SN Partial Differ. Equ. Appl. (2021)2:23
https://doi.org/10.1007/s42985-020-00045-9

ORIGINAL PAPER

On averaged exponential integrators for semilinear wave equations with solutions of low-regularity

Simone Buchholz[1] · Benjamin Dörich[1] · Marlis Hochbruck[1]

Received: 19 March 2020 / Accepted: 5 October 2020

Abstract
In this paper we introduce a class of second-order exponential schemes for the time integration of semilinear wave equations. They are constructed such that the established error bounds only depend on quantities obtained from a well-posedness result of a classical solution. To compensate missing regularity of the solution the proofs become considerably more involved compared to a standard error analysis. Key tools are appropriate filter functions as well as the integration-by-parts and summation-by-parts formulas. We include numerical examples to illustrate the advantage of the proposed methods.

Keywords Highly oscillatory problems · Error bounds · Order-reduction · Time-integration · Second-order evolution equations · Filter functions · Summation-by-parts formula

Mathematics Subject Classification Primary 65J08 · 65M12 · 65M15 · Secondary 35L05

1 Introduction

In this paper we are interested in solving abstract wave equations of the form

$$q''(t) = -Lq(t) + G(t, q(t)), \quad t \in [0, t_{\text{end}}], \quad q(0) = q_0, \quad q'(0) = q'_0, \tag{1.1}$$

in some Hilbert space H where L is a positive, self-adjoint operator and G is a sufficiently

This article is part of the topical collection "Waves 2019 invited papers" edited by Manfred Kaltenbacher and Markus Melenk.

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

✉ Benjamin Dörich
benjamin.doerich@kit.edu

Marlis Hochbruck
marlis.hochbruck@kit.edu

[1] Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, 76149 Karlsruhe, Germany

regular nonlinearity (e.g., Fréchet-differentiable). Such equations arise in many physical models. A prominent example is the cubic wave equation

$$\partial_t^2 q(t,x) = \partial_x^2 q(t,x) + q(t,x)^3, \qquad (t,x) \in [0, t_{\text{end}}] \times I$$

posed on some interval $I \subseteq \mathbb{R}$ and equipped with appropriate initial and boundary conditions.

Our aim is to construct and investigate time integration schemes for (1.1) under physically realistic assumptions, in particular finite energy conditions. Hence, the solution will in general be of low regularity and we thus restrict ourselves to second-order schemes. Clearly, a standard time integrator (e.g., a Runge–Kutta scheme or an exponential integrator) can only be applied to an abstract evolution equation if it is unconditionally stable due to the unbounded operator L.

In the finite dimensional case ($\dim H < \infty$), unconditionally stable integrators (in the sense that L does not cause any restriction on the time step) for this equation were already considered in [8, 12, 17, 24]. Such exponential (or trigonometric) integrators were shown to be second-order convergent while only assuming a finite-energy condition. This was somewhat surprising since usually, second-order (exponential) schemes need two bounded time derivatives of the solution in the error analysis. The key ingredient are certain matrix functions that act as filters. The effect of these filters is that they remove resonances in the local error, which, in contrast to a standard error analysis, enforce cancellation effects in the global error. In fact one can prove that local and global error are of the same order if the filters are chosen appropriately.

Recently, in [2, 3] we presented a completely new technique to prove related results for *ordinary differential equations* by reformulating a trigonometric integrator as a Strang splitting applied to a modified problem. Using ideas from [20, 21], a specific representation of the local error was derived, which allowed us to separate terms of order three (which can be treated in a standard way) and the leading local error term, which is of order two only. A carefully adapted Lady Windermere's fan argument is employed to treat these terms in the global error accumulation.

In this paper we prove error bounds for different classes of exponential integrators applied to an *evolution equation* (1.1) in a unified way. More precisely, we characterize the structure of the defects and the properties of filter functions which allow second-order convergence under a finite-energy condition in different abstract frameworks (i.e., in different function spaces), which define the assumptions on L, G, and the initial data. Within this framework we can handle various boundary conditions.

We point out that our results are not restricted to globally Lipschitz continuous functions G, but also apply to locally Lipschitz ones that satisfy certain growth conditions. Our analysis thus covers the class of nonlinearities for which the existence of a classical solution can be guaranteed. In particular, this includes polynomial nonlinearities up to a certain degree which is determined by the spatial dimension and the corresponding Sobolev embeddings. In the one-dimensional case such equations with periodic boundary conditions and arbitrary high polynomial degree have been studied in [9] for nonlinearities with Lipschitz properties on a whole scale of Sobolev spaces. However, this rich structure is not available in our general framework and, in contrast to our work, well-posedness cannot be guaranteed.

Further work on exponential integration schemes for the time-integration of semilinear wave equations was conducted in [1] where a sine-Gordon equation is studied and the difficulties arise in the proper treatment of a single constant $c \to \infty$ which induces high

oscillations in time. In the paper [10] the approach from [9] was extended and in the one-dimensional case a quasilinear wave equation with periodic boundary conditions was studied. However, they assume smooth coefficients and high regularity for the analysis. Exponential splitting schemes for linear evolution equations have been analyzed in [15]. The error estimates depend on commutator bounds that are not available in our scenario. Finally, we point out that we do not address long-term behavior as in [4, 6].

The paper is organized as follows. In Sect. 2, we give an informal overview over the methods of interest, the main concepts, and the main results and also present numerical examples illustrating the main results of our work. In particular, the necessity of using averaging techniques in the regime of low-regularity is shown.

The informal overview is made rigorous in Sect. 3, where we introduce the analytic framework and a functional calculus which allows us to define the operator-valued filters and ensures well-posedness of the problem as well as of the schemes.

We further state the assumptions on the operator L, the nonlinearity G, and the initial data that on the one hand will guarantee the well-posedness of (1.1) and on the other hand allow to carry out the error analysis.

In Sect. 4 we characterize filter functions which allow to prove that the exact solution of the original problem and the solution of the averaged problem only differ up to terms of order τ^2, where $\tau > 0$ denotes the step size. Section 5 provides a characterization of numerical methods in terms of the structure of their defects, which are necessary to derive error bounds.

Finally, Sects. 6 and 7 contain our main results the error bounds for one-step and for multistep methods, respectively.

2 Informal overview of methods, concepts and results

Before we present the analytical framework necessary to formulate our results rigorously, we first give an informal overview of the methods of interest, the main concepts, and the main results. In the finite dimensional case $\dim H < \infty$ (which is not the situation of interest in this paper), all the approximations presented are well-defined and the statements valid. However, for evolution equations posed in appropriate function spaces, this is no longer true unless additional assumptions are imposed. Since some of them are rather technical, we postpone them to Sect. 3.

2.1 Problem statement: Second-order differential equation

Let L be a linear, self-adjoint, and positive-definite operator on H and $G : [0, t_{\text{end}}] \times H \to H$. We consider the differential equation

$$q''(t) = -Lq(t) + G(t, q(t)), \quad t \in [0, t_{\text{end}}], \quad q(0) = q_0, \quad q'(0) = q'_0,$$

and assume that the solution q satisfies the finite-energy condition

$$\langle Lq(t), q(t)\rangle_H + \langle q'(t), q'(t)\rangle_H \leq K^2 \qquad \text{for } t \in [0, t_{\text{end}}]\,, \tag{2.1}$$

where $\langle \cdot, \cdot \rangle_H$ denotes the inner product on H. In first-order formulation, the differential equation can be written as

$$u'(t) = Au(t) + f(t, u(t)), \qquad u = \begin{pmatrix} q \\ q' \end{pmatrix}, \tag{2.2}$$

with

$$A = \begin{pmatrix} 0 & I \\ -L & 0 \end{pmatrix}, \qquad f(t,u) = \begin{pmatrix} 0 \\ G(t,q) \end{pmatrix},$$

and inner product

$$\langle u_1, u_2 \rangle = \langle q_1, q_2 \rangle_H + \langle L^{-1} q_1', q_2' \rangle_H .$$

Obviously, A is skew-adjoint with respect to $\langle \cdot, \cdot \rangle$ and hence has a purely imaginary point spectrum.

2.2 Methods

In the following we shortly present four different types of methods to discretize equation (2.2) in time with a constant stepsize $\tau > 0$.

2.2.1 Strang splitting

The exact flows φ_τ^A and φ_τ^f of the two subproblems

$$\begin{pmatrix} t' \\ u' \end{pmatrix} = \begin{pmatrix} 1 \\ Au \end{pmatrix}, \qquad \begin{pmatrix} t' \\ u' \end{pmatrix} = \begin{pmatrix} 0 \\ f(t,u) \end{pmatrix},$$

are given explicitly by

$$\varphi_\tau^A \begin{pmatrix} t_0 \\ u_0 \end{pmatrix} = \begin{pmatrix} t_0 + \tau \\ e^{\tau A} u_0 \end{pmatrix}, \qquad \varphi_\tau^f \begin{pmatrix} t_0 \\ u_0 \end{pmatrix} = \begin{pmatrix} t_0 \\ u_0 + \tau f(t_0, u_0) \end{pmatrix}.$$

We consider the Strang splitting in the variants (A,f,A) and (f,A,f) given by

$$\begin{pmatrix} t_{n+1} \\ u_{n+1} \end{pmatrix} = \varphi_{\tau/2}^A \circ \varphi_\tau^f \circ \varphi_{\tau/2}^A \begin{pmatrix} t_n \\ u_n \end{pmatrix}, \tag{2.3a}$$

$$\begin{pmatrix} t_{n+1} \\ u_{n+1} \end{pmatrix} = \varphi_{\tau/2}^f \circ \varphi_\tau^A \circ \varphi_{\tau/2}^f \begin{pmatrix} t_n \\ u_n \end{pmatrix}, \tag{2.3b}$$

respectively. Note that the (f,A,f)variant (2.3b) is equivalent to a trigonometric integrator without filter functions, see, e.g., [13, XIII.2.2].

2.2.2 Corrected Lie Splitting

Next we consider the second-order corrected Lie splitting given by

$$u_{n+1} = e^{\tau A}\big(u_n + \tau f\big(t_{n+1/2}, u_n\big) + \frac{\tau^2}{2} r\big(t_{n+1/2}, u_n\big)\big) \tag{2.4}$$

with the correction term

$$r(t,u) := J_f(t,u) \begin{pmatrix} 0 \\ Au \end{pmatrix} - Af(t,u).$$

It is inspired by a fourth-order method of this type proposed in [22, 4.9.3 (c)]. Note that in the linear case, where $f(t,u) = Fu$, the correction term reduces to the commutator

$$r(t,u) = FAu - AFu = [F,A]u\,.$$

Hence, one can consider (2.4) as an approximation to the method

$$u_{n+1} = e^{\tau A} e^{\tau F} e^{\frac{\tau^2}{2}[F,A]} u_n\,,$$

which was considered in [26, (3.37)].

2.2.3 Exponential Runge–Kutta methods

General two-stage exponential Runge–Kutta methods are of the form

$$\begin{aligned} U_n &= e^{c_2\tau A}u_n + c_2\tau\varphi_1(c_2\tau A)f\big(t_n,u_n\big),\\ u_{n+1} &= e^{\tau A}u_n + \tau b_1(\tau A)f\big(t_n,u_n\big) + \tau b_2(\tau A)f\big(t_n+c_2\tau,U_n\big)\,, \end{aligned} \tag{2.5}$$

where $c_2 \in (0,1]$ is a given quadrature node. Recall that the φ-functions are defined as

$$\varphi_{k+1}(z) := \int_0^1 e^{(1-s)z}\frac{s^k}{k!}\,ds, \quad k \geq 0.$$

If the coefficient functions b_1, b_2 satisfy

$$b_1(z) + b_2(z) = \varphi_1(z), \qquad c_2 b_2(0) = \frac{1}{2}\,,$$

the method is second-order convergent for parabolic problems, see [18, Theorem 4.3.]. Popular choices are $c_2 = \frac{1}{2}$, $b_1 = 0$ or $c_2 = 1$, $b_2(z) = \varphi_2(z)$. All our results also apply to the symmetric, but implicit exponential Runge–Kutta scheme from [5, Example 2.1] and to ERKN methods, e.g., those considered in [28]. The necessary modifications are straightforward so that we omit the details.

2.2.4 Exponential multistep methods

The two-step exponential multistep method from [19, (2.7)]

$$\begin{aligned} u_{n+1} &= e^{\tau A}u_n + \tau\varphi_1(\tau A)f(t_n,u_n)\\ &\quad + \tau\varphi_2(\tau A)(f(t_n,u_n) - f(t_{n-1},u_{n-1})), \quad n \geq 1\,,\\ u_1 &= e^{\tau A}(u_0 + \tau f(t_0,u_0))\,, \end{aligned} \tag{2.6}$$

is derived from the variation-of-constants formula for the exact solution of (1.1) by approximating the nonlinearity f in the integral term by an interpolation polynomial using the last two approximations u_{n-1}, u_n.

In a similar manner we consider a method that was used in [7, (B 4)], namely

$$\begin{aligned} u_{n+1} &= e^{2\tau A}u_{n-1} + 2\tau e^{\tau A}f(t_n,u_n)\,,\\ u_1 &= e^{\tau A}\big(u_0 + \tau f(t_0,u_0)\big)\,. \end{aligned} \tag{2.7}$$

For $A = 0$ it reduces to an explicit Nyström method, cf. method (1.13') in [14].

2.3 Averaged differential equation

Let $\chi = \phi, \psi : i\mathbb{R} \to \mathbb{R}$ be even (i.e., $\chi(-z) = \chi(z)$) and analytic functions satisfying $\chi(0) = 1$. Then we define

$$\widetilde{\chi} = \chi(i\tau L^{1/2})$$

and an averaged nonlinearity

$$\widetilde{G}(t, q) := \widetilde{\psi} G(t, \widetilde{\phi} q)\,.$$

Using the block diagonal operators

$$\Phi = \begin{pmatrix} \widetilde{\phi} & 0 \\ 0 & \widetilde{\phi} \end{pmatrix}, \qquad \Psi = \begin{pmatrix} \widetilde{\psi} & 0 \\ 0 & \widetilde{\psi} \end{pmatrix},$$

we consider the *averaged* differential equation

$$\widetilde{u}'(t) = A\widetilde{u}(t) + \widetilde{f}(t, \widetilde{u}(t)), \qquad \widetilde{f}(t, \widetilde{u}) = \Psi f(t, \Phi\widetilde{u}) = \begin{pmatrix} 0 \\ \widetilde{G}(t, \widetilde{q}) \end{pmatrix}. \tag{2.8}$$

The averaging is done such that the solution $\widetilde{u}$ of (2.8) also satisfies a finite-energy condition (2.1) (with a modified constant $\widetilde{K}$, which is independent of τ and n, cf., Lemma 4.2 below). In Theorem 4.1, we provide sufficient conditions on ψ, ϕ such that

$$\|u(t) - \widetilde{u}(t)\| \le C\tau^2, \qquad t \in [0, t_{\text{end}}],$$

where $\|\cdot\|$ denotes the norm induced by $\langle \cdot, \cdot \rangle$.

Remark 2.1 We only consider fixed stepsizes in this paper. Introducing variable stepsizes would require the investigation of a whole family of averaged problems (2.8). In addition, our analysis below would involve a much higher technical effort.

2.4 Averaged methods

The main idea is to apply one of the numerical methods to the averaged equation (2.8) instead of the original one (2.2). Equivalently, one could modify the numerical scheme in an appropriate way using filter functions. This is illustrated in Fig. 1.

Since the solutions of (2.2) and (2.8) only differ by terms of order τ^2, one might hope for second-order accuracy if the method is of order two at least. In fact we will later see that in

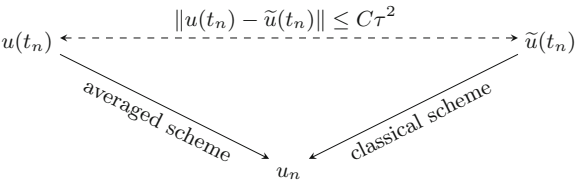

Fig. 1 Different ways to construct an approximation u_n of the solution $u(t_n)$ of the original equation (2.2) and the solution $\widetilde{u}(t_n)$ of the averaged equation (2.8)

the case of evolution equations, this intuition is *not* always justified, i.e., order reduction might appear. The main goal in this paper is to characterize the numerical methods, the assumptions on L and G, and the choice of the filter functions which lead to second-order error bounds.

2.5 Main results

Our main results, which are detailed in Theorem 6.2 for exponential one-step methods and in Sect. 7 for exponential multistep methods, are the following error bounds.

(a) The Strang splitting, the exponential Runge–Kutta, and the exponential multistep methods applied to the original equation (2.2) satisfy

$$\|u(t_n) - u_n\| \le C_1 \tau \,.$$

(b) All methods of Sect. 2.2 applied to the averaged equation (2.8) with appropriate filters ϕ, ψ satisfy

$$\|u(t_n) - u_n\| \le C_2 \tau^2 \,.$$

The constants C_1, C_2 only depend on the initial value u_0, the finite energy K, properties of G, and t_{end}, but not on n and τ.

The strategy to prove these bounds is to split the error into two terms, namely

$$\|u(t_n) - u_n\| \le \|u(t_n) - \widetilde{u}(t_n)\| + \|\widetilde{u}(t_n) - u_n\|. \tag{2.9}$$

The first term is bounded by Theorem 4.1, the second by Theorem 6.1 or Corollaries 7.1, and 7.2, respectively. A crucial step is to show that the averaged solution inherits the regularity of the original solution, which is done in Lemma 4.2.

2.6 Numerical example

In this section we illustrate the effect of averaging within numerical methods by approximating the solution of a variant of the sine-Gordon equation given on the torus $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ by

$$q''(t) = \Delta q(t) - q(t) + m_a \sin(m_i\, q(t))\, q(t), \tag{2.10}$$

with $t \in [0, 1]$ and $m_i, m_a \in L^\infty(\mathbb{T})$. Note that for $q \in L^2(\mathbb{T})$ and

$$G(q)(x) := m_a(x) \sin(m_i(x)\, q)\, q \,,$$

we have $G(q)$ in $L^2(\mathbb{T})$, but even for $q \in H^1(\mathbb{T})$ we cannot expect $G(q) \in H^\epsilon(\mathbb{T})$ for any $\epsilon > 0$. Hence, the analysis of [9, 10] does not apply to such non-smooth nonlinearities. For the spatial discretization, we used a Fourier spectral method in order to control the regularity of the solution. The initial values $(q_0, v_0) \in H^1(\mathbb{T}) \times L^2(\mathbb{T})$ are constructed such that

$$(q_0, v_0) \in H^1(\mathbb{T}) \times L^2(\mathbb{T}) \setminus H^{1+\epsilon}(\mathbb{T}) \times H^\epsilon(\mathbb{T})$$

for $\epsilon = 10^{-6}$, see [16] for details.

In Fig. 2 we computed the approximate solution with the Strang splitting variant (2.3a),

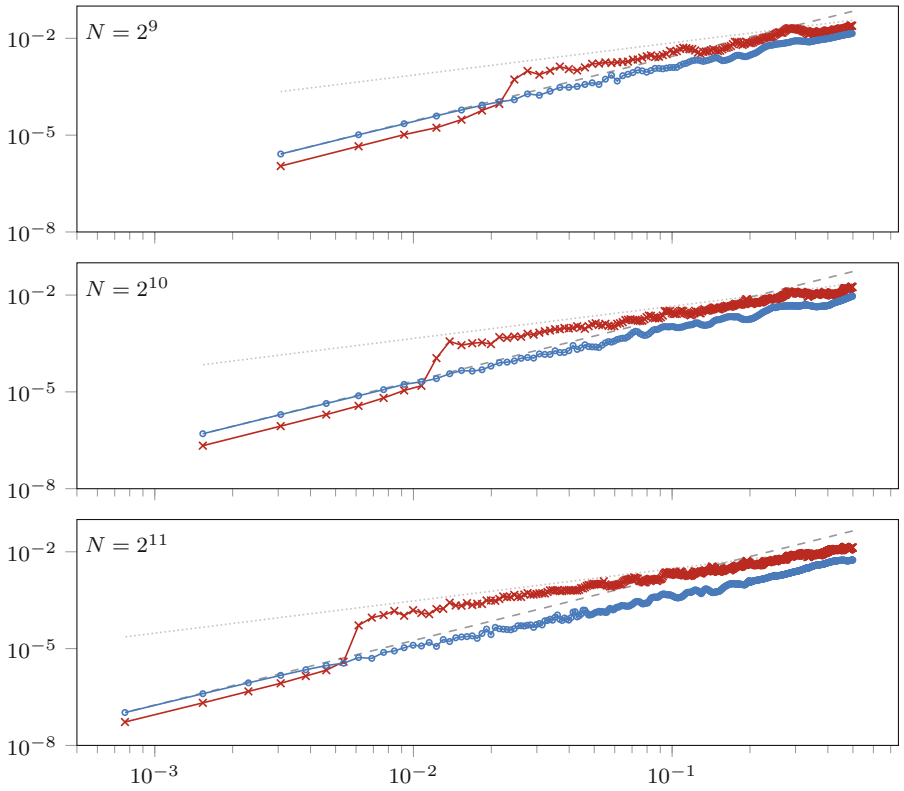

Fig. 2 Discrete $L^\infty\Big([0,1], L^2(\mathbb{T}) \times H^{-1}(\mathbb{T})\Big)$ error (on the y-axis) of the numerical solution of (2.10) plotted against the step size τ (on the x-axis) with N grid points. The gray lines indicate order one (dotted) and two (dashed)

i.e., $\big(A, \widetilde{f}, A\big)$, with filters (blue, dots)

$$\phi(z) = \psi(z) = \operatorname{sinhc}(\tfrac{z}{2}) = \frac{1}{2}\Big(\varphi_1(\tfrac{z}{2}) + \varphi_1(-\tfrac{z}{2})\Big) \tag{2.11}$$

and without filters, i.e., $\phi = \psi = 1$, (red, crosses) with $N = 2^j$, $j = 9, 10, 11$, spatial grid points. The codes are available from the authors on request. We observe order reduction of the non-averaged scheme to order one in the stiff regime, while in the non-stiff regime, the two errors of both schemes are quite close. The non-stiff regime is characterized by time steps τ for which $\varphi_1(\tau A)$ is invertible for all $\tau < \tau_0$. Since $\|A\| \approx N/2$, this is true for $\tau_0 \approx 4\pi/N$. For abstract evolution equations, only the stiff regime is relevant, i.e., the limit $N \to \infty$.

3 Analytical framework

We fix some notation for the rest of the paper. For Hilbert spaces X, Y, $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_X$ denotes the scalar product on X and $\mathcal{B}(X, Y)$ the set of all bounded operators $T : X \to Y$ equipped with the standard operator norm $\|T\|_{Y \leftarrow X}$. Further, $C^k(X, Y)$ is the space of all

k-times Fréchet-differentiable functions from X to Y. We write $W^{k,p}(\Omega)$, $k \in \mathbb{N}_0$, $1 \le p \le \infty$, for the Sobolev space of order k with all (weak) derivatives in $L^p(\Omega)$ and abbreviate $H^k(\Omega) := W^{k,2}(\Omega)$. For multi-indices $\alpha, \beta \in \mathbb{N}^\ell$ we write $\alpha \le \beta$ if $\alpha_i \le \beta_i$ for all $i = 1, \dots, \ell$.

3.1 Second-order equation

Let H be a real, separable Hilbert space and $L : \mathcal{D}(L) \subseteq H \to H$ be a positive, self-adjoint operator with compact resolvent. We consider the abstract second-order evolution equation (1.1) in H. To reformulate it as a first-order system we use the intermediate space $V = \mathcal{D}(L^{1/2})$ with

$$\mathcal{D}(L) \hookrightarrow V \hookrightarrow H, \qquad \|v\|_V = \left\|L^{1/2} v\right\|_H,$$

with dense and compact embeddings, in particular, there is a constant C_{emb} such that

$$\|v\|_H \le C_{\mathrm{emb}} \|v\|_V, \ v \in V, \qquad \|q\|_V \le C_{\mathrm{emb}} \|q\|_{\mathcal{D}(L)}, \ q \in \mathcal{D}(L)\,. \tag{3.1}$$

We exemplify the abstract framework considered in the rest of the paper by a class of semilinear wave equations.

Example 3.1 We consider the semilinear evolution equation (1.1) in the following setting:

(a) $\emptyset \ne \Omega \subseteq \mathbb{R}^d$ is a convex, bounded Lipschitz domain with $d \in \{1, 2, 3\}$.
(b) $L = -\operatorname{div}(\mathbf{A}\nabla)$ with uniformly positive definite $\mathbf{A} \in L^\infty(\Omega)^{d \times d}$.
(c) For $g : [0, t_{\mathrm{end}}] \times \Omega \times \mathbb{R} \to \mathbb{R}$ there is some $\alpha = (\alpha_t, \alpha_x, \alpha_y) \in \mathbb{N}^3$ such that all partial derivatives $\partial^\beta g$, $\beta \le \alpha$, exist, are continuous in t and y and bounded in x.
(d) There is $\gamma > 1$ and a constant $C_{\mathrm{g}} > 0$ such that for all $(t, x, y) \in [0, t_{\mathrm{end}}] \times \Omega \times \mathbb{R}$ we have

$$\begin{aligned} |g(t,x,y)|, |\partial_t g(t,x,y)| &\le C_{\mathrm{g}} \left(1 + |y|^\gamma\right), \\ |\partial_y g(t,x,y)| &\le C_{\mathrm{g}} \left(1 + |y|^{\gamma-1}\right). \end{aligned} \tag{3.2a}$$

For the corrected Lie Splitting (2.4) we assume in addition

$$|\partial_{yy} g(t,x,y)| \le C_{\mathrm{g}} \left(1 + |y|^{\gamma-1}\right). \tag{3.2b}$$

For $(t, x) \in [0 t_{\mathrm{end}}] \times \Omega$ and $q \in V$ we define

$$G(t,q)(x) := g(t, x, q(x)).$$

In Table 1 at the end of this section, details for different choices of H are presented for these problems.

In the following we recall sufficient conditions on the nonlinearity G to guarantee well-posedness of the equation and to establish the error analysis presented in Sects. 4, 5, 6, and 7.

Assumption 3.2 (Well-posedness) For G we have $G \in C^1([0, t_{\mathrm{end}}] \times V, H)$, i.e., G is Fréchet-differentiable with Fréchet-derivative $J_G(t, q) \in \mathcal{B}\left([0, t_{\mathrm{end}}] \times V, H\right)$ for all $q \in V, t \in [0, t_{\mathrm{end}}]$.

Table 1 Overview on examples

H d	$H^{-1}(\Omega)$ $d=1$	$L^2(\Omega)$ $d=1,2,3$	$H_0^1(\Omega)$ $d=1,2,3$
$\mathbf{A}$	–	$W^{1,\infty}(\Omega)^{d\times d}$	$C^{1,1}(\Omega)^{d\times d}\cap W^{2,\infty}(\Omega)^{d\times d}$ or $H^4(\Omega)^{d\times d}$
Ω	–	–	$\partial\Omega$ of class C^3
$\mathcal{D}(L)$	$H_0^1(\Omega)$	$H^2(\Omega)\cap H_0^1(\Omega)$	$\{q\in H^3(\Omega)\cap H_0^1(\Omega)\mid Lq\in H_0^1(\Omega)\}$
V	$L^2(\Omega)$	$H_0^1(\Omega)$	$H^2(\Omega)\cap H_0^1(\Omega)$
α	$(2,0,2)$	$(2,1,2)$	$(3,2,3)$
g	–	$g(t,\cdot,0)=0$ on $\partial\Omega$	$g(t,\cdot,0)=0$ on $\partial\Omega$
Growth bound	$\gamma\le 2$	$\gamma\begin{cases}<\infty & d=1,2\\ \le 3 & d=3\end{cases}$	–

The most subtle assumption is given now. It states the necessary regularity for G evaluated at a sufficiently smooth function.

Assumption 3.3 (Regularity of G evaluated at a smooth function) For $q\in C^1([0,t_{\text{end}}],V)\cap C([0,t_{\text{end}}],\mathcal{D}(L))$ we have

$$\begin{aligned} &t\mapsto G(t,q(t))\in C^1([0,t_{\text{end}}],V) \text{ with } \frac{d}{dt}G(t,q(t))=J_G(t,q(t))\begin{pmatrix}1\\ q'(t)\end{pmatrix},\\ &t\mapsto J_G(t,q(t))\in C^1([0,t_{\text{end}}],\mathcal{B}([0,t_{\text{end}}]\times V,H)) \text{ with } C>0 \text{ such that}\end{aligned} \tag{A1}$$

$$\left\|\frac{d}{dt}J_G(t,q(t))\right\|_{H\leftarrow[0,t_{\text{end}}]\times V}\le C,\quad C=C\Big(\|q(t)\|_{\mathcal{D}(L)},\|q'(t)\|_V\Big) \tag{A2}$$

The next assumption states bounds of G and J_G. We point out that the dependency of the constants arising from different radii is crucial for the error analysis.

Assumption 3.4 (Regularity of G) There are constants $C=C(r)$ such that for given $r_V,r_L>0$ and q with $\|q\|_V\le r_V$, $\|q\|_{\mathcal{D}(L)}\le r_L$, $p\in V$, and $t\in[0,t_{\text{end}}]$ the following inequalities are satisfied:

$$\|G(t,q)\|_V\le C(r_L), \tag{A3}$$

$$\left\|J_G(t,q)\begin{pmatrix}s\\ p\end{pmatrix}\right\|_H\le C(r_V)\big(|s|+\|p\|_V\big), \tag{A4a}$$

$$\left\|J_G(t,q)\begin{pmatrix}s\\ p\end{pmatrix}\right\|_V\le C(r_L)\big(|s|+\|p\|_V\big). \tag{A4b}$$

For the corrected Lie Splitting (2.4) we assume in addition for $\|p_i\|_V\le r_V$, $i=1,2$,

$$\left\|(J_G(t,p_1 - J_G(t,p_2)\begin{pmatrix}0\\q\end{pmatrix}\right\|_H \leq C(r_L,r_V)\|p_1 - p_2\|_V.$$

Remark 3.5 Note that shifting G to $G + cI$ for some $c \in \mathbb{R}$ does not affect the validity of Assumptions 3.2 to 3.4. Hence, we can also treat positive semidefinite operators L by applying a shift.

In Table 1 the main example 3.1 is specified more precisely. We collected three examples where we stated for a given Hilbert space H the dimension d of the domain Ω and additional assumptions on the data such that Assumptions 3.2 to 3.4 are satisfied. All examples are posed with homogeneous Dirichlet boundary conditions. By possibly shifting L, we can also treat Neumann, Robin, or periodic boundary conditions.

Higher order Sobolev spaces $H = H^k(\Omega)$, $k \geq 2$, can be handled as well but the spaces and conditions for the operators and parameters become more complicated.

Remark 3.6 Note that from Assumption 3.2 and the chain rule one can only conclude that

$$t \mapsto G(t,q(t)) \in C^1([0,t_{\text{end}}],H)$$

instead of (A1).

(a) In Example 3.1 the additional regularity $q \in C([0,t_{\text{end}}],\mathcal{D}(L))$ is sufficient to verify the Assumption (A1).

(b) For $G \in C^1([0,t_{\text{end}}] \times V,V)$ the chain rule immediately yields Assumption (A1). However, in Example 3.1 with $H = H^{-1}(\Omega)$ and $V = L^2(\Omega)$, this would imply that G is already affine-linear, see [11, Sect. 3]. Hence, not even the function $q \mapsto \sin(q)$ would be covered by the analysis.

3.2 First-order equation

We consider the first-order formulation (2.2) of equation (1.1) on the separable Hilbert space $X = V \times H$. The skew-adjoint operator A is given on its domain $\mathcal{D}(A) = \mathcal{D}(L) \times V$. Hence, A is the generator of a unitary group $(e^{tA})_{t\in\mathbb{R}}$. We call u a classical solution of (2.2) on $[0,t^*)$ if u solves (2.2), $u(0) = u_0$, and

$$u \in C^1([0,t_{\text{end}}],X) \cap C([0,t_{\text{end}}],\mathcal{D}(A)) \tag{3.3}$$

for any $t_{\text{end}} < t^*$. The Assumptions 3.2 to 3.4 are translated into this setting by means of the following three lemmas. The first one provides a classical solution of (2.2) by standard semigroup theory. All statements in the lemmas directly follow from the special structure of f and the assumptions in Sect. 3.1.

Lemma 3.7 (Well-posedness) *Let G satisfy Assumption 3.2. Then $f : [0,t_{\text{end}}] \times X \to X$ defined in (2.2) satisfies $f \in C^1([0,t_{\text{end}}] \times X,X)$ with Fréchet derivative $J_f(t,u) \in \mathcal{B}([0,t_{\text{end}}] \times X,X)$ for all $u \in X$ and $t \in [0,t_{\text{end}}]$.*

The following lemma shows differentiability of f in the stronger $\mathcal{D}(A)$ norm.

Lemma 3.8 (Regularity of f evaluated at a smooth function) *Let G satisfy Assumption 3.3 and u satisfy (3.3). Then we have*

$$t \mapsto f\big(t,u(t)\big) \in C^1([0,t_{\text{end}}],\mathcal{D}(A)) \text{ with } \frac{d}{dt} f\big(t,u(t)\big) = J_f\big(t,u(t)\big)\begin{pmatrix}1\\ u'(t)\end{pmatrix} \tag{A1$'$}$$

$$\begin{aligned} &t \mapsto J_f\big(t,u(t)\big) \in C^1([0,t_{\text{end}}],\mathcal{B}([0,t_{\text{end}}]\times X,X)) \text{ with } C>0 \text{ such that}\\ &\Big\|\frac{d}{dt}J_f\big(t,u(t)\big)\Big\|_{X\leftarrow[0,t_{\text{end}}]\times X} \le C(\|Au(t)\|,\|u'(t)\|)\,. \end{aligned} \tag{A2$'$}$$

The next Lemma contains two Lipschitz properties of f which easily follow from the corresponding bound on the derivative. They are crucial for the forthcoming error analysis.

Lemma 3.9 (Regularity of f) *Let G satisfy Assumption 3.4. Then there are constants $C=C(r)$ such that for given $r_X, r_A>0$ and u_i with $\|u_i\|\le r_X$, $\|u_i\|_{\mathcal{D}(A)}\le r_A$, $i=1,2$, $v\in X$, and $t\in[0,t_{\text{end}}]$ the following inequalities are satisfied*:

$$\|f(t,u_1)\|_{\mathcal{D}(A)} \le C(r_A), \tag{A3$'$}$$

$$\Big\|J_f(t,u_1)\begin{pmatrix}s\\ v\end{pmatrix}\Big\| \le C(r_X)(|s|+\|v\|), \tag{A4a$'$}$$

$$\Big\|J_f(t,u_1)\begin{pmatrix}s\\ v\end{pmatrix}\Big\|_{\mathcal{D}(A)} \le C(r_A)(|s|+\|v\|), \tag{A4b$'$}$$

$$\Big\|f(t,u_1)-f(t,u_2)\Big\| \le C(r_X)\|u_1-u_2\|, \tag{A5a$'$}$$

$$\|f(t,u_1)-f(t,u_2)\|_{\mathcal{D}(A)} \le C(r_A)\|u_1-u_2\|. \tag{A5b$'$}$$

For the corrected Lie Splitting (2.4) we further have for $\|v_i\|\le r_X$, $i=1,2$,

$$\Big\|\big(J_f(t,v_1)-J_f(t,v_2)\big)\begin{pmatrix}0\\ u_1\end{pmatrix}\Big\| \le C(r_A,r_X)\|v_1-v_2\|.$$

Lemma 3.7 guarantees local well-posedness of (2.2), see [23, Thm. 6.1.5]. Our error analysis only requires assumptions on the data, which then implies the following regularity of the solution.

Proposition 3.10 *Let Assumption 3.2 be satisfied and take an initial value $u_0\in\mathcal{D}(A)$. Then there exists a time $t^*>0$ and a classical solution of (2.2) on $[0,t^*)$ satisfying (3.3). Hence, for every $0<t_{\text{end}}<t^*$ there exists a constant $K>0$ with*

$$\max\{\|Au(t)\|,\|u'(t)\|\}\le K, \qquad t\in[0,t_{\text{end}}]. \tag{3.4}$$

In the following we refer to (3.4) as the generalized finite-energy condition.

Remark 3.11

(a) Note that for $u = (q, q')$ in the situation of Example 3.1 with $H = H^{-1}(\Omega)$, the generalized finite energy condition implies

$$\|Au(t)\|^2 = \|q\|^2_{\mathcal{D}(L)} + \|q'\|^2_V = \left\|\mathbf{A}^{1/2}\nabla q\right\|^2_{L^2} + \|q'\|^2_{L^2} \le K^2 .$$

This corresponds to the finite energy condition used in [8, 12, 17, 24].

(b) The bound (3.4) also implies

$$\|u'(t)\|^2 = \|q'\|^2_V + \|q''\|^2_H \le K^2 .$$

3.3 Filter

From the compact resolvent property of L and the compact embeddings we can infer that also A has a compact resolvent. Hence, A admits an orthonormal basis of eigenvectors

$$(\phi_k)_{k\in M}, \quad A\phi_k = i\lambda_k\phi_k, \quad \phi_k \in \bigcap_{j\in\mathbb{N}} \mathcal{D}(A^j) ,$$

where $M \subseteq \mathbb{N}$ and $\lambda_k \in \mathbb{R}$. Any $x \in X$ can thus be represented as

$$x = \sum_{k\in M} \alpha_k\phi_k, \quad \alpha_k = \langle x, \phi_k\rangle_X ,$$

with the equivalence

$$x \in \mathcal{D}(A) \iff \sum_{k\in M} |\lambda_k\alpha_k|^2 < \infty .$$

This enables us to define the following functional calculus on the set

$$\mathcal{C}_b(i\mathbb{R}) := \{h : i\mathbb{R} \to \mathbb{C} \mid h \text{ is continuous and } \|h\|_\infty < \infty\} ,$$

see [25, Theorem 5.9]. It leads to the following properties of operator functions.

Theorem 3.12 *Let $A : \mathcal{D}(A) \to H$ be a skew-adjoint operator on a separable Hilbert space X with compact resolvent. Then the map $\Psi_A : \mathcal{C}_b(i\mathbb{R}) \to \mathcal{L}(X)$,*

$$h \mapsto \begin{cases} h(A) : X \to X \\ x = \sum_{k\in M} \alpha_k\phi_k \mapsto h(A)x = \sum_{k\in M} h(i\lambda_k)\alpha_k\phi_k \end{cases}$$

satisfies the following properties:

(a) Ψ_A *is linear*
(b) $\|h(A)\|_{X\leftarrow X} \le \|h\|_\infty$
(c) $(gh)(A) = g(A)h(A)$
(d) *For $x \in \mathcal{D}(A)$ it holds $h(A)x \in \mathcal{D}(A)$ and $Ah(A)x = h(A)Ax$*

For the construction of the integrators we make use of filter functions.

Definition 3.13 Let $\chi \in \mathcal{C}_b(i\mathbb{R})$. We call χ a filter of order m, $m = 1, 2$, if the following properties are satisfied. There exist $\vartheta, \Theta \in \mathcal{C}_b(i\mathbb{R})$ such that for all $z \in i\mathbb{R}$

$$|\chi(z)| \leq 1\,, \tag{F1}$$

$$1 - \chi(z) = z^m \vartheta(z)\,, \tag{F2}$$

$$z\chi(z) = (e^z - 1)\Theta(z)\,. \tag{F3}$$

In addition, for $m = 2$, χ is symmetric, i.e.

$$\chi(z) = \chi(-z)\,. \tag{F4}$$

Note that (F3) is equivalent to $\chi(z) = \varphi_1(z)\Theta(z)$.

By Theorem 3.12 we can define a corresponding class of filter operators that we later use in the averaged schemes.

Theorem 3.14 *Let $\tau > 0$ and $\chi \in \mathcal{C}_b(i\mathbb{R})$ be a filter of order m with ϑ, Θ from Definition 3.13. Then we have*

Boundedness :	$\|\chi(\tau A)\|_{X\leftarrow X} \leq 1$	(OF1)
	$\|\vartheta(\tau A)\|_{X\leftarrow X} \leq \|\vartheta\|_\infty, \quad \|\Theta(\tau A)\|_{X\leftarrow X} \leq \|\Theta\|_\infty$	
Smoothing :	$\chi(\tau A) : X \to \mathcal{D}(A)$ *is continuous with*	(OF2)
	$\|\tau A\chi(\tau A)\|_{X\leftarrow X} \leq 2\,\|\Theta\|_\infty$	
Consistency :	$\vartheta(\tau A) : X \to \mathcal{D}(A^m)$,	
	$I - \chi(\tau A) = (\tau A)^m \vartheta(\tau A)$	(OF3)
Cancelation :	$(\tau A)\chi(\tau A) = (e^{\tau A} - I)\Theta(\tau A)$	(OF4)
Block structure :	*For $m = 2$ and $i \in \{1, 2\}$,*	
	$\pi_i x = 0$ implies $\pi_i \chi(\tau A)x = 0$.	(OF5)

Here, $\pi_i : X \to X$ denotes the projection onto the i-th component.

Proof The properties (OF1), (OF3), (OF4) directly follow from the functional calculus and (OF2) is a direct consequence of (OF4). To prove (OF5), we use the fact that we can approximate χ uniformly on $i\mathbb{R}$ by even rational functions as $\lim_{x\to\pm\infty} \chi(ix) = 0$, see [27, Sect. 1.6]. Hence, the assertion is true since it is easily verified for functions of the type

$$z \mapsto \frac{z^2}{z^2 - \delta}, \qquad z \mapsto \frac{1}{z^2 - \delta}$$

with some $\delta > 0$. □

Remark 3.15

(a) An example for $m = 2$ is the short average filter proposed in [8] that we used in (2.11). We note that in this example $\chi(ix) = \operatorname{sinc}(\frac{x}{2})$ holds for all $x \in \mathbb{R}$, which relates our filters to the ones considered in [13, Chapter XIII.].

(b) We further obtain $\|\tau A\vartheta(\tau A)\|^2_{X\leftarrow X} \le 2\|\vartheta\|_\infty$ for $m = 2$ as

$$|z\vartheta(z)|^2 = |z^2\vartheta(z)|\,|\vartheta(z)| \le 2\|\vartheta\|_\infty \quad \text{for all} \quad z \in i\mathbb{R}\,.$$

4 Averaged problem

In this section we bound the difference between the solution $\widetilde{u}$ of the averaged equation (2.8) and the solution u of (2.2). Note that by Proposition 3.10, a unique classical solution $\widetilde{u}$ of (2.8) exists since the assumptions on f also hold for $\widetilde{f}$.

In order to apply (A5a′) we define r_X via

$$\max_{t\in[0,t_{\text{end}}]} \|u(t)\| \le C_{\text{emb}}K =: \frac{1}{2}r_X$$

with C_{emb} defined in (3.1) and K in (2.1).

Theorem 4.1 *Let Assumptions 3.2, to 3.4 be valid and consider the averaged nonlinearity $\widetilde{f}$ defined in (2.8) with second-order filters. Then there is a $\tau_0 > 0$ and a constant $C_{\text{av}} > 0$ such that for all $\tau \le \tau_0$*

$$\|u(t) - \widetilde{u}(t)\| \le C_{\text{av}}\tau^2, \qquad 0 \le t \le t_{\text{end}}\,. \tag{4.1}$$

The constant C_{av} and τ_0 depend on r_X, u_0, t_{end}, the finite energy K defined in (2.1), the filter functions, and the embedding constant C_{emb}, but not on τ.

In particular, $\widetilde{u}$ exists on $[0, t_{\text{end}}]$ and is bounded by

$$\max_{t\in[0,t_{\text{end}}]} \|\widetilde{u}(t)\| \le \frac{3}{4}r_X\,.$$

Proof Let $\widetilde{t^*} > 0$ be the maximal existence time of $\widetilde{u}$ and define

$$t_0 := \sup\{s \in (0, \widetilde{t^*}) \mid \max_{t\in[0,s]} \|\widetilde{u}(t)\| \le r_X\}\,.$$

We first observe that for $t \le \min\{t_0, t_{\text{end}}\}$ the variation-of-constants formula yields

$$\begin{aligned} u(t) - \widetilde{u}(t) &= \int_0^t e^{(t-s)A}\Big(f\big(s,u(s)\big) - \widetilde{f}\big(s,\widetilde{u}(s)\big)\Big)\,ds \\ &= I_1(t) + I_2(t) + \int_0^t e^{(t-s)A}\Big(\widetilde{f}\big(s,u(s)\big) - \widetilde{f}\big(s,\widetilde{u}(s)\big)\Big)\,ds \end{aligned} \tag{4.2}$$

with

$$\begin{aligned} I_1(t) &= \int_0^t e^{(t-s)A}(I-\Psi)f\big(s,u(s)\big)\,ds, \\ I_2(t) &= \int_0^t e^{(t-s)A}\Psi\big(f\big(s,u(s)\big) - f\big(s,\Phi u(s)\big)\big)\,ds. \end{aligned}$$

By Assumption (A5a′) and since $t \le t_0$, the third term in (4.2) is bounded by

$$\Big\| \int_0^t e^{(t-s)A} \Big(\widetilde{f}\big(s, u(s)\big) - \widetilde{f}\big(s, \widetilde{u}(s)\big) \Big)\, ds \Big\| \le C\big(r_X\big) \int_0^t \|u(s) - \widetilde{u}(s)\|\, ds.$$

It remains to prove

$$\big\|I_j(t)\big\| \le C\tau^2, \qquad j = 1, 2, \tag{4.3}$$

since these bounds are sufficient to apply a Gronwall lemma which shows the assertion for all $t \le \min\{t_0, t_{\text{end}}\}$.

To bound I_1 we use integration-by-parts and (OF3) to obtain

$$\begin{aligned} I_1(t) &= \tau^2 \int_0^t e^{(t-s)A} A^2 \vartheta(\tau A) f\big(s, u(s)\big)\, ds \\ &= \tau^2 \Big[-e^{(t-s)A} A \vartheta(\tau A) f\big(s, u(s)\big) \Big]_0^t \\ &\quad + \tau^2 \int_0^t e^{(t-s)A} A \vartheta(\tau A) J_f\big(s, u(s)\big) \begin{pmatrix} 1 \\ u'(s) \end{pmatrix} ds, \end{aligned}$$

where we used that $f\big(s, u(s)\big)$ is differentiable in X. By Assumptions (A3′), (A4b′), and the bound (3.4) on u' we have

$$\big\|Af\big(s, u(s)\big)\big\| \le C(K), \qquad \Big\|AJ_f\big(s, u(s)\big) \begin{pmatrix} 1 \\ u'(s) \end{pmatrix}\Big\| \le C(K).$$

This proves (4.3) for $j = 1$.

Using the notation $\mathbf{u}(s, \sigma) = \sigma u(s) + (1 - \sigma)\Phi u(s)$ and the differentiability (A1′) of f we get

$$\begin{aligned} I_2(t) &= \int_0^t e^{(t-s)A} \Psi\big(f\big(s, u(s)\big) - f\big(s, \Phi u(s)\big)\big)\, ds \\ &= \int_0^t \int_0^1 e^{(t-s)A} \Psi \frac{d}{d\sigma} f\big(s, \mathbf{u}(s, \sigma)\big)\, d\sigma\, ds \\ &= \int_0^t \int_0^1 e^{(t-s)A} \Psi J_f\big(s, \mathbf{u}(s, \sigma)\big) \begin{pmatrix} 0 \\ (I - \Phi)u(s) \end{pmatrix} d\sigma\, ds \\ &= \int_0^t \int_0^1 e^{(t-s)A} \Psi J_f\big(s, \mathbf{u}(s, \sigma)\big) \begin{pmatrix} 0 \\ (I - \Phi)e^{sA} u_0 \end{pmatrix} d\sigma\, ds \\ &\quad + \int_0^t \int_0^1 e^{(t-s)A} \Psi J_f\big(s, \mathbf{u}(s, \sigma)\big) \begin{pmatrix} 0 \\ (I - \Phi) \int_0^s e^{(s-\theta)A} f\big(\theta, u(\theta)\big) \end{pmatrix} d\theta\, d\sigma\, ds \\ &= I_{2,1}(t) + I_{2,2}(t). \end{aligned}$$

By (OF3) and integration-by-parts, the first term can be rewritten as

$$
\begin{aligned}
I_{2,1}(t) =& \tau^2 \left[\int_0^1 e^{(t-s)A} \Psi J_f\big(s, \mathbf{u}(s,\sigma)\big) \begin{pmatrix} 0 \\ \vartheta(\tau A) e^{sA} A u_0 \end{pmatrix} d\sigma \right]_0^t \\
&+ \tau^2 \int_0^t \int_0^1 e^{(t-s)A} A \Psi J_f\big(s, \mathbf{u}(s,\sigma)\big) \begin{pmatrix} 0 \\ \vartheta(\tau A) e^{sA} A u_0 \end{pmatrix} d\sigma\, ds \\
&- \tau^2 \int_0^t \int_0^1 e^{(t-s)A} \Psi \frac{d}{ds} J_f\big(s, \mathbf{u}(s,\sigma)\big) \begin{pmatrix} 0 \\ \vartheta(\tau A) e^{sA} A u_0 \end{pmatrix} d\sigma\, ds\,.
\end{aligned}
$$

Hence, we have $\big\|I_{2,1}(t)\big\| \le C\tau^2$ by (A2′), (A4a′), and (A4b′).

By assumption (A1′) we also have

$$
\int_0^s e^{(s-\theta)A} f\big(\theta, u(\theta)\big) d\theta \in \mathcal{D}(A),
$$

$$
A \int_0^s e^{(s-\theta)A} f\big(\theta, u(\theta)\big) d\theta = \int_0^s e^{(s-\theta)A} A f\big(\theta, u(\theta)\big) d\theta.
$$

Again integration-by-parts and Assumptions (A1′) and (A4b′) yield the desired bound (4.3). Using (4.1) for $t \le \min\{t_0, t_{\text{end}}\}$ we obtain for $\tau \le \tau_0$

$$
\max_{s\in[0,t]} \|\widetilde{u}(s)\| \le \max_{s\in[0,t]} \|u(s)\| + C_{\text{av}}\tau^2 \le \frac{3}{4} r_X\,.
$$

This proves $t_0 \ge t_{\text{end}}$ and hence (4.1) holds on $[0, t_{\text{end}}]$ for all $\tau \le \tau_0$. □

In the next lemma we show that $\widetilde{u}$ inherits the regularity of u uniformly in τ.

Lemma 4.2 *Let Assumptions 3.2 to 3.4 be valid. Then there is a $\tau_0 > 0$ and a constant $\widehat{C_{\text{av}}} > 0$ such that for all $\tau \le \tau_0$*

$$
\|Au(t) - A\widetilde{u}(t)\| \le \widehat{C_{\text{av}}}\,\tau, \quad 0 \le t \le t_{\text{end}}. \tag{4.4}
$$

In particular, $\widetilde{u}$ satisfies the generalized finite-energy condition uniformly in $\tau \le \tau_0$, i.e.,

$$
\max\big\{\|A\widetilde{u}(t)\|, \big\|\widetilde{u}'(t)\big\|\big\} \le \widetilde{K}, 0 \le t \le t_{\text{end}}, \tag{4.5}
$$

where τ_0 and the constants $\widehat{C_{\text{av}}}$ and $\widetilde{K}$ depend on r_X, u_0, t_{end}, the finite energy K defined in (2.1), the filter functions, and the embedding constant C_{emb}, but not on τ.

Proof We proceed as in the proof of Theorem 4.1 and define t_0 by

$$
t_0 := \sup\{s \in (0, t_{\text{end}}] \mid \max_{t\in[0,s]} \|A\widetilde{u}(t)\| \le 2K\}\,.
$$

For $0 \le t \le t_0$, (4.1), (4.2), and (A5b′) imply

$$\begin{aligned}\|Au(t) - A\widetilde{u}(t)\| &= \Big\| \int_0^t Ae^{(t-s)A}\Big(f\big(s,u(s)\big) - \widetilde{f}\big(s,\widetilde{u}(s)\big)\Big)\,ds\Big\| \\ &\le \|AI_1(t)\| + \|AI_2(t)\| + C(2K)\int_0^t \|u(s) - \widetilde{u}(s)\|\,ds \\ &\le \|AI_1(t)\| + \|AI_2(t)\| + \tau^2 tC(2K)C_{\mathrm{av}}.\end{aligned}$$

With Remark 3.15, similar arguments as before yield $\mathcal{O}(\tau)$ bounds for $\|AI_1(t)\|$ and $\|AI_2(t)\|$. By possibly reducing τ_0 we obtain the result for $0 \le t \le t_{\mathrm{end}}$. This immediately implies the first bound in (4.5) and the second bound is then obtained from (2.8). □

Remark 4.3 Note that Theorem 4.1 and Lemma 4.2 remain true for $\Psi = I$ as for this choice $I_1(t) = 0$ and the proof does not require (F3). This case is of interest for methods (2.5) and (2.6).

5 Abstract assumptions on the methods

In this section we characterize the classes of methods which are covered by our error analysis.

We recall that u denotes the solution of the original problem (2.2) and $\widetilde{u}$ the solution of the averaged problem (2.8). Further, we denote the numerical flow by S_τ and the defect by δ_n, i.e., a one-step method is given by

$$u_{n+1} = S_\tau(t_n, u_n), \qquad \delta_n = S_\tau\big(t_n, \widetilde{u}(t_n)\big) - \widetilde{u}(t_{n+1}). \tag{5.1}$$

We start with an assumption on the stability of the method.

Assumption 5.1 (Stability) The method applied to (2.8) is stable in the sense that for all $v \in \mathcal{D}(A),\ w \in X,\ t \ge 0$,

$$S_\tau(t,v) - S_\tau(t,w) = e^{\tau A}(v-w) + \tau \mathcal{J}(t,v,w), \tag{5.2}$$

where $\mathcal{J} : \mathbb{R} \times \mathcal{D}(A) \times X \to X$ is bounded by

$$\|\mathcal{J}(t,v,w)\| \le C_{\mathcal{J}}\Big(\|v\|_{\mathcal{D}(A)}, \|w\|\Big)\|v-w\|, \qquad t \in [0, t_{\mathrm{end}}]. \tag{5.3}$$

Next, we consider the consistency.

Assumption 5.2 (Consistency for order one) The method applied to the original equation (2.2) satisfies Assumption 5.1 (with $\phi = \psi = 1$) and its defect (5.1) satisfies

$$\|\delta_n\| \le C\tau^2\,,$$

where $C > 0$ is independent of τ and n.

For second-order methods, our analysis requires a particular structure of the defect. Before we state this in an abstract way, we briefly motivate it. Most of the methods we consider are constructed from the variation-of-constants formula

$$\widetilde{u}(t_{n+1}) = e^{\tau A}\widetilde{u}(t_n) + \tau \int_0^1 e^{(1-s)\tau A}\widetilde{f}(t_n + \tau s, \widetilde{u}(t_n + \tau s))\, ds\,, \tag{5.4}$$

where only the integral term is approximated. Hence, this defect can be expressed as some quadrature error that contains the second derivative in s of

$$f_1(s) = \tau\widetilde{f}(t_n + \tau s, \widetilde{u}(t_n + \tau s)) \qquad \text{or} \qquad f_2(s) = e^{(1-s)\tau A} f_1(s)\,, \tag{5.5}$$

depending on the precise method. Terms of order τ^3 do not cause any difficulties. However, terms of lower order exist which, in general, require more careful treatment. From f_1 we obtain the second-order term

$$\tau^2 J_{\widetilde{f}}\big(t_n + \tau s, \widetilde{u}(t_n + \tau s)\big)\begin{pmatrix} 0 \\ (\tau A\Phi)A\widetilde{u}(t_n + \tau s) \end{pmatrix} \tag{5.6}$$

and f_2 gives in addition the term

$$\tau^2(\tau A\Psi)e^{(1-s)\tau A} A f(t_n + \tau s, \Phi\widetilde{u}(t_n + \tau s)). \tag{5.7}$$

For these terms property (OF4) needs to be used in order to carry over the local convergence order to the global error. Similar terms are obtained for the defect of the splitting scheme (2.4). Together with the integral in (5.4), equations (5.7) and (5.6) give rise to the following general structure of δ_n.

Assumption 5.3 (Structure of defects for order two) The defect δ_n defined in (5.1) of a numerical method applied to the averaged equation (2.8) is of the form

$$\delta_n = \delta_n^{(1)} + \delta_n^{(2)} + D_n \tag{5.8}$$

with $\|D_n\| \le C\tau^3$, where the constant $C > 0$ is independent of τ and n. In addition, one of the following sets of conditions is satisfied:

(a) If ϕ, ψ are filters of order 2, then there exist $w_n \in X$ and a linear map $W_n : X \to \mathcal{D}(A)$ which satisfy

$$\|w_n\| \le C, \tag{5.9a}$$

$$\|W_n\|_{X\leftarrow X} \le C, \tag{5.9b}$$

$$\|AW_n\|_{X\leftarrow X} \le C, \tag{5.9c}$$

with a constant C which is independent of τ and n such that $\delta_n^{(i)}$ can be written as

$$\delta_n^{(1)} = \tau^2\big(\tau A\Psi\big)w_n\,, \qquad \delta_n^{(2)} = \tau^2 W_n\big(\tau A\Phi\big)A\widetilde{u}(t_n)\,, \tag{5.10}$$

(b) If $\psi = 1$ and ϕ is a filter of order 2, then (5.9) and (5.10) hold with $w_n = 0$ for all n.

Remark 5.4 From (5.9) and (OF2) one can directly derive $\|\delta_n\| \le C\tau^2$. However, this would only yield a suboptimal first-order bound in the global error.

The following proposition embeds the methods presented in Sect. 2.2 in the abstract framework.

Proposition 5.5 *Let Assumptions 3.2 to 3.4 be satisfied.*

(a) The Strang splitting methods (2.3a) and (2.3b) applied to the averaged equation (2.8) satisfy Assumptions 5.1, 5.2, and 5.3 (a).
(b) The second-order variant of the Lie splitting (2.4) applied to the averaged equation (2.8) satisfies Assumptions 5.1 and 5.3 (a).
(c) The exponential Runge–Kutta method (2.5) applied to the averaged equation (2.8) satisfies Assumptions 5.1, 5.2, and 5.3 (b).

Proof Assumption 5.1 is easily verified for all schemes. We only prove part (*a*) for the Strang splitting $(A,\widetilde{f},A)$ as the statement for the $(\widetilde{f},A,\widetilde{f})$ variant and part (*c*) can be adapted from this proof. We comment on part (*b*) below.

Let $t_{n+\xi} := t_n + \tau\xi$, $\widetilde{u}_{n+\xi} := \widetilde{u}(t_{n+\xi})$, and $\widetilde{f}_{n+\xi} := \widetilde{f}(t_{n+\xi}, \widetilde{u}_{n+\xi})$. Since we can write the scheme as

$$S_\tau(t_n, \widetilde{u}_n) = e^{\tau A}\widetilde{u}_n + \tau e^{\frac{\tau}{2}A}\widetilde{f}\big(t_{n+1/2}, e^{\frac{\tau}{2}A}\widetilde{u}_n\big),$$

the defect is given by

$$\begin{aligned}\delta_n &= e^{\tau A}\widetilde{u}_n + \tau e^{\frac{\tau}{2}A}\widetilde{f}\big(t_{n+1/2}, e^{\frac{\tau}{2}A}\widetilde{u}_n\big) - \widetilde{u}_{n+1}\\ &= \tau e^{\frac{\tau}{2}A}\widetilde{f}\big(t_{n+1/2}, e^{\frac{\tau}{2}A}\widetilde{u}_n\big) - \int_0^\tau e^{(\tau-\xi)A}\widetilde{f}_{n+\xi}\,d\xi\\ &= \widehat{I}_1 + \widehat{I}_2,\end{aligned}$$

where

$$\begin{aligned}\widehat{I}_1 &= \tau e^{\frac{\tau}{2}A}\widetilde{f}_{n+1/2} - \int_0^\tau e^{(\tau-\xi)A}\widetilde{f}_{n+\xi}\,d\xi\\ \widehat{I}_2 &= \tau e^{\frac{\tau}{2}A}\big(\widetilde{f}\big(t_{n+1/2}, e^{\frac{\tau}{2}A}\widetilde{u}_n\big) - \widetilde{f}_{n+1/2}\big).\end{aligned}$$

$\widehat{I}_1$ is the quadrature error of the midpoint rule. It can be written in terms of the Peano kernel κ_2 as

$$\begin{aligned}\widehat{I}_1 &= \tau\int_0^1 \kappa_2(\xi)\frac{d^2}{d\xi^2}\Big(e^{(1-\xi)\tau A}\widetilde{f}_{n+\xi}\Big)\,d\xi\\ &= \tau^3\int_0^1 \kappa_2(\xi)e^{(1-\xi)\tau A}A^2\Psi f(t_{n+\xi}, \Phi\widetilde{u}_{n+\xi})\,d\xi\\ &\quad + \tau^3\int_0^1 \kappa_2(\xi)e^{(1-\xi)\tau A}\Psi J_f(t_{n+\xi}, \Phi\widetilde{u}_{n+\xi})\begin{pmatrix}0\\ A\Phi A\widetilde{u}_{n+\xi}\end{pmatrix}d\xi + \widehat{D}_n^{(1)}\end{aligned}$$

with $\big\|\widehat{D}_n^{(1)}\big\| \le C\tau^3$. Again using the variation-of-constants formula, (OF2), (A5a′), and (A5b′) we obtain

$$\begin{aligned}\widehat{I}_1 &= \tau^3 \int_0^1 \kappa_2(\xi) e^{(1-\xi)\tau A} A^2 \Psi f(t_{n+\xi}, \Phi e^{\xi\tau A}\widetilde{u}_n)\, d\xi \\ &\quad + \tau^3 \int_0^1 \kappa_2(\xi) e^{(1-\xi)\tau A} \Psi J_f\big(t_{n+\xi}, \Phi\widetilde{u}_{n+\xi}\big) \begin{pmatrix} 0 \\ A\Phi A e^{\xi\tau A}\widetilde{u}_n \end{pmatrix} d\xi + \widehat{D}_n \\ &=: \tau^3 A\Psi w_n + \tau^3 W_n A\Phi A\widetilde{u}_n + \widehat{D}_n\,,\end{aligned}$$

with $\|\widehat{D}_n\| \le C\tau^3$.

To bound $\widehat{I}_2$ recall that $\widetilde{f}$ only depends on the first component of $\widetilde{u}$. Using (A5a′), the variation-of-constants formula, and $\pi_1\widetilde{f}_{n+1/2} = 0$, we have

$$\begin{aligned}\Big\|\widetilde{f}\big(t_{n+1/2}, e^{\frac{\tau}{2}A}\widetilde{u}_n\big) - \widetilde{f}_{n+1/2}\Big\| &= \Big\|\widetilde{f}\big(t_{n+1/2}, \pi_1 e^{\frac{\tau}{2}A}\widetilde{u}_n\big) - \widetilde{f}\big(t_{n+1/2}, \pi_1\widetilde{u}_{n+1/2}\big)\Big\| \\ &\le C(r_X)\Big\|\pi_1\Big(e^{\frac{\tau}{2}A}\widetilde{u}_n - \widetilde{u}_{n+1/2}\Big)\Big\| \\ &= C(r_X)\Big\|\pi_1\Big(\frac{\tau}{2}\widetilde{f}_{n+1/2} - \int_0^{\tau/2} e^{(\tau/2-\xi)A}\widetilde{f}_{n+\xi}\, d\xi\Big)\Big\| \\ &\le C\tau^2\,,\end{aligned}$$

since this is just a quadrature error of the (right) rectangular rule.

The properties (5.9a) to (5.9c) follow directly from Lemma 3.9. Using the first order Peano kernel κ_1, Assumption 5.2 is verified by writing

$$\widehat{I}_1 = \tau \int_0^1 \kappa_1(\xi)\frac{d}{d\xi}\Big(e^{(1-\xi)\tau A}\widetilde{f}_{n+\xi}\Big)\, d\xi$$

as this yields $\|\widehat{I}_1\| \le C\tau^2$ for $\psi = \phi = 1$ by (A1′) and (A3′).

We briefly comment on the scheme (2.4). The defect can be written as

$$\begin{aligned}\delta_n &= e^{\tau A}\Big(\widetilde{u}_n + \tau\widetilde{f}\big(t_{n+1/2}, \widetilde{u}_n\big) + \frac{\tau^2}{2} r\big(t_{n+1/2}, \widetilde{u}_n\big)\Big) - \widetilde{u}_{n+1} \\ &= \int_0^\tau \frac{d}{d\xi}\Big(e^{\xi A}\big(\widetilde{u}_{n+1-\xi} + \xi\widetilde{f}\big(t_{n+1/2}, \widetilde{u}_{n+1-\xi}\big) + \frac{\xi^2}{2} r\big(t_{n+1/2}, \widetilde{u}_{n+1-\xi}\big)\big)\Big)\, d\xi \\ &= \int_0^\tau e^{\xi A}\Big(\widetilde{f}\big(t_{n+1/2}, \widetilde{u}_{n+1-\xi}\big) - \widetilde{f}_{n+1-\xi}\Big)\, d\xi \\ &\quad + \int_0^\tau \frac{\xi^2}{2} e^{\xi A}\Big(\frac{d}{d\xi} r\big(t_{n+1/2}, \widetilde{u}_{n+1-\xi}\big) + A r\big(t_{n+1/2}, \widetilde{u}_{n+1-\xi}\big)\Big)\, d\xi \\ &= \widehat{I}_3 + \widehat{I}_4\,.\end{aligned}$$

In the first term $\widehat{I}_3$ we add and subtract $\tau e^{\tau/2A}\widetilde{f}_{n+1/2}$ and get the quadrature error of the

midpoint rule. The term $\widehat{I}_4$ admits a similar structure as in $\widehat{I}_1$ and hence Assumption 5.3 can be verified as before. □

Remark 5.6 We note that method (2.4) applied to the original equation (2.2) does not satisfy Assumption 5.2.

6 Main result for exponential one-step methods

The following result is the last step towards our main Theorem 6.2. It states the global error of a numerical integrator applied to the averaged equation (2.8) with suitable filters satisfying our assumptions (e.g., all the methods of Sect. 2.2) is second order accurate. As before, u denotes the solution of the original problem (2.2) and $\widetilde{u}$ the solution of the averaged problem (2.8).

Theorem 6.1 (Global error of the averaged problem) *Let Assumptions 3.2 to 3.4 be fulfilled. Moreover, let $(u_n)_n$ be the numerical approximations of a scheme applied to the averaged equation (2.8) that satisfies Assumptions 5.1 and 5.3. Then there is a $\tau_0 > 0$ and a constant $C_\mathrm{e} > 0$ such that for all $\tau \le \tau_0$*

$$\|u_n - \widetilde{u}(t_n)\| \le C_\mathrm{e}\,\tau^2, \qquad 0 \le t_n = n\tau \le t_\mathrm{end}.$$

The constant C_e and τ_0 depend on u_0, t_end, the finite energy K defined in (2.1), the filter functions, and the embedding constant C_emb, but are independent of τ and n.

Proof The proof makes use of the error recursion from [12] and adapts techniques from Theorem 5.3 in [2].

Due to definition (5.1) of the defect δ_n, the global error $\widetilde{e}_n = \widetilde{u}(t_n) - u_n$ can be written as

$$\widetilde{e}_{n+1} = S_\tau(t_n, \widetilde{u}(t_n)) - S_\tau(t_n, u_n) - \delta_n.$$

By Assumption 5.1, the global error satisfies

$$\widetilde{e}_{n+1} = e^{(n+1)\tau A}\widetilde{e}_0 + \tau \sum_{j=0}^{n} e^{(n-j)\tau A}\mathcal{J}\big(t_j, \widetilde{u}(t_j), u_j\big) - \sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j. \tag{6.1}$$

The error bound follows from a discrete Gronwall lemma, once we established the bound

$$\Big\| \sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j \Big\| \le C_\delta \tau^2 \tag{6.2}$$

with a constant C_δ being independent of τ and n.

The proof is done by induction on n. For $n = 0$, the statement is obviously true. Hence we assume that for all $0 \le k \le n$ it holds

$$\|u_k\| \le r_X, \qquad \|u_k - \widetilde{u}(t_k)\| \le C_\mathrm{e}\,\tau^2, \qquad C_\mathrm{e} := C_\delta\, e^{C_\mathcal{J}(\widetilde{K}, r_X)t_\mathrm{end}}.$$

By Assumption 5.3, the defect is split into three parts, which motivates to write

$$\sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j = \widetilde{e}_{n+1}^{(1)} + \widetilde{e}_{n+1}^{(2)} + \widetilde{e}_{n+1}^{(D)},$$

where

$$\widetilde{e}_{n+1}^{(\ell)} = \sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j^{(\ell)}, \quad \ell = 1,2, \qquad \widetilde{e}_{n+1}^{(D)} = \sum_{j=0}^{n} e^{(n-j)\tau A} D_j.$$

Since $\|D_j\| \le C\tau^3$ and $n\tau \le t_{\text{end}}$ we easily see

$$\left\|\widetilde{e}_{n+1}^{(D)}\right\| = \left\|\sum_{j=0}^{n} e^{(n-j)\tau A} D_j\right\| \le C\tau^2.$$

To bound $\widetilde{e}_{n+1}^{(\ell)}$, $\ell = 1,2$, we define

$$E_n = \sum_{j=0}^{n} e^{j\tau A} \qquad \text{and} \qquad F_n = \sum_{j=0}^{n} \widetilde{u}(t_j).$$

Summation-by-parts, Assumption 5.3, and (OF4) yield

$$\begin{aligned}
\sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j^{(1)} &= E_n\delta_0^{(1)} + \sum_{j=0}^{n-1} E_{n-j-1}\big(\delta_{j+1}^{(1)} - \delta_j^{(1)}\big) \\
&= \tau^3 E_n A\Psi w_0 \ + \tau^3 \sum_{j=0}^{n-1} E_{n-j-1} A\Psi\big(w_{j+1} - w_j\big) \\
&= \tau^2 E_n (e^{\tau A} - I)\Theta_\Psi w_0 \\
&\quad + \tau^2\Big(\tau \sum_{j=0}^{n-1} E_{n-j-1}(e^{\tau A} - I)\Theta_\Psi \frac{1}{\tau}\big(w_{j+1} - w_j\big)\Big).
\end{aligned}$$

To bound $E_j(e^{\tau A} - I)$ we exploit a telescopic sum to get

$$\big\|E_j(e^{\tau A} - I)\big\| = \Big\|\sum_{k=0}^{j} e^{k\tau A}(e^{\tau A} - I)\Big\| = \big\|e^{(j+1)\tau A} - I\big\| \le 2.$$

Together with (5.9a) and (OF4) this yields (6.2) for $\delta_j^{(1)}$ instead of δ_j.

Next we consider $\widetilde{e}_{n+1}^{(2)}$. Again, Assumption 5.3, summation-by-parts, and (OF4) with $\chi = \Phi$ yield

$$\begin{aligned}
\sum_{j=0}^{n} e^{(n-j)\tau A}\delta_j^{(2)} &= \tau^3 W_n A\Phi A F_n + \tau^3 \sum_{j=0}^{n-1} e^{(n-j)\tau A}\big(W_j - e^{-\tau A} W_{j+1}\big) A\Phi A F_j \\
&= \tau^2 W_n \Theta_\Phi (e^{\tau A} - I) A F_n \\
&\quad + \tau^2\Big(\tau \sum_{j=0}^{n-1} e^{(n-j)\tau A}\frac{1}{\tau}\big(W_j - e^{-\tau A} W_{j+1}\big)\Theta_\Phi (e^{\tau A} - I) A F_j\Big).
\end{aligned}$$

Here, we have

$$\frac{1}{\tau}\left(W_j - e^{-\tau A}W_{j+1}\right) = \frac{1}{\tau}e^{-\tau A}\left(W_j - W_{j+1}\right) - \frac{1}{\tau}\left(e^{-\tau A} - I\right)W_j.$$

The terms can be estimated by (5.9b) and (5.9c)

$$\left\|\frac{1}{\tau}e^{-\tau A}\left(W_j - W_{j+1}\right)\right\|_{X\leftarrow X} = \left\|\frac{1}{\tau}\left(W_j - W_{j+1}\right)\right\|_{X\leftarrow X} \le C,$$
$$\left\|\frac{1}{\tau}\left(e^{-\tau A} - I\right)W_j\right\|_{X\leftarrow X} = \left\|\varphi_1(-\tau A)AW_j\right\|_{X\leftarrow X} \le C,$$

since $|\varphi_1(z)| \le 1$ for $z \in i\mathbb{R}$.

Next we consider $(e^{\tau A} - I)AF_j$ for $j \le n$. After adding the exact solution we apply the variation-of-constants formula, (A3′), and (4.5), which gives

$$\begin{aligned}\left\|(e^{\tau A} - I)AF_j\right\| &= \left\|A\sum_{k=0}^{j}(e^{\tau A}\widetilde{u}(t_k) - \widetilde{u}(t_k+\tau)) + A\sum_{k=0}^{j}(\widetilde{u}(t_k+\tau) - \widetilde{u}(t_k))\right\| \\ &= \left\|\sum_{k=0}^{j}\int_0^\tau e^{(\tau-s)A}A\widetilde{f}(\widetilde{u}(t_k+s))\,ds + A(\widetilde{u}(t_{j+1}) - \widetilde{u}_0)\right\| \\ &\le t_{\text{end}}C(\widetilde{K}) + 2\widetilde{K}.\end{aligned}$$

This yields (6.2) for $\delta_j^{(2)}$ instead of δ_j and together with the results above proves (6.2).

Finally, (5.3), (6.1), (6.2), and $\widetilde{e}_0 = 0$ give

$$\begin{aligned}\|\widetilde{e}_{n+1}\| &= \left\|\tau\sum_{j=0}^{n}e^{(n-j)\tau A}\mathcal{J}\left(t_j, \widetilde{u}(t_j), u_j\right) - \sum_{j=0}^{n}e^{(n-j)\tau A}\delta_j\right\| \\ &\le C_\delta\tau^2 + \tau\sum_{j=1}^{n}C_{\mathcal{J}}(\widetilde{K}, r_X)\left\|\widetilde{e}_j\right\|.\end{aligned}$$

A discrete Gronwall Lemma thus yields

$$\|\widetilde{e}_{n+1}\| \le \tau^2\, C_\delta\, e^{C_{\mathcal{J}}(\widetilde{K}, r_X)t_{\text{end}}} = C_{\mathrm{e}}\tau^2,$$
$$\|u_{n+1}\| \le \|\widetilde{u}(t_{n+1})\| + \|\widetilde{e}_{n+1}\| \le \frac{3}{4}r_X + C_{\mathrm{e}}\tau^2 \le r_X$$

for $\tau \le \tau_0 \le \frac{1}{2}\left(\frac{r_X}{C_{\mathrm{e}}}\right)^{1/2}$ and the induction is closed. □

Our main result is the following theorem.

Theorem 6.2 *Let Assumptions 3.2 to 3.4 be fulfilled. Further let $(u_n)_n$ be the numerical approximations of a scheme that satisfies Assumptions 5.1 and 5.3.*

(a) If the method also satisfies Assumptions 5.2 and is applied to the original equation (2.2), then there is a $\tau_0 > 0$ and a constant $C_1 > 0$ such that for all $\tau \le \tau_0$

$$\|u_n - u(t_n)\| \le C_1\tau, \qquad 0 \le t_n = n\tau \le t_{\text{end}}.$$

(b) Let ϕ, ψ such that Assumption 5.3 is satisfied. Then there is a $\tau_0 > 0$ and a constant $C_2 > 0$ such that for all $\tau \le \tau_0$

$$\|u_n - u(t_n)\| \le C_2\tau^2, \qquad 0 \le t_n = n\tau \le t_{\text{end}},$$

if the method is applied to the averaged equation (2.8).

The constants C_1, C_2 and τ_0 depend on u_0, t_{end}, the finite energy K defined in (2.1), *the filter functions, and the embedding constant C_{emb}, but are independent of τ and n.*

Proof Part (*a*) follows directly from Assumption 5.2 and equation (6.1). For part (*b*), we simply combine Theorem 4.1 and Theorem 6.1 by the triangle inequality (2.9). □

7 Main result for exponential multistep methods

We briefly indicate how to extend the developed theory to the exponential multistep methods of Sect. 2.2.4. The first-order convergence as in part (*a*) of Theorem 6.2 is easily shown. To get second order, Assumption 5.1 needs to be modified.

For method (2.6), we denote the numerical flow by $S_\tau(t, v_n, v_{n-1})$ and obtain

$$S_\tau(t, v_n, v_{n-1}) - S_\tau(t, w_n, w_{n-1}) = e^{\tau A}\big(v_n - w_n\big) + \tau \mathcal{J}_n,$$

where $\mathcal{J}_n = \mathcal{J}\big(t, v_n, v_{n-1}, w_n, w_{n-1}\big)$ is bounded by

$$\begin{aligned}\|\mathcal{J}_n\| \le\ & C_{\mathcal{J}}\big(\|v_n\|, \|w_n\|\big)\|v_n - w_n\| \\ & + C_{\mathcal{J}}\big(\|v_{n-1}\|, \|w_{n-1}\|\big)\|v_{n-1} - w_{n-1}\|, \quad t \in [0, t_{\text{end}}].\end{aligned}$$

This yields the following convergence result.

Corollary 7.1 *Let Assumptions 3.2 to 3.4 be valid. Consider the numerical approximations $(u_n)_n$ from* (2.6) *applied to the averaged equation* (2.8) *with $\psi = 1$ and a filter ϕ of order 2. Then there is a $\tau_0 > 0$ and a constant $C > 0$ such that for all $\tau \le \tau_0$*

$$\|u(t_n) - u_n\| \le C\tau^2, \qquad 0 \le t_n = n\tau \le t_{\text{end}},$$

where C and τ_0 depend on u_0, t_{end}, the finite energy K defined in (2.1), *the filter functions, and the embedding constant C_{emb}, but are independent of τ and n.*

Proof We first employ Theorem 4.1 and Lemma 4.2, so again it remains to prove the error in approximating the filtered solution. As in the proof of [19, Thm. 4.3] the defect stems from a quadrature error that yields the dominant terms as in (5.6). Considering the defect

$$\delta_n = S_\tau\big(t_n, \widetilde{u}(t_n), \widetilde{u}(t_{n-1})\big) - \widetilde{u}(t_{n+1}),$$

Assumption 5.3 (*b*) is satisfied and a slight modification of the proof of Theorem 6.1 yields the assertion. □

For method (2.7) we have

$$S_\tau(t, v_n, v_{n-1}) - S_\tau(t, w_n, w_{n-1}) = e^{2\tau A}\big(v_{n-1} - w_{n-1}\big) + \tau \mathcal{J}_n$$

where $\mathcal{J}_n = \mathcal{J}(t, v_n, w_n)$ is bounded by

$$\|\mathcal{J}_n\| \le C_{\mathcal{J}}\big(\|v_n\|, \|w_n\|\big)\|v_n - w_n\|, \quad \forall t \in [0, t_{\text{end}}].$$

In order to apply the techniques from above we define the modification

$$\chi_2 : \mathcal{C}_b\big(i\mathbb{R}\big) \to \mathcal{C}_b\big(i\mathbb{R}\big), \quad \chi(\cdot) \mapsto \chi(2\cdot)\,,$$

and can state the following result.

Corollary 7.2 *Let Assumptions 3.2 to 3.4 be valid and u be the classical solution of (2.2).*

Consider the numerical approximations $(u_n)_n$ *from (2.7) applied to the averaged equation (2.8) with filters* $\chi_2\psi, \chi_2\phi$ *where* ψ, ϕ *are filters of order* 2. *Then there is a* $\tau_0 > 0$ *and a constant* $C > 0$ *such that for all* $\tau \le \tau_0$

$$\|u(t_n) - u_n\| \le C\tau^2, \qquad 0 \le t_n = n\tau \le t_{\text{end}},$$

where C and τ_0 *depend on* u_0, t_{end}, *the finite energy K defined in (2.1), the filter functions, and the embedding constant* C_{emb}, *but are independent of* τ *and n.*

Proof Since the method stems from a midpoint rule applied to the variation-of-constants formula the defect is again given with dominant terms similar to (5.6) and (5.7). If we resolve the error recursion, we only obtain every second defect and the propagation is driven by $e^{2\tau A}$. As e^z in (F3) is replaced by e^{2z}, this can be combined to conclude the assertion similar to the proof of Theorem 6.1. □

Acknowledgements The authors thank Ludwig Gauckler for helpful discussions on handling the global error in Theorem 6.1 and Jan Leibold for his careful reading of this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Baumstark, S., Faou, E., Schratz, K.: Uniformly accurate exponential-type integrators for Klein-Gordon equations with asymptotic convergence to the classical NLS splitting. Math. Comp. **87**(311), 1227–1254 (2018). https://doi.org/10.1090/mcom/3263
2. Buchholz, S., Gauckler, L., Grimm, V., Hochbruck, M., Jahnke, T.: Closing the gap between trigonometric integrators and splitting methods for highly oscillatory differential equations. IMA J. Numer. Anal. **38**(1), 57–74 (2018). https://doi.org/10.1093/imanum/drx007
3. Buchholz, S.F.: Fehleranalyse von auf trigonometrischen Integratoren basierenden Splittingverfahren für hochoszillatorische, semilineare Probleme. Ph.D. thesis, Karlsruher Institut für Technologie (KIT) (2019), https://doi.org/10.5445/IR/1000088935
4. Cano, B.: Conservation of invariants by symmetric multistep cosine methods for second-order partial differential equations. BIT **53**(1), 29–56 (2013). https://doi.org/10.1007/s10543-012-0393-1

5. Celledoni, E., Cohen, D., Owren, B.: Symmetric exponential integrators with an application to the cubic Schrödinger equation. Found. Comput. Math. **8**(3), 303–317 (2008). https://doi.org/10.1007/s10208-007-9016-7
6. Cohen, D., Hairer, E., Lubich, C.: Conservation of energy, momentum and actions in numerical discretizations of non-linear wave equations. Numer. Math. **110**(2), 113–143 (2008). https://doi.org/10.1007/s00211-008-0163-9
7. Frisch, U., She, Z.S., Thual, O.: Viscoelastic behaviour of cellular solutions to the Kuramoto–Sivashinsky model. J. Fluid Mech. **168**, 221–240 (1986). https://doi.org/10.1017/S0022112086000356
8. García-Archilla, B., Sanz-Serna, J.M., Skeel, R.D.: Long-time-step methods for oscillatory differential equations. SIAM J. Sci. Comput. **20**(3), 930–963 (1999). https://doi.org/10.1137/S1064827596313851
9. Gauckler, L.: Error analysis of trigonometric integrators for semilinear wave equations. SIAM J. Numer. Anal. **53**(2), 1082–1106 (2015). https://doi.org/10.1137/140977217
10. Gauckler, L., Lu, J., Marzuola, J.L., Rousset, F., Schratz, K.: Trigonometric integrators for quasilinear wave equations. Math. Comp. **88**(316), 717–749 (2019). https://doi.org/10.1090/mcom/3339
11. Goldberg, H., Kampowsky, W., Tröltzsch, F.: On Nemytskij operators in L_p-spaces of abstract functions. Math. Nachr. **155**, 127–140 (1992). https://doi.org/10.1002/mana.19921550110
12. Grimm, V., Hochbruck, M.: Error analysis of exponential integrators for oscillatory second-order differential equations. J. Phys. A **39**(19), 5495–5507 (2006). https://doi.org/10.1088/0305-4470/39/19/S10
13. Hairer, E., Lubich, C., Wanner, G.: Geometric numerical integration: Structure-preserving algorithms for ordinary differential equations. Springer Series in Computational Mathematics, vol. 31, 2nd edn. Springer-Verlag, Berlin (2006)
14. Hairer, E., Nørsett, S.P., Wanner, G.: Solving ordinary differential equations I: Nonstiff problems. Springer Series in Computational Mathematics, vol. 8, 2nd edn. Springer-Verlag, Berlin (1993)
15. Hansen, E., Ostermann, A.: Exponential splitting for unbounded operators. Math. Comp. **78**(267), 1485–1496 (2009). https://doi.org/10.1090/S0025-5718-09-02213-3
16. Hochbruck, M., Leibold, J., Ostermann, A.: On the convergence of Lawson methods for semilinear stiff problems. Numer. Math. **145**(3), 553–580 (2020). https://doi.org/10.1007/s00211-020-01120-4
17. Hochbruck, M., Lubich, C.: A Gautschi-type method for oscillatory second-order differential equations. Numer. Math. **83**(3), 403–426 (1999). https://doi.org/10.1007/s002110050456
18. Hochbruck, M., Ostermann, A.: Explicit exponential Runge–Kutta methods for semilinear parabolic problems. SIAM J. Numer. Anal. **43**(3), 1069–1090 (2005). https://doi.org/10.1137/040611434
19. Hochbruck, M., Ostermann, A.: Exponential multistep methods of Adams-type. BIT **51**(4), 889–908 (2011). https://doi.org/10.1007/s10543-011-0332-6
20. Jahnke, T., Lubich, C.: Error bounds for exponential operator splittings. BIT **40**(4), 735–744 (2000). https://doi.org/10.1023/A:10223965196567
21. Lubich, C.: On splitting methods for Schrödinger–Poisson and cubic nonlinear Schrödinger equations. Math. Comp. **77**(264), 2141–2153 (2008). https://doi.org/10.1090/S0025-5718-08-02101-7
22. McLachlan, R.I., Quispel, G.R.W.: Splitting methods. Acta Numer. **11**, 341–434 (2002). https://doi.org/10.1017/S0962492902000053
23. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences, vol. 44. Springer, New York (1983), https://doi.org/10.1007/978-1-4612-5561-1
24. Sanz-Serna, J.M.: Mollified impulse methods for highly oscillatory differential equations. SIAM J. Numer. Anal. **46**(2), 1040–1059 (2008). https://doi.org/10.1137/070681636
25. Schmüdgen, K.: Unbounded self-adjoint operators on Hilbert space. Graduate Texts in Mathematics, vol. 265. Springer, Dordrecht (2012), https://doi.org/10.1007/978-94-007-4753-1
26. Suzuki, M.: Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. Comm. Math. Phys. **51**(2), 183–190 (1976), https://doi.org/10.1007/BF01609348
27. Timan, A.F.: Theory of approximation of functions of a real variable. Translated from the Russian by J. Berry. English translation edited and editorial preface by J. Cossar. International Series of Monographs in Pure and Applied Mathematics, Vol. 34. A Pergamon Press Book. The Macmillan Co., New York (1963)
28. Wang, B., Wu, X.: Global error bounds of one-stage extended RKN integrators for semilinear wave equations. Numer. Algorithms **81**(4), 1203–1218 (2019). https://doi.org/10.1007/s11075-018-0585-0