting but not considered in this paper are the rational functions

$$\Psi(z) = \frac{z}{1 + \frac{\nu}{4} z}, \quad \nu > 1, \tag{2.10}$$

or the trigonometric functions

$$\Psi(z) = \begin{cases} 2 - \frac{2}{\cosh(\eta)} \cosh\Big(\big(\eta^2 - \frac{z}{\alpha}\big)^{1/2} \Big), & 0 \le z < \alpha\eta^2 \\ 2 - \frac{2}{\cosh(\eta)} \cos\Big(\big(\frac{z}{\alpha} - \eta^2\big)^{1/2} \Big), & z \ge \alpha\eta^2 \end{cases}, \quad \alpha = \frac{\operatorname{sinhc}(\eta)}{\cosh(\eta)}, \tag{2.11}$$

where $\operatorname{sinhc}(x) = \sinh(x)/x$ and $\eta > 0$. Note that the choice (2.10) in (2.4) leads to a scheme, which is implicit on the stiff and explicit on the nonstiff components of (1.1).

We conclude this section by stating the general scheme (2.4) in an equivalent one-step formulation and some geometric properties. The one-step scheme of (2.4) is given by

$$\mathbf{p}_{n+1/2} = \mathbf{p}_n + \tfrac{1}{2}\tau \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\big(-\mathbf{L}\mathbf{q}_n + \mathbf{g}_n\big), \tag{2.12a}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \tau \mathbf{p}_{n+1/2}, \qquad n = 0, 1, 2, \dots, \tag{2.12b}$$

$$\mathbf{p}_{n+1} = \mathbf{p}_{n+1/2} + \tfrac{1}{2}\tau \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\big(-\mathbf{L}\mathbf{q}_{n+1} + \mathbf{g}_{n+1}\big). \tag{2.12c}$$

Here, $\mathbf{p}_n$ can be interpreted as an approximation to $\dot{\mathbf{q}}(t_n)$.

COROLLARY 2.6. *The scheme* (2.4a) *and thus also the equivalent one-step version* (2.12a)–(2.12c) *are symmetric and symplectic.*

Proof. The scheme (2.4a) is equivalent to the LF scheme (2.1a) applied to the modified equation

$$\ddot{\mathbf{q}} = \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\big(-\mathbf{L}\mathbf{q} + g(\cdot, \mathbf{q})\big).$$

Hence, it inherits the properties of the LF method. ∎

3. Stability analysis. In this section we show stability of the scheme (2.4) for linear and semilinear problems under some general conditions for the function Ψ.

3.1. Properties and further assumptions on Ψ. So far, the only restrictions on Ψ are the consistency conditions (2.6). For stability, it was already shown in [4] that $0 \le \Psi(z) \le 4$ for z in a suitable interval $[0, \beta]$ is not sufficient to ensure stability for multirate methods. In fact, we need $\Psi(z) \in (0, 4)$ for $z \in (0, \beta)$. The precise conditions are stated in the following definition.

DEFINITION 3.1. *For given $m_1, \widetilde{m}_1, \widetilde{m}_2 \in (0, 1)$ with $\widetilde{m}_1 \le 1 - m_1$ we define $\widehat{\beta} = \widehat{\beta}(m_1, \widetilde{m}_1, \widetilde{m}_2) \in (0, \infty)$ as the maximal value such that*

$$\min\{\widetilde{m}_2 z, 4\widetilde{m}_1\} \le \Psi(z) \le 4(1 - m_1) \qquad \text{for all } z \in [0, \widehat{\beta}^2], \tag{3.1}$$

and $\widehat{\beta} = \infty$, if (3.1) holds for all $z \ge 0$. Moreover, we define m_3 as the smallest constant such that for $\mathcal{X}$ defined in (2.5) holds

$$|\mathcal{X}(z)| \le \tfrac{1}{2} m_3 \qquad \text{for all } z \in [0, \widehat{\beta}^2] \cap \mathbb{R}. \tag{3.2}$$

The existence of such a $\widehat{\beta} > 0$ and m_3 is guaranteed by (2.6) and (2.5), respectively. If $\widehat{\beta}$ is finite, the lower bound in (3.1) can be replaced by $\Psi(z) \geq \widetilde{m}_2 z$ for all $z \in [0, \widehat{\beta}^2]$.

In practice, the values for the constants $m_1, \widetilde{m}_1, \widetilde{m}_2$ for a specific function Ψ are chosen such that $\widehat{\beta}$ is as large as possible, since $\widehat{\beta}$ (but also m_1) will enter the step-size restriction for the scheme (2.4). Moreover, m_1 also enters the error constant. For the polynomials (2.9) explicit values for $m_1, \widetilde{m}_1, \widetilde{m}_2$, and $\widehat{\beta}$ are given in Section 5.

As a direct consequence of (3.1) we have

$$\widehat{\Psi}(z) \geq \min\{\widetilde{m}_2, 4\widetilde{m}_1/z\} > 0 \qquad \text{for all } z \in [0, \widehat{\beta}^2] \cap \mathbb{R}. \tag{3.3}$$

Hence, the inverse $\widehat{\Psi}(z)^{-1}$ exists for all $z \in [0, \widehat{\beta}^2] \cap \mathbb{R}$ and the following bound holds.

LEMMA 3.2. *There exists a constant $c_{\mathrm{cs}} > 0$ such that*

$$|\widehat{\Psi}(z)^{-1}\mathcal{X}(z)| \leq c_{\mathrm{cs}} \qquad \textit{for all } z \in [0, \widehat{\beta}^2] \cap \mathbb{R}. \tag{3.4}$$

Proof. The statement follows directly from (3.3) and Definition 2.3. ∎

Additionally to the consistency conditions (2.6) we assume for Ψ the following.

ASSUMPTION 3.3. *For Ψ satisfying (2.6) we have*

$$\Psi(z) \leq z \qquad \textit{for all } z \in \left[0, \widehat{\beta}^2\right] \cap \mathbb{R}. \tag{3.5}$$

From this assumption we immediately obtain with (2.5)

$$|\widehat{\Psi}(z)| \leq 1, \qquad \mathcal{X}(z) \leq 0 \qquad \text{for all } z \in \left[0, \widehat{\beta}^2\right] \cap \mathbb{R}. \tag{3.6}$$

We note that Assumption 3.3 could theoretically be weakened to $\Psi(z) \leq c_* z$ for all $z \in \left[0, \widehat{\beta}^2\right] \cap \mathbb{R}$ and a constant $c_* > 1$. However, this would lead to a stronger step-size restriction, since then $|\widehat{\Psi}(z)| \leq c_*$ for all z and $\mathcal{X} > 0$ for some z; cf. the proof of Lemma 3.6. In particular, (3.5) is fulfilled for the polynomials (2.9); see Lemma 5.1.

In the remaining part of this paper, let Assumptions 2.1, 2.2, and 3.3 hold without mentioning it explicitly everywhere. Moreover, we define

$$\mathbf{L}_{\Psi,\tau} = \widehat{\Psi}(\tau^2 \mathbf{L}\mathbf{R})\mathbf{L}, \tag{3.7}$$

and, if $\mathbf{L}$ is positive definite, we have a constant $c_{\mathrm{inv}} > 0$ such that

$$\|\mathbf{L}^{-1}\| \leq c_{\mathrm{inv}}^2. \tag{3.8}$$

3.2. Stability estimates. In this section we show some bounds on the matrix $\mathbf{L}_{\Psi,\tau}$ under a step-size restriction which are necessary for proving stability of the scheme (2.4). More precisely, our step-size restriction depends on $\mathbf{S}$, $\mathbf{N}$, κ defined in Assumption 2.1, and m_1 given in Definition 3.1.

DEFINITION 3.4 (CFL condition). *For fixed $\vartheta \in (0, 1]$ let $\tau_{\mathrm{CFL}}(\vartheta)$ be the maximal step size $\tau > 0$ such that the step-size restriction(s)*

$$\tau^2 \|\mathbf{S}\| \leq \widehat{\beta}^2, \tag{3.9a}$$

$$\tau^2 \|\mathbf{N}\| \leq 4\gamma\vartheta^2, \qquad \gamma = \frac{2}{1 + (1 + 4\kappa^2 m_1^{-1})^{1/2}}, \tag{3.9b}$$

hold for all $0 \leq \tau \leq \tau_{\mathrm{CFL}}(\vartheta)$. Further, we define $\tau_{\mathrm{CFL}} = \tau_{\mathrm{CFL}}(1)$.

Observe that $\gamma \leq 1$ for all κ and $m_1 > 0$. Further, the second CFL condition (3.9b) becomes stronger with increasing κ. On the other hand, for $\kappa = 0$ (implying $\mathbf{K} = 0$) we have $\gamma = 1$. In this case, (3.9b) corresponds to the (standard) CFL condition for the LF scheme applied to (1.1) with $\mathbf{S} = \mathbf{K} = 0$ as one would expect. We emphasize that in our numerical experiments the CFL condition (3.9b) turns out to be rather pessimistic; cf. Section 6.

We start by deriving an explicit block formula for $\mathbf{L}_{\Psi,\tau}$. In there we make use of the following identity

$$f(\tau^2\mathbf{LR})\,\mathbf{LR} = f(\tau^2\mathbf{LR}^2)\,\mathbf{LR}^2 = \mathbf{LR}\,f(\tau^2\mathbf{RLR})\,\mathbf{R} = \mathbf{LR}\,f(\tau^2\widetilde{\mathbf{S}})\mathbf{R} \tag{3.10}$$

with $\widetilde{\mathbf{S}}$ defined in (2.3), which holds for a sufficiently smooth function f due to $\mathbf{R} = \mathbf{R}^2$.

LEMMA 3.5. *For $\mathbf{L}_{\Psi,\tau}$ defined in (3.7) with $\mathbf{L}$ defined in (2.2) we have*

$$\tau^2\mathbf{L}_{\Psi,\tau} = \begin{pmatrix} \Psi(\tau^2\mathbf{S}) & \tau^2\widehat{\Psi}(\tau^2\mathbf{S})\mathbf{K}^T \\ \tau^2\mathbf{K}\widehat{\Psi}(\tau^2\mathbf{S}) & \tau^2\mathbf{N} + \tau^4\mathbf{K}\mathcal{X}(\tau^2\mathbf{S})\mathbf{K}^T \end{pmatrix}. \tag{3.11}$$

Moreover, $\mathbf{L}_{\Psi,\tau}$ is symmetric.

Proof. The symmetry immediately follows from the definition of $\mathbf{L}_{\Psi,\tau}$. Further, we get from Definition 2.3 and (3.10)

$$\begin{aligned} \widehat{\Psi}(\tau^2\mathbf{LR}) &= \mathbf{I}_d + \tau^2\mathcal{X}(\tau^2\mathbf{LR})\,\mathbf{LR} = \mathbf{I}_d + \tau^2\mathbf{LR}\,\mathcal{X}(\tau^2\widetilde{\mathbf{S}})\mathbf{R} \\ &= \mathbf{I}_d + \tau^2 \begin{pmatrix} \mathbf{S} & 0 \\ \mathbf{K} & 0 \end{pmatrix} \begin{pmatrix} \mathcal{X}(\tau^2\mathbf{S}) & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \widehat{\Psi}(\tau^2\mathbf{S}) & 0 \\ \tau^2\mathbf{K}\mathcal{X}(\tau^2\mathbf{S}) & \mathbf{I}_{d-s} \end{pmatrix}. \end{aligned} \tag{3.12}$$

Using (3.7) and again Definition 2.3 completes the proof. ∎

We emphasize that the symmetry of $\mathbf{L}_{\Psi,\tau}$ is crucial for our stability analysis and used at several points in the following. For the polynomials (2.9) a different proof of (3.11) was given in [7]. In there, similar estimates as in the following lemma are shown for $\kappa = 1$, however, under a stronger step-size restriction than ours.

LEMMA 3.6. *Let $\vartheta \in (0,1]$ and $\tau \leq \tau_{\mathrm{CFL}}(\vartheta)$. Then we have for all $\mathbf{q} \in \mathbb{R}^d$*

$$0 \leq \tau^2\big(\mathbf{L}_{\Psi,\tau}\mathbf{q}, \mathbf{q}\big) \leq 4\big(1 - m_1 + m_1\vartheta^2\big)\|\mathbf{q}\|^2. \tag{3.13}$$

In particular, we have $\tau^2\|\mathbf{L}_{\Psi,\tau}\| \leq 4\big(1 - m_1 + m_1\vartheta^2\big) \leq 4$.

Proof. We start with the upper bound. For this we write

$$\mathbf{q} = \begin{pmatrix} \mathbf{q}_S \\ \mathbf{q}_N \end{pmatrix} \in \mathbb{R}^d \qquad \text{with} \quad \mathbf{q}_S \in \mathbb{R}^s,\ \mathbf{q}_N \in \mathbb{R}^{d-s}. \tag{3.14}$$

By Lemma 3.5 we then have

$$\begin{aligned} \tau^2\big(\mathbf{L}_{\Psi,\tau}\mathbf{q}, \mathbf{q}\big) &= \big(\Psi(\tau^2\mathbf{S})\mathbf{q}_S, \mathbf{q}_S\big) + \tau^2\big(\widehat{\Psi}(\tau^2\mathbf{S})\mathbf{K}^T\mathbf{q}_N, \mathbf{q}_S\big) \\ &\quad + \tau^2\big(\mathbf{K}\widehat{\Psi}(\tau^2\mathbf{S})\mathbf{q}_S, \mathbf{q}_N\big) + \tau^2\big((\mathbf{N} + \tau^2\mathbf{K}\mathcal{X}(\tau^2\mathbf{S})\mathbf{K}^T)\mathbf{q}_N, \mathbf{q}_N\big). \end{aligned} \tag{3.15}$$

For the first term, the CFL condition (3.9a) together with (3.1) yields

$$\big(\Psi(\tau^2\mathbf{S})\mathbf{q}_S, \mathbf{q}_S\big) \leq 4(1 - m_1)\|\mathbf{q}_S\|^2.$$

For the second and third term in (3.15) we have $\|\widehat{\Psi}(\tau^2\mathbf{S})\| \leq 1$ because of Assumption 3.3 under the CFL condition (3.9a). Hence, we get with the Cauchy-Schwarz inequality, the scaled Young's inequality and Assumption 2.1

$$\begin{aligned}\tau^2\big(\widehat{\Psi}(\tau^2\mathbf{S})\mathbf{K}^T\mathbf{q}_N, \mathbf{q}_S\big) &\leq \tau^2\|\mathbf{K}\|\|\mathbf{q}_N\|\|\mathbf{q}_S\| \\ &\leq \tfrac{1}{2}\tau^2\kappa\|\mathbf{N}\|\big(\gamma_*\|\mathbf{q}_N\|^2 + \gamma_*^{-1}\|\mathbf{q}_S\|^2\big)\end{aligned}$$

with a parameter $\gamma_* > 0$, which is yet to be determined.

The last term in (3.15) can be bounded by

$$\tau^2\big(\mathbf{N}\mathbf{q}_N, \mathbf{q}_N\big) + \tau^4\big(\mathcal{X}(\tau^2\mathbf{S})\mathbf{K}^T\mathbf{q}_N, \mathbf{K}^T\mathbf{q}_N\big) \leq \tau^2\|\mathbf{N}\|\|\mathbf{q}_N\|^2,$$

since by (3.9a) the second bound in (3.6) holds.

Combining these estimates yields with $\gamma_* = \kappa m_1^{-1}\gamma$ and (3.9b)

$$\begin{aligned}\tau^2\big(\mathbf{L}_{\Psi,\tau}\mathbf{q}, \mathbf{q}\big) &\leq 4\big(1 - m_1 + m_1\vartheta^2\big)\|\mathbf{q}_S\|^2 + (1 + \kappa^2 m_1^{-1}\gamma)4\gamma\vartheta^2\|\mathbf{q}_N\|^2 \\ &\leq 4\big(1 - m_1 + m_1\vartheta^2\big)\big(\|\mathbf{q}_S\|^2 + \|\mathbf{q}_N\|^2\big) \\ &= 4\big(1 - m_1 + m_1\vartheta^2\big)\|\mathbf{q}\|^2,\end{aligned}$$

where we used that $1 + \kappa^2 m_1^{-1}\gamma = \gamma^{-1}$ and $\vartheta^2 \leq 1 - m_1 + m_1\vartheta^2$.

For the lower bound in (3.13) we exploit that a symmetric, positive semidefinite matrix $\mathbf{L}$ admits a block decomposition of the form

$$\mathbf{L} = \mathbf{C}\mathbf{C}^T \qquad \text{with} \qquad \mathbf{C} = \begin{pmatrix} \mathbf{S}^{\frac{1}{2}} & 0 \\ \mathbf{K}\mathbf{S}^{+}\mathbf{S}^{\frac{1}{2}} & \mathbf{A}^{\frac{1}{2}} \end{pmatrix},$$

where $\mathbf{S}^+$ denotes the Moore-Penrose inverse of $\mathbf{S}$, $\mathbf{A} = \mathbf{N} - \mathbf{K}\mathbf{S}^{+}\mathbf{K}^T$ is a positive semidefinite matrix, and $\mathbf{K}\mathbf{S}^{+}\mathbf{S} = \mathbf{K}$; see [1] and also [12, Theorems 1.19, 1.20]. Then we obtain

$$\mathbf{L}_{\Psi,\tau} = \widehat{\Psi}(\tau^2\mathbf{C}\mathbf{C}^T\mathbf{R})\mathbf{C}\mathbf{C}^T = \mathbf{C}\widehat{\Psi}(\tau^2\mathbf{C}^T\mathbf{R}\mathbf{C})\mathbf{C}^T = \mathbf{C}\widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})\mathbf{C}^T.$$

Thus, we get with (3.3) and the CFL condition (3.9a) for all $\mathbf{q} \in \mathbb{R}^d$

$$\tau^2\big(\mathbf{L}_{\Psi,\tau}\mathbf{q}, \mathbf{q}\big) = \tau^2\big(\widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})\mathbf{C}^T\mathbf{q}, \mathbf{C}^T\mathbf{q}\big) \geq 0,$$

which finishes the proof. ∎

Remark 3.7. If $\widehat{\beta}^2 \geq 4$ and $\tau^2\|\mathbf{L}\| \leq 4$, which is the step-size restriction of the LF scheme (2.1), we have with (3.6)

$$\tau^2\big(\mathbf{q}, \mathbf{L}_{\Psi,\tau}\mathbf{q}\big) = \tau^2\big(\mathbf{C}^T\mathbf{q}, \widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})\mathbf{C}^T\mathbf{q}\big) \leq \tau^2\big(\mathbf{C}^T\mathbf{q}, \mathbf{C}^T\mathbf{q}\big) = \tau^2\big(\mathbf{q}, \mathbf{L}\mathbf{q}\big) \leq 4\|\mathbf{q}\|^2,$$

where $\mathbf{L} = \mathbf{C}\mathbf{C}^T$ is given as in the previous proof. Thus, with a sensible choice of Ψ the scheme (2.4) is stable for at least all step sizes for which the LF scheme is.

Further, we need properties of $\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})$ for the stability of the scheme (2.4) and for proving the positive definiteness of $\mathbf{L}_{\Psi,\tau}$, if $\mathbf{L}$ is positive definite. Note that we cannot directly employ (3.6) because $\mathbf{L}\mathbf{R}$ is non-symmetric.

LEMMA 3.8. *Let $\vartheta \in (0, 1]$ and $\tau \leq \tau_{\mathrm{CFL}}(\vartheta)$. Then the inverse of $\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})$ exists and*

$$\big\|\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\big\| \leq c_{\widehat{\Psi}}, \qquad c_{\widehat{\Psi}} = 1 + 2m_3\kappa\gamma\vartheta^2. \tag{3.16}$$

Proof. By (3.3) and the CFL condition (3.9a) all eigenvalues of $\widehat{\Psi}(\tau^2\mathbf{S})$ are positive. Hence, the inverse exists by the block structure of (3.12).

For the estimate of $\widehat{\Psi}(\tau^2\mathbf{LR})$ we use (3.12), the Cauchy-Schwarz inequality, the CFL condition (3.9a), (3.6), and Young's inequality to obtain for all $\mathbf{q} \in \mathbb{R}^d$

$$\begin{aligned}\big\|\widehat{\Psi}(\tau^2\mathbf{LR})\mathbf{q}\big\|^2 &= \big\|\widehat{\Psi}(\tau^2\mathbf{S})\mathbf{q}_S\big\|^2 + \big\|\tau^2\mathbf{K}\mathcal{X}(\tau^2\mathbf{S})\mathbf{q}_S\big\|^2 \\ &\quad + 2\big(\tau^2\mathbf{K}\mathcal{X}(\tau^2\mathbf{S})\mathbf{q}_S, \mathbf{q}_N\big) + \|\mathbf{q}_N\|^2 \\ &\le \|\mathbf{q}_S\|^2 + \rho^2\|\mathbf{q}_S\|^2 + 2\rho\|\mathbf{q}_S\|\|\mathbf{q}_N\| + \|\mathbf{q}_N\|^2 \\ &\le (1+\rho)^2\|\mathbf{q}\|^2,\end{aligned}$$

where $\rho \le \frac{1}{2}m_3\tau^2\|\mathbf{K}\| \le 2m_3\kappa\gamma\vartheta^2$ by (3.9), (3.2), and Assumption 2.1. ∎

With this lemma we are able to show the positive definiteness of $\mathbf{L}_{\Psi,\tau}$ for a positive definite $\mathbf{L}$.

LEMMA 3.9. *Let $\tau \le \tau_{\mathrm{CFL}}$ and $\mathbf{L}$ be positive definite. Then the inverse of $\mathbf{L}_{\Psi,\tau}$ exists and we have*

$$\big(\mathbf{L}_{\Psi,\tau}\mathbf{q}, \mathbf{q}\big) \ge \big(c_{\mathrm{inv}}^2 + \tau^2 c_{\mathrm{cs}}\big)^{-1}\|\mathbf{q}\|^2 \qquad \text{for all } \mathbf{q} \in \mathbb{R}^d. \tag{3.17}$$

Proof. The existence of the inverse of $\mathbf{L}_{\Psi,\tau}$ follows from positive definiteness of $\mathbf{L}$ and Lemma 3.8. Further, we obtain with the definition of $\mathcal{X}$ in (2.5) and (3.10)

$$\begin{aligned}\mathbf{L}_{\Psi,\tau}^{-1} &= \mathbf{L}^{-1}\widehat{\Psi}(\tau^2\mathbf{LR})^{-1} = \mathbf{L}^{-1} + \mathbf{L}^{-1}\Big(\widehat{\Psi}(\tau^2\mathbf{LR})^{-1} - \mathbf{I}_d\Big) \\ &= \mathbf{L}^{-1} - \tau^2\mathbf{L}^{-1}\widehat{\Psi}(\tau^2\mathbf{LR})^{-1}\,\mathcal{X}(\tau^2\mathbf{LR})\,\mathbf{LR} \\ &= \mathbf{L}^{-1} - \tau^2\mathbf{R}\,\widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})^{-1}\,\mathcal{X}(\tau^2\widetilde{\mathbf{S}})\,\mathbf{R}.\end{aligned}$$

Using this identity we get with (3.8) and (3.4) under the CFL condition (3.9a)

$$\begin{aligned}\big(\mathbf{L}_{\Psi,\tau}^{-1}\mathbf{q}, \mathbf{q}\big) &= \big(\mathbf{L}^{-1}\mathbf{q}, \mathbf{q}\big) + \tau^2\big(-\widehat{\Psi}^{-1}(\tau^2\widetilde{\mathbf{S}})\mathcal{X}(\tau^2\widetilde{\mathbf{S}})\mathbf{Rq}, \mathbf{Rq}\big) \\ &\le c_{\mathrm{inv}}^2\|\mathbf{q}\|^2 + \tau^2 c_{\mathrm{cs}}\|\mathbf{Rq}\|^2 \\ &\le (c_{\mathrm{inv}}^2 + \tau^2 c_{\mathrm{cs}})\|\mathbf{q}\|^2,\end{aligned}$$

which yields (3.17). ∎

3.3. Stability of the scheme. After we have shown estimates for $\mathbf{L}_{\Psi,\tau}$ in the last section, we will use these for showing stability. We first start with a representation formula of the numerical solution of the scheme (2.4).

THEOREM 3.10. *Let $\tau \le \tau_{\mathrm{CFL}}$. Then the approximations of (2.4) are given by*

$$\mathbf{q}_n = \cos(n\mathbf{\Phi})\mathbf{q}_0 + \tau\boldsymbol{\mathcal{S}}_n\mathbf{p}_0 + \tau^2\sum_{\ell=0}^{n-1}\chi_\ell\,\boldsymbol{\mathcal{S}}_{n-\ell}\widehat{\Psi}(\tau^2\mathbf{LR})\,\mathbf{g}_\ell, \quad \boldsymbol{\mathcal{S}}_k = \frac{\sin(k\mathbf{\Phi})}{\sin\mathbf{\Phi}}, \tag{3.18a}$$

for $n = 0, 1, 2, \ldots$, where $\chi_0 = \frac{1}{2}$ and $\chi_\ell = 1$, $\ell \ge 1$, and the symmetric matrix $\mathbf{\Phi} \in \mathbb{R}^{d\times d}$ with spectrum in $[0, \pi]$ is uniquely defined by

$$\cos\mathbf{\Phi} = \mathbf{I}_d - \tfrac{1}{2}\tau^2\mathbf{L}_{\Psi,\tau} \qquad \text{and} \qquad \sin\mathbf{\Phi} = \tau\big(\mathbf{L}_{\Psi,\tau}(\mathbf{I}_d - \tfrac{1}{4}\tau^2\mathbf{L}_{\Psi,\tau})\big)^{1/2}. \tag{3.18b}$$

Proof. The proof can be done as in [4, Theorem 3.3], since $\mathbf{L}_{\Psi,\tau}$ is symmetric and the spectrum of $\mathbf{L}_{\Psi,\tau}$ is contained in $[0,4]$ under the CFL conditions (3.9) by Lemma 3.6. Hence, we obtain for the two-step scheme (2.4a)

$$\mathbf{q}_n = \cos(n\mathbf{\Phi})\mathbf{q}_0 + \boldsymbol{\mathcal{S}}_n\big(\mathbf{q}_1 - \cos\mathbf{\Phi}\,\mathbf{q}_0\big) + \tau^2 \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\,\mathbf{g}_\ell. \tag{3.19}$$

Inserting the definition of the starting value (2.4b) leads to (3.18a). ∎

Finally, in order to prove stability of the numerical solution $\mathbf{q}_n$, $n \in \mathbb{N}$, we have to provide bounds for the trigonometric functions occurring in (3.18a).

LEMMA 3.11. *(a) Let $\tau \le \tau_{\mathrm{CFL}}$. Then we have for $n \in \mathbb{N}$*

$$\|\cos(n\mathbf{\Phi})\| \le 1, \qquad \|\sin(n\mathbf{\Phi})\| \le 1, \qquad \|\boldsymbol{\mathcal{S}}_n\| \le n. \tag{3.20a}$$

(b) Let $\vartheta \in (0,1)$ and $\tau \le \tau_{\mathrm{CFL}}(\vartheta)$. Then we have for $\mathbf{L}$ positive definite

$$\tau\,\|(\sin\mathbf{\Phi})^{-1}\| \le c_{\mathrm{stb}} \qquad \text{with } c_{\mathrm{stb}} = \Big(\frac{c_{\mathrm{inv}}^2 + \tau^2 c_{\mathrm{cs}}}{m_1(1-\vartheta^2)}\Big)^{1/2}. \tag{3.20b}$$

For $\vartheta = 1$ or $\mathbf{L}$ positive semidefinite we formally set $c_{\mathrm{stb}} = \infty$.

Proof. (a) Since the spectrum of $\mathbf{\Phi}$ is in $[0,\pi]$ and $|\sin(n\zeta)/\sin\zeta| \le n$ for $\zeta \in \mathbb{R}$ and $n \in \mathbb{N}$, the estimates follow immediately.
(b) From (3.18b) and the symmetry of $\mathbf{L}_{\Psi,\tau}$ we obtain with (3.17), (3.13) for $\mathbf{q} \in \mathbb{R}^d$

$$\big\|(\sin\mathbf{\Phi})\,\mathbf{q}\big\|^2 = \big(\tau^2\mathbf{L}_{\Psi,\tau}(\mathbf{I}_d - \tfrac{1}{4}\tau^2\mathbf{L}_{\Psi,\tau})\mathbf{q},\mathbf{q}\big) \ge \tau^2\big(c_{\mathrm{inv}}^2 + \tau^2 c_{\mathrm{cs}}\big)^{-1} m_1(1-\vartheta^2)\|\mathbf{q}\|^2.$$

Hence, the inverse of $\sin\mathbf{\Phi}$ exists for $\tau > 0$ and replacing $\mathbf{q}$ by $(\sin\mathbf{\Phi})^{-1}\mathbf{q}$ completes the proof. ∎

We are now in the position to state the stability results. For the semilinear case this additionally requires Lipschitz continuity of the function g.

ASSUMPTION 3.12. *The function $g\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ is (globally) Lipschitz continuous in the second argument, i.e.,*

$$\|g(t,\mathbf{q}) - g(t,\widehat{\mathbf{q}})\| \le L_g\|\mathbf{q}-\widehat{\mathbf{q}}\| \qquad \text{for all } \mathbf{q},\widehat{\mathbf{q}} \in \mathbb{R}^d,\ t \in [0,T]. \tag{3.21}$$

It is well-known that under this assumption the exact solution of (1.1) is unique and exists for all $t \ge 0$.

Remark 3.13. It is sufficient to require that g is locally Lipschitz continuous in a strip around the exact solution, i.e., in a neighborhood of $\{\mathbf{q}(t) : 0 \le t \le T\}$ with T smaller than the maximal existence time of the exact solution. This can be seen from the error bounds in Theorem 4.4 below. For the sake of presentation we omit the details.

THEOREM 3.14. *Let Assumptions 2.1, 2.2, and 3.3 and Assumption 3.12 on g hold. Further, let $\vartheta \in (0,1]$, $\tau \le \tau_{\mathrm{CFL}}(\vartheta)$, and denote by $\mathbf{q}_n$ and $\widehat{\mathbf{q}}_n$ the approximations obtained by (2.4) with initial values $\mathbf{q}_0,\mathbf{p}_0$ and $\widehat{\mathbf{q}}_0,\widehat{\mathbf{p}}_0$. Then we have for $t_n \le T$*

$$\|\mathbf{q}_n - \widehat{\mathbf{q}}_n\| \le \Big(\|\mathbf{q}_0 - \widehat{\mathbf{q}}_0\| + \min\{T, c_{\mathrm{stb}}\}\|\mathbf{p}_0 - \widehat{\mathbf{p}}_0\|\Big)\mathrm{e}^{(c_{\widehat{\Psi}}L_g)^{1/2}T}. \tag{3.22}$$

Clearly, for $\vartheta = 1$ or $\mathbf{L}$ positive semidefinite we have $\min\{t_n, c_{\mathrm{stb}}\} = t_n$, since then $c_{\mathrm{stb}} = \infty$; cf. Lemma 3.11.

Proof. From the representation formula (3.18a) and the Lipschitz condition (3.21) we obtain

$$\begin{aligned}\|\mathbf{q}_n - \widehat{\mathbf{q}}_n\| \le \|\cos(n\mathbf{\Phi})\|\|\mathbf{q}_0 - \widehat{\mathbf{q}}_0\| + \tau\|\boldsymbol{\mathcal{S}}_n\|\|\mathbf{p}_0 - \widehat{\mathbf{p}}_0\| \\ + \tau^2 L_g \sum_{\ell=0}^{n-1} \|\boldsymbol{\mathcal{S}}_{n-\ell}\|\big\|\widehat{\Psi}(\tau^2\mathbf{LR})\big\|\|\mathbf{q}_\ell - \widehat{\mathbf{q}}_\ell\|.\end{aligned}$$

Employing Lemmas 3.8 and 3.11 yields

$$\|\mathbf{q}_n - \widehat{\mathbf{q}}_n\| \le \|\mathbf{q}_0 - \widehat{\mathbf{q}}_0\| + \min\{t_n, c_{\mathrm{stb}}\}\|\mathbf{p}_0 - \widehat{\mathbf{p}}_0\| + \tau^2 L_g c_{\widehat{\Psi}} \sum_{\ell=0}^{n-1} (n-\ell)\|\mathbf{q}_\ell - \widehat{\mathbf{q}}_\ell\|,$$

which finishes the proof by using $t_n \le T$ and applying the Gronwall-type inequality in [4, Lemma 3.8]. ∎

In the linear case, i.e., $g(t, \mathbf{q}) = g(t)$, we obtain the following result.

THEOREM 3.15. *Let Assumptions 2.1, 2.2, and 3.3 hold as well as $g(t, \mathbf{q}) = g(t)$. Further, let $\vartheta \in (0, 1]$ and $\tau \le \tau_{\mathrm{CFL}}(\vartheta)$. Then we have for the approximation $\mathbf{q}_n$ obtained by the scheme (2.4)*

$$\|\mathbf{q}_n\| \le \|\mathbf{q}_0\| + \min\{t_n, c_{\mathrm{stb}}\}\|\mathbf{p}_0\| + \min\{t_n, c_{\mathrm{stb}}\}\, c_{\widehat{\Psi}}\tau \sum_{\ell=0}^{n-1} \|\mathbf{g}_\ell\|. \tag{3.23}$$

Proof. Similarly to the proof for the semilinear case, we obtain the assertion by the representation formula (3.18a) and Lemmas 3.8 and 3.11. ∎

4. Error analysis. The aim of this section is to provide an error analysis for the scheme (2.4). More precisely, we will show a convergence result in the standard norm $\|\cdot\|$ for the linear as well as for the semilinear problem.

Let us denote the error of the scheme (2.4) by

$$\mathbf{e}_n = \widetilde{\mathbf{q}}_n - \mathbf{q}_n, \qquad \widetilde{\mathbf{q}}_n = \mathbf{q}(t_n), \tag{4.1}$$

where $\mathbf{q}(t)$ is the exact solution of (1.1) at time t. We denote bounds on the kth derivative of $\mathbf{q}$ by

$$B_n^{(k)} = \max_{0\le t\le t_n} \|\mathbf{q}^{(k)}(t)\|. \tag{4.2}$$

Moreover, it is well-known that for $\mathbf{q} \in C^k([0, T])$ the remainder terms

$$\boldsymbol{\delta}_{n,\pm}^{(k)} = \tau^{k-1} \int_{t_n}^{t_{n\pm1}} \kappa_{n,\pm}^{(k-1)}(t)\, \mathbf{q}^{(k)}(t)\, dt, \qquad \kappa_{n,\pm}^{(\ell)}(t) = (t_{n\pm1} - t)^\ell/(\ell!\tau^\ell), \tag{4.3}$$

of the $(k-1)$st-order Taylor expansion of $\widetilde{\mathbf{q}}_{n\pm1}$ at t_n are bounded by

$$\|\boldsymbol{\delta}_{n,+}^{(k)}\| \le \tau^k \tfrac{1}{k!} \max_{t_n\le t\le t_{n+1}} \|\mathbf{q}^{(k)}(t)\|, \qquad \|\boldsymbol{\delta}_{n,-}^{(k)}\| \le \tau^k \tfrac{1}{k!} \max_{t_{n-1}\le t\le t_n} \|\mathbf{q}^{(k)}(t)\|. \tag{4.4}$$

As for the stability analysis, we will utilize a representation formula for $\mathbf{e}_n$ to show the error bounds. For this we first derive an error recursion for the scheme (2.4).

LEMMA 4.1. *For* $\mathbf{q} \in C^4([0,T])$ *the error* $\mathbf{e}_n$ *of the scheme* (2.4a) *satisfies for* $n \geq 1$ *the recursion*

$$\mathbf{e}_{n+1} - 2\mathbf{e}_n + \mathbf{e}_{n-1} = \widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\big(-\tau^2\mathbf{L}\mathbf{e}_n + \mathbf{r}_n\big) + \mathbf{d}_n, \tag{4.5a}$$

where

$$\mathbf{d}_n = \boldsymbol{\Delta}_n + \boldsymbol{\delta}_n^{(4)}, \qquad \boldsymbol{\Delta}_n = -\tau^4 \mathcal{X}(\tau^2\mathbf{L}\mathbf{R})\mathbf{L}\mathbf{R}\,\ddot{\mathbf{q}}(t_n), \quad \boldsymbol{\delta}_n^{(4)} = \boldsymbol{\delta}_{n,+}^{(4)} + \boldsymbol{\delta}_{n,-}^{(4)}\,, \tag{4.5b}$$

and

$$\mathbf{r}_n = \tau^2\big(g(t_n, \widetilde{\mathbf{q}}_n) - g(t_n, \mathbf{q}_n)\big). \tag{4.5c}$$

Proof. Inserting the exact solution $\widetilde{\mathbf{q}}_n$ into the scheme (2.4a) yields

$$\widetilde{\mathbf{q}}_{n+1} - 2\widetilde{\mathbf{q}}_n + \widetilde{\mathbf{q}}_{n-1} = \tau^2\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\big(-\mathbf{L}\widetilde{\mathbf{q}}_n + g(t_n, \widetilde{\mathbf{q}}_n)\big) + \mathbf{d}_n \tag{4.6}$$

with a defect $\mathbf{d}_n$. Subtracting the recursion (2.4a) from (4.6) leads to (4.5a). In order to determine $\mathbf{d}_n$ we use Taylor expansion to obtain

$$\widetilde{\mathbf{q}}_{n+1} - 2\widetilde{\mathbf{q}}_n + \widetilde{\mathbf{q}}_{n-1} = \tau^2\ddot{\mathbf{q}}(t_n) + \boldsymbol{\delta}_{n,+}^{(4)} + \boldsymbol{\delta}_{n,-}^{(4)}.$$

Subtracting this relation from (4.6), using the differential equation (1.1), and the definition (2.5) of $\mathcal{X}$ completes the proof. ∎

Next, we compute the error of the starting value $\mathbf{q}_1$ defined in (2.4b).

LEMMA 4.2. *For* $\mathbf{q} \in C^3([0,T])$ *the error* $\mathbf{e}_1$ *of the scheme* (2.4b) *satisfies*

$$\mathbf{e}_1 = \tfrac{1}{2}\boldsymbol{\Delta}_0 + \boldsymbol{\delta}_{0,+}^{(3)}, \tag{4.7}$$

with $\boldsymbol{\Delta}_0$ *given in* (4.5b) *and* $\boldsymbol{\delta}_{0,+}^{(3)}$ *in* (4.3).

Proof. Similarly to the previous lemma we first insert the exact solution into (2.4b) and obtain

$$\widetilde{\mathbf{q}}_1 = \mathbf{q}_0 + \tau\mathbf{p}_0 + \tfrac{1}{2}\tau^2\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\big(-\mathbf{L}\mathbf{q}_0 + \mathbf{g}_0\big) + \mathbf{d}_0 = \mathbf{q}_1 + \mathbf{d}_0$$

with a defect $\mathbf{d}_0 = \mathbf{e}_1$. A Taylor expansion of $\widetilde{\mathbf{q}}_1$ shows (4.7) with the same arguments as before. ∎

With the previous two lemmas we are now in the position to state a representation formula for the error $\mathbf{e}_n$, $n \in \mathbb{N}$. Assume $\tau \leq \tau_{\mathrm{CFL}}$. Then we can show for (4.5a) as in the proof of Theorem 3.10, cf. (3.19),

$$\mathbf{e}_n = \cos(n\boldsymbol{\Phi})\mathbf{e}_0 + \boldsymbol{\mathcal{S}}_n\big(\mathbf{e}_1 - \cos\boldsymbol{\Phi}\,\mathbf{e}_0\big) + \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\Big(\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\,\mathbf{r}_\ell + \mathbf{d}_\ell\Big).$$

Since $\mathbf{e}_0 = 0$, we have with (4.7)

$$\mathbf{e}_n = \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\Big(\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\,\mathbf{r}_\ell + \boldsymbol{\delta}_\ell^{(4)}\Big) + \boldsymbol{\mathcal{S}}_n\boldsymbol{\delta}_{0,+}^{(3)} + \tfrac{1}{2}\boldsymbol{\mathcal{S}}_n\boldsymbol{\Delta}_0 + \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\boldsymbol{\Delta}_\ell\,. \tag{4.8}$$

Before we continue with the error analysis, we first have a closer look at $\boldsymbol{\Delta}_n$ defined in (4.5b). We obtain $\|\boldsymbol{\Delta}_n\| \leq \tau^4 c\,\|\mathbf{L}\mathbf{R}\,\ddot{\mathbf{q}}(t_n)\|$, since similarly to the proof of

Lemma 3.8 we have $\|\mathcal{X}(\tau^2\mathbf{LR})\| \le c$. However, to avoid the loss of consistency if $\mathbf{L}$ is a discretized differential operator, we want the bounds to depend only on derivatives of $\mathbf{q}$ or $\mathbf{Lq}$; cf. [11, Lemma 2.8].

This can be achieved by combining the defects $\boldsymbol{\Delta}_\ell$ of three successive time steps. A similar trick is used in the error analysis of one-step methods for (spatially discretized) partial differential equations; see, e.g., [3] or [13, Lemma II.2.3]. In the context of locally implicit schemes this approach was employed in [11, 16] for Maxwell's equation.

LEMMA 4.3. *Let $\tau \le \tau_{\mathrm{CFL}}$. Then we have*

$$\begin{aligned}\tfrac{1}{2}\boldsymbol{\mathcal{S}}_n\boldsymbol{\Delta}_0 + \sum_{\ell=1}^{n-1}\boldsymbol{\mathcal{S}}_{n-\ell}\boldsymbol{\Delta}_\ell &= \sum_{\ell=1}^{n-1}\boldsymbol{\mathcal{S}}_{n-\ell}\big(\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\widetilde{\boldsymbol{\Delta}}_\ell + \widetilde{\boldsymbol{\Delta}}_{\ell-1}\big)\\ &\quad - \widetilde{\boldsymbol{\Delta}}_n + \boldsymbol{\mathcal{S}}_n(\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0) + \cos(n\boldsymbol{\Phi})\widetilde{\boldsymbol{\Delta}}_0\end{aligned} \tag{4.9}$$

with

$$\widetilde{\boldsymbol{\Delta}}_n = \tau^2\mathbf{R}\,\widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})^{-1}\mathcal{X}(\tau^2\widetilde{\mathbf{S}})\mathbf{R}\,\ddot{\mathbf{q}}(t_n). \tag{4.10}$$

Proof. Let $\ell \in \{0, 1, \dots, n-1\}$. For $\boldsymbol{\Delta}_\ell$ given in (4.5b) we have with Lemma 3.8, (3.10), and definition (3.7) for $\mathbf{L}_{\Psi,\tau}$

$$\begin{aligned}\boldsymbol{\Delta}_\ell &= -\tau^4\,\widehat{\Psi}(\tau^2\mathbf{LR})\,\widehat{\Psi}(\tau^2\mathbf{LR})^{-1}\mathcal{X}(\tau^2\mathbf{LR})\,\mathbf{LR}\,\ddot{\mathbf{q}}(t_\ell)\\ &= -\tau^4\,\widehat{\Psi}(\tau^2\mathbf{LR})\mathbf{L}\,\mathbf{R}\,\widehat{\Psi}(\tau^2\widetilde{\mathbf{S}})^{-1}\mathcal{X}(\tau^2\widetilde{\mathbf{S}})\mathbf{R}\,\ddot{\mathbf{q}}(t_\ell)\\ &= -\tau^2\mathbf{L}_{\Psi,\tau}\widetilde{\boldsymbol{\Delta}}_\ell.\end{aligned}$$

Hence, we obtain with (3.18b)

$$\tfrac{1}{2}\boldsymbol{\mathcal{S}}_n\boldsymbol{\Delta}_0 = -\boldsymbol{\mathcal{S}}_n\tfrac{1}{2}\tau^2\mathbf{L}_{\Psi,\tau}\widetilde{\boldsymbol{\Delta}}_0 = \boldsymbol{\mathcal{S}}_n\cos\boldsymbol{\Phi}\,\widetilde{\boldsymbol{\Delta}}_0 - \boldsymbol{\mathcal{S}}_n\widetilde{\boldsymbol{\Delta}}_0. \tag{4.11}$$

Further, again (3.18b) and a trigonometric identity yield

$$\boldsymbol{\mathcal{S}}_{n-\ell}\boldsymbol{\Delta}_\ell = \boldsymbol{\mathcal{S}}_{n-\ell}\,2\,(\cos\boldsymbol{\Phi} - \mathbf{I}_d)\,\widetilde{\boldsymbol{\Delta}}_\ell = \big(\boldsymbol{\mathcal{S}}_{n-\ell+1} - 2\boldsymbol{\mathcal{S}}_{n-\ell} + \boldsymbol{\mathcal{S}}_{n-\ell-1}\big)\widetilde{\boldsymbol{\Delta}}_\ell,$$

which implies

$$\begin{aligned}\sum_{\ell=1}^{n-1}\boldsymbol{\mathcal{S}}_{n-\ell}\boldsymbol{\Delta}_\ell &= \sum_{\ell=0}^{n-2}\boldsymbol{\mathcal{S}}_{n-\ell}\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\sum_{\ell=1}^{n-1}\boldsymbol{\mathcal{S}}_{n-\ell}\widetilde{\boldsymbol{\Delta}}_\ell + \sum_{\ell=2}^{n}\boldsymbol{\mathcal{S}}_{n-\ell}\widetilde{\boldsymbol{\Delta}}_{\ell-1}\\ &= \sum_{\ell=1}^{n-1}\boldsymbol{\mathcal{S}}_{n-\ell}\big(\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\widetilde{\boldsymbol{\Delta}}_\ell + \widetilde{\boldsymbol{\Delta}}_{\ell-1}\big) - \widetilde{\boldsymbol{\Delta}}_n + \boldsymbol{\mathcal{S}}_n\widetilde{\boldsymbol{\Delta}}_1 - \boldsymbol{\mathcal{S}}_{n-1}\widetilde{\boldsymbol{\Delta}}_0.\end{aligned}$$

Combining this equation with (4.11) and applying a trigonometric identity finishes the proof. ∎

We can now state the error bound for the semilinear problems.

THEOREM 4.4. *Let Assumptions 2.1, 2.2, 3.3, and 3.12 hold and let $\vartheta \in (0, 1]$, $\tau \le \tau_{\mathrm{CFL}}(\vartheta)$. Further, the solution $\mathbf{q}$ of (1.1) satisfies $\mathbf{q} \in C^4([0, T])$. Then we have for the error $\mathbf{e}_n = \mathbf{q}_n - \mathbf{q}(t_n)$ of the scheme (2.4) and $t_n \le T$*

$$\|\mathbf{e}_n\| \le \big(\min\{T, c_{\mathrm{stb}}\}(C_1T + C_2) + C_3\big)\mathrm{e}^{(L_g c_{\widehat{\Psi}})^{1/2}T}\tau^2, \tag{4.12}$$

with

$$C_1 = (\tfrac{1}{12} + c_{\mathrm{cs}})B_n^{(4)}, \quad C_2 = (\tfrac{1}{6} + c_{\mathrm{cs}})B_1^{(3)}, \quad C_3 = c_{\mathrm{cs}}\big(\|\ddot{\mathbf{q}}(0)\| + \|\ddot{\mathbf{q}}(t_n)\|\big).$$

Proof. With Lemma 4.3 the representation formula (4.8) can be written as

$$\begin{aligned}\mathbf{e}_n &= \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\Big(\widehat{\Psi}(\tau^2\mathbf{L}\mathbf{R})\,\mathbf{r}_\ell + \boldsymbol{\delta}_\ell^{(4)}\Big) + \boldsymbol{\mathcal{S}}_n\boldsymbol{\delta}_{0,+}^{(3)} \\ &\quad + \sum_{\ell=1}^{n-1} \boldsymbol{\mathcal{S}}_{n-\ell}\big(\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\widetilde{\boldsymbol{\Delta}}_\ell + \widetilde{\boldsymbol{\Delta}}_{\ell-1}\big) - \widetilde{\boldsymbol{\Delta}}_n + \boldsymbol{\mathcal{S}}_n(\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0) + \cos(n\boldsymbol{\Phi})\widetilde{\boldsymbol{\Delta}}_0\,.\end{aligned} \tag{4.13}$$

Lemmas 3.8 and 3.11, (4.5c), and the Lipschitz continuity of g then give

$$\begin{aligned}\|\mathbf{e}_n\| &\le L_g\tau^2\sum_{\ell=1}^{n-1}(n-\ell)c_{\widehat{\Psi}}\|\mathbf{e}_\ell\| + \min\{t_n, c_{\mathrm{stb}}\}\Big(\sum_{\ell=1}^{n-1}\tfrac{1}{\tau}\|\boldsymbol{\delta}_\ell^{(4)}\| + \tfrac{1}{\tau}\|\boldsymbol{\delta}_{0,+}^{(3)}\|\Big) \\ &\quad + \min\{t_n, c_{\mathrm{stb}}\}\Big(\sum_{\ell=1}^{n-1}\tfrac{1}{\tau}\big\|\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\widetilde{\boldsymbol{\Delta}}_\ell + \widetilde{\boldsymbol{\Delta}}_{\ell-1}\big\| + \tfrac{1}{\tau}\|\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0\|\Big) + \|\widetilde{\boldsymbol{\Delta}}_0\| + \|\widetilde{\boldsymbol{\Delta}}_n\|.\end{aligned}$$

We bound the single terms separately.
(a) The remainder terms of the Taylor expansion arising in (4.5b) and (4.7) can be bounded with (4.4) by

$$\sum_{\ell=1}^{n-1}\tfrac{1}{\tau}\|\boldsymbol{\delta}_\ell^{(4)}\| + \tfrac{1}{\tau}\|\boldsymbol{\delta}_{0,+}^{(3)}\| \le \tau^2 t_n \tfrac{1}{12}B_n^{(4)} + \tau^2\tfrac{1}{6}B_1^{(3)}.$$

(b) For $\widetilde{\boldsymbol{\Delta}}_0$ and $\widetilde{\boldsymbol{\Delta}}_n$ defined in (4.10) we have by (3.4) and the CFL condition (3.9a)

$$\|\widetilde{\boldsymbol{\Delta}}_0\| + \|\widetilde{\boldsymbol{\Delta}}_n\| \le \tau^2 c_{\mathrm{cs}}\big(\|\ddot{\mathbf{q}}(0)\| + \|\ddot{\mathbf{q}}(t_n)\|\big).$$

(c) For the central difference quotient of $\widetilde{\boldsymbol{\Delta}}_\ell$ we first use again (3.4) and (3.9a) and afterwards Taylor expansion to obtain

$$\begin{aligned}\sum_{\ell=1}^{n-1}\tfrac{1}{\tau}\|\widetilde{\boldsymbol{\Delta}}_{\ell+1} - 2\widetilde{\boldsymbol{\Delta}}_\ell + \widetilde{\boldsymbol{\Delta}}_{\ell-1}\| &\le \sum_{\ell=1}^{n-1}\tau c_{\mathrm{cs}}\|\ddot{\mathbf{q}}(t_{\ell+1}) - 2\ddot{\mathbf{q}}(t_\ell) + \ddot{\mathbf{q}}(t_{\ell-1})\| \\ &\le \sum_{\ell=1}^{n-1}\tau^3 c_{\mathrm{cs}}\max_{t_{\ell-1}\le t\le t_{\ell+1}}\|\mathbf{q}^{(4)}(t)\| \\ &\le \tau^2 c_{\mathrm{cs}}\,t_n B_n^{(4)}.\end{aligned}$$

(d) Finally, with the same arguments as in the previous step we have

$$\tfrac{1}{\tau}\|\widetilde{\boldsymbol{\Delta}}_1 - \widetilde{\boldsymbol{\Delta}}_0\| \le \tau^2 c_{\mathrm{cs}} B_1^{(3)}.$$

Collecting the estimates and using the Gronwall-type inequality from [4, Lemma 3.8] finishes the proof. ∎

For linear problems we have $L_g = 0$ (and $\mathbf{r}_\ell = 0$ for all $\ell = 1,\dots,n$), which leads to the following result.

COROLLARY 4.5. *Let Assumptions 2.1, 2.2, and 3.3 hold and let $\vartheta \in (0,1]$, $\tau \le \tau_{\mathrm{CFL}}(\vartheta)$. Further, let $g(t,\mathbf{q}) = g(t)$ and the solution $\mathbf{q}$ of (1.1) satisfies $\mathbf{q} \in C^4([0,T])$. Then we have for the the error $\mathbf{e}_n = \mathbf{q}_n - \mathbf{q}(t_n)$ of the scheme (2.4) and $t_n \le T$*

$$\|\mathbf{e}_n\| \le \big(\min\{t_n, c_{\mathrm{stb}}\}(C_1 t_n + C_2) + C_3\big)\tau^2, \tag{4.14}$$

with C_1, C_2, and C_3 defined as in Theorem 4.4.

5. sLFC schemes – constants, implementation, and efficiency. After we have derived abstract stability and error results for the scheme (2.4) in the last two sections, we now focus on the sLFC scheme, i.e., scheme (2.4) equipped with the polynomials (2.9). We first show that these polynomials satisfy Assumption 3.3 and state explicit values for the constants arising in Section 3.1. Afterwards we show how the sLFC scheme can be implemented efficiently and make a heuristic comparison of the efficiency with the LF scheme.

5.1. Constants for sLFC schemes. For the polynomials (2.9), most of the constants have already been computed in [4, 7] for a specific choice of $\widehat{\beta}$. Adopting this choice yields the following.

LEMMA 5.1. *For the LFC polynomials* (2.9) *we have for* $\nu > 1$ *and*

$$m_1 = \widetilde{m}_1 = \frac{1}{2}\Big(1 - \frac{1}{T_p(\nu)}\Big), \quad \widetilde{m}_2 = \frac{4\widetilde{m}_1}{\alpha_p(\nu-1)} = \frac{T_p(\nu)-1}{T_p'(\nu)(\nu-1)} \tag{5.1a}$$

that $\widehat{\beta}^2 = \widehat{\beta}_p^2 = \alpha_p(\nu+1) \le 4p^2$. *Moreover, Assumption 3.3 is satisfied and we have*

$$m_3 = -\Psi_p''(0) = 2\,\frac{T_p''(\nu)}{\alpha_p^2 T_p(\nu)}, \qquad c_{\mathrm{cs}} = \frac{1}{4\widetilde{m}_1}. \tag{5.1b}$$

Proof. Assumption 3.3 and the constants for m_1, m_3 arising in Definition 3.1 are proven in [4, Theorems 5.1 and 5.2(b)].

Next, we show the lower bounds in (3.1) and abbreviate $\sigma_{p,\nu} = \alpha_p(\nu-1)$. For $\sigma_{p,\nu} \le z \le \widehat{\beta}^2$ the constant $\widetilde{m}_1$ in (3.1) can be shown similarly to m_1; see also [7, Lemma A.4]. For $0 \le z \le \sigma_{p,\nu}$ the function Ψ_p is concave and monotonically increasing. Thus, we have

$$\Psi_p(z) \ge \frac{\Psi_p(\sigma_{p,\nu})}{\sigma_{p,\nu}} z = \widetilde{m}_2 z.$$

It remains to show (3.4). First, we note that $\widehat{\Psi}(z)^{-1}\mathcal{X}(z) \le 0$ for $z \in [0, \widehat{\beta}^2]$. Further, we have for $\sigma_{p,\nu} \le z \le \widehat{\beta}^2$

$$\widehat{\Psi}(z)^{-1}\mathcal{X}(z) = \frac{1}{z} - \frac{1}{\Psi(z)} > -\frac{1}{\Psi(z)} \ge -\frac{1}{4\widetilde{m}_1}.$$

For $0 \le z \le \sigma_{p,\nu}$ a tedious calculation shows that $\widehat{\Psi}(\cdot)^{-1}\mathcal{X}(\cdot)$ is monotonically decreasing. This yields

$$\widehat{\Psi}(z)^{-1}\mathcal{X}(z) \ge \widehat{\Psi}(\sigma_{p,\nu})^{-1}\mathcal{X}(\sigma_{p,\nu}) \ge -\frac{1}{\Psi(\sigma_{p,\nu})} = -\frac{1}{4\widetilde{m}_1},$$

which concludes the proof. ∎

If we set $\nu = 1+\eta^2/(2p^2)$ with a parameter $\eta > 0$, then all constants in Lemma 5.1 except $\widehat{\beta}^2$ can be chosen independent of the polynomial degree p. The choice is motivated by stabilized/damped Runge–Kutta–Chebyshev methods [13, 17], where a similar scaling of the parameter also leads to bounds which are independent of p.

COROLLARY 5.2. *If* $\nu = \nu_p = 1 + \eta^2/(2p^2)$ *with* $\eta > 0$, *we have for the LFC polynomials* (2.9)

$$\frac{4p^2}{\cosh(\eta)} \le \widehat{\beta}^2 \le 4p^2 \qquad \text{for} \quad m_1 = \widetilde{m}_1 = \frac{\eta^2}{4+2\eta^2}, \quad \widetilde{m}_2 = \frac{1}{\operatorname{sinhc}(\eta)}, \tag{5.2a}$$

and

$$m_3 = \frac{\cosh(\eta)}{2}\,\frac{\cosh(\eta)-\operatorname{sinhc}(\eta)}{\eta^2}, \qquad c_{\mathrm{cs}} = \frac{2+\eta^2}{2\eta^2}. \tag{5.2b}$$

Note that the bounds hold for all $\eta > 0$ and $p \in \mathbb{N}$.

Proof. The proofs of these constants rely on Lemma 5.1 and the estimates

$$T_p^{(k)}(\nu_p) \geq T_p^{(k)}(1) + T_p^{(k+1)}(1)(\nu_p - 1), \qquad \frac{\mathrm{d}^k}{\mathrm{d}x^k} T_p\Big(1 + \frac{x}{2p^2}\Big) \leq \frac{\mathrm{d}^k}{\mathrm{d}x^k}\cosh(\sqrt{x})$$

for $p \in \mathbb{N}$, $k \in \mathbb{N}_0$, and $x \geq 0$, which can be seen by Taylor expansion. ∎

5.2. Implementation. Next, we turn towards the implementation of the sLFC scheme. For the sake of readability we focus on the implementation of the two-step scheme (2.4). The same strategy can be applied to the corresponding one-step formulation (2.12). Moreover, the dominant parts of the computational cost coincide for both schemes. We emphasize that the algorithm can be easily adapted to the case $\mathbf{M} \neq \mathbf{I}_d$, if we are in the setting of Remark 2.5.

An efficient implementation of one time step of the sLFC scheme is based on the representation (3.12) for $\widehat{\Psi}(\tau^2\mathbf{LR})$. Details are given in Algorithm 5.1, in which the notation from (3.14) is used, i.e., for $\mathbf{b} \in \mathbb{R}^d$ we denote by $\mathbf{b}_S \in \mathbb{R}^s$ and $\mathbf{b}_N \in \mathbb{R}^{d-s}$ the subvectors of $\mathbf{b}$ belonging to the stiff and the nonstiff part of the differential equation. The starting value (2.4b) can be computed via a similar strategy.

Algorithm 5.1 One time step of the sLFC scheme ((2.4a) with (2.9))

1: $\mathbf{b} = -\mathbf{L}\mathbf{q}_n + \mathbf{g}_n$
2: $\widetilde{\mathbf{b}}_S = \mathcal{X}(\tau^2\mathbf{S})\mathbf{b}_S$
3: $\widehat{\mathbf{b}}_S = \mathbf{b}_S + \tau^2\mathbf{S}\widetilde{\mathbf{b}}_S$
4: $\widehat{\mathbf{b}}_N = \mathbf{b}_N + \tau^2\mathbf{K}\mathbf{b}_S$
5: $\mathbf{q}_{n+1} = 2\mathbf{q}_n - \mathbf{q}_{n-1} + \tau^2\widehat{\mathbf{b}}$

For the computation of $\mathcal{X}\big(\tau^2\mathbf{S}\big)\mathbf{b}_S$ one can employ the recurrence relation in Lemma 5.3, because $\mathcal{X} = \mathcal{X}_{p,p}$. This is advantageous over a computation via Horner's method, since it is more stable and the factors of the polynomial $\mathcal{X}$ (which change with varying p and ν) need not to be precomputed. Only the scalar values $T_k(\nu)$, $T_k'(\nu)$, and α_k for $k = 1, \dots, p$ have to be known in advance, but they are easy to compute. An algorithm for the computation of $\mathcal{X}\big(\tau^2\mathbf{S}\big)\mathbf{b}_S$ can be derived from [4, Algorithm 6.1], in which a similar recursion is computed.

LEMMA 5.3. *The polynomials*

$$\mathcal{X}_{k,p}(z) = \frac{R_{k,p}(z)}{z^2}, \quad R_{k,p}(z) = 2 - \frac{2}{T_k(\nu)} T_k\Big(\nu - \frac{z}{\alpha_p}\Big) - \frac{\alpha_k}{\alpha_p} z, \qquad 1 \leq k \leq p,$$

satisfy the recursion

$$\mathcal{X}_{1,p}(z) = 0, \qquad \mathcal{X}_{2,p}(z) = -\frac{4}{\alpha_p^2 T_2(\nu)},$$
$$T_{k+1}(\nu)\mathcal{X}_{k+1,p}(z) = -2T_k(\nu)\frac{\alpha_k}{\alpha_p^2} + 2\Big(\nu - \frac{z}{\alpha_p}\Big)T_k(\nu)\mathcal{X}_{k,p}(z) - T_{k-1}(\nu)\mathcal{X}_{k-1,p}(z),$$

for $k = 2, \dots, p-1$.

Table 5.1: Effort in matrix-vector multiplications (MVM) and evaluations of g for the LF scheme (2.1a) with $\mathbf{M} = \mathbf{I}_d$ and the sLFC scheme (2.4a),(2.9). p denotes the degree of the polynomial Ψ_p.

LF scheme (2.1a)	sLFC scheme (2.4a) + (2.9)
1 MVM with $\mathbf{L}$	1 MVM with $\mathbf{L}$
1 evaluation of g	1 evaluation of g
	$p-1$ MVMs with $\mathbf{S}$
	1 MVM with $\mathbf{K}$

Proof. The result follows from the recursion of Chebyshev polynomials and its derivatives. ∎

5.3. Efficiency. The efficiency of the sLFC scheme (Algorithm 5.1) is now compared to the classical LF scheme (2.1). The main cost in terms of matrix-vector multiplications and evaluations of g for one time step of each of the schemes are given in Table 5.1. As one observes, the additional (main) effort for the sLFC scheme consists of $p-1$ matrix-vector multiplications with $\mathbf{S}$ and one with $\mathbf{K}$. For $s \ll d$ this is comparatively cheap.

For the total cost we now compare the maximal step sizes for which the schemes are stable. Under Assumption 2.1 we have for the LF scheme

$$\tau_{\mathrm{CFL,LF}}^2 = \frac{4}{\|\mathbf{L}\|} \approx \frac{4}{\|\mathbf{S}\|} = \frac{4}{r\|\mathbf{N}\|}.$$

For the step-size restriction of the sLFC scheme we recall (3.9) with $\vartheta = 1$

$$\tau_{\mathrm{CFL}}^2 = \min\Big\{\frac{\widehat{\beta}^2}{\|\mathbf{S}\|}, \frac{4\gamma}{\|\mathbf{N}\|}\Big\} = \frac{1}{\|\mathbf{N}\|}\min\Big\{\frac{\widehat{\beta}^2}{r}, 4\gamma\Big\}, \quad \gamma = \frac{2}{1+(1+4\kappa m_1^{-1})^{1/2}}.$$

With Corollary 5.2 we then obtain for $\kappa = 1$

$$\min\Big\{\tfrac{4p^2}{r\cosh(\eta)}, \tfrac{8\eta}{\eta+\sqrt{9\eta^2+16}}\Big\} \le \tau_{\mathrm{CFL,sLFC}}^2\|\mathbf{N}\| < \min\Big\{\tfrac{4p^2}{r}, 2\Big\}. \tag{5.3}$$

If we now choose $p \in \mathbb{N}$ such that $p^2 \approx r$ we get a step-size restriction which depends only on η and the submatrix $\mathbf{N}$ but is independent of the stiff part. Hence, with an appropriately chosen η we can use almost as large step sizes as for the LF scheme applied to the nonstiff problem (1.1) with $\mathbf{S} = \mathbf{K} = 0$.

Remark 5.4 (Choice of η). For $p^2 = r$ the maximal value of the lower bound in (5.3) is attained for $\eta \approx 1.556$ and takes an approximate value of 1.6. However, our numerical experiments indicate that the choice $\eta \in [0.4, 1]$ is sufficient if $\kappa \le 1$. For smaller values of η instabilities can occur for certain step sizes; cf. Section 6.1. If η is chosen too large, the value for $\widehat{\beta}$ deteriorates rapidly; see Corollary 5.2 and also [4, Fig. 5.2].

We finally note that in our applications we observed the following weaker step-size restriction

$$\tau_{\mathrm{CFL,sLFC}}^2\|\mathbf{N}\| \approx \min\Big\{\tfrac{\widehat{\beta}^2}{r}, 4\Big\} \approx \min\Big\{\tfrac{4p^2}{r\cosh(\eta)}, 4\Big\}, \tag{5.4}$$

if $\eta \in [0.4, 1]$; see, e.g., Sections 6.2 and 6.3. Besides the fact that κ is often smaller than 1 (which increases γ), the influence of the coupling matrix $\mathbf{K}$ to the (largest) eigenvalues of $\mathbf{S}$, $\mathbf{N}$ is often – although not neglectable – rather small.

6. Examples. In this last section we present some examples confirming our theoretical results and showing possible applications. We start with a simple example for $d = 2$ verifying the necessity of the bounds (3.1) in Definition 3.1 and, hence, the use of stabilization parameters in (2.9). Afterwards we turn towards more realistic numerical examples. The codes for reproducing the numerical results are available on https://doi.org/10.5445/IR/1000133907.

6.1. A two-dimensional problem. We consider the simple linear problem

$$\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t) = -\begin{pmatrix} r & \kappa \\ \kappa & 1 \end{pmatrix} \mathbf{q}(t) \tag{6.1}$$

with $r \gg 1$ and $0 \le \kappa \le r^{1/2}$, i.e., $\mathbf{S} = r$, $\mathbf{N} = 1$, and $\mathbf{K} = \kappa$ in (2.2).

Applying the scheme (2.4) yields with Lemma 3.5 and the definitions of $\widehat{\Psi}$ and $\mathcal{X}$

$$\tau^2\mathbf{L}_{\Psi,\tau} = \begin{pmatrix} \Psi(\tau^2 r) & \Psi(\tau^2 r)\rho \\ \Psi(\tau^2 r)\rho & \tau^2(1-\kappa\rho) + \Psi(\tau^2 r)\rho^2 \end{pmatrix}, \qquad \rho = \kappa r^{-1} \in [0, r^{-1/2}]. \tag{6.2}$$

For the eigenvalues of $\tau^2\mathbf{L}_{\Psi,\tau}$

$$\begin{aligned} \lambda_\pm &= \tfrac{1}{2}\Psi(\tau^2 r)(1+\rho^2) + \tfrac{1}{2}\tau^2(1-\kappa\rho) \\ &\quad \pm \tfrac{1}{2}\Big(\Psi(\tau^2 r)^2(1+\rho^2)^2 - 2\Psi(\tau^2 r)\tau^2(1-\rho^2)(1-\kappa\rho) + \tau^4(1-\kappa\rho)^2\Big)^{1/2} \end{aligned}$$

an easy computation shows that the larger eigenvalue is bounded by

$$\lambda_+ \ge \Psi(\tau^2 r)(1+\rho^2) \qquad \text{if } \tau^2 \le \Psi(\tau^2 r)\frac{1+\rho^2}{1-\kappa\rho}.$$

Hence, the existence of a $\beta > \widehat{\beta}$ such that $\Psi(z) \le 4$ for all $z \in [0, \beta^2]$ is in general not sufficient to guarantee $\lambda_+ \le 4$ for $\tau^2 \le \min\{\beta^2/r, 4\}$ as it would be true for $\mathbf{K} = 0$. This confirms that condition (3.1) with $m_1 > 0$ is indeed necessary to ensure $\lambda_+ \le 4$ and, thus, at least linear stability. A similar behavior occurs for (multirate) LFC schemes; cf. [4]. Moreover, note that the stronger the coupling, i.e., the greater κ, the greater λ_+ can become, since $\rho = \kappa/r$.

In Figure 6.1 we illustrate this by plotting the eigenvalues of (6.2) with the LFC polynomials (2.9) for different stabilization parameters $\eta \ge 0$. We choose $\kappa = 2$, $r = 9$, and, hence, $p = 3$ as polynomial degree for the LFC polynomial; cf. Section 5.3.

As one can clearly see the eigenvalues λ_+ with the unstabilized polynomial Ψ_3 are larger than 4, if the polynomial is equal or too close to 4. At the roots of Ψ_3 the polynomial leads only to a linear growth of the approximations $\mathbf{q}_n$ in time, whereas the exact solution is uniformly bounded for all $t \ge 0$. We further observe that with a sufficiently large η the eigenvalues are bounded away from 0 and 4. The price to pay is a slightly smaller $\widehat{\beta}$, hence, a (slightly) stronger step-size restriction.

6.2. A modified Fermi–Pasta–Ulam–Tsingou problem. As a second example we choose a modification of the famous Fermi–Pasta–Ulam–Tsingou (FPUT) β-problem [6], where a chain of $d + 2$ mass points (with mass 1 for all points) are connected via nonlinear springs. The first and the last point at the end are fixed.

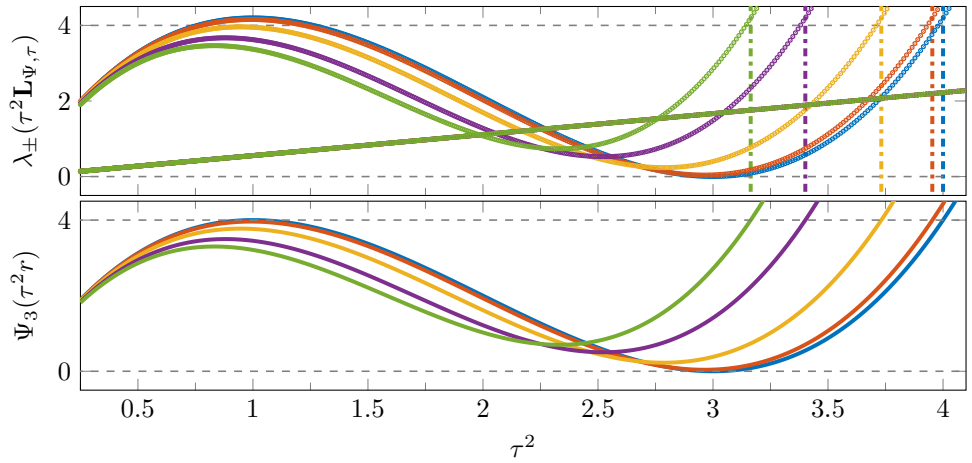


Fig. 6.1: Eigenvalues $\lambda_\pm$ of (6.2) with $r = 9$, $\kappa = 2$, and $\Psi = \Psi_p$ (LFC polynomials) plotted over step sizes τ^2. For Ψ_p we use polynomial degree $p = r^{1/2} = 3$ and stabilization parameters $\eta = 0$, $\eta = 0.2$, $\eta = 0.5$, $\eta = 0.8$, and $\eta = 1$. The dash-dotted lines indicate, where the polynomials $\Psi_3(\tau^2 r)$ leave the interval $[0, 4]$.

In contrast to the original problem we choose springs whose elastic constants can be different for each spring. By q_i we denote the displacement of the ith mass point from its equilibrium and with $p_i = \dot{q}_i$ the velocities, $i = 1, \dots, d$. The Hamiltonian describing the motion of the mass points is then given by

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = \frac{1}{2}\sum_{i=1}^{d} p_i^2 + \frac{1}{2}\sum_{i=0}^{d} \omega_{i+1}^2 (q_{i+1} - q_i)^2 + \frac{\beta_{\mathrm{FPU}}}{4}\sum_{i=0}^{d} (q_{i+1} - q_i)^4,$$

where $\mathbf{q} = (q_1, \dots, q_d)$, $\mathbf{p} = (p_1, \dots, p_d)$, and $q_0 = q_{d+1} = 0$. Deriving the differential equations leads to (1.1) with $\mathbf{M} = \mathbf{I}_d$, a tridiagonal matrix

$$\mathbf{L} = \mathrm{tridiag}\big(\boldsymbol{\omega}_-, \boldsymbol{\omega}_+, \boldsymbol{\omega}_-\big) \in \mathbb{R}^{d\times d}, \quad \boldsymbol{\omega}_- = (-\omega_i^2)_{i=2}^{d},\ \boldsymbol{\omega}_+ = (\omega_i^2 + \omega_{i+1}^2)_{i=1}^{d},$$

and $g = (g_1, \dots, g_d)$ with

$$g_i(\mathbf{q}) = \beta_{\mathrm{FPU}}\big((q_{i+1} - q_i)^3 - (q_i - q_{i-1})^3\big), \qquad i = 1, \dots, d.$$

In our example we set $d = 100$, $\beta_{\mathrm{FPU}} = 2$,

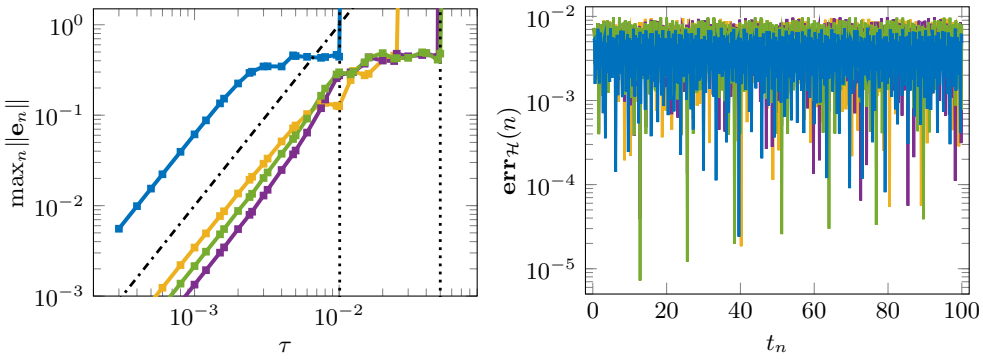
$$\omega_i = 110, \quad i = 1, 2, 3, \qquad \omega_i = 20, \quad i = 4, \dots, d+1,$$

as well as the starting values

$$\mathbf{q}_0 = \begin{cases} 0.25, & i \in \{1, 8\}, \\ 0, & \text{else}, \end{cases} \qquad \mathbf{p}_0 = \begin{cases} -0.1, & i \in \{1, 8\}, \\ 0, & \text{else}. \end{cases}$$

By choosing $\mathbf{S} = (\mathbf{L}_{i,j})_{i,j=1}^{3}$ we have $\|\mathbf{S}\| \approx 39332$, $\|\mathbf{N}\| \approx 1599.6$, and $\|\mathbf{K}\| = 400$. Thus, we obtain $r \approx 24.6$ and $\kappa \approx 0.25$.

In Figure 6.2 we apply the LF scheme (2.1) and the sLFC scheme (2.4), (2.9) with $\eta = 0.5$ and $p = 3, 5, 6$ to the FPUT β-problem. As reference solution for computing errors we apply the LF scheme with step size $\tau = 10^{-5}$ to the problem.



(a) Error for $t = 1.2$.

(b) Relative error of $\mathcal{H}$ for $\tau = 0.001$.

Fig. 6.2: Error (left) and relative error of Hamiltonian (right) for the numerical solution of the FPUT β-problem. For $\Psi = \Psi_p$ we use polynomial degrees $p = 3$, $p = 5$, $p = 6$ and as stabilization parameter $\eta = 0.5$. The blue lines represents the LF scheme. In the left plot the dash-dotted line indicates order two, the dotted lines correspond to the "maximal stable" step size for the LF scheme applied to the stiff system (1.1) and to the the nonstiff problem (1.1) with $\mathbf{S} = \mathbf{K} = 0$. In the right plot the relative error of the Hamiltonian is plotted only at every 200th time step for the sake of clarity.

In the error plot on the left one can observe that with the sLFC scheme the step size can be chosen approximately p times larger than for the LF scheme until $p = 5$. A further increase of the polynomial degree has no positive effect on the step size, which is consistent with Section 5.3 since $r^{1/2} \approx 5$. Moreover, for the same step size the error for the sLFC scheme is significantly smaller than the error for the LF scheme. This is due to the fact that the polynomials (2.9) are for small z a better approximation to the exact solution than the LF scheme; cf. [4]. In general, the difference in the errors of the LF and sLFC schemes strongly depends on the initial values and in many cases there is no visible difference between them.

The plot on the right confirms that the Hamiltonian is nearly conserved for long times, which is due to the symplecticity of the scheme (2.4); cf. Corollary 2.6. The relative error of the Hamiltonian is defined via

$$\mathbf{err}_{\mathcal{H}}(n) = \frac{|\mathcal{H}(\mathbf{p}_n, \mathbf{q}_n) - \mathcal{H}(\mathbf{p}_0, \mathbf{q}_0)|}{\mathcal{H}(\mathbf{p}_0, \mathbf{q}_0)}.$$

6.3. Wave equation. Last, we consider a differential equation stemming from the spatial discretization of the inhomogeneous wave equation with homogeneous Dirichlet boundary conditions on the unit square $\Omega = (0, 1)^2$,

$$\begin{aligned} &\ddot{q}(t,x) = \nabla \cdot \big(c(x)\nabla q(t,x)\big) + f(t,x), && x \in \Omega,\ t \in [0,T], \\ &q(t,x) = 0, && x \in \partial\Omega,\ t \in [0,T], \\ &q(0,x) = q_0(x), \quad \dot{q}(0,x) = \dot{q}_0(x), && x \in \Omega, \end{aligned} \tag{6.3a}$$

with

$$c(x) = \begin{cases} 8.5, & x \in [0.75, 1]^2, \\ 0.73, & x \in [0,1]^2 \setminus [0.75, 1]^2. \end{cases}$$

As initial data and for the inhomogeneity we choose the smooth functions

$$q_0(x) = h(x; 3, 0.25, \left(\begin{smallmatrix}0.4\\0.4\end{smallmatrix}\right)), \quad \dot{q}_0(x) = 0, \qquad f(t,x) = h(x; 4, 0.1, \left(\begin{smallmatrix}0.875\\0.875\end{smallmatrix}\right))\,\mathrm{e}^{-t}, \tag{6.3b}$$

where

$$h(x; a, r_0, x_0) = \mathbb{1}_{\|x-x_0\|\le r_0}\, a\, \exp\Big(-\big(1 - \|x-x_0\|^2/r_0^2\big)^{-1}\Big).$$

For the spatial discretization of (6.3) we employ a symmetric interior penalty discontinuous Galerkin method [2, 8] using piecewise polynomials of degree three on the unstructured mesh illustrated in Figure 6.3(a). This results in (1.1) with g only depending on time t and a symmetric, positive definite $\mathbf{L}$. The mass matrix $\mathbf{M}$ is block diagonal, where the blocks contain the degrees of freedom (dofs) of one mesh element. Thus, solving with or inversion of $\mathbf{M}$ can be done at low cost. For the implementation, we used (2.8) instead of (2.4a); cf. Remark 2.5.

The stiffest part of the differential equation consists of these dofs belonging to the mesh elements, where c is larger, and its adjacent elements because of the flux terms at faces; cf. dotted area in Figure 6.3(a). Hence, after a possible reordering, $\mathbf{S}$ corresponds to the part of $\mathbf{L}$ with the dofs belonging to the dotted area. A numerical computation of the largest eigenvalues yields $r \approx 9.036$ and $\kappa \approx 0.215$.

In Figure 6.3(b) we apply the LF scheme (2.1) and the sLFC scheme (2.4), (2.9) with $p = 2, 3, 4$, $\eta = 0.5$ as well as $p = 3$, $\eta = 0.1$ to this equation. The errors are computed at time $t_N = 2.7$, where we use the LF scheme with step size $\tau = 10^{-5}$ on the same mesh as reference solution. We clearly observe second-order convergence for all applied schemes. Moreover, if stable, the errors between the LF scheme and the sLFC schemes are almost identical. Further, one can observe that, in accordance with Section 5.3, $p = 3 \approx r^{1/2}$ is optimal in the sense of efficiency, since $p = 4$ leads to (almost) no further increase of the maximal step size, where the scheme is stable. In the magnified image section we again see that an insufficient stabilization leads to a stricter step-size restriction than for a larger value for η.

Acknowledgments. We thank Benjamin Dörich for his careful reading of this manuscript.

REFERENCES

[1] A. Albert, *Conditions for positive and nonnegative definiteness in terms of pseudoinverses*, SIAM J. Appl. Math., 17 (1969), pp. 434–440, https://doi.org/10.1137/0117041.

[2] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760, https://doi.org/10.1137/0719052.

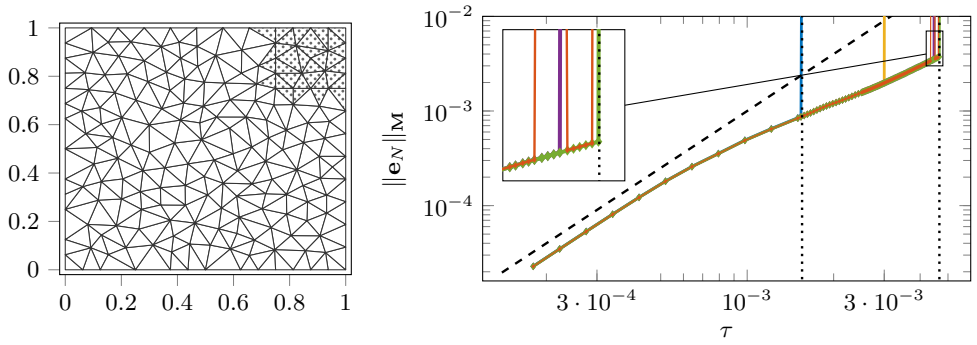
[3] P. Brenner, M. Crouzeix, and V. Thomée, *Single-step methods for inhomogeneous linear differential equations in Banach space*, RAIRO Anal. Numér., 16 (1982), pp. 5–26, www.numdam.org/item/M2AN_1982__16_1_5_0/.

[4] C. Carle, M. Hochbruck, and A. Sturm, *On leapfrog-Chebyshev schemes*, SIAM J. Numer. Anal., 58 (2020), pp. 2404–2433, https://doi.org/10.1137/18M1209453.

[5] J. Diaz and M. J. Grote, *Energy conserving explicit local time stepping for second-order wave equations*, SIAM J. Sci. Comput., 31 (2009), pp. 1985–2014, https://doi.org/10.1137/070709414.

[6] E. Fermi, J. Pasta, S. Ulam, and M. Tsingou, *Studies of the nonlinear problems*, Tech. Report LA-1940, Los Alamos Scientific Lab., N. Mex., 1955, https://doi.org/10.2172/4376203. later published in *E. Fermi: Collected Papers*, Vol. II, E. Segrè (Ed.), University of Chicago Press (1965), pp. 978-988; and *Nonlinear wave motion*, A. C. Newell (Ed.), Lect. Appl. Math. 15, AMS, Providence, RI (1974), pp. 143-156.

[7] M. J. Grote, S. Michel, and S. A. Sauter, *Stabilized leapfrog based local time-stepping method for the wave equation*, 2020, https://arxiv.org/abs/2005.13350. to appear in Math. Comp.



(a) Triangulation of Ω.

(b) Error for $t = 2.7$.

Fig. 6.3: Mesh and error of the numerical solution of the (spatially discretized) wave equation (6.3). For $\Psi = \Psi_p$ we use polynomial degrees $p = 2$, $p = 3$, $p = 4$ with stabilization parameter $\eta = 0.5$ as well as $p = 3$,$\eta = 0.1$. The blue lines represents the LF scheme. The dash-dotted line indicates order two, the dotted lines correspond to the 'maximal stable' step size for the LF scheme applied to the stiff system (1.1) and to the the nonstiff problem (1.1) with $\mathbf{S} = \mathbf{K} = 0$.

[8] M. J. Grote, A. Schneebeli, and D. Schötzau, *Discontinuous Galerkin finite element method for the wave equation*, SIAM J. Numer. Anal., 44 (2006), pp. 2408–2431, https://doi.org/10.1137/05063194X.

[9] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration illustrated by the Störmer-Verlet method*, Acta Numer., 12 (2003), pp. 399–450, https://doi.org/10.1017/S0962492902000144.

[10] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration*, vol. 31 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 2006, https://doi.org/10.1007/3-540-30666-8. Structure-preserving algorithms for ordinary differential equations.

[11] M. Hochbruck and A. Sturm, *Error analysis of a second-order locally implicit method for linear Maxwell's equations*, SIAM J. Numer. Anal., 54 (2016), pp. 3167–3191, https://doi.org/10.1137/15M1038037.

[12] R. A. Horn and F. Zhang, *Basic properties of the Schur complement*, in The Schur complement and its applications, F. Zhang, ed., vol. 4 of Numerical Methods and Algorithms, Springer-Verlag, New York, 2005, pp. xvi+295, https://doi.org/10.1007/b105056.

[13] W. Hundsdorfer and J. Verwer, *Numerical solution of time-dependent advection-diffusion-reaction equations*, vol. 33 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2003, https://doi.org/10.1007/978-3-662-09017-6.

[14] B. Leimkuhler and C. Matthews, *Molecular dynamics*, vol. 39 of Interdisciplinary Applied Mathematics, Springer, Cham, 2015, https://doi.org/10.1007/978-3-319-16375-8. With deterministic and stochastic numerical methods.

[15] B. Leimkuhler and S. Reich, *Simulating Hamiltonian dynamics*, vol. 14 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2004, https://doi.org/10.1017/CBO9780511614118.

[16] J. G. Verwer, *Component splitting for semi-discrete Maxwell equations*, BIT, 51 (2011), pp. 427–445, https://doi.org/10.1007/s10543-010-0296-y.

[17] J. G. Verwer, W. H. Hundsdorfer, and B. P. Sommeijer, *Convergence properties of the Runge-Kutta-Chebyshev method*, Numer. Math., 57 (1990), pp. 157–178, https://doi.org/10.1007/BF01386405.