SIAM J. NUMER. ANAL.
Vol. 58, No. 4, pp. 2404–2433

ON LEAPFROG-CHEBYSHEV SCHEMES*

CONSTANTIN CARLE†, MARLIS HOCHBRUCK†, AND ANDREAS STURM†

Abstract. This paper is dedicated to the improvement of the efficiency of the leapfrog method for linear and semilinear second-order differential equations. In numerous situations the strict CFL condition of the leapfrog method is the main bottleneck that thwarts its performance. Based on Chebyshev polynomials new methods have been constructed for linear problems that exhibit a much weaker CFL condition than the leapfrog method (at a higher cost). However, these methods fail to produce the correct long-time behavior of the exact solution which can result in a bad approximation quality. In this paper we introduce a new class of leapfrog-Chebyshev methods for semilinear problems. For the linear part, we use Chebyshev polynomials while the nonlinearity is treated by the standard leapfrog method. The method can be viewed as a multirate scheme because the nonlinearity is evaluated only once in each time step whereas the number of evaluations of the linear part corresponds to the degree of the Chebyshev polynomial. In contrast to existing literature (which is restricted to linear problems), we suggest to stabilize the scheme and we introduce a new starting value required for the two-step method. A new representation formula for the approximations obtained by using generating functions allows us to fully understand the stability and the long-time behavior of the stabilized and the unstabilized scheme. In particular, for linear problems we prove that these new schemes approximately preserve a discrete energy norm over arbitrarily long times. The stability analysis shows that stabilization is essential to guarantee a favorable CFL condition for the multirate scheme, which is closely related to local time-stepping schemes. We also show error bounds of order two for semilinear problems and that a special choice of the stabilization yields order four for linear problems. Finally, we discuss the efficient implementation of the new schemes and give generalizations to fully nonlinear equations.

Key words. time integration, leapfrog method, Chebyshev polynomials, stability analysis, error analysis, second-order ODE, Hamiltonian systems, wave equations, CFL condition, generating functions

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

DOI. 10.1137/18M1209453

1. Introduction.

In this paper we are concerned with the second-order differential equation in $\mathbb{R}^d$,

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

with a symmetric and positive definite matrix $\mathbf{L} \in \mathbb{R}^{d\times d}$ (w.r.t. a given inner product) with large norm and a "nice" function g. Such equations are used to model a plurality of phenomena. Among others, Hamiltonian problems and (spatially discretized) wave-type problems are described by (1.1).

The most natural approach to discretize (1.1) is to replace the second-order time derivative by a centered second-order difference quotient—the well-known leapfrog (LF) scheme. Thanks to a variety of nice features such as symplecticity, symmetry [10, 11], and an easy implementation, the LF scheme serves as the standard time integrator for problems of the type (1.1).

*Received by the editors August 23, 2018; accepted for publication (in revised form) May 18, 2020; published electronically August 20, 2020.
https://doi.org/10.1137/18M1209453
Funding: The work of the authors was supported by the Deutsche Forschungsgemeinschaft (DFG) grant 258734477 - SFB 1173.
†Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany (constantin.carle@kit.edu, marlis.hochbruck@kit.edu, andreas.sturm@kit.edu).

However, its efficiency can be severely limited by the step-size restriction (CFL condition) arising from the large norm of $\mathbf{L}$. This forces a large number of evaluations of the nonlinear function g. In many situations such an evaluation is costly which renders the LF method prohibitively expensive.

The same issue arises for first-order parabolic problems and explicit Runge–Kutta (RK) methods. In this setting Runge-Kutta–Chebyshev (RKC) methods [13, 19, 20, 21] have been found to be a remedy. First-order RKC methods are constructed by using a scaled and shifted Chebyshev polynomial as the stability function. This choice maximizes the stability region and thus alleviates the CFL condition compared to standard RK methods. Based on this idea higher-order methods and further extensions, as, e.g., the ROCK family [1, 2], have been proposed.

In [7, 15] the authors applied analogous ideas to the linear problem (i.e., $g = 0$ in (1.1)) and the LF method. Unfortunately, these schemes fail to reproduce the long-time behavior of the exact solution for certain step sizes, which can result in a poor approximation quality. This is proved rigorously in our analysis and confirmed in numerical examples below.

To improve the methods of [7, 15] for linear problems such that they generate approximations with the correct long-time behavior and a good approximation grade, we propose the following two remedies for the aforementioned problem:

- We replace the standard starting value required for the two-step method, which is based on a Taylor expansion of the exact solution, by one involving the Chebyshev polynomial and its derivative.
- Motivated by stabilized RKC methods [13, 19, 20, 21] we construct a stabilized version of the Chebyshev polynomial.

We show for both of these modifications that the new schemes nearly conserve a discrete energy and lead to uniformly bounded approximations over arbitrarily long times. The analysis is based on the generating functions technique. It provides a characterization of the polynomials to ensure stability and a correct long-time behavior which can be checked easily. Thus, the error analysis is not restricted to methods of order two and four as presented here but it can also be generalized to higher-order methods, e.g., the methods proposed in [7, 15], without further difficulties.

Having these methods at hand, they can be combined, e.g., with the LF scheme for g to integrate the semilinear problem (1.1). The resulting class of methods will be called Leapfrog-Chebyshev (LFC) schemes. As we will show in the course of this paper, this multirate method can be employed with (approximately) p times larger step sizes than the LF method. This renders the method considerably more efficient than the LF scheme since it requires p times fewer evaluations of the nonlinearity g. An interesting special case is splitting the right-hand side into a stiff and a nonstiff part, where important variants are the local time-stepping schemes [6, 9].

Surprisingly, even for linear problems, where the p times larger CFL constant is compromised by a p times higher cost per time step (in terms of matrix-vector multiplications with $\mathbf{L}$), the stabilized methods together with the new starting value outperform the LF in terms of efficiency because of a smaller error constant.

Our paper is organized as follows: In section 2 we present a general two-step time integration method for (1.1) which comprises among others the LF and the LFC scheme. Section 3 deals with the stability and long-time behavior of this general class of methods. A new representation of the numerical solution is a key result for our analysis. It allows us to derive conditions which guarantee the stability of the scheme both in the standard and in the energy norm. It also provides characteristic properties of the schemes (i.e., the polynomials) such that they nearly conserve a discrete energy

for $g = 0$. Moreover, we construct a new special starting value which significantly improves the geometric properties of the scheme. In section 4 we present the error analysis. We show that the general scheme is of order two and can be adapted to converge with fourth order in the linear case. Then, in section 5 we prove explicit formulas for all relevant constants arising in the stability and error analysis of LFC methods. Subsequently, we discuss in section 6 the efficiency and the implementation of the LFC method and also generalize it to fully nonlinear problems. We conclude our paper in section 7 with numerical examples. In particular, we show that our new schemes with stabilization and the new starting value overcome the problems of the LFC methods proposed in [7, 15].

2. A general class of two-step schemes. The LF scheme for the semilinear problem (1.1) is given by

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

$$\mathbf{q}_1 = \big(\mathbf{I} - \tfrac{1}{2}\tau^2 \mathbf{L}\big)\mathbf{q}_0 + \tau \dot{\mathbf{q}}_0 - \tfrac{1}{2}\tau^2 \mathbf{g}_0, \tag{2.1b}$$

where $\tau > 0$ is the time step size and $\mathbf{g}_n = g(\mathbf{q}_n)$. $\mathbf{q}_n$ approximates the exact solution $\mathbf{q}(t_n)$ at time $t_n = n\tau$.

Our aim is to modify the "linear part" of the LF method such that the resulting scheme remains stable for larger time step sizes than the standard LF method (2.1). For this purpose we use a polynomial P of degree $p \geq 1$ satisfying

$$P(0) = 0, \qquad P'(0) = 1. \tag{2.2}$$

These two conditions are required for second-order consistency. We then propose the scheme

$$\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1} = -P(\tau^2 \mathbf{L})\mathbf{q}_n - \tau^2 \mathbf{g}_n, \qquad n = 1, 2, \dots, \tag{2.3a}$$

$$\mathbf{q}_1 = \big(\mathbf{I} - \tfrac{1}{2}P(\tau^2 \mathbf{L})\big)\mathbf{q}_0 + \tau P'(\tau^2 \mathbf{L})\dot{\mathbf{q}}_0 - \tfrac{1}{2}\tau^2 \mathbf{g}_0. \tag{2.3b}$$

Each time step requires p multiplications with $\mathbf{L}$ and only one evaluation of g. Hence, the scheme can be viewed as a *multirate method*, where the stiff linear part is integrated by a p times smaller time step size than the nonstiff nonlinear part. Note also that $\|\mathbf{q}_1 - \mathbf{q}(\tau)\| \leq C\tau^3$ by the consistency conditions (2.2).

Remark 2.1. The scheme (2.3a), (2.3b) can be interpreted as a particular implementation (in the sense of a particular approximation of the matrix functions) of Gautschi-type methods [11, 12] given by

$$\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1} = -2\big(1 - \cos(\tau\mathbf{\Lambda})\big)\mathbf{q}_n - \tau^2 \psi(\tau\mathbf{\Lambda}) g\big(\phi(\tau\mathbf{\Lambda})\mathbf{q}_n\big),$$

where $\mathbf{\Lambda} = \mathbf{L}^{1/2}$. In fact, we have $P(\tau^2\mathbf{L}) \approx 2(1 - \cos(\tau\mathbf{\Lambda}))$ due to (2.2). Moreover, we can set $\mathbf{g}_n = P_\psi(\tau^2\mathbf{L})g(P_\phi(\tau^2\mathbf{L})\mathbf{q}_n)$ for polynomials P_ψ and P_ϕ (again based on Chebyshev polynomials) approximating the even (trigonometric) filter functions ψ and ϕ. In contrast to Krylov subspace methods for the approximation of these matrix functions, this implementation uses *fixed* polynomials and it permits an implementation based on three-term recurrences with known coefficients. However, we would like to stress that the scheme (2.3a) is fully explicit (and thus has a bounded stability region) while Gautschi-type methods with exact evaluation of the matrix functions are unconditionally stable *and* even provide the exact solution for the special case that g is constant and a suitably chosen function ψ.

In this paper we examine the general scheme (2.3a), (2.3b) with particular attention to the choice

$$P(z) = P_p(z) = 2 - \frac{2}{T_p(\nu_p)} T_p\Big(\nu_p - \frac{z}{\alpha_p}\Big), \qquad \alpha_p = 2\frac{T_p'(\nu_p)}{T_p(\nu_p)}. \tag{2.3c}$$

Here, T_p denotes the pth Chebyshev polynomial of the first kind, so that P_p is a polynomial of degree $p \geq 1$, and $\nu_p \geq 1$ is a stabilization parameter whose choice will be discussed later. Obviously, P_p satisfies (2.2) for arbitrary p and $\nu_p \geq 1$. We note that (2.3c) is motivated by the construction of (stabilized) RKC methods [13, 19, 20, 21] and accordingly we name methods from the class (2.3) LFC schemes.

For $\nu_p = 1$ (unstabilized case) and $g = 0$, the method (2.3a), (2.3c) without specifying $\mathbf{q}_1$ has been constructed in [7, 15]. There, it was shown that the unstabilized polynomials (2.3c) yield a method which is optimal in the following sense: Among all polynomials of a fixed degree satisfying the consistency conditions (2.2), the method (2.3a) is stable for $\tau \leq \tau_{\mathrm{CFL}}$ with the maximal τ_{CFL}; see also [14, Thm. 5.1]. More precisely, the largest time step size yielding a stable scheme (2.3a), (2.3c) is p times larger than that for the LF scheme.

Unfortunately, as we will show in the next section, these methods do not show the correct long-time behavior for certain time step sizes. Additionally, the stability of the scheme for linear problems, where $g(\mathbf{q}) = \mathbf{G}\mathbf{q}$ with $\|\mathbf{G}\| \ll \|\mathbf{L}\|$, is only guaranteed under a CFL condition, which is only slightly weaker than that of the LF scheme. In particular, this CFL condition does not improve by increasing the polynomial degree p.

Our remedy to these problems consists of using the starting value (2.3b) which also involves P and its derivative P'. A second crucial modification is the use of a stabilization parameter $\nu_p > 1$. In the next section, we show that these modifications lead to stable schemes with the desired long-time behavior.

Remark 2.2. For $p = 1$ the general scheme (2.3a), (2.3b) reduces to the standard LF method, since $P(z) = z$ is uniquely defined by (2.2). Moreover, the definition (2.3c) is independent of ν_1.

Furthermore, the general scheme also comprises (for $g = 0$) the modified equation LF (modified LF) method [18] with

$$P_{\mathrm{modLF}}(z) = z - \tfrac{1}{12} z^2. \tag{2.4}$$

We conclude this section by stating the general scheme (2.3a), (2.3b) in an equivalent one-step formulation and giving some geometric properties.

LEMMA 2.3. *The scheme* (2.3a), (2.3b) *with* P *satisfying* (2.2) *can be written in the equivalent form*

$$\mathbf{v}_{n+1/2} = \mathbf{v}_n - \frac{\tau}{2}\widehat{P}(\tau^2\mathbf{L})\mathbf{L}\mathbf{q}_n - \frac{\tau}{2}\mathbf{g}_n, \qquad \widehat{P}(z) = \frac{P(z)}{z}, \tag{2.5a}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \tau\mathbf{v}_{n+1/2}, \tag{2.5b}$$

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} - \frac{\tau}{2}\widehat{P}(\tau^2\mathbf{L})\mathbf{L}\mathbf{q}_{n+1} - \frac{\tau}{2}\mathbf{g}_{n+1}, \tag{2.5c}$$

$n = 0, 1, 2, \ldots$, *with starting value*

$$\mathbf{v}_0 = P'(\tau^2\mathbf{L})\dot{\mathbf{q}}_0. \tag{2.5d}$$

Proof. The proof is done by induction on n. ∎

In this one-step formulation, $\mathbf{v}_n$ can be interpreted as an approximation of $\dot{\mathbf{q}}(t_n)$.

COROLLARY 2.4. *The scheme* (2.3a) *and thus also the equivalent one-step version* (2.5a)–(2.5c) *is symmetric and symplectic.*

Proof. The scheme (2.3a) is equivalent to the LF scheme (2.1a) applied to the modified equation

$$\ddot{\mathbf{q}} = -\widehat{P}(\tau^2\mathbf{L})\mathbf{L}\mathbf{q} - g(\mathbf{q}).$$

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

3. Stability and long-time behavior. In this section, we first derive a representation formula for the numerical solution which allows us to characterize properties of P to ensure stability and favorable long-time behavior of the scheme (2.3a), (2.3b). Our various stability results are summarized in section 3.5.

We start by considering the exact solution of (1.1). By the variation-of-constants formula the solution is given by

$$\mathbf{q}(t) = \cos(t\mathbf{L}^{1/2})\mathbf{q}_0 + \mathbf{L}^{-1/2}\sin(t\mathbf{L}^{1/2})\dot{\mathbf{q}}_0 - \int_0^t \mathbf{L}^{-1/2}\sin\big((t-s)\mathbf{L}^{1/2}\big)g\big(\mathbf{q}(s)\big)ds.$$

Recall that $\mathbf{L} \in \mathbb{R}^{d\times d}$ is a symmetric, positive definite matrix w.r.t. a given inner product $(\cdot,\cdot)$, i.e., $\mathbf{L}$ satisfies

$$\big(\mathbf{L}\mathbf{q},\mathbf{p}\big) = \big(\mathbf{q},\mathbf{L}\mathbf{p}\big), \qquad \big(\mathbf{L}\mathbf{q},\mathbf{q}\big) > 0, \qquad \text{for all } \mathbf{q},\mathbf{p} \in \mathbb{R}^d.$$

Further, there exists a constant $c_{\mathrm{inv}} > 0$ such that

$$\|\mathbf{L}^{-1/2}\| \le c_{\mathrm{inv}}. \tag{3.1}$$

For $g = 0$, the solution to the linear problem

$$\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0, \tag{3.2}$$

satisfies

$$\|\mathbf{q}(t)\| \le \|\mathbf{q}_0\| + \min\{t, c_{\mathrm{inv}}\}\|\dot{\mathbf{q}}_0\| \qquad \text{and} \qquad |||\mathbf{q}(t)||| = |||\mathbf{q}(0)||| \tag{3.3}$$

for all $t \ge 0$. Here, we denoted the *standard norm* by $\|\cdot\|^2 = (\cdot,\cdot)$ and the *energy norm* by

$$|||\mathbf{q}(t)|||^2 = \|\dot{\mathbf{q}}(t)\|^2 + \|\mathbf{q}(t)\|_{\mathbf{L}}^2, \qquad \|\mathbf{q}\|_{\mathbf{L}}^2 = \big(\mathbf{L}\mathbf{q},\mathbf{q}\big). \tag{3.4}$$

Remark 3.1. If $\mathbf{L}$ is only positive semidefinite, one can formally set $c_{\mathrm{inv}} = \infty$. In (3.3) we no longer get a uniform bound since $\min\{t, c_{\mathrm{inv}}\} = t$. Moreover, $\|\cdot\|_{\mathbf{L}}$ and thus also $|||\cdot|||$ are only seminorms. All our results remain valid for this situation, but in this paper we are mainly interested in uniform bounds, which require $\mathbf{L}$ to be positive definite.

For a numerical scheme it is desirable to exhibit similar properties. In particular, we will show that our scheme approximately preserves the *discrete energy norm* defined as

$$|||\mathbf{q}_{n+\frac12}|||_\tau^2 = \left\|\frac{[\mathbf{q}_{n+\frac12}]}{\tau}\right\|^2 + \big\|\{\mathbf{q}_{n+\frac12}\}\big\|_{\mathbf{L}}^2 \approx |||\mathbf{q}(t_{n+\frac12})|||^2 \tag{3.5}$$

with

$$[\mathbf{q}_{n+\frac12}] = \mathbf{q}_{n+1} - \mathbf{q}_n, \qquad \{\mathbf{q}_{n+\frac12}\} = \tfrac12(\mathbf{q}_{n+1} + \mathbf{q}_n). \tag{3.6}$$

3.1. Representation of the numerical solution. For the representation of the numerical solution we apply the generating functions technique.

DEFINITION 3.2. *For a polynomial P satisfying* (2.2) *we define $\beta > 0$ as the maximal value such that*

$$0 \le P(z) \le 4 \qquad \text{for all } z \in \big[0, \beta^2\big] \tag{3.7}$$

and $\tau_{\mathrm{CFL}} > 0$ via

$$\tau_{\mathrm{CFL}}^2 = \frac{\beta^2}{\|\mathbf{L}\|}. \tag{3.8}$$

It was already shown in [4, 7, 15] that (3.7) is necessary to ensure stability of the scheme (2.3a) for $\tau \le \tau_{\mathrm{CFL}}$.

THEOREM 3.3. *Let $\tau \le \tau_{\mathrm{CFL}}$. Then, the approximations of the scheme* (2.3a) *are given by*

$$\mathbf{q}_n = \cos\big(n\mathbf{\Phi}\big)\mathbf{q}_0 + \frac{\sin\big(n\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\Big(\mathbf{q}_1 - \cos\mathbf{\Phi}\,\mathbf{q}_0\Big) - \tau^2 \sum_{\ell=1}^{n-1} \frac{\sin\big((n-\ell)\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\mathbf{g}_\ell \tag{3.9a}$$

for $n = 0, 1, 2, \ldots$, where $\mathbf{\Phi}$ with spectrum in $[0, \pi]$ is uniquely defined by

$$\cos\mathbf{\Phi} = \mathbf{I} - \tfrac{1}{2}\mathbf{P} \qquad \text{and} \qquad \sin\mathbf{\Phi} = \big(\mathbf{P}(\mathbf{I} - \tfrac{1}{4}\mathbf{P})\big)^{1/2}, \qquad \mathbf{P} = P(\tau^2\mathbf{L}). \tag{3.9b}$$

Proof. Following the generating functions technique we define the formal power series

$$\mathbf{q}(\zeta) = \sum_{n=0}^{\infty} \mathbf{q}_n \zeta^n, \qquad \mathbf{g}(\zeta) = \sum_{n=0}^{\infty} \mathbf{g}_n \zeta^n.$$

Multiplying the recursion (2.3a) by ζ^{n+1} and summing over $n \ge 1$ we obtain

$$\boldsymbol{\varrho}(\zeta)\mathbf{q}(\zeta) = \mathbf{q}_0 + \zeta\mathbf{q}_1 - \zeta\big(2\mathbf{I} - \mathbf{P}\big)\mathbf{q}_0 - \tau^2\zeta\big(\mathbf{g}(\zeta) - \mathbf{g}_0\big), \tag{3.10a}$$

$$\boldsymbol{\varrho}(\zeta) = \zeta^2\mathbf{I} - \zeta(2\mathbf{I} - \mathbf{P}) + \mathbf{I}. \tag{3.10b}$$

The matrix-valued roots $\boldsymbol{\zeta}_\pm$ of $\boldsymbol{\varrho}$ are given by

$$\boldsymbol{\zeta}_\pm = \mathbf{I} - \tfrac{1}{2}\mathbf{P} \pm \mathrm{i}\big(\mathbf{P}(\mathbf{I} - \tfrac{1}{4}\mathbf{P})\big)^{1/2} = \cos\mathbf{\Phi} \pm \mathrm{i}\sin\mathbf{\Phi},$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. The second identity holds because by (3.7) we have $\|\boldsymbol{\zeta}_\pm\| = 1$ so that we can write $\boldsymbol{\zeta}_\pm = \mathrm{e}^{\pm\mathrm{i}\mathbf{\Phi}}$ with a matrix $\mathbf{\Phi}$ whose spectrum is contained in $[0, \pi]$. Clearly, this yields $\boldsymbol{\zeta}_+ = \boldsymbol{\zeta}_-^{-1}$ and thus

$$\boldsymbol{\varrho}(\zeta) = (\zeta\mathbf{I} - \boldsymbol{\zeta}_+)(\zeta\mathbf{I} - \boldsymbol{\zeta}_-) = (\mathbf{I} - \zeta\boldsymbol{\zeta}_-)(\mathbf{I} - \zeta\boldsymbol{\zeta}_+) = (\mathbf{I} - \zeta\mathrm{e}^{-\mathrm{i}\mathbf{\Phi}})(\mathbf{I} - \zeta\mathrm{e}^{\mathrm{i}\mathbf{\Phi}}).$$

Employing the Neumann series and the Cauchy product we have for $|\zeta| < 1$

$$\boldsymbol{\varrho}(\zeta)^{-1} = \sum_{n=0}^{\infty} \mathrm{e}^{-\mathrm{i}n\mathbf{\Phi}}\zeta^n \sum_{\ell=0}^{n} \mathrm{e}^{2\mathrm{i}\ell\mathbf{\Phi}} = \sum_{n=0}^{\infty} \frac{\sin\big((n+1)\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\zeta^n.$$

Here, the second equality follows with the geometric sum identity. Using this in (3.10) we deduce by comparing the coefficients of ζ^n

$$\mathbf{q}_n = \frac{\sin\big((n+1)\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\mathbf{q}_0 + \frac{\sin\big(n\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\Big(\mathbf{q}_1 - 2\cos\mathbf{\Phi}\,\mathbf{q}_0\Big) - \tau^2\sum_{\ell=1}^{n-1}\frac{\sin\big((n-\ell)\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\mathbf{g}_\ell. \tag{3.11}$$

A trigonometric identity completes the proof. ∎

The representation (3.9) motivates us to choose the starting value (2.3b). In particular, using $P'(\tau^2\mathbf{L})\dot{\mathbf{q}}_0$ instead of $\dot{\mathbf{q}}_0$ in (2.3b) is based on the observation that $\sin\mathbf{\Phi} = (\mathbf{P}(\mathbf{I} - \frac{1}{4}\mathbf{P}))^{1/2}$ becomes singular if $\mathbf{P}$ has eigenvalues 0 or 4. However, $P(z) \in \{0, 4\}$ for some $z \in (0, \beta^2)$ means that z is a stationary point of P, i.e., $P'(z) = 0$. Our choice of the starting value thus removes all singularities in the interior of the interval. Later, in Theorem 3.11 we will see how this choice affects the long-time behavior for linear problems.

Another reason for choosing P' in (2.3b) is given in section 5.2, where we show that for $g = 0$ one step of the scheme (2.3) with $\nu_p = 1$ is equivalent to p steps of the standard LF scheme (2.1) with step size τ/p. However, this equivalence does not hold for any other starting value $\mathbf{q}_1$.

To study the effect of $\mathbf{q}_1$ further note that by (2.2) an arbitrary starting value satisfying $\|\mathbf{q}_1 - \mathbf{q}(\tau)\| \le C\tau^3$ can be written in the form

$$\mathbf{q}_1 = (\mathbf{I} - \tfrac{1}{2}\mathbf{P})\,\mathbf{q}_0 + \tau\,\mathbf{P}'\dot{\mathbf{q}}_0 - \tfrac{1}{2}\tau^2\mathbf{g}_0 + \tau^3\boldsymbol{\delta}_0, \tag{3.12}$$

for a bounded perturbation $\boldsymbol{\delta}_0$. We have $\boldsymbol{\delta}_0 = 0$ for $\mathbf{q}_1$ defined in (2.3b) and

$$\tau^3\boldsymbol{\delta}_0 = \tfrac{1}{2}(\mathbf{P} - \tau^2\mathbf{L})\mathbf{q}_0 + \tau(\mathbf{I} - \mathbf{P}')\dot{\mathbf{q}}_0$$

for the starting value (2.1b) derived from Taylor expansion.

Corollary 3.4. *Let $\tau \le \tau_{\mathrm{CFL}}$. For the scheme* (2.3a) *with general starting value* (3.12) *we have*

$$\mathbf{q}_n = \cos\big(n\mathbf{\Phi}\big)\mathbf{q}_0 + \tau\frac{\sin\big(n\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\big(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\big) - \tau^2\sum_{\ell=0}^{n-1}\frac{\sin\big((n-\ell)\mathbf{\Phi}\big)}{\sin\mathbf{\Phi}}\chi_\ell\mathbf{g}_\ell, \tag{3.13}$$

where $\chi_0 = 1/2$ and $\chi_\ell = 1$, $\ell \ge 1$.

Proof. The proof is a direct consequence of (3.9). ∎

For the quantities arising in the discrete energy norm (3.5) we also need a representation of differences and means.

Lemma 3.5. *Let $\tau \le \tau_{\mathrm{CFL}}$ and $g = 0$. Then the scheme* (2.3a) *with general starting value* (3.12) *satisfies*

$$\frac{1}{\tau}[\mathbf{q}_{n+\frac{1}{2}}] = -\frac{2}{\tau}\sin\big((n+\tfrac{1}{2})\mathbf{\Phi}\big)\sin\big(\tfrac{1}{2}\mathbf{\Phi}\big)\mathbf{q}_0 + \frac{\cos\big((n+\frac{1}{2})\mathbf{\Phi}\big)}{\cos\big(\frac{1}{2}\mathbf{\Phi}\big)}\big(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\big), \tag{3.14a}$$

$$\{\mathbf{q}_{n+\frac{1}{2}}\} = \cos\big((n+\tfrac{1}{2})\mathbf{\Phi}\big)\cos\big(\tfrac{1}{2}\mathbf{\Phi}\big)\mathbf{q}_0 + \frac{\tau}{2}\frac{\sin\big((n+\frac{1}{2})\mathbf{\Phi}\big)}{\sin\big(\frac{1}{2}\mathbf{\Phi}\big)}\big(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\big). \tag{3.14b}$$

Proof. The proof follows directly from (3.13) and trigonometric identities. ∎

We are now in a position to study the stability of the recursion (2.3a).

3.2. Stability for semilinear problems. In the following sections we assume $g\colon \mathbb{R}^d \to \mathbb{R}^d$ to be a Lipschitz continuous function, i.e.,

$$\|g(\mathbf{q}) - g(\mathbf{p})\| \le L_g \|\mathbf{q} - \mathbf{p}\| \qquad \text{for all } \mathbf{q}, \mathbf{p} \in \mathbb{R}^d. \tag{3.15}$$

Remark 3.6. It is sufficient to require that g is locally Lipschitz continuous in a strip around the exact solution, i.e., in a neighborhood of $\{\mathbf{q}(t) : 0 \le t \le T\}$. This can be seen from the error bounds in Theorem 4.3 below. For the sake of presentation we omit the details.

We first present a general stability result for the scheme (2.3a).

THEOREM 3.7. *Let $\tau \le \tau_{\mathrm{CFL}}$ and denote by $\mathbf{q}_n$ and $\mathbf{p}_n$ the approximations obtained by* (2.3a) *with initial values $\mathbf{q}_0, \dot{\mathbf{q}}_0$ and $\mathbf{p}_0, \dot{\mathbf{p}}_0$, and starting values $\mathbf{q}_1, \mathbf{p}_1$ given by* (3.12) *with $\boldsymbol{\delta}_0, \widehat{\boldsymbol{\delta}}_0$, respectively.*

For $t_n \le T$ and $P'_{\max} = \max_{z\in[0,\beta^2]} |P'(z)|$, we have

$$\|\mathbf{q}_n - \mathbf{p}_n\| \le \Big(\|\mathbf{q}_0 - \mathbf{p}_0\| + T\left(P'_{\max} \|\dot{\mathbf{q}}_0 - \dot{\mathbf{p}}_0\| + \tau^2 \|\boldsymbol{\delta}_0 - \widehat{\boldsymbol{\delta}}_0\| \right) \Big) \mathrm{e}^{L_g^{1/2} T}. \tag{3.16}$$

For the LFC scheme (2.3) we have $P'_{\max} = 1$; see Theorem 5.1.

Proof. From (3.13) we get

$$\begin{aligned} \|\mathbf{q}_n - \mathbf{p}_n\| \le \|\mathbf{q}_0 - \mathbf{p}_0\| + T\big(P'_{\max} \|\dot{\mathbf{q}}_0 - \dot{\mathbf{p}}_0\| + \tau^2 \|\boldsymbol{\delta}_0 - \widehat{\boldsymbol{\delta}}_0\| \big) \Big) \\ + \tau^2 L_g \sum_{\ell=0}^{n-1} (n-\ell) \|\mathbf{q}_\ell - \mathbf{p}_\ell\|, \end{aligned}$$

where we used that $|\sin(n\zeta)/\sin\zeta| \le n$ for $\zeta \in \mathbb{R}$ and the Lipschitz condition (3.15). Application of Lemma 3.8 finishes the proof. ∎

The proof of the previous theorem makes use of the following Gronwall-type lemma.

LEMMA 3.8. *Let $\kappa, \gamma \ge 0$. If the nonnegative sequence $\{\varepsilon_n\}_{n\ge 0}$ satisfies*

$$\varepsilon_n \le \kappa + (\gamma\tau)^2 \sum_{\ell=0}^{n-1} (n-\ell)\varepsilon_\ell, \tag{3.17}$$

then

$$\varepsilon_n \le \kappa\, \mathrm{e}^{\gamma T} \qquad \text{for} \qquad n\tau = t_n \le T. \tag{3.18}$$

Proof. Let

$$\rho_n = \kappa + (\gamma\tau)^2 \sum_{\ell=0}^{n-1} (n-\ell)\rho_\ell, \qquad n \ge 0. \tag{3.19}$$

Obviously, we have $\varepsilon_n \le \rho_n$ for all $n \ge 0$ and ρ_n satisfies the linear recurrence relation

$$\rho_{n+1} - 2\rho_n + \rho_{n-1} = (\gamma\tau)^2 \rho_n, \qquad n \ge 1.$$

Solving this recursion yields for $n \ge 0$

$$\rho_n = c_1 \eta_+^n + c_2 \eta_-^n, \qquad \eta_\pm = f_\pm(\gamma\tau), \qquad f_\pm(x) = 1 + \tfrac{1}{2}x^2 \pm x\big(1 + \tfrac{1}{4}x^2\big)^{1/2},$$

where $c_1, c_2 \in \mathbb{R}$ are given by

$$c_1 + c_2 = \kappa \qquad \text{and} \qquad c_1\eta_+ + c_2\eta_- = \kappa\big(1 + (\gamma\tau)^2\big).$$

It is easily verified that c_1, c_2 are both nonnegative. To bound $\eta_\pm^n$, we have, by using $(1+y)^{1/2} \le 1 + y/2$ for $y \ge 0$,

$$0 < f_+(x) \le 1 + \tfrac{1}{2}x^2 + x\big(1 + \tfrac{1}{8}x^2\big) \le \mathrm{e}^x, \qquad x \ge 0.$$

Further, we obtain for $x \ge 0$

$$f_-(x) \le 1 \qquad \text{and} \qquad f_-(x) = \frac{f_-(x)f_+(x)}{f_+(x)} = \frac{1}{f_+(x)} > 0.$$

Thus, we have $\eta_+ \le \mathrm{e}^{\gamma\tau}$ and $0 < \eta_- \le 1$ which implies $\rho_n \le c_1\mathrm{e}^{\gamma t_n} + c_2 \le \kappa\mathrm{e}^{\gamma t_n}$. ∎

3.3. Stability and long-time behavior for linear problems. We next investigate conditions for which the scheme (2.3a) yields uniformly bounded approximations for the linear problem (3.2) in the standard norm $\|\cdot\|$ as well as in the discrete energy norm $|||\cdot|||_\tau$. It will be shown that we have to choose $\mathbf{q}_1$ as in (2.3b), i.e., with $\boldsymbol{\delta}_0 = 0$ in (3.12), *or* to employ a stronger CFL condition than $\tau \le \tau_{\mathrm{CFL}}$.

To simplify the presentation we assume for the remaining paper that in addition to (2.2), the polynomial P satisfies

$$0 \le P(z) \le z \qquad \text{for all } z \in \big[0, \beta^2\big] \tag{3.20}$$

with $\beta > 0$ given in Definition 3.2. This is a natural assumption if we aim at schemes with a larger stability region than the LF scheme. In particular, (3.20) is fulfilled for the polynomials in (2.3c) for all $p \in \mathbb{N}$ and $\nu_p \ge 1$; cf. Theorem 5.1.

The stronger CFL condition is defined as follows.

DEFINITION 3.9. *For given $m_1, m_2 \in (0,1)$ we define $\widehat{\beta} = \widehat{\beta}(m_1, m_2) > 0$ as the maximal value such that*

$$m_1 \le 1 - \tfrac{1}{4}P(z) \le 1, \qquad m_2 z \le P(z) \le z, \qquad z \in [0, \widehat{\beta}^2], \tag{3.21}$$

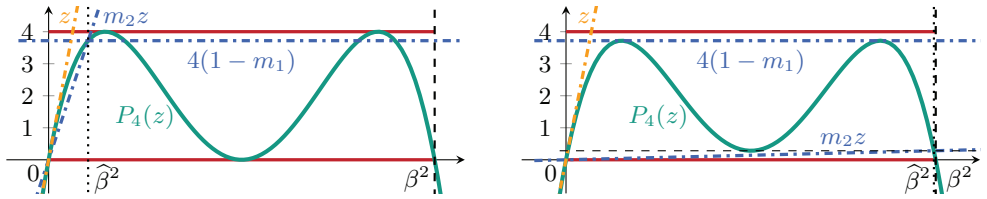
and $\widehat{\tau}_{\mathrm{CFL}} > 0$ via

$$\widehat{\tau}_{\mathrm{CFL}}^2 = \frac{\widehat{\beta}^2}{\|\mathbf{L}\|}. \tag{3.22}$$

Obviously, we have $\widehat{\beta} < \beta$ and thus $\widehat{\tau}_{\mathrm{CFL}} < \tau_{\mathrm{CFL}}$. More precisely, the ratio between $\widehat{\beta}$ and β strongly depends on the polynomial P. We are interested in choosing P and the constants m_1, m_2 in such a way that $\widehat{\beta} \approx \beta$. For the LFC polynomials (2.3c) we show in section 5.1 that one can only achieve $\widehat{\beta} \approx \beta$ if $\nu_p > 1$, while for $\nu_p = 1$ we have $\widehat{\beta} \sim \beta/p$ and thus $\widehat{\tau}_{\mathrm{CFL}} \sim \tau_{\mathrm{CFL}}/p$; cf. Figure 3.1.

Example 3.10. The LF method satisfies the second condition in (3.21) for every $m_2 \le 1$. For the first condition in (3.21) let $m_1 = 1 - \vartheta^2$ for some $\vartheta \in (0,1)$. Then, we obtain the CFL condition (3.22) with $\widehat{\beta}^2 = \widehat{\beta}_{\mathrm{LF}}^2 = 4\vartheta^2 < 4 = \beta^2$.

THEOREM 3.11. *Let $g = 0$ and $n \ge 2$. Then, the approximations obtained by* (2.3a) *with general starting value* (3.12) *and either*



(a) $\nu_p = 1$. (b) $\nu_p = 1.01$.

FIG. 3.1. *Illustration of Definitions* 3.2 *and* 3.9 *for LFC polynomials* P_4 *from* (2.3c).

(a) $\tau \le \vartheta\tau_{\mathrm{CFL}}$ *for arbitrary* $\vartheta \in (0,1)$ *if* $\boldsymbol{\delta}_0 = 0$, *i.e., if* $\mathbf{q}_1$ *is chosen as in* (2.3b), *or*

(b) $\tau \le \widehat{\tau}_{\mathrm{CFL}}$ *and* $\vartheta = \widehat{\tau}_{\mathrm{CFL}}/\tau_{\mathrm{CFL}}$ *if* $\boldsymbol{\delta}_0 \ne 0$

satisfy

$$\|\mathbf{q}_n\| \le \|\mathbf{q}_0\| + c_0\|\dot{\mathbf{q}}_0\| + \widehat{c}_0\tau^2\|\boldsymbol{\delta}_0\|, \tag{3.23a}$$

$$c_0 = \min\{P'_{\max}t_n, c_{\mathrm{inv}}C_0(\vartheta,P)\}, \quad \widehat{c}_0 = \min\{t_n, c_{\mathrm{inv}}(m_1m_2)^{-1/2}\}, \tag{3.23b}$$

and

$$\|\!|\mathbf{q}_{n+\frac12}|\!\|_\tau \le 2^{1/2}\Big(\|\mathbf{q}_0\|_{\mathbf{L}} + C_1(\vartheta,P)\|\dot{\mathbf{q}}_0\| + \frac{\tau^2}{\min\{m_1,m_2\}^{1/2}}\|\boldsymbol{\delta}_0\|\Big). \tag{3.23c}$$

The constants $C_k(\vartheta,P)$, $k = 0,1$, *only depend on* ϑ *and on* P *but not on* $\mathbf{L}$ *or* τ.

Moreover, if $\tau \le \tau_{\mathrm{CFL}}$, *then* (3.23a) *holds for arbitrary* $\boldsymbol{\delta}_0$ *with* $c_0 = P'_{\max}t_n$ *and* $\widehat{c}_0 = t_n$.

Proof. The bounds involving t_n follow from Theorem 3.7 by setting $\mathbf{p}_0 = \dot{\mathbf{p}}_0 = \widehat{\boldsymbol{\delta}}_0 = 0$.

(i) To prove the remaining bounds we start with (3.23a). From (3.13) we get with (3.9b) and (3.1)

$$\begin{aligned}\|\mathbf{q}_n\| &\le \|\mathbf{q}_0\| + \tau\Big\|\big(\mathbf{P}(\mathbf{I} - \tfrac14\mathbf{P})\big)^{-1/2}\big(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\big)\Big\| \\ &\le \|\mathbf{q}_0\| + c_{\mathrm{inv}}\Big\|\psi_0(\tau^2\mathbf{L})\big(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\big)\Big\|,\end{aligned}$$

where

$$\psi_0(z) = \psi_1(z)\psi_2(z), \qquad \psi_1(z) = \Big(\frac{1}{1-\frac14P(z)}\Big)^{1/2}, \quad \psi_2(z) = \Big(\frac{z}{P(z)}\Big)^{1/2}. \tag{3.24}$$

By Definition 3.9 and (3.20) we have

$$\psi_j^2(z) \le \frac{1}{m_j}, \quad z \in [0,\widehat{\beta}^2], \quad \text{and} \quad \psi_2(z) \ge 1, \quad z \in [0,\beta^2]. \tag{3.25}$$

Next, we prove that there is a constant $C_0 = C_0(\vartheta,P)$ such that

$$|\widetilde{\psi}_0(z)| \le C_0 \quad \text{for } z \in [0,\beta^2], \quad \text{where} \quad \widetilde{\psi}_0(z) = \psi_0(z)P'(z). \tag{3.26}$$

$\widetilde{\psi}_0$ is continuous in the interior of $[0,\beta^2]$ since if $z_* \in (0,\beta^2)$ satisfies $P(z_*) \in \{0,4\}$, then z_* is a stationary point of P and thus $P'(z_*) = 0$. In fact, L'Hôpital's rule applied to $\widetilde{\psi}_0^2$ yields

$$|\widetilde{\psi}_0(z_*)| = |2z_*P''(z_*)|^{1/2}.$$

Moreover, $\widetilde{\psi}_0(0) = 1$ because of $P(0) = 0$ and $P'(0) = 1$. Finally, since β^2 is the maximum length of the interval, we have $P(\beta^2) \in \{0, 4\}$. Hence, $\widetilde{\psi}_0$ is continuous on $[0, \vartheta^2\beta^2]$ for all $\vartheta \in (0, 1)$ and we conclude $\|\widetilde{\psi}_0(\tau^2\mathbf{L})\| \le C_0$ for $\tau \le \vartheta\tau_{\mathrm{CFL}}$. This proves (3.23a) for $\boldsymbol{\delta}_0 = 0$.

For $\boldsymbol{\delta}_0 \neq 0$, (3.23a) follows directly from (3.25).

(ii) For the proof of (3.23c) we use Lemma 3.5 and

$$\sin\bigl(\tfrac{1}{2}\boldsymbol{\Phi}\bigr) = \tfrac{1}{2}\mathbf{P}^{1/2}, \qquad \cos\bigl(\tfrac{1}{2}\boldsymbol{\Phi}\bigr) = \bigl(\mathbf{I} - \tfrac{1}{4}\mathbf{P}\bigr)^{1/2}. \tag{3.27}$$

This yields

$$\begin{aligned} |||\mathbf{q}_{n+\frac{1}{2}}|||_\tau^2 \le &\Bigl(\|\psi_2(\tau^2\mathbf{L})^{-1}\mathbf{q}_0\|_{\mathbf{L}} + \|\psi_1(\tau^2\mathbf{L})\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr)\|\Bigr)^2 \\ &+ \Bigl(\|\mathbf{q}_0\|_{\mathbf{L}} + \|\psi_2(\tau^2\mathbf{L})\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr)\|\Bigr)^2. \end{aligned} \tag{3.28}$$

As in (i), one can show that for all $\vartheta \in (0, 1)$ the functions $\widetilde{\psi}_j = \psi_j P'$, $j = 1, 2$, are continuous on $[0, \vartheta^2\beta^2]$. Hence, we conclude $\|\widetilde{\psi}_j(\tau^2\mathbf{L})\| \le C_1(\vartheta, P)$ for $\tau \le \vartheta\tau_{\mathrm{CFL}}$. The bounds for the remaining terms again follow from (3.25). ∎

An additional positive effect of stabilization manifests when linear problems with inhomogeneities are considered.

COROLLARY 3.12. *Consider the linear, inhomogeneous problem*

$$\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t) - g(t), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0.$$

Then, the approximations obtained by (2.3a) *with* $\mathbf{g}_n = g(t_n)$ *and* $\tau \le \widehat{\tau}_{\mathrm{CFL}}$ *satisfy*

$$\|\mathbf{q}_n\| \le \|\mathbf{q}_0\| + c_0\|\dot{\mathbf{q}}_0\| + \widehat{c}_0\tau^2\|\boldsymbol{\delta}_0\| + \widehat{c}_0\tau\sum_{\ell=0}^{n-1}\|\mathbf{g}_\ell\| \tag{3.29}$$

with c_0, $\widehat{c}_0$ *given in* (3.23b).

If $\tau \le \tau_{\mathrm{CFL}}$, *then* (3.29) *holds with* $c_0 = P'_{\max}t_n$ *and* $\widehat{c}_0 = t_n$.

Proof. This follows directly from Corollary 3.4 and the proof of Theorem 3.11. ∎

Remark 3.13. If there exists $z \in (\widehat{\beta}, \beta)$ such that $P(z) \in \{0, 4\}$, then the factor t_n appearing in $\widehat{c}_0$ multiplying $\|\boldsymbol{\delta}_0\|$ in Theorem 3.11 and $\|\mathbf{g}_\ell\|$ in Corollary 3.12 are sharp. This can be seen from Corollary 3.4 by choosing $\boldsymbol{\delta}_0$ and $\mathbf{g}_\ell$ as eigenvectors of $\mathbf{L}$ corresponding to an eigenvalue λ and the step size τ such that $P(\tau\lambda) \in \{0, 4\}$.

Next, we prove that the scheme (2.3a) nearly preserves the discrete energy norm $|||\cdot|||_\tau$ by showing that it is order two close to a preserved quantity. This also reflects the behavior of the exact solution which is energy conserving; see (3.3).

LEMMA 3.14. *Let* $g = 0$. *The approximations obtained by* (2.3a) *satisfy*

$$\mathcal{M}_{\mathbf{q},n+\frac{1}{2}} \equiv \mathcal{M}_{\mathbf{q},\frac{1}{2}} \qquad \textit{for all } n = 0, 1, \ldots, \tag{3.30a}$$

where

$$\mathcal{M}_{\mathbf{q},n+\frac{1}{2}} = \bigl(\bigl(\mathbf{I} - \tfrac{1}{4}\mathbf{P}\bigr)[\mathbf{q}_{n+\frac{1}{2}}], [\mathbf{q}_{n+\frac{1}{2}}]\bigr) + \bigl(\mathbf{P}\{\mathbf{q}_{n+\frac{1}{2}}\}, \{\mathbf{q}_{n+\frac{1}{2}}\}\bigr). \tag{3.30b}$$

Proof. We write the recursion (2.3a) in the equivalent form

$$\big(\mathbf{I} - \tfrac{1}{4}\mathbf{P}\big)\big(\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1}\big) + \tfrac{1}{4}\mathbf{P}\big(\mathbf{q}_{n+1} + 2\mathbf{q}_n + \mathbf{q}_{n-1}\big) = 0. \tag{3.31}$$

The statement follows by taking the inner product of (3.31) with $\mathbf{q}_{n+1} - \mathbf{q}_{n-1}$. ∎

THEOREM 3.15. *Let the assumptions of Theorem* 3.11 *be satisfied. Then, we have*

$$\Big| \|\!|\mathbf{q}_{n+\frac{1}{2}}|\!\|_\tau - \frac{\mathcal{M}_{\mathbf{q},\frac{1}{2}}}{\tau^2} \Big| \le C\tau^2, \qquad n = 0, 1, 2, \dots,$$

with a constant C which is independent of $\mathbf{L}, \tau$, n, and c_{inv}.

This theorem shows that the approximations of the scheme (2.3a) do not have a drift in the discrete energy for arbitrarily long simulation times.

Proof. By Lemma 3.14 we have $\mathcal{M}_{\mathbf{q},n+\frac{1}{2}} = \mathcal{M}_{\mathbf{q},\frac{1}{2}}$ and thus

$$\tau^2 \|\!|\mathbf{q}_{n+\frac{1}{2}}|\!\|_\tau^2 - \mathcal{M}_{\mathbf{q},n+\frac{1}{2}} = \tfrac{1}{4}\big(\mathbf{P}[\mathbf{q}_{n+\frac{1}{2}}], [\mathbf{q}_{n+\frac{1}{2}}]\big) - \big((\mathbf{P} - \tau^2\mathbf{L})\{\mathbf{q}_{n+\frac{1}{2}}\}, \{\mathbf{q}_{n+\frac{1}{2}}\}\big).$$

Using (3.20) we can bound this by

$$\begin{aligned} 0 \le \|\!|\mathbf{q}_{n+\frac{1}{2}}|\!\|_\tau^2 - \frac{\mathcal{M}_{\mathbf{q},n+\frac{1}{2}}}{\tau^2} &\le \tfrac{1}{4}\tau^2 \left\| \frac{[\mathbf{q}_{n+\frac{1}{2}}]}{\tau} \right\|_{\mathbf{L}}^2 + c\tau^2 \big\|\mathbf{L}\{\mathbf{q}_{n+\frac{1}{2}}\}\big\|^2 \\ &\le \big(\tfrac{1}{4} + c\big)\tau^2 \Big(\|\mathbf{L}\mathbf{q}_0\| + C_1(\vartheta, P)\|\dot{\mathbf{q}}_0\|_{\mathbf{L}} + \tfrac{1}{\min\{m_1, m_2\}^{1/2}} \|\boldsymbol{\delta}_0\|_{\mathbf{L}} \Big)^2, \end{aligned}$$

where $c > 0$ is a constant such that $|P(z) - z| \le cz^2$ for $z \in [0, \beta^2]$. Such a constant exists because of (2.2). The last inequality follows as in the proof of Theorem 3.11. ∎

3.4. Stability of the multirate method for linear problems. We now focus on the stability of the multirate method (2.3a) applied to the linear differential equation

$$\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t) - \mathbf{G}\mathbf{q}(t), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0. \tag{3.32}$$

Here, $\mathbf{G} \in \mathbb{R}^{d\times d}$ is a symmetric and positive semidefinite matrix (w.r.t. the given inner product) with $\|\mathbf{G}\| \ll \|\mathbf{L}\|$. The multirate method is closely related to local time-stepping methods for wave equations [6, 9], where the polynomials (2.3c) are used without stabilization, i.e., with $\nu_p = 1$. Roughly speaking, the matrices $\mathbf{L}$ and $\mathbf{G}$ then correspond to the space discretization of the differential operator on the fine and the coarse part of the mesh, respectively.

THEOREM 3.16. *Let the CFL conditions*

$$\tau^2\|\mathbf{L}\| \le \widehat{\beta}^2, \qquad \tau^2\|\mathbf{G}\| \le 4\vartheta^2, \qquad \vartheta^2 \in (0, m_1), \tag{3.33}$$

be satisfied and let $m_{1,\vartheta} = m_1 - \vartheta^2$. Then, the recursion (2.3a) *with general starting value* (3.12) *applied to* (3.32)*, i.e., $\mathbf{g}_n = \mathbf{G}\mathbf{q}_n$, is stable with bounds*

$$\|\mathbf{q}_n\| \le \|\mathbf{q}_0\| + \min\Big\{t_n, \frac{c_{\mathrm{inv}}}{(m_{1,\vartheta}m_2)^{1/2}}\Big\}\big(P'_{\max}\|\dot{\mathbf{q}}_0\| + \tau^2\|\boldsymbol{\delta}_0\|\big), \tag{3.34a}$$

$$\|\!|\mathbf{q}_{n+\frac{1}{2}}|\!\|_{\tau,\star} \le \big(\tfrac{2}{m_2}\big)^{\frac{1}{2}}\|\mathbf{q}_0\|_{\mathbf{L}+\mathbf{G}} + \Big(\frac{2}{\min\{m_{1,\vartheta}, m_2\}}\Big)^{\frac{1}{2}}\big(P'_{\max}\|\dot{\mathbf{q}}_0\| + \tau^2\|\boldsymbol{\delta}_0\|\big), \tag{3.34b}$$

where $\|\!|\cdot|\!\|_{\tau,\star}$ is defined analogously to (3.5) *with $\|\cdot\|_{\mathbf{L}}$ replaced by $\|\cdot\|_{\mathbf{L}+\mathbf{G}}$.*

Remark 3.17. Note that the CFL conditions (3.33) require the stronger condition (3.21) with $\widehat{\beta} < \beta$. For LFC methods, $\widehat{\beta} \approx \beta$ and a reasonable value of m_1 can only be achieved for sufficiently large stabilization parameters $\nu_p > 1$; cf. Figure 3.1 and section 5.

Proof. The first part of the proof closely follows the proof of Theorem 3.3. Replacing $\mathbf{P}$ in (3.10) by the symmetric, positive semidefinite matrix $\mathbf{P_G} = \mathbf{P} + \tau^2\mathbf{G}$ we get analogously to (3.9a)

$$\mathbf{q}_n = \cos\bigl(n\widetilde{\mathbf{\Phi}}\bigr)\mathbf{q}_0 + \frac{\sin\bigl(n\widetilde{\mathbf{\Phi}}\bigr)}{\sin\widetilde{\mathbf{\Phi}}}\Bigl(\mathbf{q}_1 - \cos\widetilde{\mathbf{\Phi}}\,\mathbf{q}_0\Bigr),$$

where $\widetilde{\mathbf{\Phi}}$ with spectrum in $[0,\pi]$ is defined by

$$\cos\widetilde{\mathbf{\Phi}} = \mathbf{I} - \tfrac{1}{2}\mathbf{P_G} \qquad \text{and} \qquad \sin\widetilde{\mathbf{\Phi}} = \bigl(\mathbf{P_G}(\mathbf{I} - \tfrac{1}{4}\mathbf{P_G})\bigr)^{1/2}. \tag{3.35}$$

Note that $\mathbf{P_G}(\mathbf{I} - \frac{1}{4}\mathbf{P_G})$ is symmetric and positive semidefinite since

$$\Bigl(\mathbf{P_G}(\mathbf{I} - \tfrac{1}{4}\mathbf{P_G})\mathbf{q},\mathbf{q}\Bigr) \geq m_{1,\vartheta}\bigl(\mathbf{P_G}\,\mathbf{q},\mathbf{q}\bigr) \geq m_{1,\vartheta}m_2\tau^2 c_{\mathrm{inv}}^{-2}\|\mathbf{q}\|^2 \geq 0, \tag{3.36}$$

where we used the CFL conditions (3.33), and (3.21) and (3.1). Inserting the general starting value (3.12) yields the representation formula

$$\mathbf{q}_n = \cos\bigl(n\widetilde{\mathbf{\Phi}}\bigr)\mathbf{q}_0 + \tau\frac{\sin\bigl(n\widetilde{\mathbf{\Phi}}\bigr)}{\sin\widetilde{\mathbf{\Phi}}}\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr). \tag{3.37}$$

(i) To prove (3.34a) we employ in (3.37) once more $|\sin(n\zeta)/\sin\zeta| \leq n$ for $\zeta \in \mathbb{R}$. Moreover, from (3.35) and (3.36) we obtain

$$\tau\|(\sin\widetilde{\mathbf{\Phi}}\,)^{-1}\mathbf{q}\| \leq \bigl(m_{1,\vartheta}m_2\bigr)^{-1/2}c_{\mathrm{inv}}\|\mathbf{q}\|.$$

(ii) From (3.37) we get analogously to Lemma 3.5

$$\frac{1}{\tau}[\mathbf{q}_{n+\frac{1}{2}}] = -\frac{2}{\tau}\sin\bigl((n+\tfrac{1}{2})\widetilde{\mathbf{\Phi}}\bigr)\sin\bigl(\tfrac{1}{2}\widetilde{\mathbf{\Phi}}\bigr)\mathbf{q}_0 + \frac{\cos\bigl((n+\frac{1}{2})\widetilde{\mathbf{\Phi}}\bigr)}{\cos\bigl(\frac{1}{2}\widetilde{\mathbf{\Phi}}\bigr)}\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr), \tag{3.38a}$$

$$\{\mathbf{q}_{n+\frac{1}{2}}\} = \cos\bigl((n+\tfrac{1}{2})\widetilde{\mathbf{\Phi}}\bigr)\cos\bigl(\tfrac{1}{2}\widetilde{\mathbf{\Phi}}\bigr)\mathbf{q}_0 + \frac{\tau}{2}\frac{\sin\bigl((n+\frac{1}{2})\widetilde{\mathbf{\Phi}}\bigr)}{\sin\bigl(\frac{1}{2}\widetilde{\mathbf{\Phi}}\bigr)}\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr) \tag{3.38b}$$

with

$$\sin\bigl(\tfrac{1}{2}\widetilde{\mathbf{\Phi}}\bigr) = \tfrac{1}{2}\mathbf{P_G}^{1/2}, \qquad \cos\bigl(\tfrac{1}{2}\widetilde{\mathbf{\Phi}}\bigr) = (\mathbf{I} - \tfrac{1}{4}\mathbf{P_G})^{1/2}.$$

Thus, we obtain for (3.38a) with (3.21) and (3.33)

$$\begin{aligned}\bigl\|\tfrac{1}{\tau}[\mathbf{q}_{n+\frac{1}{2}}]\bigr\| &\leq \|\tfrac{1}{\tau}\mathbf{P_G}^{1/2}\mathbf{q}_0\| + \|(\mathbf{I} - \tfrac{1}{4}\mathbf{P_G})^{-1/2}\bigl(\mathbf{P}'\dot{\mathbf{q}}_0 + \tau^2\boldsymbol{\delta}_0\bigr)\| \\ &\leq \|\mathbf{q}_0\|_{\mathbf{L}+\mathbf{G}} + m_{1,\vartheta}^{-1/2}\bigl(P'_{\max}\|\dot{\mathbf{q}}_0\| + \tau^2\|\boldsymbol{\delta}_0\|\bigr).\end{aligned}$$

To bound (3.38b) in $\|\cdot\|_{\mathbf{L}+\mathbf{G}}$ note that, in general, $\mathbf{L}+\mathbf{G}$ does neither commute with $\mathbf{P_G}$ nor with $\widetilde{\mathbf{\Phi}}$ while $\mathbf{P_G}$ and $\widetilde{\mathbf{\Phi}}$ do. We thus have by (3.21)

$$\|\mathbf{q}\|_{\mathbf{L}+\mathbf{G}}^2 \leq \tfrac{1}{m_2\tau^2}\bigl(\mathbf{P}\mathbf{q},\mathbf{q}\bigr) + \bigl(\mathbf{G}\mathbf{q},\mathbf{q}\bigr) \leq \tfrac{1}{m_2\tau^2}\|\mathbf{P_G}^{1/2}\mathbf{q}\|^2,$$

TABLE 3.1
Summary of stability bounds for the scheme (2.3a), (2.3b) *applied to* (1.1) *with general* g, $g = 0$, *and* $g(\mathbf{q}) = \mathbf{G}\mathbf{q}$, *respectively. Details can be found in the theorems given in the last column.*

g	$\mathbf{q}_1$	$\tau \le$	norm	growth in t_n	reference
g	(3.12)	τ_{CFL}	$\|\cdot\|$	exponential	Theorem 3.7
0	(3.12)	τ_{CFL}	$\|\cdot\|$	linear	Theorem 3.11
	(2.3b)	$\vartheta\tau_{\mathrm{CFL}}, \vartheta \in (0,1)$	$\|\cdot\|, \|\|\|\cdot\|\|\|_\tau$	no growth	
	(3.12)	$\widehat{\tau}_{\mathrm{CFL}}$			
$\mathbf{G}$	(3.12)	$\min\{\widehat{\tau}_{\mathrm{CFL}}, \frac{2\vartheta}{\|\mathbf{G}\|^{1/2}}\}$, $\vartheta^2 \in (0, m_1)$	$\|\cdot\|, \|\|\|\cdot\|\|\|_\tau$	no growth	Theorem 3.16

which yields

$$\begin{aligned}\big\|\cos\big((n+\tfrac12)\widetilde{\mathbf{\Phi}}\big)\cos\big(\tfrac12\widetilde{\mathbf{\Phi}}\big)\mathbf{q}_0\big\|_{\mathbf{L}+\mathbf{G}}^2 &\le \tfrac{1}{m_2\tau^2}\big\|\cos\big((n+\tfrac12)\widetilde{\mathbf{\Phi}}\big)\cos\big(\tfrac12\widetilde{\mathbf{\Phi}}\big)\mathbf{P}_{\mathbf{G}}^{1/2}\mathbf{q}_0\big\|^2\\ &\le \tfrac{1}{m_2\tau^2}\big\|\mathbf{P}_{\mathbf{G}}^{1/2}\mathbf{q}_0\big\|^2\\ &\le \tfrac{1}{m_2}\big\|\mathbf{q}_0\big\|_{\mathbf{L}+\mathbf{G}}^2\end{aligned}$$

by the same argument. Similarly, we have

$$\left\|\frac{\tau}{2}\frac{\sin\big((n+\frac12)\widetilde{\mathbf{\Phi}}\big)}{\sin\big(\frac12\widetilde{\mathbf{\Phi}}\big)}\big(\mathbf{P}'\dot{\mathbf{q}}_0+\tau^2\boldsymbol{\delta}_0\big)\right\|_{\mathbf{L}+\mathbf{G}}^2 \le \tfrac{1}{m_2}\big(P'_{\max}\|\dot{\mathbf{q}}_0\|+\tau^2\|\boldsymbol{\delta}_0\|\big)^2.$$

Combining these estimates yields (3.34b). ∎

3.5. Summary of stability results. We conclude this section by summarizing all our stability bounds in Table 3.1. It states how the stability of the scheme depends on the problem ($g = 0$, or g linear or nonlinear), the choice of the starting value, the norm, the CFL condition, and where to find the detailed bound.

4. Error analysis. In the previous section we established the stability and long-time behavior of the general scheme (2.3a). The aim of this section is to provide its error analysis. We will show a convergence result in the standard norm $\|\cdot\|$ for semilinear problems as well as convergence results in both the standard and energy norm $|||\cdot|||_\tau$ for linear problems where $g = 0$.

Let us denote the error of the scheme (2.3a) 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). We denote bounds on derivatives of $\mathbf{q}(t)$ by

$$B_n^{(k)} = \max_{0\le t\le t_n}\|\mathbf{q}^{(k)}(t)\|, \qquad k = 1, 2, \dots. \tag{4.2}$$

Further, our error analysis requires the following definition.

DEFINITION 4.1. *We define* m_3 *as the smallest constant such that*

$$\big|P(z)-z\big| \le \tfrac12 m_3 z^2, \qquad \big|P'(z)-1\big| \le m_3 z, \qquad z \in [0,\beta^2]. \tag{4.3}$$

The existence of m_3 is guaranteed by (2.2). For the LFC polynomial (2.3c) we have $m_3 = -P_p''(0) > 0$; cf. Theorem 5.1 below.

4.1. Error analysis for semilinear problems. We show an error bound for the scheme (2.3a) for semilinear problems (1.1). First, we prove the following error recursion.

LEMMA 4.2. *For* $\mathbf{q} \in C^4(0,T)$ *the error* $\mathbf{e}_n$, $n \geq 1$, *of the scheme* (2.3a) *satisfies the recursion*

$$[\![\mathbf{e}_n]\!] + \mathbf{P}\mathbf{e}_n = \mathbf{d}_n + \mathbf{r}_n, \qquad \mathbf{d}_n = \big(\mathbf{P} - \tau^2\mathbf{L}\big)\widetilde{\mathbf{q}}_n + \boldsymbol{\delta}_n^{(4)}, \tag{4.4a}$$

where $[\![\mathbf{e}_n]\!] = \mathbf{e}_{n+1} - 2\mathbf{e}_n + \mathbf{e}_{n-1}$ *and*

$$\mathbf{r}_n = -\tau^2(g(\widetilde{\mathbf{q}}_n) - g(\mathbf{q}_n)), \tag{4.4b}$$

$$\boldsymbol{\delta}_n^{(k)} = \tau^{k-1}\int_{t_n}^{t_{n+1}} \kappa_{n,+}^{(k-1)}(t)\mathbf{q}^{(k)}(t)\,dt - \tau^{k-1}\int_{t_{n-1}}^{t_n} \kappa_{n,-}^{(k-1)}(t)\mathbf{q}^{(k)}(t)\,dt \tag{4.4c}$$

with $\kappa_{n,\pm}^{(\ell)}(t) = (t_{n\pm1} - t)^\ell/(\ell!\tau^\ell)$.

Proof. Inserting the exact solution $\widetilde{\mathbf{q}}_n$ into the scheme (2.3a) yields

$$[\![\widetilde{\mathbf{q}}_n]\!] + \mathbf{P}\widetilde{\mathbf{q}}_n = \mathbf{d}_n - \tau^2 g(\widetilde{\mathbf{q}}_n),$$

and thus (4.4a) with a defect $\mathbf{d}_n$. In order to determine $\mathbf{d}_n$ we use the Taylor expansion and the differential equation (1.1) to obtain

$$[\![\widetilde{\mathbf{q}}_n]\!] = \tau^2\ddot{\mathbf{q}}(t_n) + \boldsymbol{\delta}_n^{(4)} = -\tau^2\mathbf{L}\widetilde{\mathbf{q}}_n - \tau^2 g(\widetilde{\mathbf{q}}_n) + \boldsymbol{\delta}_n^{(4)}.$$

Subtracting this relation from the previous equation completes the proof. ∎

The error recursion (4.4) leads to the following error bound.

THEOREM 4.3. *Let* $\mathbf{q} \in C^4(0,T)$ *and consider the scheme* (2.3a) *with general starting value* (3.12). *Then, for* $\tau \leq \tau_{\mathrm{CFL}}$ *and* $t_n \leq T$ *we have*

$$\|\mathbf{e}_n\| \leq \big(C_1 T + \tfrac{1}{2}C_d T^2\big)\mathrm{e}^{\sqrt{L_g}T}\,\tau^2, \tag{4.5}$$

where the constants C_1, C_d *are independent of* $\|\mathbf{L}\|$, n, *and* τ.

Proof. As in the proof of Theorem 3.3 we again use the generating functions technique. From Lemma 4.2 we get

$$\mathbf{e}_n = \frac{\sin(n\boldsymbol{\Phi})}{\sin\boldsymbol{\Phi}}\mathbf{e}_1 + \sum_{\ell=1}^{n-1}\frac{\sin\big((n-\ell)\boldsymbol{\Phi}\big)}{\sin\boldsymbol{\Phi}}\big(\mathbf{d}_\ell + \mathbf{r}_\ell\big), \tag{4.6}$$

where we used that $\mathbf{e}_0 = 0$. Taylor expansion of $\mathbf{q}(\tau)$ and subtracting (3.12) shows

$$\mathbf{e}_1 = \tfrac{1}{2}(\mathbf{P} - \tau^2\mathbf{L})\mathbf{q}_0 + \tau(\mathbf{I} - \mathbf{P}')\dot{\mathbf{q}}_0 - \tau^3\boldsymbol{\delta}_0 + \tau^2\int_0^\tau \kappa_{0,+}^{(2)}(t)\mathbf{q}^{(3)}(t)\,dt.$$

By Definition 4.1 we get for $\tau \leq \tau_{\mathrm{CFL}}$

$$\|\mathbf{e}_1\| \leq C_1\tau^3, \qquad C_1 = \tfrac{1}{4}\tau\, m_3\|\mathbf{L}^2\mathbf{q}_0\| + m_3\|\mathbf{L}\dot{\mathbf{q}}_0\| + \tfrac{1}{6}B_1^{(3)} + \|\boldsymbol{\delta}_0\|, \tag{4.7}$$

and for the defects

$$\|\mathbf{d}_\ell\| \leq C_d\tau^4, \qquad C_d = \tfrac{1}{2}m_3 \max_{0\leq t\leq t_n}\|\mathbf{L}^2\mathbf{q}(t)\| + \tfrac{1}{12}B_n^{(4)}. \tag{4.8}$$

Inserting these bounds and $\|\mathbf{r}_n\| \le \tau^2 L_g \|\mathbf{e}_n\|$ in (4.6) yields

$$\|\mathbf{e}_n\| \le \big(C_1 t_n + \tfrac{1}{2} C_d t_n^2\big)\tau^2 + \tau^2 L_g \sum_{\ell=1}^{n-1} (n-\ell)\|\mathbf{e}_\ell\|. \tag{4.9}$$

Since $t_n \le T$, the claim follows from Lemma 3.8. ∎

4.2. Error analysis for linear problems. We now restrict ourselves to the linear problem (3.2). As an immediate consequence of Theorem 4.3 we get the following result.

COROLLARY 4.4. *Let the assumptions of Theorem* 4.3 *be satisfied and* $g = 0$. *Then, for* $\tau \le \tau_{\mathrm{CFL}}$ *and* $t_n \le T$ *we have*

$$\|\mathbf{e}_n\| \le \big(C_1 t_n + \tfrac{1}{2} C_d t_n^2\big)\tau^2, \tag{4.10}$$

where $C_1 = \frac{1}{4}\tau m_3 B_0^{(4)} + m_3 B_0^{(3)} + \frac{1}{6} B_1^{(3)} + \|\boldsymbol{\delta}_0\|$ *and* $C_d = \frac{1}{12}\big(6m_3 + 1\big) B_n^{(4)}$.

Proof. The proof follows from (4.7), (4.8), and (4.9) since $L_g = 0$ and $\mathbf{L}\mathbf{q} = -\ddot{\mathbf{q}}$. ∎

A similar result has already be proven in [4] for the exact starting value $\mathbf{q}_1 = \mathbf{q}(\tau)$.

Next, we show that the scheme (2.3a) with general starting value (3.12) also converges with order two in the discrete energy norm $|||\cdot|||_\tau$. Here, we use the stronger CFL condition (3.22).

THEOREM 4.5. *Let the assumptions of Theorem* 4.3 *be satisfied and* $g = 0$. *Then, for* $\tau \le \widehat{\tau}_{\mathrm{CFL}}$ *and* $t_{n+1} \le T$ *we have*

$$|||\mathbf{e}_{n+\frac{1}{2}}|||_\tau \le \big(\tfrac{1}{m_1} + \tfrac{1}{m_2}\big)^{1/2}\big(C_1 + t_n C_d\big)\tau^2, \tag{4.11}$$

where C_1 *and* C_d *are given as in Corollary* 4.4 *(with* $B_n^{(4)}$ *replaced by* $B_{n+1}^{(4)}$*).*

One can also show second-order convergence under the weaker CFL condition (3.8), however, this requires a more regular solution.

Proof. Similarly to Lemma 3.5 we get from (4.6) with $g = 0$

$$[\mathbf{e}_{n+\frac{1}{2}}] = \frac{\cos\big((n+\frac{1}{2})\boldsymbol{\Phi}\big)}{\cos\big(\frac{1}{2}\boldsymbol{\Phi}\big)}\mathbf{e}_1 + \sum_{\ell=1}^{n} \frac{\cos\big((n-\ell+\frac{1}{2})\boldsymbol{\Phi}\big)}{\cos\big(\frac{1}{2}\boldsymbol{\Phi}\big)}\mathbf{d}_\ell,$$

$$\{\mathbf{e}_{n+\frac{1}{2}}\} = \frac{\sin\big((n+\frac{1}{2})\boldsymbol{\Phi}\big)}{2\sin\big(\frac{1}{2}\boldsymbol{\Phi}\big)}\mathbf{e}_1 + \sum_{\ell=1}^{n} \frac{\sin\big((n-\ell+\frac{1}{2})\boldsymbol{\Phi}\big)}{2\sin\big(\frac{1}{2}\boldsymbol{\Phi}\big)}\mathbf{d}_\ell.$$

By using (3.27) and (3.21) we obtain

$$\big\|\tfrac{1}{\tau}[\mathbf{e}_{n+\frac{1}{2}}]\big\|^2 + \|\{\mathbf{e}_{n+\frac{1}{2}}\}\|_{\mathbf{L}}^2 \le \big(\tfrac{1}{m_1} + \tfrac{1}{m_2}\big)\Big(\big\|\tfrac{1}{\tau}\mathbf{e}_1\big\| + \sum_{\ell=1}^{n}\big\|\tfrac{1}{\tau}\mathbf{d}_\ell\big\|\Big)^2. \tag{4.12}$$

Inserting (4.7) and (4.8) into this estimate completes the proof. ∎

We conclude this section by showing in a refined analysis that under additional assumptions the scheme (2.3a) with special starting value (2.3b) converges with order four.

DEFINITION 4.6. *Let* $m_3^* = -P''(0)$. *Then we define* m_4 *as the smallest constant such that*

$$\big|P(z) - z + \tfrac{1}{2}m_3^* z^2\big| \le m_4 z^3, \quad \big|P'(z) - 1 + m_3^* z\big| \le 3m_4 z^2, \qquad z \in [0, \beta^2]. \tag{4.13}$$

Note that $m_3^* > 0$ because of (3.20). With this definition, we can state the error bound for sufficiently smooth solutions.

THEOREM 4.7. *Let $\mathbf{q} \in C^6(0,T)$ be the solution of* (3.2). *We consider the scheme* (2.3a), (2.3b) *with a polynomial P satisfying* (2.2) *and*

$$m_3^* = -P''(0) = \frac{1}{6}. \tag{4.14}$$

Then, for $t_{n+1} \le T$ we have

$$\|\mathbf{e}_n\| \le t_n C_4 \tau^4, \qquad \tau \le \tau_{\mathrm{CFL}}, \tag{4.15a}$$

$$|||\mathbf{e}_{n+\frac{1}{2}}|||_\tau \le \left(\tfrac{1}{m_1} + \tfrac{1}{m_2}\right)^{1/2} C_4 \tau^4, \qquad \tau \le \widehat{\tau}_{\mathrm{CFL}}, \tag{4.15b}$$

where C_4 only depends on t_n, the bounds $B_1^{(5)}$, $B_{n+1}^{(6)}$, and m_4.

Proof. The statement follows mainly as in the proof of Lemma 4.2 and Theorem 4.3 with two minor changes. Definition 4.6 implies for all $z \in \left[0, \beta^2\right]$

$$|Q(z)| \le m_4, \qquad Q(z) = \frac{P(z) - z + \frac{1}{2} m_3^* z^2}{z^3}, \tag{4.16a}$$

$$|\widehat{Q}(z)| \le 3m_4, \qquad \widehat{Q}(z) = \frac{P'(z) - 1 + m_3^* z}{z^2}. \tag{4.16b}$$

We can then write the defect in (4.4) as $\mathbf{d}_n = \boldsymbol{\Delta}_n + \boldsymbol{\delta}_n^{(6)}$ with $\boldsymbol{\delta}_n^{(6)}$ defined in (4.4c) and

$$\boldsymbol{\Delta}_n = \left(\mathbf{P} - \tau^2 \mathbf{L} + \tfrac{1}{12}\tau^4 \mathbf{L}^2\right) \widetilde{\mathbf{q}}_n = \tfrac{1}{12}\left(1 - 6m_3^*\right)\tau^4 \mathbf{q}^{(4)}(t_n) - \tau^6 Q(\tau^2 \mathbf{L}) \mathbf{q}^{(6)}(t_n),$$

and the error $\mathbf{e}_1$ as

$$\begin{aligned}
\mathbf{e}_1 &= \tfrac{1}{2}\left(\mathbf{P} - \tau^2 \mathbf{L} + \tfrac{1}{12}\tau^4 \mathbf{L}^2\right)\mathbf{q}_0 + \tau\left(\mathbf{I} - \tfrac{1}{6}\tau^2 \mathbf{L} - \mathbf{P}'\right)\dot{\mathbf{q}}_0 + \tau^4 \int_0^\tau \kappa_{0,+}^{(4)}(t) \mathbf{q}^{(5)}(t)\, dt \\
&= \tfrac{1}{24}\left(1 - 6m_3^*\right)\tau^4 \mathbf{q}^{(4)}(0) + \tfrac{1}{6}\left(1 - 6m_3^*\right)\tau^3 \mathbf{q}^{(3)}(0) \\
&\quad - \tfrac{1}{2} Q(\tau^2 \mathbf{L})\tau^6 \mathbf{q}^{(6)}(0) - \widehat{Q}(\tau^2 \mathbf{L})\tau^5 \mathbf{q}^{(5)}(0) + \tau^4 \int_0^\tau \kappa_{0,+}^{(4)}(t) \mathbf{q}^{(5)}(t)\, dt.
\end{aligned}$$

This yields

$$\|\mathbf{d}_n\| \le 2M_3 B_n^{(4)} \tau^4 + \left(m_4 + \tfrac{1}{360}\right) B_{n+1}^{(6)} \tau^6, \qquad M_3 = \tfrac{1}{24}\left|1 - 6m_3^*\right|,$$

and

$$\|\mathbf{e}_1\| \le 4M_3 B_0^{(3)} \tau^3 + M_3 B_0^{(4)} \tau^4 + 3m_4 B_0^{(5)} \tau^5 + \tfrac{1}{120} B_1^{(5)} \tau^5 + \tfrac{1}{2} m_4 B_0^{(6)} \tau^6.$$

Inserting these bounds into (4.6) (with $g = 0$) and into (4.12) proves

$$\|\mathbf{e}_n\| \le t_n\left((C_2 + C_2')\tau^2 + C_3 \tau^3 + C_4 \tau^4\right), \qquad \tau \le \tau_{\mathrm{CFL}}, \tag{4.17a}$$

$$|||\mathbf{e}_{n+\frac{1}{2}}|||_\tau \le \left(\tfrac{1}{m_1} + \tfrac{1}{m_2}\right)^{1/2}\left((C_2 + 2C_2')\tau^2 + C_3 \tau^3 + C_4 \tau^4\right), \quad \tau \le \widehat{\tau}_{\mathrm{CFL}}, \tag{4.17b}$$

where

$$C_2 = 4M_3 B_0^{(3)}, \qquad C_2' = t_n M_3 B_n^{(4)}, \qquad C_3 = M_3 B_0^{(4)}. \tag{4.17c}$$

The claim follows from $P''(0) = -m_3^* = -1/6$. ∎

The proof of Theorem 4.7 shows that the scheme applied to a linear problem is of order two unless (4.14) is satisfied, when it is of order four. Moreover, for $0 < m_3^* < \frac{1}{3}$, the error constant M_3 arising in (4.17) is smaller than the one of the LF scheme, where $m_3^* = 0$. This results in smaller errors, as will be confirmed in our numerical examples in section 7.

For the LFC method (2.3) there exists for every $p \geq 2$ a stabilization parameter $\nu_p > 1$ such that the method is of order four; cf. Theorem 5.1 and Remark 5.3. We further note that the analysis in this section can easily be generalized to higher-order schemes if the polynomial P satisfies additional consistency properties.

5. LFC methods. In this section we focus on the LFC method which arises if we use the polynomial (2.3c) for the scheme (2.3a), (2.3b). To be more precise, we give all relevant constants arising in the error and stability analysis of the two previous sections explicitly.

Moreover, we show a relation of LFC methods without stabilization ($\nu_p = 1$) and the LF method for linear problems.

5.1. Explicit bounds for LFC methods. We explicitly state the constants given in sections 3 and 4 for the LFC method (2.3).

First, we show the result for the weaker CFL condition (3.8).

THEOREM 5.1. *Let $p \geq 1$ and $\nu_p \geq 1$. For the polynomial P_p defined in* (2.3c) *the constants in Definitions* 3.2, 4.1 *and* 4.6 *are explicitly given by*

$$\beta^2 = \beta_p^2 = 2\alpha_p\nu_p, \qquad m_3 = m_3^* = -P_p''(0) = 2\,\frac{T_p''(\nu_p)}{\alpha_p^2 T_p(\nu_p)}, \quad m_4 = \frac{T_p'''(\nu_p)}{3\alpha_p^3 T_p(\nu_p)}. \tag{5.1}$$

Moreover, $P'_{\max} = \max_{z\in[0,\beta_p^2]} |P_p'(z)| = 1$ and (3.20) *is satisfied for all $z \in [0, \beta_p^2]$.*

By definition (2.3c) of α_p we have $\beta_p^2 = 4p^2$ for $\nu_p = 1$.

Proof. Throughout this proof we change between the coordinates

$$x = \nu_p - \frac{z}{\alpha_p} \in [-\nu_p, \nu_p] \qquad \text{and} \qquad z = \alpha_p(\nu_p - x) \in \big[0, \beta_p^2\big]. \tag{5.2}$$

(i) We have to prove that the inequalities (3.7) hold true. It is well known that for $\nu_p \geq 1$ we have

$$-T_p(\nu_p) \leq T_p(x) \leq T_p(\nu_p) \qquad \text{for } x \in [-\nu_p, \nu_p],$$

which is equivalent to

$$0 \leq P_p(z) \leq 4 \qquad \text{for } z \in [0, \beta_p^2].$$

(ii) Next, we show that the upper bound of (3.20) is satisfied for the LFC polynomials (2.3c). We use that

$$T_p'(x) \leq T_p'(1) \qquad \text{for } x \in [-1, 1];$$

see, e.g., [8, Theorem 2.1] or the original work [16]. Since T_p' is monotonically increasing on $[1, \infty)$ and because of the symmetry properties of the Chebyshev polynomials we deduce that

$$T_p'(x) \leq T_p'(\nu_p) \qquad \text{for } x \in [-\nu_p, \nu_p]. \tag{5.3}$$

Integrating from x to ν_p gives

$$T_p(\nu_p) - T_p(x) \le T_p'(\nu_p)(\nu_p - x),$$

which yields

$$P_p(z) \le 2\frac{T_p'(\nu_p)}{T_p(\nu_p)}(\nu_p - x) = \alpha_p(\nu_p - x) = z \qquad \text{for } z \in [0, \beta_p^2].$$

(iii) From (5.3) we obtain again by the symmetry of the Chebyshev polynomials

$$|T_p'(x)| \le T_p'(\nu_p) \qquad \text{for } x \in [-\nu_p, \nu_p].$$

Thus, we have $P'_{\max} \le 1$ by using

$$P_p'(z) = \frac{2}{\alpha_p T_p(\nu_p)} T_p'(x) = \frac{1}{T_p'(\nu_p)} T_p'(x). \tag{5.4}$$

(iv) We have to show the first inequalities in (4.3) and (4.13) with constants $m_3 = m_3^*$ and m_4 in (5.1). Markov brothers' inequality (see, e.g., [8, Theorem 2.2] or the original work [17]) states that

$$T_p^{(k)}(x) \le T_p^{(k)}(1) \qquad \text{for } x \in [-1, 1],\ k \in \mathbb{N}.$$

From this one can again deduce

$$T_p^{(k)}(x) \le T_p^{(k)}(\nu_p) \qquad \text{for } x \in [-\nu_p, \nu_p],\ k \in \mathbb{N}. \tag{5.5}$$

Using $k = 2$ in this inequality and integrating it twice from x to ν_p yields

$$T_p(\nu_p) - T_p(x) \ge T_p'(\nu_p)(\nu_p - x) - \frac{T_p''(\nu_p)}{2}(\nu_p - x)^2.$$

Choosing $k = 3$ and integrating three times we get

$$T_p(\nu_p) - T_p(x) \le T_p'(\nu_p)(\nu_p - x) - \frac{T_p''(\nu_p)}{2}(\nu_p - x)^2 + \frac{T_p'''(\nu_p)}{6}(\nu_p - x)^3.$$

From these two inequalities we conclude

$$P_p(z) \ge z - \tfrac{1}{2} m_3 z^2, \qquad P_p(z) \le z - \tfrac{1}{2} m_3^* z^2 + m_4 z^3.$$

Together with the second bound of (3.20) we have

$$0 \ge P_p(z) - z \ge -\tfrac{1}{2} m_3 z^2, \qquad 0 \le P_p(z) - z + \tfrac{1}{2} m_3^* z^2 \le m_4 z^3.$$

(v) It remains to show the second bounds in (4.3) and (4.13) with constants (5.1). First, we get by (5.3) and (5.4) that $P_p'(z) - 1 \le 0$. Integrating (5.5) with $k = 2$ once from x to ν_p we obtain

$$T_p'(\nu_p) - T_p'(x) \le T_p''(\nu_p)(\nu_p - x).$$

Moreover, by integrating (5.5) with $k = 3$ twice we get

$$T_p'(\nu_p) - T_p'(x) \ge T_p''(\nu_p)(\nu_p - x) - \frac{T_p'''(\nu_p)}{2}(\nu_p - x)^2.$$

Hence, we can conclude

$$m_3 z \le P_p'(z) - 1 \le 0, \qquad 0 \le P_p'(z) - 1 + m_3^* z \le 3m_4 z^2,$$

which finishes the proof. ∎

Next, we give a result for the stronger CFL condition in (3.21).

THEOREM 5.2. *Let $p > 1$. The polynomial P_p in* (2.3c) *satisfies* (3.21)
(a) *for $\nu_p = 1$ and every $m_1, m_2 > 0$ with $\widehat{\beta}^2 < 2p^2\big(1 - \cos\frac{\pi}{p}\big) < \pi^2$;*
(b) *for $\nu_p > 1$ and*

$$m_1 = \frac{1}{2}\Big(1 - \frac{1}{T_p(\nu_p)}\Big), \qquad m_2 = 4\frac{m_1}{\alpha_p(\nu_p + 1)} \tag{5.6a}$$

with

$$\widehat{\beta}^2 = \widehat{\beta}_p^2 = \alpha_p(\nu_p + 1). \tag{5.6b}$$

Note that it is possible to slightly increase the stability bound $\widehat{\beta}_p^2$ given in (5.6b) for $\nu_p > 1$. However, we have to degrade either m_1 or m_2 depending on whether the polynomial degree p is odd or even.

Proof. Similarly to the previous proof we change between the coordinates

$$x = \nu_p - \frac{z}{\alpha_p} \in [-1, \nu_p] \qquad \text{and} \qquad z = \alpha_p(\nu_p - x) \in \big[0, \widehat{\beta}_p^2\big].$$

(a) Let $\nu_p = 1$. Thus, we have $\alpha_p = 2p^2$ and $\beta_p^2 = 4p^2$. Since for $x \in (-1, 1)$ the local extrema of T_p are given by $x_k = \cos(k\pi/p)$, $k = 1, \dots, n-1$, we get

$$P_p(z_k) \in \{0, 4\} \qquad \text{for } z_k = 2p^2\big(1 - \cos(k\pi/p)\big), \quad k = 1, \dots, n-1.$$

In particular, z_1 is the first maximum point of P_p in $z \in (0, \beta_p^2)$. Thus, because of $m_1, m_2 > 0$ we obtain

$$\widehat{\beta}^2 < z_1 = 2p^2\big(1 - \cos(\pi/p)\big) < \pi^2.$$

(b) Let $\nu_p > 1$. We have to prove that the inequalities (3.21) hold true with constants (5.6a). First inequality: We have

$$-1 \le T_p(x) \le T_p(\nu_p) \qquad \text{for } x \in [-1, \nu_p];$$

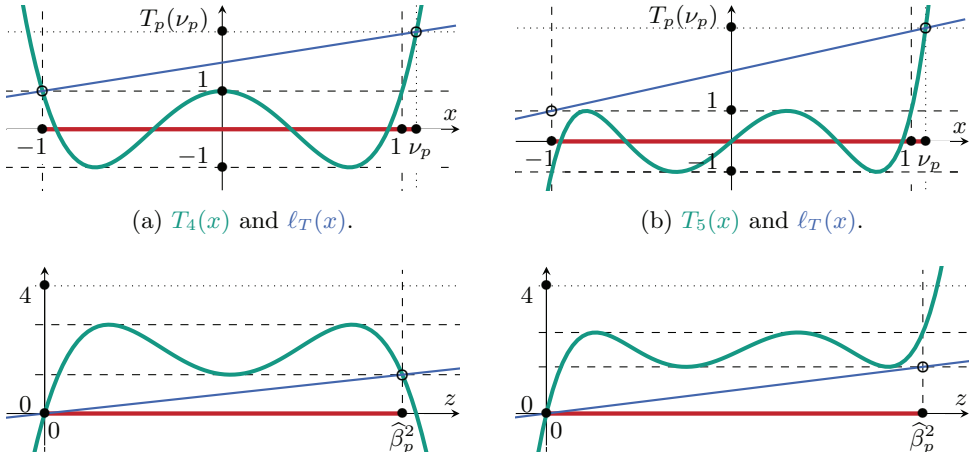
see also Figures 5.1(a) and 5.1(b). This is equivalent to

$$\frac{1}{2}\Big(1 - \frac{1}{T_p(\nu_p)}\Big) \le 1 - \frac{1}{4}P_p(z) \le 1 \qquad \text{for } z \in [0, \widehat{\beta}_p^2],$$

which is the desired bound with m_1 given in (5.6a).

Second inequality: The upper bound was already shown in the previous theorem. For the lower bound note that T_p is bounded by the line ℓ_T through $(-1, 1)$ and $(\nu_p, T_p(\nu_p))$ (the blue line in Figures 5.1(a) and 5.1(b)), i.e., for $x \in [-1, \nu_p]$, we have

$$\begin{aligned} T_p(x) \le \ell_T(x) &= T_p(\nu_p) + \frac{1 - T_p(\nu_p)}{1 + \nu_p}(\nu_p - x) \\ &= T_p(\nu_p) + \frac{1 - T_p(\nu_p)}{\beta_p^2}\alpha_p(\nu_p - x). \end{aligned}$$



(a) $T_4(x)$ and $\ell_T(x)$. (b) $T_5(x)$ and $\ell_T(x)$.

(c) $P_4(z)$ and $m_2 z$. (d) $P_5(z)$ and $m_2 z$.

FIG. 5.1. *Illustration of the Chebyshev polynomial $T_p(x)$ and the line $\ell_T(x)$ (top) and of the LFC polynomial $P_p(z)$ and the line $m_2 z$ (bottom) for $p = 4, 5$.*

From this we obtain

$$P_p(z) \geq \frac{2}{\beta_p^2}\Big(1 - \frac{1}{T_p(\nu_p)}\Big)z = m_2 z \qquad \text{for } z \in [0, \widehat{\beta}_p^2],$$

which is the claimed bound (see also the blue line in Figures 5.1(c) and 5.1(d)). ∎

Remark 5.3. (i) For $p = 2, \dots, 5$ the following choices of ν_p fulfill (4.14):

$$\nu_2 = \frac{\sqrt{6}}{2} \approx 1.224745, \quad \nu_3 \approx 1.029086, \quad \nu_4 \approx 1.008261, \quad \nu_5 \approx 1.003233,$$

and thus give fourth-order schemes for linear problems. In the case of $p = 2$ and $\nu_2 = \frac{\sqrt{6}}{2}$ we retrieve the modified LF method with $\beta_p^2 = 12$; see (2.4).

(ii) For $\nu_p > 1$ we have the following limits:

$$\lim_{\nu_p \to 1} m_1 = \lim_{\nu_p \to 1} m_2 = 0, \qquad \lim_{\nu_p \to 1} \widehat{\beta}_p^2 = 4p^2,$$
$$\lim_{\nu_p \to 1} m_3 = \frac{p^2 - 1}{6p^2}, \qquad \lim_{\nu_p \to 1} m_4 = \frac{(p^2 - 1)(p^2 - 4)}{360p^4}.$$

From Figure 5.2 we see that the stability constants m_1, m_2 in (5.6a) improve and $\widehat{\beta}_p$ in (5.6b) degrades with increasing ν_p. Further, we observe that the limits of $m_1, m_2, \widehat{\beta}$ do not coincide with the values in Theorem 5.2 for $\nu_p = 1$. Moreover, we see that the error constant $M_3 = \frac{1}{24}|1 - 6m_3^*|$ defined in the proof of Theorem 4.7 depends on ν_p.

5.2. LFC schemes without stabilization. In the following we prove that the unstabilized LFC and the LF scheme are closely related for linear problems (3.2).

THEOREM 5.4. *Let $g = 0$. For $k, m \in \mathbb{N}$ we denote by*

(a) $\mathbf{q}_k$ *the solution of the LFC scheme* (2.3) *with $\nu_p = 1$ and polynomial degree $p \in \mathbb{N}$ after k time steps with step size τ and by*

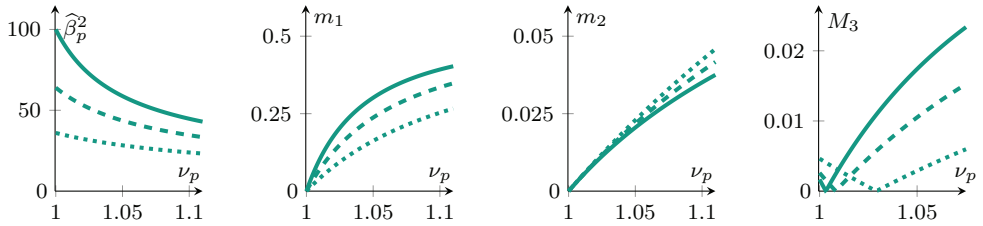


FIG. 5.2. *Dependence of* $\widehat{\beta}_p^2$, m_1, m_2 *in* (5.6) *and* $M_3 = \frac{1}{24}|1 - 6m_3^*|$ *on* ν_p *for* $p = 3, 4, 5$ *(dotted, dashed, solid).*

(b) $\widehat{\mathbf{q}}_m$ *the solution of the LF scheme* (2.1) *after* m *time steps with step size* $\widehat{\tau}$. *If* $\tau \le \tau_{\mathrm{CFL}}$ *and* $\widehat{\tau} = \tau/p$, *we have*

$$\mathbf{q}_n = \widehat{\mathbf{q}}_{np}, \qquad n = 1, 2, \dots.$$

Note that the equivalence only holds because of the special choice of the starting value (2.3b) for the LFC scheme. In fact, for other starting values for both the LFC and the LF scheme the equivalence does not hold true.

Proof. The proof is mainly based on the representation of the numerical solution in Theorem 3.3 and Corollary 3.4, respectively. First, note that by the CFL condition $\tau^2 \le \tau_{\mathrm{CFL}}^2 = 4p^2/\|\mathbf{L}\|$ for the unstabilized LFC scheme we get

$$\widehat{\tau}^2 = \frac{\tau^2}{p^2} \le \frac{4}{\|\mathbf{L}\|},$$

which is the CFL condition of the LF scheme. Since the scheme (2.3a), (2.3b) reduces for $p = 1$ to the LF method (2.1), Corollary 3.4 holds with $\boldsymbol{\delta}_0 = 0$ and we get for the LF scheme

$$\widehat{\mathbf{q}}_m = \cos(m\widehat{\boldsymbol{\Phi}})\mathbf{q}_0 + \widehat{\tau}\frac{\sin(m\widehat{\boldsymbol{\Phi}})}{\sin(\widehat{\boldsymbol{\Phi}})}\dot{\mathbf{q}}_0, \qquad \text{where} \quad \cos(\widehat{\boldsymbol{\Phi}}) = \mathbf{I} - \tfrac{1}{2}\widehat{\tau}^2\mathbf{L}$$

with a matrix $\widehat{\boldsymbol{\Phi}}$ with spectrum in $[0, \pi]$.

By (3.9b) and definition (2.3c) of $\mathbf{P}$ for $\nu_p = 1$ we have

$$\cos(\boldsymbol{\Phi}) = \mathbf{I} - \tfrac{1}{2}\mathbf{P} = T_p\Big(\mathbf{I} - \tfrac{1}{2p^2}\tau^2\mathbf{L}\Big) = \cos\Big(p\arccos\big(\mathbf{I} - \tfrac{1}{2}\widehat{\tau}^2\mathbf{L}\big)\Big) = \cos(p\widehat{\boldsymbol{\Phi}}).$$

From this we get on the one hand

$$\cos(n\boldsymbol{\Phi}) = \cos(np\widehat{\boldsymbol{\Phi}})$$

by using $\cos(n\zeta) = 2\cos((n-1)\zeta)\cos(\zeta) - \cos((n-2)\zeta)$ for $\zeta \in \mathbb{R}, n \in \mathbb{N}$, and an induction argument. On the other hand we have by using an angle sum identity and again an induction argument

$$\frac{\sin(n\boldsymbol{\Phi})}{\sin(\boldsymbol{\Phi})} = \frac{\sin(np\widehat{\boldsymbol{\Phi}})}{\sin(p\widehat{\boldsymbol{\Phi}})}, \qquad n \in \mathbb{N}.$$

Inserting these identities into (3.13) for the LFC scheme and using

$$\mathbf{P}' = P_p'(\tau^2\mathbf{L}) = \frac{1}{p^2}T_p'\left(\mathbf{I} - \tfrac{1}{2}\widehat{\tau}^2\mathbf{L}\right) = \frac{\sin(p\widehat{\boldsymbol{\Phi}})}{p\sin(\widehat{\boldsymbol{\Phi}})}$$

yields

$$\mathbf{q}_n = \cos(n\mathbf{\Phi})\mathbf{q}_0 + \tau\frac{\sin(n\mathbf{\Phi})}{\sin(\mathbf{\Phi})}\mathbf{P}'\dot{\mathbf{q}}_0 = \cos(np\widehat{\mathbf{\Phi}})\mathbf{q}_0 + \frac{\tau}{p}\frac{\sin(np\widehat{\mathbf{\Phi}})}{\sin(\widehat{\mathbf{\Phi}})}\dot{\mathbf{q}}_0 = \widehat{\mathbf{q}}_{np},$$

which completes the proof. ∎

As an immediate consequence of the previous theorem we can state the constants $C_k(\vartheta, P_p)$, $k = 0, 1$, in Theorem 3.11 for the LFC polynomials with $\nu_p = 1$ explicitly.

COROLLARY 5.5. *For $p \geq 1$ and $\nu_p = 1$, Theorem* 3.11 *holds for the LFC polynomial* (2.3c) *with $C_k(\vartheta, P_p) = \left(1 - \vartheta^2\right)^{-1/2}$, $k = 0, 1$.*

Proof. By Theorem 5.4, the LFC scheme with $\nu_p = 1$ and the LF scheme are equivalent. Hence, it is sufficient to prove the claim for the LF scheme, i.e., $p = 1$ and $P_1(z) = z$. In the proof of Theorem 3.11 we thus have $\widetilde{\psi}_j(z) = (1 - z/4)^{-1/2}$ for $j = 0, 1$ and $\widetilde{\psi}_2(z) = 1$. Moreover, we have $\tau_{\mathrm{CFL}} = 4$ and $\widehat{\tau}_{\mathrm{CFL}} = 4\vartheta^2$, $\vartheta \in (0, 1)$, for $p = 1$. This proves the bound for all $n \in \mathbb{N}$ and thus in particular for multiples of p, where the LF approximations coincide with the LFC approximations. ∎

6. Efficiency, implementation, and generalizations of the LFC method. In this section we discuss the efficiency and the implementation of the LFC method (2.3) for semilinear differential equations and we generalize it to fully nonlinear problems.

6.1. Implementation and efficiency of LFC methods. In Algorithm 6.1 we present an efficient implementation of the nth time step of the LFC method (2.3) to integrate the semilinear problem (1.1). This requires a recursion for P_p inspired by RKC methods [19] which we provide in the next lemma.

LEMMA 6.1. *The polynomial*

$$P_{k,p}(z) = 2 - \frac{2}{T_k(\nu_p)}T_k\Big(\nu_p - \frac{z}{\alpha_p}\Big)$$

satisfies the recursion

$$\begin{aligned} P_{0,p}(z) &= 0, \\ P_{1,p}(z) &= \frac{2}{\alpha_p\nu_p}z, \\ T_k(\nu_p)P_{k,p}(z) &= 2\nu_p T_{k-1}(\nu_p)P_{k-1,p}(z) \\ &\quad + \frac{2}{\alpha_p}T_{k-1}(\nu_p)z\big(2 - P_{k-1,p}(z)\big) - T_{k-2}(\nu_p)P_{k-2,p}(z) \end{aligned}$$

for $k = 2, \ldots, p$.

Proof. The result easily follows from the recursion of Chebyshev polynomials. ∎

Lemma 6.1 and $P_p(z) = P_{p,p}(z)$ imply the following algorithm to implement one time step of the LFC scheme (2.3) for the semilinear problem (1.1).

The parameters α_p and $T_0(\nu_p), \ldots, T_p(\nu_p)$ have to be precomputed only once by means of the Chebyshev recursions. Hence, each time step requires p matrix-vector multiplications with $\mathbf{L}$ and one evaluation of g. As we show below this makes the algorithm attractive in applications where on the one hand the evaluation of g is

Algorithm 6.1 LFC scheme for semilinear problems (1.1).

$\widetilde{P}_0 = 0$, $\widetilde{P}_1 = \frac{2}{\alpha_p \nu_p}\tau^2 \mathbf{L}\mathbf{q}_n$
for $k = 2, \dots, p$ **do**
  $\widetilde{P}_k = 2\nu_p \frac{T_{k-1}(\nu_p)}{T_k(\nu_p)}\widetilde{P}_{k-1} + \frac{2}{\alpha_p}\frac{T_{k-1}(\nu_p)}{T_k(\nu_p)}\tau^2 \mathbf{L}\big(2\mathbf{q}_n - \widetilde{P}_{k-1}\big) - \frac{T_{k-2}(\nu_p)}{T_k(\nu_p)}\widetilde{P}_{k-2}$
end for
$\mathbf{q}_{n+1} = 2\mathbf{q}_n - \mathbf{q}_{n-1} - \widetilde{P}_p - \tau^2 \mathbf{g}_n$

expensive compared to a matrix-vector multiplication by $\mathbf{L}$ but on the other hand the time step size is restricted by a CFL condition dominated by $\mathbf{L}$.

We compare the CFL conditions of the standard LF scheme and the general multirate recursion (2.3a) for the linear problem (3.32), where $\|\mathbf{L}\| = r\|\mathbf{G}\|$ with a factor $r \gg 1$. In Theorem 3.16 the stability of the multirate scheme was shown under the CFL condition (3.33). For $p^2 \lesssim r$ the first CFL condition in (3.33) limits the time step size, whereas for $p^2 \gtrsim r$ the second CFL condition applies. This means that a larger polynomial degree p of P_p in (2.3a) improves the CFL condition until $p \approx \sqrt{r}$. A further increase of the polynomial degree does not alleviate the CFL condition anymore. So, let $p^2 \lesssim r$. Then the CFL condition of the recursion (2.3a) and of the LF method are

$$\tau^2 \lesssim \frac{4p^2}{r\|\mathbf{G}\|} \qquad \text{and} \qquad \tau^2 \lesssim \frac{4}{(r+1)\|\mathbf{G}\|},$$

respectively. The fraction is $\frac{r+1}{r}p^2 \sim p^2$ since we assume $r \gg 1$. Thus, the recursion (2.3a) allows an (approximately) p times larger time step size than the LF method.

In summary, we conclude that N time steps of the LFC method (via Algorithm 6.1) cost

$$pN \text{ matrix-vector multiplications with } \mathbf{L} \quad + \quad N \text{ evaluations of } g.$$

Due to its stricter CFL condition the LF method has to perform pN time steps with costs

$$pN \text{ matrix-vector multiplications with } \mathbf{L} \quad + \quad pN \text{ evaluations of } g.$$

We see that the effort on the linear part corresponding to $\mathbf{L}$ are equal for the LFC and the LF method, but the evaluations of the nonlinearity g can be (considerably) reduced by using the LFC method.

6.2. LFC methods for fully nonlinear problems. In this section we show that the LFC method can be generalized to solve fully nonlinear problems of the form

$$\ddot{\mathbf{q}}(t) = -f(\mathbf{q}(t)) - g(\mathbf{q}(t)), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0. \tag{6.2}$$

Here, f is a function with a large Lipschitz constant, which results in a severe step-size restriction for explicit schemes. Hence, we propose a scheme which requires p evaluations of f but only one evaluation of g in each time step.

The construction of the scheme is based on the recursion given in Lemma 6.1, where we use $P_p(z) = P_{p,p}(z)$. It is inspired by RKC methods for nonlinear first-order ODEs [21] (without a multirate character).

Algorithm 6.2 LFC scheme for nonlinear problems (6.2).

1: $\widetilde{P}_0 = 0$, $\widetilde{P}_1 = \frac{2}{\alpha_p \nu_p}\tau^2 f(\mathbf{q}_n)$
2: **for** $k = 2, \dots, p$ **do**
3: $\quad \widetilde{P}_k = 2\nu_p \frac{T_{k-1}(\nu_p)}{T_k(\nu_p)}\widetilde{P}_{k-1} + \frac{2}{\alpha_p}\frac{T_{k-1}(\nu_p)}{T_k(\nu_p)}\tau^2\big(2f(\mathbf{q}_n) - f(\widetilde{P}_{k-1})\big) - \frac{T_{k-2}(\nu_p)}{T_k(\nu_p)}\widetilde{P}_{k-2}$
4: **end for**
5: $\mathbf{q}_{n+1} = 2\mathbf{q}_n - \mathbf{q}_{n-1} - \widetilde{P}_p - \tau^2 \mathbf{g}_n$

The scheme is attractive, if the stiffness stems from f, whose evaluation is cheap while the evaluation of g is costly. Such applications arise in molecular dynamics simulations, for instance.

7. Numerical examples. In our last section we illustrate our theoretical findings on LFC schemes by numerical examples. It turns out that already the most simple examples show the lack of (uniform) stability for general starting values or for the unstabilized case $\nu_p = 1$. All implementations have been performed in Python. The codes will be made available by the authors upon request.

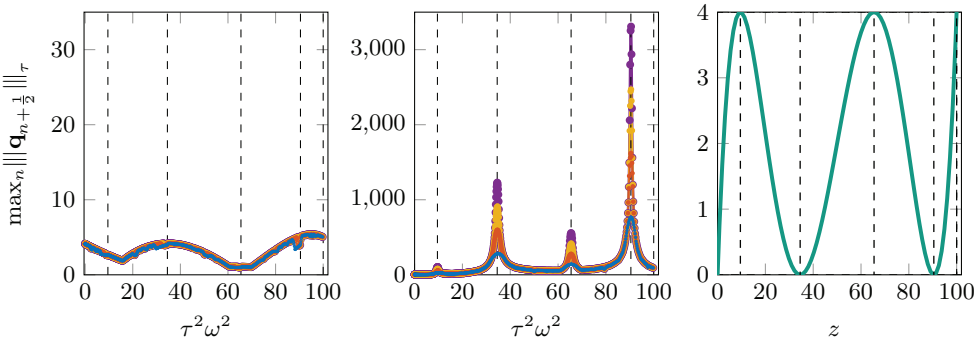
7.1. Harmonic oscillator. We consider the harmonic oscillator

$$\ddot{q}(t) = -\omega^2 q(t), \qquad q(0) = q_0, \qquad \dot{q}(0) = \dot{q}_0, \tag{7.1}$$

where $\omega > 0$ is a fixed frequency. Recall that the solution q satisfies (3.3) and, in particular, it preserves the energy.

Now, we examine the LFC method (2.3a), (2.3c) with the standard starting values (2.1b) obtained from the Taylor expansion and the new ones we proposed in (2.3b).

In Figure 7.1 we present the results for $\omega^2 = 4$, $q_0 = 2$, and $\dot{q}_0 = 1$. We used the fifth-order polynomial P_5 in (2.3c) without stabilization ($\nu_p = 1$) and employ a range of time step sizes τ with $0 \le \tau^2\omega^2 \le \beta_5^2 = 100$. In Figure 7.1(a) we depict the discrete energy norm of the approximations q_n obtained with starting value (2.3b). For this choice the energy norm stays bounded independent of the simulation time. In contrary, for the standard choice (2.1b) illustrated in Figure 7.1(b) we observe resonance effects appearing at $z = \tau^2\omega^2$ where $P_p(z) = 4$ or $P_p(z) = 0$. This fits perfectly with our



(a) Starting values (2.3b). (b) Starting values (2.1b). (c) Polynomial $P_5(z)$.

FIG. 7.1. *Time integration of the harmonic oscillator with the LFC recursion* (2.3a), (2.3c) *and different starting values* q_1. *We ran the simulation for* $N = 5, 10, 15$, *and* 20 *time steps.*

analysis because these values force $\widehat{\beta}^2 < \pi^2 \ll \beta_5^2$ in (3.21) (because of $m_1, m_2 > 0$) to obtain the stability result in Theorem 3.11.

7.2. Wave equation. Next, we consider the homogeneous wave equation with homogeneous Dirichlet boundary conditions in the unit square $\Omega = (0,1)^2$,

$$\begin{aligned} &\ddot{q}(t,x,y) = \Delta q(t,x,y) - g\big(q(t,x,y)\big), && (x,y) \in \Omega,\ t \in [0,T], \\ &q(t,x,y) = 0, && (x,y) \in \partial\Omega,\ t \in [0,T], \\ &q(0,x,y) = q_0(x,y), \quad \dot{q}(0,x,y) = \dot{q}_0(x,y), && (x,y) \in \Omega. \end{aligned} \tag{7.2a}$$

As initial data we choose

$$q_0(x,y) = \sin(\pi x)\sin(\pi y), \qquad \dot{q}_0(x,y) = \sqrt{2\pi^2 + \gamma}\,\sin(\pi x)\sin(\pi y) \tag{7.2b}$$

with a parameter $\gamma \geq 0$.

We discretize (7.2) with a symmetric interior penalty discontinuous Galerkin method [3], [5, Chapter 4] using piecewise polynomials of degree three on an unstructured mesh with 312 triangles with smallest and largest diameter 0.0301 and 0.0744, respectively. This results in the following system of ODEs:

$$\mathbf{M}\ddot{\mathbf{q}}(t) = -\mathbf{A}\mathbf{q}(t) - \mathbf{M}g\big(\mathbf{q}(t)\big), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0. \tag{7.3}$$

The block-diagonal mass matrix $\mathbf{M}$ and the stiffness matrix $\mathbf{A}$ are symmetric (w.r.t. the standard Euclidean inner product) and positive definite. The boundary condition in (7.2a) is enforced through $\mathbf{A}$.

Because the mass matrix is block diagonal it can be inverted at low costs. Thus, (7.2) can be written in the form (1.1) with $\mathbf{L} = -\mathbf{M}^{-1}\mathbf{A}$. Note that $\mathbf{L}$ is symmetric w.r.t. the inner product $\big(\mathbf{q}, \widehat{\mathbf{q}}\big) = \mathbf{q}^T\mathbf{M}\widehat{\mathbf{q}}$.

In the following we distinguish two cases for the semilinearity g:

$$g(q) = \gamma q \qquad \text{or} \qquad g(q) = \sin(q). \tag{7.4}$$

In the first case, the solution of (7.2) is given by

$$q(t,x,y) = \sin(\pi x)\sin(\pi y)\big(\cos(t\sqrt{2\pi^2+\gamma}\,) + \sin(t\sqrt{2\pi^2+\gamma}\,)\big). \tag{7.5}$$

In the second case we obtain the sine-Gordon equation. Here, we set $\gamma = 0$ in (7.2b).

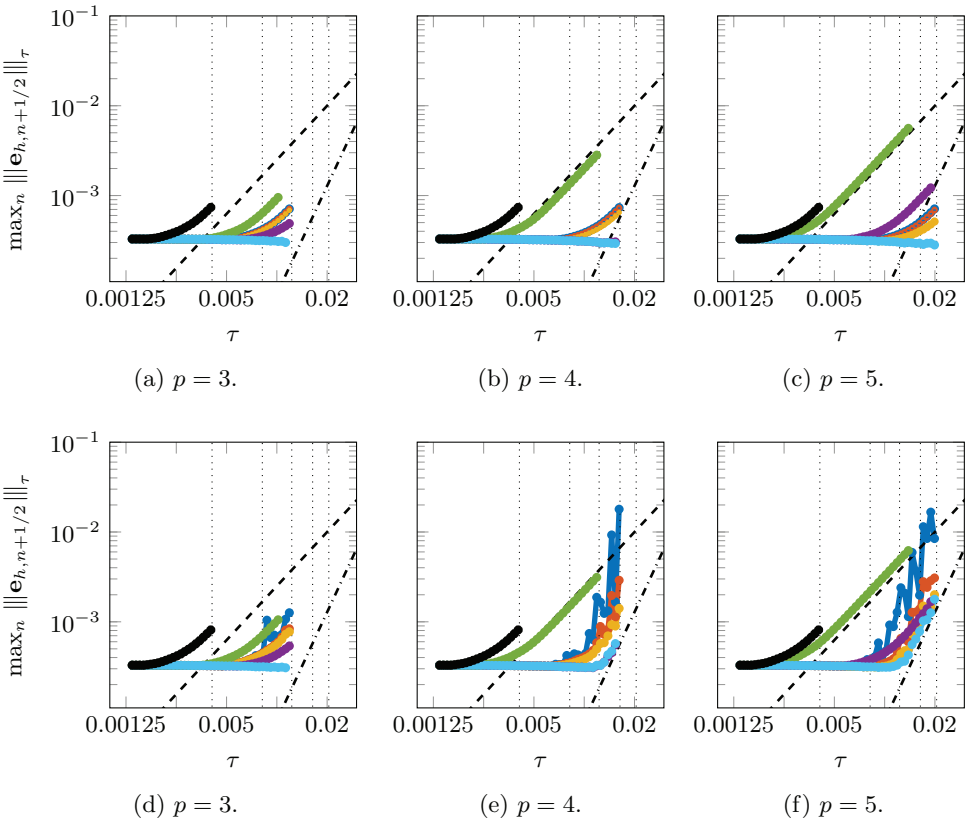
7.2.1. Linear wave equation. For $g(\mathbf{q}) = \gamma\mathbf{q}$ we integrate (7.3) with the LFC method (2.3) until the final time $T = 4.2$ and consider the error

$$\mathbf{e}_{h,n} = \mathbf{q}_h(t_n) - \mathbf{q}_n \tag{7.6}$$

between the $L^2(\Omega)$-orthogonal projection $\mathbf{q}_h(t)$ of the exact solution (7.5) onto the discontinuous Galerkin space and the approximation $\mathbf{q}_n$ of the LFC scheme. We distinguish between $\gamma = 0$ and $\gamma > 0$.

Wave equation with $\gamma = 0$. We are in the situation of section 4.2 and in particular of Theorems 4.5 and 4.7. To show the validity of these elaborations we plot in Figures 7.2(a)–(c) the error (7.6) of the LFC method for polynomial degrees $p = 3, 4, 5$ and different choices of the stabilization parameter ν_p.

We observe that the LFC method allows us to choose an approximately p times larger time step size compared to the LF method (see the dashed lines which mark integer multiples of the maximum stable time step size of the LF method). If we use



(a) $p = 3$. (b) $p = 4$. (c) $p = 5$.

(d) $p = 3$. (e) $p = 4$. (f) $p = 5$.

FIG. 7.2. *Error of the LFC method* (2.3a), (2.3c) *applied to* (7.3) *with* $g = 0$ *plotted over time step sizes* τ *for starting value* (2.3b) *(top) and starting value* (7.7) *(bottom) for polynomial degree* $p = 3, 4, 5$. *For* ν_p *we used the following choices:* $\nu_p = 1$, $\nu_p = 1.0001$, $\nu_p = 1.001$, $\nu_p = 1.01$, $\nu_p = 1.1$, ν_p *from Remark* 5.3 *(fourth-order scheme). The solid black line stems from the LF method. The black dashed lines indicate orders two and four. The dotted lines correspond to integer multiples of the maximum stable time step size* τ_{LF} *of the LF method, i.e.,* $m\tau_{\mathrm{LF}}$ *with* $m = 1, \dots, 5$.

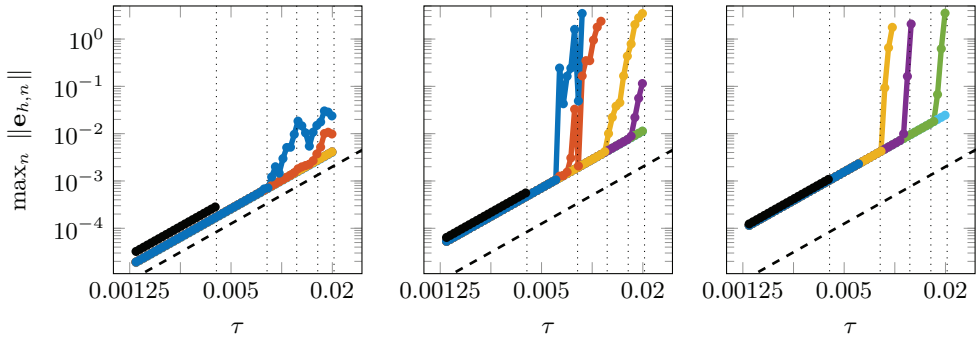
more stabilization the maximum stable step size gets smaller, since $\widehat{\beta}_p^2$ is a monotonically decreasing function of the stabilization parameter ν_p; see also Figure 5.2. Moreover, one can clearly see the effects of the value of ν_p on the error constant. In particular, we observe that a choice of ν_p near the value which gives a fourth-order scheme (see (4.14) and Remark 5.3) yields a remarkably better error constant compared to the LF method and consequently clearly smaller errors.

We can confirm the second-order convergence rate of the general LFC method. However, the fourth order achieved via (4.14) is not visible in this example since the time discretization error is so small that it is already dominated by the space discretization error.

As a comparison we give in Figures 7.2(d)–7.2(f) the error of the LFC recursion (2.3a), (2.3c) supplemented with the standard fifth-order starting value

$$\mathbf{q}_1 = \mathbf{q}_0 + \tau\dot{\mathbf{q}}_0 - \tfrac{1}{2}\tau^2\mathbf{L}\mathbf{q}_0 - \tfrac{1}{6}\tau^3\mathbf{L}\dot{\mathbf{q}}_0 + \tfrac{1}{24}\tau^4\mathbf{L}^2\mathbf{q}_0. \tag{7.7}$$

We clearly see larger errors compared to the LFC method (2.3). In particular, the unstabilized case $\nu_p = 1$ suffers from stability problems. However, with enough stabi-



(a) $\gamma = 10$. (b) $\gamma = 25$. (c) $\gamma = 50$.

FIG. 7.3. *Error of the LFC method* (2.3) *applied to* (7.3) *with* $g(\mathbf{q}) = \gamma\mathbf{q}$ *plotted over time step sizes* τ *for different values of* γ. *We used polynomial degree* $p = 5$ *and stabilization parameters* $\nu_p = 1$, $\nu_p = 1.00001$, $\nu_p = 1.00005$, $\nu_p = 1.0001$, $\nu_p = 1.00025$, $\nu_p \approx 1.003233$ *(fourth-order scheme). The solid black line stems from the LF method and the black dashed line indicates order two. The dotted lines correspond to integer multiples of the maximum stable time step size* τ_{LF} *of the LF method, i.e.,* $m\tau_{\mathrm{LF}}$ *with* $m = 1, \ldots, 5$.

lization this can be controlled and we even can confirm the fourth-order convergence rate achieved by the choice (4.14) of ν_p.

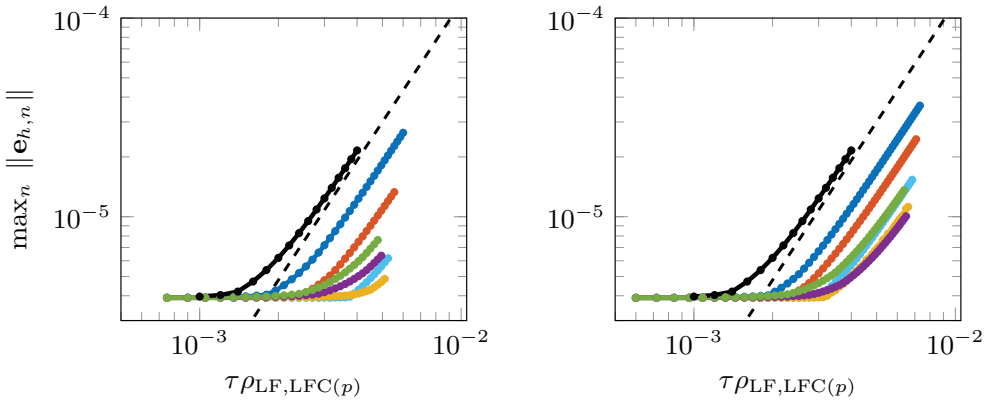
Wave equation with $\gamma > 0$. We consider the case $\gamma > 0$ as a model problem for the linear problem (3.32) to show the effects of the multirate LFC method (2.3) discussed in section 3.4. In Figure 7.3 we plot the error (7.6) for $d = 10, 25, 50$, polynomial degree $p = 5$, and different values of ν_p. As stated in Theorem 3.16 and Remark 3.17 we observe that without enough stabilization the LFC method cannot achieve a p times larger time step size than the LF scheme. The larger d is the more stabilization we have to use. However, if ν_p is sufficiently large we observe an almost p times larger maximum stable time step size and second-order convergence in accordance with Theorem 4.3.

7.2.2. Sine-Gordon equation. Finally, we apply the LFC method (2.3) to (7.3) with $g(\mathbf{q}) = \sin(\mathbf{q})$. As the final time we set $T = 1$. To approximate the error we compute a reference solution on a twice refined mesh with the LF scheme with step size $\tau = 0.0002$. In Figure 7.4 we plot the errors of the LFC method (2.3) for polynomial degree $p = 2, 3$ and for different values of ν_p. Again, we observe second-order convergence confirming our theoretical findings. Moreover, we see that the errors strongly depend on the choice of the stabilization parameter ν_p.

To compare the efficiency of the schemes, let r be the ratio of the costs of the evaluation of g and the matrix-vector multiplication with $\mathbf{L}$. Since one time step with the LF scheme requires one multiplication with $\mathbf{L}$ while the LFC scheme requires p such multiplications, and both schemes require only one evaluation of g, the ratio of the computational efforts is given by

$$\rho_{\mathrm{LF,LFC}(p)} = \frac{1+r}{p+r}. \tag{7.8}$$

It is beyond the scope of this paper to present advanced applications, e.g., from molecular dynamics. Nevertheless, to get an idea of the potential of the LFC scheme, we show in Figure 7.4 errors plotted over scaled step sizes $\tau\rho_{\mathrm{LF,LFC}(p)}$ for $r = 2$, i.e.,



(a) $p = 2$. (b) $p = 3$.

FIG. 7.4. *Error of the LFC method* (2.3) *applied to* (7.3) *with* $g(\mathbf{q}) = \sin(\mathbf{q})$ *plotted over scaled time step sizes* $\tau\rho_{LF,LFC(p)}$ *for polynomial degree* 2 *and* 3. *We used stabilization parameters depending on* p *as* $\nu_p = \nu_{\mathrm{opt}} + \kappa(\nu_{\mathrm{opt}} - 1)$ *with* $\kappa = -1$ *(*$\nu_p = 1$*),* $\kappa = -0.5$*,* $\kappa = 0$ *(*$\nu_p = \nu_{\mathrm{opt}}$*),* $\kappa = 0.5$*,* $\kappa = 1$*,* $\kappa = 1.5$ *and* ν_{opt} *chosen such that the linear scheme is of order four. The scaling factor* $\rho_{LF,LFC(p)}$ *is given as in* (7.8) *with* $r = 2$. *The solid black line stems from the LF method (with* $\rho_{LF,LFC(1)} = 1$*) and the black dashed line indicates order two.*

the evaluation of g is only twice as expensive than the multiplication with $\mathbf{L}$. One clearly observes that even in this case the LFC schemes outperform the LF schemes. It also shows that a reasonable choice of the stabilization parameter is $\nu_p \approx \nu_{\mathrm{opt}}$.

Acknowledgments. We thank Sébastien Imperiale for interesting conversations on using Chebyshev polynomials in local time-stepping methods and our colleagues Benjamin Dörich and Jan Leibold for their careful reading of this manuscript. Moreover, we thank an anonymous referee for suggesting a variant of Lemma 3.8 which led to smaller bounds in Theorems 3.7 and 4.3.

REFERENCES

[1] A. Abdulle, *Fourth order Chebyshev methods with recurrence relation*, SIAM J. Sci. Comput., 23 (2002), pp. 2041–2054, https://doi.org/10.1137/S1064827500379549.

[2] A. Abdulle and A. A. Medovikov, *Second order Chebyshev methods based on orthogonal polynomials*, Numer. Math., 90 (2001), pp. 1–18, https://doi.org/10.1007/s002110100292.

[3] 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.

[4] J. Chabassier and S. Imperiale, *Space/time convergence analysis of a class of conservative schemes for linear wave equations*, C. R. Math. Acad. Sci. Paris, 355 (2017), pp. 282–289, https://doi.org/10.1016/j.crma.2016.12.009.

[5] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl. (Berlin) 69, Springer, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-22980-0.

[6] 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.

[7] J. C. Gilbert and P. Joly, *Higher order time stepping for second order hyperbolic problems and optimal CFL conditions*, in Partial Differential Equations, Comput. Methods Appl. Sci. 16, Springer, Dordrecht, The Netherlands, 2008, pp. 67–93, https://doi.org/10.1007/978-1-4020-8758-5_4.

[8] N. K. Govil and R. N. Mohapatra, *Markov and Bernstein type inequalities for polynomials*, J. Inequal. Appl., 3 (1999), pp. 349–387, https://www.emis.de/journals/HOA/JIA/Volume3_4/156027.pdf.

[9] M. J. Grote, M. Mehlin, and S. A. Sauter, *Convergence analysis of energy conserving explicit local time-stepping methods for the wave equation*, SIAM J. Numer. Anal., 56 (2018), pp. 994–1021, https://doi.org/10.1137/17M1121925.

[10] 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.

[11] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration*, Springer Ser. Comput. Math. 31, 2nd ed., Springer, Berlin, 2006, https://doi.org/10.1007/3-540-30666-8.

[12] M. Hochbruck and C. Lubich, *A Gautschi-type method for oscillatory second-order differential equations*, Numer. Math., 83 (1999), pp. 403–426, https://doi.org/10.1007/s002110050456.

[13] W. H. Hundsdorfer and J. G. Verwer, *Numerical Solution of Time Dependent Advection Diffusion Reaction Equations*, Springer Ser. Comput. Math. 33, Springer, Berlin, 2007, https://doi.org/10.1007/978-3-662-09017-6.

[14] R. Jeltsch and O. Nevanlinna, *Stability of explicit time discretizations for solving initial value problems*, Numer. Math., 37 (1981), pp. 61–91, https://doi.org/10.1007/BF01396187.

[15] P. Joly and J. Rodríguez, *Optimized higher order time discretization of second order hyperbolic problems: Construction and numerical study*, J. Comput. Appl. Math., 234 (2010), pp. 1953–1961, https://doi.org/10.1016/j.cam.2009.08.046.

[16] A. A. Markov, *On a problem of D. I. Mendeleev*, Zap. Imp. Akad. Nauk, 62 (1889), pp. 1–24.

[17] V. A. Markov, *Über Polynome, die in einem gegebenen Intervalle möglichst wenig von null abweichen*, Math. Ann., 77 (1916), pp. 213–258.

[18] G. R. Shubin and J. B. Bell, *A modified equation approach to constructing fourth order methods for acoustic wave propagation*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. 135–151, https://doi.org/10.1137/0908026.

[19] P. J. van der Houwen and B. P. Sommeijer, *On the internal stability of explicit, m-stage Runge–Kutta methods for large m-values*, ZAMM Z. Angew. Math. Mech., 60 (1980), pp. 479–485, https://doi.org/10.1002/zamm.19800601005.

[20] J. G. Verwer, *A note on a Runge-Kutta-Chebyshev method*, ZAMM Z. Angew. Math. Mech., 62 (1982), pp. 561–563, https://doi.org/10.1002/zamm.19820621008.

[21] 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.