Wave phenomena
analysis and numerics

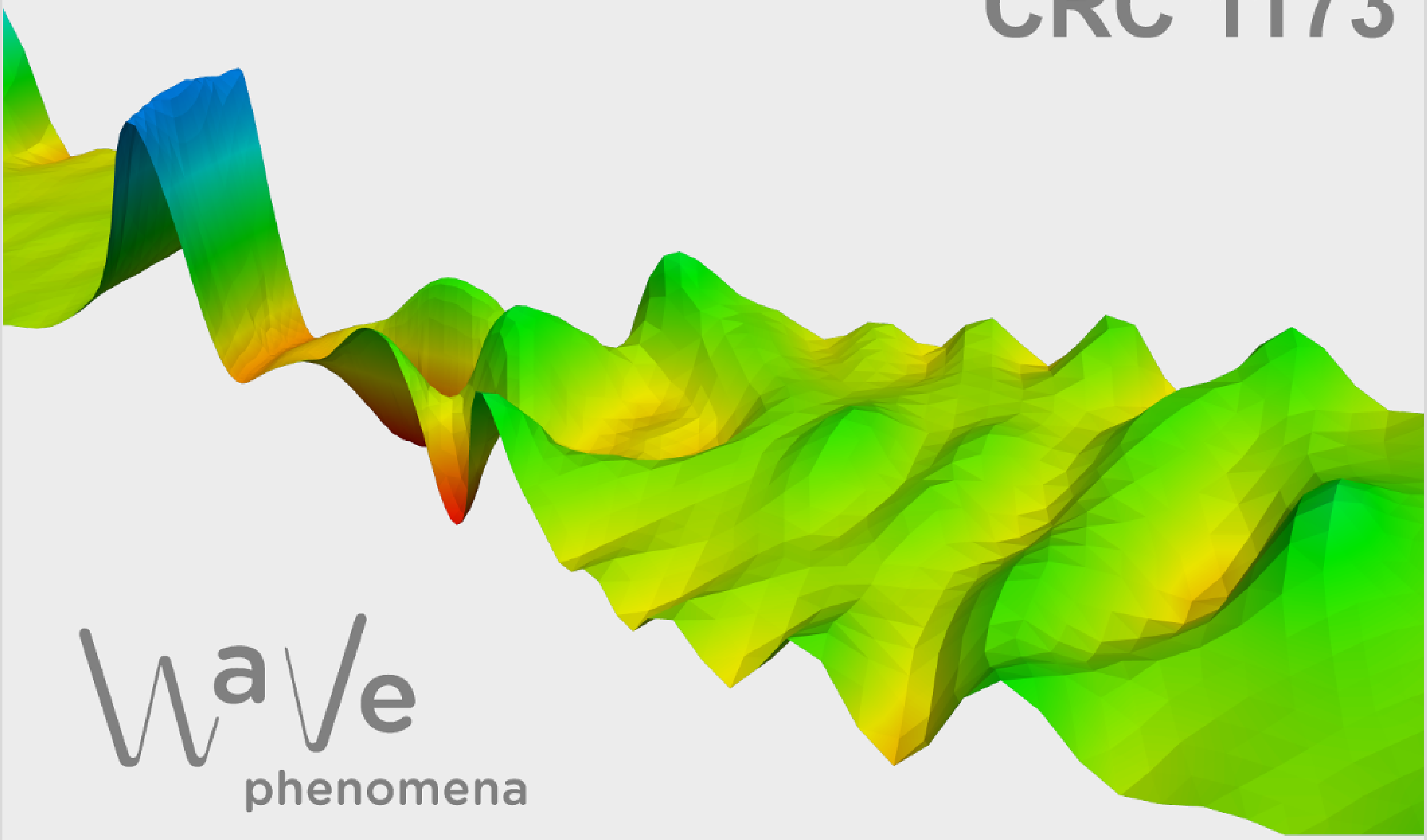
Metastable energy strata in numerical discretizations of weakly nonlinear wave equations

Ludwig Gauckler, Daniel Weiß

CRC Preprint 2016/13, May 2016

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Universität Stuttgart

Eberhard Karls Universität Tübingen

Funded by

DFG

ISSN 2365-662X

Metastable energy strata in numerical discretizations of weakly nonlinear wave equations

Ludwig Gauckler [1] Daniel Weiss [2]

Version of 28 May 2016

Abstract

The quadratic nonlinear wave equation on a one-dimensional torus with small initial values located in a single Fourier mode is considered. In this situation, the formation of metastable energy strata has recently been described and their long-time stability has been shown. The topic of the present paper is the correct reproduction of these metastable energy strata by a numerical method. For symplectic trigonometric integrators applied to the equation, it is shown that these energy strata are reproduced even on long time intervals in a qualitatively correct way.

Mathematics Subject Classification (2010): 65P10, 65P40, 65M70, 35L05.

Keywords: Nonlinear wave equation, trigonometric integrators, long-time behaviour, modulated Fourier expansion.

1 Introduction

We consider the nonlinear wave equation

$$\partial_{tt}u - \partial_{xx}u + \rho u = u^2, \qquad u = u(x,t) \in \mathbb{R}, \tag{1}$$

on a one-dimensional torus, $x \in \mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, with a positive Klein–Gordon parameter ρ. We assume that the initial value consists of a single Fourier mode and is small. The smallness of the initial value corresponds to a weakly nonlinear setting. In [12], it has been investigated how the energy, which is initially located in this single Fourier mode, is distributed among the other modes in the course of time. More precisely, the formation of energy strata has been shown that are persist on long time intervals, see Section 2 for a precise description of this result.

In the present paper, we discretize the nonlinear wave equation and answer the question whether this long-time property of the exact solution is inherited by the numerical method. As a numerical method, we consider a spectral collocation in space combined with a symplectic trigonometric integrator in time. We show that this numerical method in fact reproduces the energy strata of the exact solution even on a long time interval, provided that a certain numerical non-resonance condition is fulfilled, see Section 2 for a formulation of this main result.

[1] Institut für Mathematik, TU Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany (gauckler@math.tu-berlin.de).

[2] Institut für Angewandte und Numerische Mathematik, Karlsruher Institut für Technologie (KIT), Englerstr. 2, D-76131 Karlsruhe, Germany (daniel.weiss@kit.edu).

The considered numerical method is already known to behave well on long time intervals with respect to preservation of regularity and near-conservation of actions [5, 9], as well as near-conservation of energy and momentum [2, 3, 5, 8]. Our result adds to this list an even more sophisticated long-time property of the exact solution that is reproduced in a qualitatively correct way. In comparison to previous results, the considered situation requires less control of interactions in the nonlinearity, which allows us to exclude numerical resonances under weaker restrictions on the time step-size than needed previously.

The proof of our main result is given in Sections 3–5. As for the exact solution, it is based on a modulated Fourier expansion in time, with a multitude of additional difficulties due to the discrete setting that will be described in detail.

2 Statement of the main result

2.1 Metastable energy strata revisited

Writing the solution $u = u(x,t)$ of the nonlinear wave equation (1) as a Fourier series $\sum_{j\in\mathbb{Z}} u_j(t)\mathrm{e}^{\mathrm{i}jx}$ with Fourier coefficients u_j, the *mode energies* of the nonlinear wave equation (1) are given by

$$E_j(t) = \tfrac{1}{2}|\omega_j u_j(t)|^2 + \tfrac{1}{2}|\dot{u}_j(t)|^2, \qquad j \in \mathbb{Z}, \tag{2}$$

where ω_j are the frequencies

$$\omega_j = \sqrt{j^2+\rho}, \qquad j \in \mathbb{Z}. \tag{3}$$

Note that $E_j = E_{-j}$ for real-valued initial values (and hence real-valued solutions).

We are interested in the evolution of these mode energies. Assuming that the initial values are small and concentrated in the first mode,

$$E_1(0) \le \epsilon \ll 1, \qquad E_j(0) = 0 \qquad \text{for} \qquad |j| \ne 1,$$

an inspection of interaction of Fourier modes in the nonlinearity $g(u) = u^2$ suggests that the energy in the first mode is distributed among the other modes in a geometrically decaying way:

$$E_0(t) = \mathcal{O}(\epsilon^2), \qquad E_j(t) = \mathcal{O}(\epsilon^{|j|}) \qquad \text{for} \qquad j \ne 0,$$

at least for small times t.

The main result of [12] (Theorem 1) states that these energy strata are in fact stable on long time intervals in the following sense: for fixed but arbitrary $K \ge 2$,

$$\sum_{l=0}^{\infty} \sigma_l \epsilon^{-e(l)} E_l(t) \le C \qquad \text{for} \qquad 0 \le t \le c\epsilon^{-K/2}$$

with the energy profile

$$e(j) = \begin{cases} 2, & j = 0, \\ |j|, & 0 < |j| < K, \\ K, & |j| \ge K, \end{cases} \tag{4}$$

and with the weights

$$\sigma_j = \max\bigl(|j|,1\bigr)^{2s}, \qquad s > \tfrac{1}{2}, \tag{5}$$

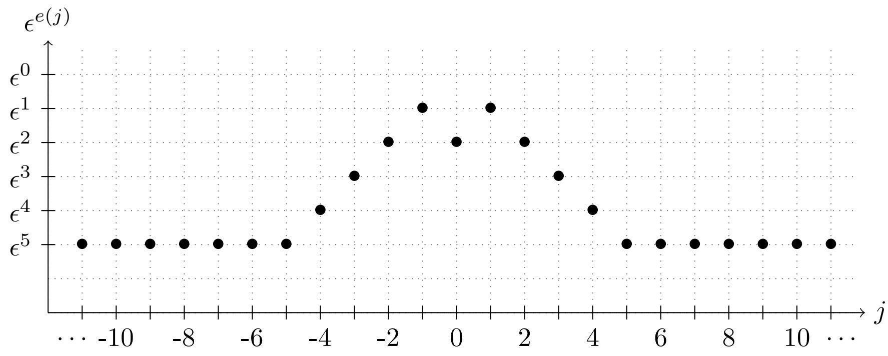


Figure 1: Illustration of the bound $\epsilon^{e(j)}$ for E_j in the case $K = 5$.

see also Figure 1 for an illustration. The constants depend on K but not on ϵ. This result is valid for all but finitely many values of the parameter $\rho \in [0, \rho_0]$ in the wave equation (1), which excludes some resonant situations. The question that we want to answer in the present paper is whether this stable behaviour on long time intervals is inherited by a typical structure-preserving numerical discretization of the nonlinear wave equation.

2.2 Trigonometric integrators

For the numerical discretization of the nonlinear wave equation (1) we consider trigonometric integrators in time applied to a spectral collocation in space, see, for example, [1–3, 5, 7–9, 11].

In these methods, the solution $u = u(x, t)$ of (1) is approximated, at discrete times $t_n = n\tau$ with the time step-size τ, by a trigonometric polynomial of degree M:

$$u(x, t_n) \approx \sum_{j=-M}^{M-1} u_j^n \mathrm{e}^{\mathrm{i}jx}, \qquad n = 0, 1, \ldots.$$

The Fourier coefficients $\mathbf{u}^n = (u_{-M}^n, \ldots, u_{M-1}^n)^{\mathrm{T}}$ of this trigonometric polynomials are computed with the trigonometric integrator

$$\mathbf{u}^{n+1} - 2\cos(\tau\boldsymbol{\Omega})\mathbf{u}^n + \mathbf{u}^{n-1} = \tau^2 \boldsymbol{\Psi}\big((\boldsymbol{\Phi}\mathbf{u}^n) * (\boldsymbol{\Phi}\mathbf{u}^n)\big), \tag{6a}$$

where $\boldsymbol{\Omega}$ denotes the diagonal matrix containing the frequencies ω_j, $j = -M, \ldots, M-1$, of (3). In addition, $*$ denotes the discrete convolution defined by

$$(\mathbf{u} * \mathbf{v})_j = \sum_{j_1 + j_2 \equiv j \bmod 2M} u_{j_1} v_{j_2}, \qquad j = -M, \ldots, M-1,$$

which can be computed efficiently using the fast Fourier transform. The method is characterized by the diagonal filter matrices $\boldsymbol{\Psi} = \psi(\tau\boldsymbol{\Omega})$ and $\boldsymbol{\Phi} = \phi(\tau\boldsymbol{\Omega})$, which are computed from filter functions ψ and ϕ. These filter functions are assumed to be real-valued, bounded and even with $\psi(0) = \phi(0) = 1$. The starting approximation is computed by

$$\mathbf{u}^1 = \cos(\tau\boldsymbol{\Omega})\mathbf{u}^0 + \tau\,\mathrm{sinc}(\tau\boldsymbol{\Omega})\dot{\mathbf{u}}^0 + \tfrac{1}{2}\tau^2 \boldsymbol{\Psi}\big((\boldsymbol{\Phi}\mathbf{u}^0) * (\boldsymbol{\Phi}\mathbf{u}^0)\big), \tag{6b}$$

and a velocity approximation by

$$2\tau\,\mathrm{sinc}(\tau\mathbf{\Omega})\dot{\mathbf{u}}^n = \mathbf{u}^{n+1} - \mathbf{u}^{n-1}. \tag{6c}$$

An error analysis of these methods has been given in [11]. We assume here that the considered trigonometric integrator is symplectic,

$$\psi(\xi) = \mathrm{sinc}(\xi)\phi(\xi), \tag{7}$$

see [15, Chap. XIII, Eq. (2.10)]. Examples for such filter functions are $\phi = 1$ and $\psi = \mathrm{sinc}$, which is the impulse method [13, 16] or method of Deuflhard [6], as well as $\phi = \mathrm{sinc}$ and $\psi = \mathrm{sinc}^2$, which is the mollified impulse method of García-Archilla, Sanz-Serna & Skeel [10]. Certain choices of filter functions leading to non-symplectic methods could also be handled using the transformation indicated in [4, Remark 3.2], but we do not pursue this here.

We are interested in the mode energies (2) along the numerical solution (6). We denote them in the following by

$$E_j^n = \tfrac{1}{2}|\omega_j u_j^n|^2 + \tfrac{1}{2}|\dot{u}_j^n|^2, \qquad j = -M, \dots, M-1. \tag{8}$$

We note that $E_j^n = E_{-j}^n$ for real-valued initial values (and hence numerical solutions that take real values in the collocation points $x_k = \pi k/M$, $k = -M, \dots, M-1$), and we set $E_M^n = E_{-M}^n$. As for the exact solution in Section 2.1, we consider initial values with

$$E_1^0 \le \epsilon, \qquad E_j^0 = 0 \qquad \text{for} \qquad |j| \ne 1. \tag{9}$$

2.3 Metastable energy strata in trigonometric integrators

We now state our main result which says, roughly speaking, that trigonometric integrators (6) integrate the metastable energy strata in nonlinear wave equations (1) qualitatively correctly.

For this result, we need a non-resonance condition on the time step-size τ and on the frequencies ω_j, $j = -M, \dots, M-1$, of (3). In the statement of this condition, we consider indices $j \in \{-M, \dots, M-1\}$ and vectors $\mathbf{k} = (k_0, \dots, k_M)^{\mathrm{T}}$ of integers k_l, and we write

$$\mathbf{k}\cdot\boldsymbol{\omega} = \sum_{l=0}^{M} k_l\omega_l$$

with the vector $\boldsymbol{\omega} = (\omega_0, \dots, \omega_M)^{\mathrm{T}}$ of frequencies. We fix $3 \le K \le M$, and we denote, corresponding to [12], by $\mathcal{K}$ the set

$$\begin{aligned}\mathcal{K} = \big\{\, (j, \mathbf{k}) : \max(|j|, \mu(\mathbf{k})) < 2K \text{ and } k_l = 0 \text{ for all } l \ge K \,\big\} \\ \cup \big\{ (j, \pm\langle (j-r) \bmod 2M\rangle + \mathbf{k}) : |(j-r) \bmod 2M| \ge K,\ |r| < K,\ \mu(\mathbf{k}) < K \,\big\},\end{aligned} \tag{10}$$

where $\langle j\rangle = (0, \dots, 0, 1, 0, \dots, 0)^{\mathrm{T}}$ is the $|j|$th unit vector and

$$\mu(\mathbf{k}) = \sum_{l=0}^{M} |k_l| e(l) = 2|k_0| + \sum_{l=1}^{K-1} |k_l| l + K\sum_{l=K}^{M} |k_l|. \tag{11}$$

In comparison to the set $\mathcal{K}$ for the exact solution (see [12, Equation (11)]), we have to consider indices modulo $2M$ due to the discretization in space. In addition, we correct here a typo in [12, Equation (11)], where $|j| \ge K$ should be replaced by $|j - r| \ge K$.

Non-resonance condition. For given $0 \le \nu < \frac{1}{2}$, we assume that there exists a constant $0 < \gamma \le 1$ such that

$$\left|\sin\left(\tfrac{1}{2}\tau(\omega_j \pm \mathbf{k}\cdot\boldsymbol{\omega})\right)\right| \ge \gamma\tau\epsilon^{\nu/2} \qquad \text{for all} \qquad (j,\mathbf{k}) \in \mathcal{K},\ \mathbf{k} \ne \mp\langle j\rangle \tag{12a}$$

and

$$\left|\sin\left(\tfrac{1}{2}\tau(\omega_j \pm \mathbf{k}\cdot\boldsymbol{\omega})\right)\right| \ge \gamma\tau \qquad \text{for all} \qquad (j,\mathbf{k}) \in \mathcal{K},\ \mathbf{k} \ne \langle j\rangle,\ \mathbf{k} \ne -\langle j\rangle,\ |j| \le K. \tag{12b}$$

Under this non-resonance condition, we prove in this paper the following discrete analogon of the analytical result [12, Theorem 1] described in Section 2.1.

Theorem 1. *Fix an integer $3 \le K \le M$ and real numbers $0 \le \nu < \frac{1}{2}$ and $s > \frac{1}{2}$, and assume that the non-resonance condition (12) holds for this ν. Then there exist $\epsilon_0 > 0$ and positive constants c and C such that the mode energies (8) along trigonometric integrators (6) for nonlinear wave equations (1) with initial data (9) with $0 < \epsilon \le \epsilon_0$ satisfy, over long times*

$$0 \le t_n = n\tau \le c\epsilon^{-K(1-2\nu)/2},$$

the bounds

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} E_l^n \le C$$

and $\epsilon^{-e(1)}|E_1^n - E_1^0| \le C\epsilon^{1-2\nu}$. The constants c and C are independent of ϵ and the discretization parameters τ and M.

Remark 2. Both, the result and the assumption of Theorem 1, get stronger for larger K. Hence, we can also cover the (not so interesting) border case $K = 2$ as considered in [12] provided that the non-resonance assumption holds for $K = 3$. Without requiring the non-resonance condition with $K = 3$, our proof shows that Theorem 1 holds for $K = 2$, if we replace $\epsilon^{-e(1)}|E_1^n - E_1^0| \le C\epsilon^{1-2\nu}$ by $\epsilon^{-e(1)}|E_1^n - E_1^0| \le C\epsilon^{(1-2\nu)/2}$.

The proof of Theorem 1 is given in Sections 3–5 below. The structure of the proof is similar to the structure of the corresponding proof for the exact solution given in [12], with a multitude of additional difficulties due to the discrete setting.

The main difference of this result in comparison with the corresponding result [12, Theorem 1] for the exact solution is the required non-resonance condition. The non-resonance condition (12) excludes two types of resonances. First, it requires that $\omega_j \pm \mathbf{k}\cdot\boldsymbol{\omega}$ is bounded away from zero for $(j,\mathbf{k}) \in \mathcal{K}$ with $\mathbf{k} \ne \mp\langle j\rangle$. This is the non-resonance condition for the exact solution, which can be shown to hold for all except finitely many values of ρ in a fixed interval, see [12, Section 3.2]. In addition, the non-resonance condition (12) requires also that products $\tau(\omega_j \pm \mathbf{k}\cdot\boldsymbol{\omega})$ are bounded away from *nonzero* integer multiples of 2π. If this latter condition is violated, we observe numerical resonances as illustrated in the following section.

Numerical resonances can be typically excluded under some CFL-type step-size restriction on the time step-size. A feature of the considered situation is that they can be excluded under less restrictive assumptions than in previous studies on the long-time behaviour of trigonometric integrators for nonlinear wave equations [5, 9]. In particular, there are no numerical resonances on the time interval $0 \le t \le \epsilon^{-K/2}$ under the CFL-type step-size restriction

$$\tau(M+K)\sqrt{1+\rho} \le c < \pi,$$

for some constant c (we may take $\nu = 0$ here). This follows from the structure of the set $\mathcal{K}$ of (10) which enters the non-resonance condition: this set allows only a single large frequency in the linear combination $\mathbf{k}\cdot\boldsymbol{\omega}$, and more precisely $\sum_{l=0}^{M}|k_l|\omega_l \le (M+2K)\sqrt{1+\rho}$. In the situation of [5, 9], the stronger CFL-type step-size restriction of the form $\tau MK \le c < \pi$ has to be used to exclude numerical resonances.

For nonzero ν, the numerical non-resonance condition (12) resembles the one used in [5], see Equations (23) and (24) therein. We note that the reduction (24) there is of no relevance in our context because of the structure of the set $\mathcal{K}$. We also note that the number of indices $(j, \mathbf{k})$ that have to satisfy the stronger condition (12b), which does not appear in [5], is bounded independently of the spatial discretization parameter M.

There are several possible extensions of Theorem 1. First, we could extend the result, as in [12], to an energy profile e with $e(l) = K + (|l| - K)(1 - \theta)$ instead of $e(l) = K$ for $|l| \ge K$, where $0 < \theta \le 1$. This shows that also the mode-energies E_l for $l > K$ decay geometrically, but only with a smaller power of ϵ close to 1. As in [12], the time-scale is then restricted to $0 \le t_n \le \epsilon^{-\theta K(1-2\nu)/2}$.

Second, the result holds for general nonlinearities $g(u)$ instead of u^2 if g is real-analytic near 0 and $g(0) = g'(0) = 0$. In addition, stronger estimates hold if further derivatives of g vanish at 0.

Finally, the result can be extended to more general initial energy profiles, for example to the situation where a whole band E_l^0, $|l| \le B$, of initial mode energies is of order ϵ.

2.4 Numerical experiment

We consider the nonlinear wave equation (1) with $\rho = 0.5$ and with initial value satisfying (9) for $\epsilon = 10^{-3}$. We apply the trigonometric integrator (6) with filter functions $\phi = 1$ and $\psi = \text{sinc}$ (the impulse method or method of Deuflhard) to the equation. We take $M = 2^5$ for the spectral collocation in space, and we use three different time step-sizes for the discretization in time[1].

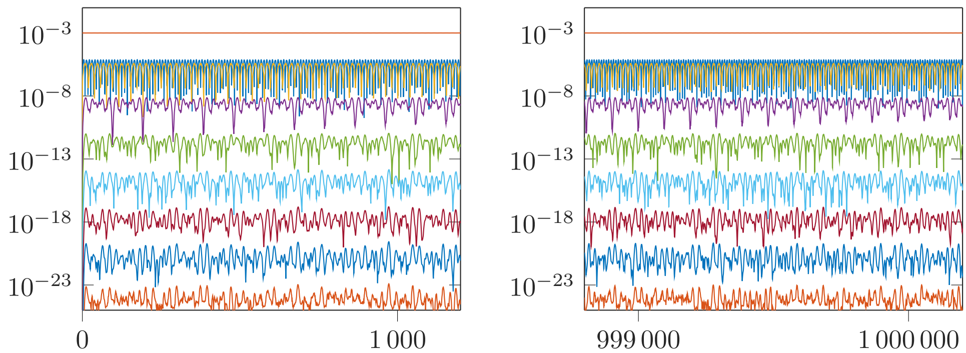


Figure 2: Mode energies E_l^n vs. time t_n for the numerical solution with time step-size $\tau = 0.45$.

The first time step-size is $\tau = 0.45$. For this time step-size, the trigonometric integrator integrates the geometrically decaying energy strata qualitatively correctly, even on long time intervals, see Figure 2. This can be explained with Theorem 1.

[1] The code is available at `http://www.waves.kit.edu/downloads/CRC1173_Preprint_2016-13_supplement.zip`

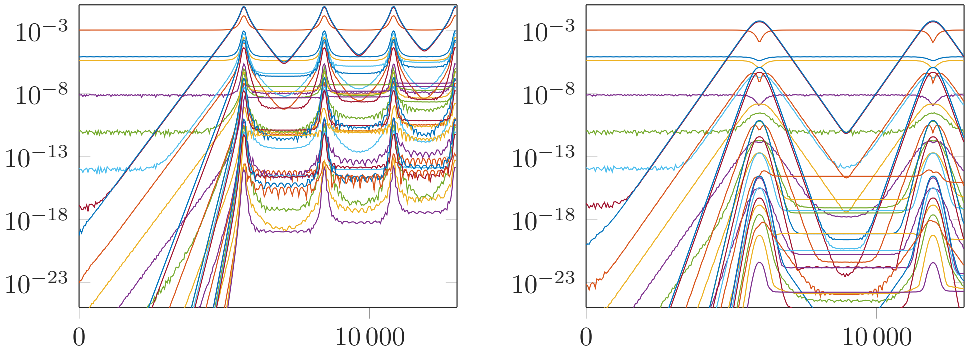


Figure 3: Mode energies E_l^n vs. time t_n for the numerical solutions with time step-size $\tau = 2\pi/(\omega_1 + \omega_6 + \omega_7)$ (left) and time step-size $\tau = 2\pi/(-\omega_1 + \omega_6 + \omega_7)$ (right).

The second time step-size is $\tau = 2\pi/(\omega_1+\omega_6+\omega_7) \approx 0.4393$. It is rather close to the first one but chosen in such a way that it does not satisfy the non-resonance condition (12). The resulting numerical resonance becomes apparent in Figure 3, where we observe in particular an exchange among the first, sixth and seventh mode. For better visibility, we have plotted $\max_{m=0,\dots,99} E_l^{n+m}$ for $n = 0, 100, 200, \dots$ instead of E_l^n for $n = 0, 1, 2, \dots$ in Figure 3. In the same figure, a similar behaviour can be observed for the resonant time step-size $\tau = 2\pi/(-\omega_1 + \omega_6 + \omega_7)$.

3 Modulated Fourier expansions: Proof of Theorem 1

We assume that the non-resonance conditions (12) holds for some $0 \le \nu < \frac{1}{2}$.

3.1 Approximation ansatz

We will approximate the numerical solution by a *modulated Fourier expansion*,

$$u_j^n \approx \widetilde{u}_j^n = \sum_{\mathbf{k}\in\mathcal{K}_j} z_j^{\mathbf{k}}(\epsilon^{\nu/2} t_n) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n}, \qquad j = -M, \dots, M-1, \tag{13}$$

where $\mathcal{K}_j = \{\mathbf{k} : (j, \mathbf{k}) \in \mathcal{K}\}$ with the set $\mathcal{K}$ of (10). We require that the *modulation functions* $z_j^{\mathbf{k}}$ are polynomials. In contrast to the modulated Fourier expansion of [12, Equation (7)] for the exact solution, the modulation functions considered here vary on a slow time-scale $\epsilon^{\nu/2} t$, where ν is the parameter from the non-resonance condition (12). With this slow time scale, derivatives of $z_j^{\mathbf{k}}(\epsilon^{\nu/2} t)$ with respect to t carry additional factors $\epsilon^{\nu/2}$, which will be used to compensate for this factor in the weak non-resonance condition (12a).

Having in mind that u_j^n, $j = -M, \dots, M-1$, are the Fourier coefficients of a trigonometric polynomial of degree M, we write in the following $u_M^n = u_{-M}^n$ and $z_M^{\mathbf{k}} = z_{-M}^{\mathbf{k}}$.

Inserting the ansatz (13) into the trigonometric integrator (6a) and comparing the coefficients of $\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t}$, yields a set of equations for the modulation functions $\mathbf{z} =$

$(z_j^{\mathbf{k}})_{(j,\mathbf{k})\in\mathcal{K}}$, for which solutions are to be constructed up to a small defect $d_j^{\mathbf{k}}$:

$$\begin{aligned} &\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} z_j^{\mathbf{k}}\big(\epsilon^{\nu/2}(t+\tau)\big) - 2\cos(\tau\omega_j) z_j^{\mathbf{k}}\big(\epsilon^{\nu/2}t\big) + \mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} z_j^{\mathbf{k}}\big(\epsilon^{\nu/2}(t-\tau)\big) \\ &\qquad = -\tau^2\psi(\tau\omega_j)\nabla_{-j}^{-\mathbf{k}}\mathcal{U}\big(\mathbf{\Phi z}(\epsilon^{\nu/2}t)\big) + \tau^2\psi(\tau\omega_j) d_j^{\mathbf{k}}\big(\epsilon^{\nu/2}t\big) \end{aligned} \tag{14}$$

with the extended potential

$$\mathcal{U}(\mathbf{z}) = -\frac{1}{3}\sum_{j_1+j_2+j_3\equiv 0}\ \sum_{\mathbf{k}^1+\mathbf{k}^2+\mathbf{k}^3=\mathbf{0}} z_{j_1}^{\mathbf{k}^1} z_{j_2}^{\mathbf{k}^2} z_{j_3}^{\mathbf{k}^3}, \tag{15}$$

with $\mathbf{\Phi z} = (\phi(\tau\omega_j) z_j^{\mathbf{k}})_{(j,\mathbf{k})\in\mathcal{K}}$ and with $\nabla_{-j}^{-\mathbf{k}}$ denoting the partial derivative with respect to $z_{-j}^{-\mathbf{k}}$. In (15) and in the following we denote by $\equiv$ the congruence modulo $2M$.

With the modulated Fourier expansion (13) at hand, the formula (6c) for the velocity approximation leads to an approximation of $\dot{u}_j^n$ by a modulated Fourier expansion:

$$\begin{aligned} \dot{u}_j^n \approx \dot{\tilde{u}}_j^n = \big(2\tau\,\mathrm{sinc}(\tau\omega_j)\big)^{-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \Big(& z_j^{\mathbf{k}}(\epsilon^{\nu/2} t_{n+1})\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} \\ & - z_j^{\mathbf{k}}(\epsilon^{\nu/2} t_{n-1})\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau}\Big)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n}. \end{aligned} \tag{16}$$

Finally, we get conditions from the fact that the ansatz (13) should satisfy the initial condition:

$$u_j^0 = \sum_{\mathbf{k}\in\mathcal{K}_j} z_j^{\mathbf{k}}(0), \tag{17a}$$

$$\dot{u}_j^0 = \big(2\tau\,\mathrm{sinc}(\tau\omega_j)\big)^{-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \Big(z_j^{\mathbf{k}}(\epsilon^{\nu/2}\tau)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} - z_j^{\mathbf{k}}(-\epsilon^{\nu/2}\tau)\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau}\Big), \tag{17b}$$

where (16) was used for the derivation of the second equation.

3.2 Norms

For vectors $\mathbf{v} = (v_{-M}, \ldots, v_{M-1})$ of Fourier coefficients of trigonometric polynomials of degree M, we consider the norm

$$\|\mathbf{v}\| = \bigg(\sum_{j=-M}^{M-1} \sigma_j \epsilon^{-2e(j)\nu} |v_j|^2\bigg)^{1/2},$$

where $e(j)$ is the energy profile of (4) and σ_j are the weights of (5). For $\nu = 0$, this is the Sobolev H^s-norm of the corresponding trigonometric polynomial $\sum_{j=-M}^{M-1} v_j e^{\mathrm{i}jx}$. The rescaling $\epsilon^{-2e(j)\nu}$ originates from the non-resonance condition (12a) that introduces $\epsilon^{\nu/2}$.

We will make frequent use of the fact that this norm behaves well with respect to the discrete convolution,

$$\|\mathbf{u} * \mathbf{v}\| \le C\|\mathbf{u}\|\,\|\mathbf{v}\|. \tag{18}$$

This follows from the corresponding property of the Sobolev H^s-norm, see, for example, [14, Lemma 4.2], and the fact that $e(j) \le e(j_1) + e(j_2)$ for $j \equiv j_1 + j_2$. Similarly, we also have

$$\|\mathbf{\Omega}(\mathbf{u} * \mathbf{v})\| \le C\|\mathbf{\Omega u}\|\,\|\mathbf{\Omega v}\|. \tag{19}$$

3.3 The modulated Fourier expansion on a short time interval

Instead of initial mode energies (9), we consider the more general choice

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} E_l^0 \le C_0 \tag{20}$$

of initial mode energies, which is the expected situation at later times (see Theorem 1). We then have the following discrete counterpart of [12, Theorem 3], whose proof will be given in Section 4.

Theorem 3. *Under the assumptions of Theorem 1, but with (20) instead of (9), the numerical solution* $\mathbf{u}^n = (u_{-M}^n, \dots, u_{M-1}^n)^{\mathrm{T}}$ *admits an expansion*

$$u_j^n = \sum_{\mathbf{k}\in\mathcal{K}_j} z_j^{\mathbf{k}}(\epsilon^{\nu/2} t_n) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n} + r_j^n \qquad \text{for} \qquad 0 \le t_n = n\tau \le 1,$$

where $\mathcal{K}_j = \{\mathbf{k} : (j,\mathbf{k}) \in \mathcal{K}\}$ *with the set* $\mathcal{K}$ *of (10), where the coefficient functions* $z_j^{\mathbf{k}}$ *are polynomials satisfying* $z_{-j}^{-\mathbf{k}} = \overline{z_j^{\mathbf{k}}}$, *and where the remainder term* $\mathbf{r}^n = (r_{-M}^n, \dots, r_{M-1}^n)^{\mathrm{T}}$ *and the corresponding remainder term* $\dot{\mathbf{r}}^n = (\dot r_{-M}^n, \dots, \dot r_{M-1}^n)^{\mathrm{T}}$ *in the velocity approximation (16) are bounded by*

$$\|\boldsymbol{\Omega}\mathbf{r}^n\| + \|\dot{\mathbf{r}}^n\| \le C\epsilon^{K(1-2\nu)} \qquad \text{for} \qquad 0 \le t_n = n\tau \le 1$$

and the defect $d_j(t) = \sum_{\mathbf{k}\in\mathcal{K}_j} |d_j^{\mathbf{k}}(\epsilon^{\nu/2}t)|$ *in (14) is bounded by*

$$\|\mathbf{d}(t)\| \le C\epsilon^{K(1-2\nu)} \qquad \text{for} \qquad 0 \le t \le 1.$$

The constant C *is independent of* ϵ, τ *and* M, *but depends on* K, *on* ν *and* γ *of (12), on* ρ *of (1), on* s *of (5) and on* C_0 *of (20).*

3.4 Almost-invariant energies

We now derive almost-invariant energies of the modulation system (14) which enable us to consider long time intervals. The derivation of these almost-invariant energies is similar as in the case of the exact solution, see [12, Section 3.5], but it now leads to discrete almost-invariant energies that involve additionally the time step-size τ and are related to those of [5].

As in the case of the exact solution, the extended potential $\mathcal{U}$ of (15) is invariant under the transformation $\mathcal{S}_{\boldsymbol{\lambda}}(\theta)\mathbf{z} := (\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda})\theta} z_j^{\mathbf{k}})_{(j,\mathbf{k})\in\mathcal{K}}$ for $\theta \in \mathbb{R}$ and real vectors $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_M)$. This invariance shows that

$$0 = \frac{\mathrm{d}}{\mathrm{d}\theta}\bigg|_{\theta=0} \mathcal{U}\big(\mathcal{S}_{\boldsymbol{\lambda}}(\theta)\boldsymbol{\Phi}\mathbf{z}\big) = -\sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\lambda})\phi(\tau\omega_j) z_{-j}^{-\mathbf{k}} \nabla_{-j}^{-\mathbf{k}} \mathcal{U}(\boldsymbol{\Phi}\mathbf{z}).$$

We multiply this equation with $\tau/2$, we replace $-\nabla_{-j}^{-\mathbf{k}}\mathcal{U}$ with the help of (14), we use the symplecticity (7) of the method, and we use

$$\sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \frac{\mathbf{k}\cdot\boldsymbol{\lambda}}{\operatorname{sinc}(\tau\omega_j)} z_{-j}^{-\mathbf{k}}\big(\epsilon^{\nu/2}t\big) \cos(\tau\omega_j) z_j^{\mathbf{k}}\big(\epsilon^{\nu/2}t\big) = 0. \tag{21}$$

Additionally, we choose $\boldsymbol{\lambda} = \langle l \rangle$ with $l = 0, \dots, M$. Altogether, this shows that

$$\mathcal{E}_l(t) = \mathcal{E}_l(t-\tau) - \frac{\mathrm{i}}{2} \sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \tau k_l \omega_l \phi(\tau\omega_j) z_{-j}^{-\mathbf{k}}(\epsilon^{\nu/2}t) d_j^{\mathbf{k}}(\epsilon^{\nu/2}t) \tag{22}$$

with

$$\mathcal{E}_l(t) = -\frac{\mathrm{i}}{2} \sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \frac{k_l\omega_l}{\tau \operatorname{sinc}(\tau\omega_j)} z_{-j}^{-\mathbf{k}}(\epsilon^{\nu/2}t) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} z_j^{\mathbf{k}}(\epsilon^{\nu/2}(t+\tau)). \tag{23}$$

Using a Taylor expansion of $z_j^{\mathbf{k}}(\epsilon^{\nu/2}(t+\tau))$ and the property (21), we can derive the following alternative form of $\mathcal{E}_l$:

$$\begin{aligned}\mathcal{E}_l(t) = \frac{1}{2} \sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \Big(& k_l\omega_l (\mathbf{k}\cdot\boldsymbol{\omega}) \frac{\operatorname{sinc}(\tau(\mathbf{k}\cdot\boldsymbol{\omega}))}{\operatorname{sinc}(\tau\omega_j)} \big| z_j^{\mathbf{k}}(\epsilon^{\nu/2}t)\big|^2 \\ & - \mathrm{i} \frac{k_l\omega_l}{\operatorname{sinc}(\tau\omega_j)} z_{-j}^{-\mathbf{k}}(\epsilon^{\nu/2}t) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})\tau} \Big(\epsilon^{\nu/2} \dot{z}_j^{\mathbf{k}}(\epsilon^{\nu/2}t) + \tfrac{1}{2}\tau\epsilon^{\nu} \ddot{z}_j^{\mathbf{k}}(\epsilon^{\nu/2}t) + \dots \Big)\Big).\end{aligned} \tag{24}$$

Using (21), we see that this quantity coincides, for $\nu = 0$ and in the limits $\tau \to 0$ and $M \to \infty$, with the almost-invariant energy of [12, Equation (23)] for the exact solution. We therefore also call $\mathcal{E}_l(t)$ an *almost-invariant energy.* In Sections 4 and 5 below, we prove the following discrete counterparts of [12, Theorems 4–7] for this new almost-invariant energy.

Theorem 4 (Almost-invariant energies controlled by mode energies)**.** *Under the conditions of Theorem 3 we have, for* $0 \le t \le 1$,

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} |\mathcal{E}_l(t)| \le \mathcal{C}_0, \tag{25}$$

where $\mathcal{C}_0$ *is independent of* ϵ, τ *and* M *and depends on the initial data only through the constant* C_0 *of* (20).

Theorem 5 (Variation of almost-invariant energies)**.** *Under the conditions of Theorem 3 we have, for* $0 \le t_n \le 1$,

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} \big|\mathcal{E}_l(t_n) - \mathcal{E}_l(0)\big| \le C\epsilon^{K(1-2\nu)/2}$$

and $\epsilon^{-e(1)}|\mathcal{E}_1(t_n) - \mathcal{E}_1(0)| \le C\epsilon^{(K-1)(1-2\nu)/2}\epsilon^{K(1-2\nu)/2}$, *where* C *is independent of* ϵ, τ *and* M *and depends on the initial data only through the constant* C_0 *of* (20).

At time $t = 1$, for which we assume without loss of generality that $t_N = N\tau = 1$ for some N, we consider a new modulated Fourier expansion leading to new almost-invariant energies.

Theorem 6 (Transitions in the almost-invariant energies)**.** *Let the conditions of Theorem 3 be fulfilled. Let* $\mathbf{z}(\epsilon^{\nu/2}t) = (z_j^{\mathbf{k}}(\epsilon^{\nu/2}t))_{(j,\mathbf{k})\in\mathcal{K}}$ *for* $0 \le t \le 1$ *be the coefficient functions as in Theorem 3 with corresponding almost-invariant energies* $\mathcal{E}_l(t)$ *for initial data* $(\mathbf{u}^0, \dot{\mathbf{u}}^0)$. *Let further* $\widetilde{\mathbf{z}}(\epsilon^{\nu/2}t) = (\widetilde{z}_j^{\mathbf{k}}(\epsilon^{\nu/2}t))_{(j,\mathbf{k})\in\mathcal{K}}$ *be the coefficient functions* *and $\widetilde{\mathcal{E}}(t)$ the corresponding almost-invariants of the modulated Fourier expansion for $0 \le t \le 1$ corresponding to initial data $(\mathbf{u}^N, \dot{\mathbf{u}}^N)$ with $t_N = N\tau = 1$, constructed as in Theorem 3. If $(\mathbf{u}^N, \dot{\mathbf{u}}^N)$ also satisfies the bound (20), then*

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} \big|\mathcal{E}_l(1) - \widetilde{\mathcal{E}}_l(0)\big| \le C\epsilon^{K(1-2\nu)/2}$$

and $\epsilon^{-e(1)}|\mathcal{E}_1(1) - \widetilde{\mathcal{E}}_1(0)| \le C\epsilon^{(K-1)(1-2\nu)/2}\epsilon^{K(1-2\nu)/2}$, where C is independent of ϵ, τ and M and depends on the initial data only through the constant C_0 of (20).

Theorem 7 (Mode energies controlled by almost-invariant energies)**.** *Let the conditions of Theorem 3 be fulfilled. If the almost-invariant energies satisfy (25) for $0 \le t_n = n\tau \le 1$, then the mode energies are bounded by*

$$\sum_{l=0}^{M} \sigma_l \epsilon^{-e(l)} E_l^n \le \mathcal{C}$$

and $\epsilon^{-e(1)}|E_1^n - \mathcal{E}_1(t_n)| \le \mathcal{C}\epsilon^{1-2\nu}$, where $\mathcal{C}$ depends on $\mathcal{C}_0$ in (25), but is independent of ϵ, τ, M and C_0 of (20) if $\epsilon^{1-2\nu}$ is sufficiently small.

3.5 From short to long time intervals: Proof of Theorem 1

Based on Theorems 3–7, the proof of Theorem 1 is the same as in the case of the exact solution, see [12, Section 3.6]: Theorems 3–5 and 7 yield the statement of Theorem 1 on a short time interval $0 \le t_n \le 1$. Theorem 6 can be used to patch many of these short time intervals together, on which the almost-invariant energies $\mathcal{E}_l$ are still well preserved (Theorems 5 and 6) and allow to control the mode energies E_l (Theorem 7).

4 Construction of a modulated Fourier expansion: Proofs of Theorems 3 and 4

Throughout this section, we work under the assumptions of Theorem 3. In addition, we let

$$\delta = \epsilon^{1/2},$$

such that $u_j^0, \dot{u}_j^0 = \mathcal{O}(\delta^{e(l)})$. Moreover, we write $\lesssim$ for an inequality up to a factor that is independent of $\delta = \epsilon^{1/2}$, τ and M.

4.1 Expansion of the modulation functions

The construction of modulation functions $z_j^{\mathbf{k}}$ of the modulated Fourier expansion (13) is based on an expansion

$$\begin{aligned} z_j^{\mathbf{k}} &= \delta^{m(j,\mathbf{k})} \sum_{m=m(j,\mathbf{k})}^{2K-1} \delta^{(m-m(j,\mathbf{k}))(1-2\nu)} z_{j,m}^{\mathbf{k}} \\ &= \delta^{2m(j,\mathbf{k})\nu} \sum_{m=m(j,\mathbf{k})}^{2K-1} \delta^{m(1-2\nu)} z_{j,m}^{\mathbf{k}} \end{aligned} \tag{26}$$

with polynomials $z_{j,m}^{\mathbf{k}} = z_{j,m}^{\mathbf{k}}(\delta^\nu t)$ and with

$$m(j,\mathbf{k}) = \max\Big(e(j), \max_{l:\ k_l \neq 0} e(l)\Big).$$

This can be motivated by the fact that on the one hand we expect $z_j^{\mathbf{k}} = \mathcal{O}(\delta^{m(j,\mathbf{k})})$ from the analysis of the exact solution in [12], but on the other hand we do not expect an expansion in true powers of δ as in [12] because of the non-resonance condition (12) involving δ^ν.

Our goal now is to derive equations for the modulation coefficient functions $z_{j,m}^{\mathbf{k}}$ in the ansatz (26).

Inserting this ansatz into (14), expanding $z_j^{\mathbf{k}}(\delta^\nu(t \pm \tau))$ in a Taylor series and requiring that powers of $\delta^{1-2\nu}$ agree on both sides yields, neglecting the defect d_j^k,

$$\begin{aligned} 4 s_{\langle j\rangle - \mathbf{k}} s_{\langle j\rangle + \mathbf{k}} z_{j,m}^{\mathbf{k}} + 2\mathrm{i}\tau\delta^\nu s_{2\mathbf{k}} \dot{z}_{j,m}^{\mathbf{k}} + \tau^2\delta^{2\nu} c_{2\mathbf{k}} \ddot{z}_{j,m}^{\mathbf{k}} + \dots \\ = \tau^2 \psi(\tau\omega_j) \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{j_1+j_2\equiv j} \delta^{2(m(j_1,\mathbf{k}^1)+m(j_2,\mathbf{k}^2)-m(j,\mathbf{k}))\nu} \\ \sum_{m_1+m_2=m} \phi(\tau\omega_{j_1}) z_{j_1,m_1}^{\mathbf{k}^1} \phi(\tau\omega_{j_2}) z_{j_2,m_2}^{\mathbf{k}^2}. \end{aligned} \tag{27}$$

Here, we use the notation $s_{\mathbf{k}} = \sin(\frac{\tau}{2}\mathbf{k}\cdot\boldsymbol{\omega})$ and $c_{\mathbf{k}} = \cos(\frac{\tau}{2}\mathbf{k}\cdot\boldsymbol{\omega})$ and the fact that $\cos x - \cos y = 2\sin(\frac{y-x}{2})\sin(\frac{y+x}{2})$. All functions in (27) are evaluated at $\delta^\nu t$ and dots denote derivatives with respect to the slow time scale $\delta^\nu t$. Note that $m(j,\mathbf{k}) \leq m(j_1,\mathbf{k}^1) + m(j_2,\mathbf{k}^2)$ if $\mathbf{k} = \mathbf{k}^1 + \mathbf{k}^2$ and $j \equiv j_1 + j_2 \bmod 2M$, and hence the power of δ on the right-hand side of (27) is small. We recall that $\equiv$ denotes the congruence modulo $2M$.

Remark 8. For $\nu = 0$ we recover in (27), after division by τ^2 and in the limit $\tau \to 0$, the system of equations which was used in the case of the exact solution, see [12, Equation (28)]. For $\nu > 0$, the above construction becomes significantly more involved than there. The reason is that the non-resonance condition (12a) only allows us to bound the factor $s_{\langle j\rangle-\mathbf{k}} s_{\langle j\rangle+\mathbf{k}}$ on the left-hand side of (27) from below by $\gamma^2\tau^2\delta^{2\nu}$. As we will see in the proof of Lemma 9 below, we can compensate this possibly small factor with an additional $\delta^{2\nu}$ on the right-hand side (together with τ^2), which we gain from the special choice (26) as ansatz for $z_j^{\mathbf{k}}$.

In addition to (27), we get from condition (17) that

$$\sum_{\mathbf{k}\in\mathcal{K}_j} \delta^{2(m(j,\mathbf{k})-e(j))\nu} z_{j,m}^{\mathbf{k}}(0) = \begin{cases} \delta^{-e(j)} u_j^0, & m = e(j), \\ 0, & \text{else}, \end{cases} \tag{28a}$$

$$\begin{aligned} \sum_{\mathbf{k}\in\mathcal{K}_j} \delta^{2(m(j,\mathbf{k})-e(j))\nu} \Big(2\mathrm{i} s_{2\mathbf{k}} z_{j,m}^{\mathbf{k}}(0) + 2\tau\delta^\nu c_{2\mathbf{k}} \dot{z}_{j,m}^{\mathbf{k}}(0) + \mathrm{i}\tau^2\delta^{2\nu} s_{2\mathbf{k}} \ddot{z}_{j,m}^{\mathbf{k}}(0) + \dots\Big) \\ = \begin{cases} 2\tau \operatorname{sinc}(\tau\omega_j)\delta^{-e(j)} \dot{u}_j^0, & m = e(j), \\ 0, & \text{else}, \end{cases} \end{aligned} \tag{28b}$$

where we use again a Taylor expansion of the modulation functions.

4.2 Construction of modulation functions

We construct polynomial modulation functions of the form (26) by solving (27) and (28) consecutively for $m = 1, 2, \dots, 2K-1$. Assuming that we have computed polynomials $z_{j,m'}^{\mathbf{k}}$ for $m' < m$ (and setting $z_{j,m'}^{\mathbf{k}} = 0$ for $m' < m(j,\mathbf{k})$, for convenience), the construction relies on the observation that only already computed functions $z_{j',m'}^{\mathbf{k}'}$ appear on the right-hand side of (27). This equation is thus, for $z(s) = z_{j,m}^{\mathbf{k}}(s)$, of the form

$$\alpha_0 z(s) - \alpha_1 \dot{z}(s) - \alpha_2 \ddot{z}(s) - \cdots - \alpha_L z^{(L)}(s) = p(s) \tag{29}$$

with coefficients $\alpha_0, \ldots, \alpha_L$ and a polynomial p.

For $\mathbf{k} \in \mathcal{K}_j$ with $\mathbf{k} \neq \pm\langle j \rangle$, the coefficient α_0 in this equation is nonzero (by the non-resonance condition (12)), and the unique polynomial solution of this equation is given by

$$z(s) = \sum_{k=0}^{\deg(p)} \left(\frac{\alpha_1}{\alpha_0} \frac{\mathrm{d}}{\mathrm{d}s} + \frac{\alpha_2}{\alpha_0} \frac{\mathrm{d}^2}{\mathrm{d}s^2} + \cdots + \frac{\alpha_L}{\alpha_0} \frac{\mathrm{d}^L}{\mathrm{d}s^L} \right)^k \frac{1}{\alpha_0} p(s). \tag{30}$$

The modulation coefficient functions constructed in this way are called off-diagonal modulation coefficient functions.

For $\mathbf{k} = \pm\langle j \rangle$, the coefficient α_0 is zero, and the polynomial solutions of (29) are given by

$$z(s) = z(0) + \int_0^s \dot{z}(\sigma) \, \mathrm{d}\sigma \tag{31a}$$

with

$$\dot{z}(s) = \sum_{k=0}^{\deg(p)} \left(\frac{\alpha_2}{\alpha_1} \frac{\mathrm{d}}{\mathrm{d}s} + \cdots + \frac{\alpha_L}{\alpha_1} \frac{\mathrm{d}^{L-1}}{\mathrm{d}s^{L-1}} \right)^k \frac{(-1)^{k+1}}{\alpha_1} p(s). \tag{31b}$$

In (31a), the initial value $z(0)$ is still a free parameter. We use (28) to fix it. Adding and subtracting the two equations of (28), after multiplying the first equation with $2\mathrm{i}\omega_j$ and the second one with $\omega_j / s_{2\langle j \rangle} = 1/(\tau \operatorname{sinc}(\tau\omega_j))$, gives us

$$\begin{aligned} 2\mathrm{i}\omega_j z_{j,m}^{\pm\langle j \rangle}(0) = -\mathrm{i} &\sum_{\substack{\mathbf{k} \in \mathcal{K}_j \\ \mathbf{k} \neq \langle j \rangle,\, \mathbf{k} \neq -\langle j \rangle}} \delta^{2(m(j,\mathbf{k}) - e(j))\nu} \left(\omega_j \pm (\mathbf{k} \cdot \boldsymbol{\omega}) \frac{\operatorname{sinc}(\tau(\mathbf{k} \cdot \boldsymbol{\omega}))}{\operatorname{sinc}(\tau\omega_j)} \right) z_{j,m}^{\mathbf{k}}(0) \\ &\mp \sum_{\mathbf{k} \in \mathcal{K}_j} \frac{\delta^{2(m(j,\mathbf{k}) - e(j))\nu}}{\operatorname{sinc}(\tau\omega_j)} \left(\delta^{\nu} c_{2\mathbf{k}} \dot{z}_{j,m}^{\mathbf{k}}(0) + \tfrac{\mathrm{i}}{2} \tau \delta^{2\nu} s_{2\mathbf{k}} \ddot{z}_{j,m}^{\mathbf{k}}(0) + \ldots \right) \\ &+ \delta^{-e(j)} \begin{cases} \mathrm{i}\omega_j u_j^0 \pm \dot{u}_j^0, & m = e(j), \\ 0, & \text{else}. \end{cases} \end{aligned} \tag{32}$$

Note again that this equation becomes for $\nu = 0$ and in the limit $\tau \to 0$ the corresponding equation for the exact solution, see [12, Equation (31)]. The modulation coefficient functions constructed with (31) and (32) are called diagonal modulation coefficient functions.

Going carefully through this construction, we verify as in [12, Section 4.1] the following properties of the constructed functions $z_{j,m}^{\mathbf{k}}$. Using $e(j) \leq e(j_1) + e(j_2)$ for $j \equiv j_1 + j_2$, $\mu(\mathbf{k}) \leq \mu(\mathbf{k}^1) + \mu(\mathbf{k}^2)$ for $\mathbf{k} = \mathbf{k}^1 + \mathbf{k}^2$ and $e(j) = \mu(\mathbf{k}) < \mu(\mathbf{k}^1) + \mu(\mathbf{k}^2)$ for $\mathbf{k} = \pm\langle j \rangle = \mathbf{k}^1 + \mathbf{k}^2$ with μ defined in (11), we see that they are polynomials of degree

$$\deg\bigl(z_{j,m}^{\mathbf{k}}\bigr) \leq m - \max(e(j), \mu(\mathbf{k})), \tag{33}$$

where a negative degree corresponds to the zero polynomial. Moreover, the polynomial $z_{j,m}^{\mathbf{k}}$ can be different from the zero polynomial only in the two cases

$$\text{case 1:} \quad |j| \leq m, \quad \mu(\mathbf{k}) \leq m, \quad k_l = 0 \text{ for } l \geq \min(m+1, K), \tag{34}$$

$$\begin{aligned} \text{case 2:} \quad m \geq K, \quad &|(j - r) \bmod 2M| \geq K, \quad |r| \leq m - K, \\ &\mathbf{k} = \pm\langle (j - r) \bmod 2M \rangle + \bar{\mathbf{k}}, \quad \mu(\bar{\mathbf{k}}) \leq m - K. \end{aligned} \tag{35}$$

This explains how the set $\mathcal{K}$ of (10) is built up. The two cases follow from the decomposition

$$2\phi(\tau\omega_{j_1})\phi(\tau\omega_{j_2})\sum_{l=1}^{m-K} z_{j_1,m-l}^{\mathbf{k}^1} z_{j_2,l}^{\mathbf{k}^2} + \phi(\tau\omega_{j_1})\phi(\tau\omega_{j_2}) \sum_{\substack{m_1+m_2=m\\ m_1<K,\, m_2<K}} z_{j_1,m_1}^{\mathbf{k}^1} z_{j_2,m_2}^{\mathbf{k}^2}$$

of the last line in (27), where the first term is only present for $m > K$: the functions $z_{j_i,m_i}^{\mathbf{k}^i}$ in the second sum and the function $z_{j_2,l}^{\mathbf{k}^2}$ in the first sum belong inductively to the first case above, whereas the function $z_{j_1,m-l}^{\mathbf{k}^1}$ in the first sum belongs either to the first or to the second case, leading to a function $z_{j,m}^{\mathbf{k}}$ belonging to the first or second case, respectively. In addition, we have $\overline{z_{j,m}^{\mathbf{k}}} = z_{-j,m}^{-\mathbf{k}}$ and

$$z_{j,e(l)}^{\pm\langle l\rangle} = 0 \qquad \text{if} \qquad |l| \neq |j|. \tag{36}$$

4.3 Bounds of the modulation functions

For polynomials $z = z(s)$, we introduce the norm

$$|||z|||_t = \sum_{l\geq 0} \frac{1}{l!}\left|\frac{\mathrm{d}^l}{\mathrm{d}s^l} z(s)\right|_{s=\delta^\nu t}$$

for $0 \leq t \leq 1$. It has the properties that $|||z \cdot w|||_t \leq |||z|||_t \cdot |||w|||_t$ and $|||\dot{v}|||_t \leq \deg(v)\, |||v|||_t$. With this norm, we have the following discrete counterpart of [12, Lemma 1].

Lemma 9. *For $m = 1, \dots, 2K-1$ and $0 \leq t \leq 1$ we have*

$$\sum_{j=-M}^{M-1} \sigma_j \Bigl(|\mathrm{sinc}(\tau\omega_j)|^{-1} \sum_{\pm\langle j\rangle \neq \mathbf{k} \in \mathcal{K}_j} \gamma_j^{\mathbf{k}} \, |||z_{j,m}^{\mathbf{k}}|||_t \Bigr)^2 \lesssim 1,$$

$$\sum_{j=-M}^{M-1} \sigma_j \Bigl(|\mathrm{sinc}(\tau\omega_j)|^{-1} \gamma_j^{\pm\langle j\rangle} \, |||\dot{z}_{j,m}^{\pm\langle j\rangle}|||_t \Bigr)^2 \lesssim 1, \qquad \sum_{j=-M}^{M-1} \sigma_j \Bigl(\gamma_j^{\pm\langle j\rangle} \, |||z_{j,m}^{\pm\langle j\rangle}|||_t \Bigr)^2 \lesssim 1$$

with the additional weight

$$\gamma_j^{\mathbf{k}} = \max(1, \omega_j, |\mathbf{k}\cdot\boldsymbol{\omega}|).$$

Proof. The statement is shown by induction on m, the case $m = 0$ being clear by notation ($z_{j,0}^{\mathbf{k}} = 0$).

(a) We first consider the case $\mathbf{k} \neq \pm\langle j\rangle$. Within this case, we first consider the case $|j| \leq K$, in which the strong non-resonance condition (12b) holds. In the notation (29) of (27) we thus have $1/|\alpha_0| \leq \gamma^{-2}\tau^{-2}$ and $|\alpha_l|/|\alpha_0| \leq \gamma^{-1}\delta^\nu$, where we have used $|s_{2\mathbf{k}}| \leq |s_{\langle j\rangle+\mathbf{k}}| + |s_{\langle j\rangle-\mathbf{k}}|$ for $l = 1$. Using the symplecticity (7), the boundedness of ϕ, $m(j_1, \mathbf{k}^1) + m(j_2, \mathbf{k}^2) \geq m(j, \mathbf{k})$ and the properties $|||z \cdot w|||_t \leq |||z|||_t \cdot |||w|||_t$ and $|||\dot{p}|||_t \leq \deg(p)\, |||p|||_t$, we get

$$|||z_{j,m}^{\mathbf{k}}|||_t \lesssim |\mathrm{sinc}(\tau\omega_j)| \sum_{j_1+j_2\equiv j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{m_1+m_2=m} |||z_{j_1,m_1}^{\mathbf{k}^1}|||_t |||z_{j_2,m_2}^{\mathbf{k}^2}|||_t$$

for the solution $z = z_{j,m}^{\mathbf{k}}$ of (29) given by (30).

The same estimate also holds in the case $\mathbf{k} \neq \pm\langle j \rangle$ if $|j| > K$. This can be seen as follows. For these indices, only the weaker non-resonance condition (12a) holds, and hence we only have $1/|\alpha_0| \leq \gamma^{-2}\tau^{-2}\delta^{-2\nu}$ and $|\alpha_l|/|\alpha_0| \leq \gamma^{-1}$. The problematic factor $\delta^{-2\nu}$, however, can be compensated with the power of δ in the polynomial p on the right-hand side of (29). In fact, we have in this case

$$m(j,\mathbf{k}) = e(j) = K < e(j_1) + e(j_2) \leq m(j_1,\mathbf{k}^1) + m(j_2,\mathbf{k}^2)$$

since $|j| > K$, and hence there is an additional factor $\delta^{2\nu}$ in the polynomial p.

With $\gamma_j^{\mathbf{k}} \lesssim \gamma_{j_1}^{\mathbf{k}^1}\gamma_{j_2}^{\mathbf{k}^2}$, the algebra property (18) and $1 \leq |\mathrm{sinc}(\tau\omega_j)|^{-1}$ we finally get the first claimed estimate.

(b) In the case $\mathbf{k} = \pm\langle j \rangle$, the absolute value of the coefficient α_1 is bounded from below by $\gamma\tau^2\delta^{2\nu}$ by the (weaker) non-resonance condition (12a). Also in this case, we have an additional factor $\delta^{2\nu}$ in the polynomial on the right-hand side of (29) since $m(j,\mathbf{k}) = e(j)$ and $m(j_1,\mathbf{k}^1) \geq e(j)$ or $m(j_2,\mathbf{k}^2) \geq e(j)$ for $\mathbf{k} = \pm\langle j \rangle = \mathbf{k}^1 + \mathbf{k}^2$. This shows that

$$|||\dot{z}_{j,m}^{\pm\langle j \rangle}|||_t \lesssim |\mathrm{sinc}(\tau\omega_j)| \sum_{j_1+j_2\equiv j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\pm\langle j \rangle} \sum_{m_1+m_2=m} |||z_{j_1,m_1}^{\mathbf{k}^1}|||_t |||z_{j_2,m_2}^{\mathbf{k}^2}|||_t$$

for the solution $\dot{z} = \dot{z}_{j,m}^{\pm\langle j \rangle}$ of (29) given by (31). As in (a), this yields the second estimate of the lemma.

(c) With the results of (a) and (b) and with the assumption (20) on the initial values, we get for the initial values $z_{j,m}^{\pm\langle j \rangle}$ constructed with (32) that

$$\sum_{j=-M}^{M-1} \sigma_j \left(\gamma_j^{\pm\langle j \rangle} \big|z_{j,m}^{\pm\langle j \rangle}(0)\big|\right)^2 \lesssim 1.$$

This yields the last estimate of the lemma. □

4.4 Proof of Theorem 4

The proof of Theorem 4 relies on the alternative form (24) of the almost-invariant energies $\mathcal{E}_l(t)$ of (23). Note that $z_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)})$ if $k_l \neq 0$ (by (26) and (33), since then $\mu(\mathbf{k}) \geq e(l)$ and $m(j,\mathbf{k}) \geq e(l)$). This shows that $\mathcal{E}_l(t) = \mathcal{O}(\delta^{2e(l)})$.

To get the precise estimate of Theorem 4, we multiply (24) with $\sigma_l\delta^{-2e(l)}$ and sum over l, we use

$$\sum_{l=0}^{M} \sigma_l |k_l| \omega_l \lesssim \sigma_j \gamma_j^{\mathbf{k}} \qquad \text{if} \qquad \mathbf{k} \in \mathcal{K}_j, \tag{37}$$

and we apply the non-resonance condition (12a), the Cauchy-Schwarz inequality and the estimates of Lemma 9.

4.5 Bounds of the defect

When constructing $z_j^{\mathbf{k}}$ with the expansion (26), the defect $d_j^{\mathbf{k}}$, $\mathbf{k} \in \mathcal{K}_j$, in (14) is given by

$$\begin{aligned} d_j^{\mathbf{k}} = -\sum_{m=2K}^{2(2K-1)} \delta^{m(1-2\nu)} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{j_1+j_2\equiv j} \delta^{2(m(j_1,\mathbf{k}^1)+m(j_2,\mathbf{k}^2))\nu} \\ \sum_{m_1+m_2=m} \phi(\tau\omega_{j_1}) z_{j_1,m_1}^{\mathbf{k}^1} \phi(\tau\omega_{j_2}) z_{j_2,m_2}^{\mathbf{k}^2}. \end{aligned} \tag{38}$$

Although we consider system (14) defining the defect $d_j^{\mathbf{k}}$ only for $\mathbf{k} \in \mathcal{K}_j$, we use this formula to define $d_j^{\mathbf{k}}$ also for $\mathbf{k} \notin \mathcal{K}_j$. This will be helpful below.

Lemma 10. *The defect $d_j^{\mathbf{k}}$ in (14) given by (38) satisfies, for $0 \leq t \leq 1$,*

$$\sum_{j=-M}^{M-1} \sigma_j \Big(\sum_{\mathbf{k}\in\mathbb{Z}^{M+1}} \gamma_j^{\mathbf{k}} \delta^{-2m(j,\mathbf{k})\nu} |||d_j^{\mathbf{k}}|||_t \Big)^2 \lesssim \delta^{4K(1-2\nu)}$$

with $\gamma_j^{\mathbf{k}}$ as in Lemma 9.

Proof. The result follows with Lemma 9 and the arguments used in its proof. □

On a higher level, the approximation $\widetilde{\mathbf{u}}^n = (\widetilde{u}_{-M}^n, \ldots, \widetilde{u}_{M-1}^n)^{\mathrm{T}}$ of (13) to the numerical solution $\mathbf{u}^n$ then has a defect $\mathbf{e}^n = (e_{-M}^n, \ldots, e_{M-1}^n)^{\mathrm{T}}$ when inserted into the numerical method (6a):

$$\widetilde{\mathbf{u}}^{n+1} - 2\cos(\tau\boldsymbol{\Omega})\widetilde{\mathbf{u}}^n + \widetilde{\mathbf{u}}^{n-1} = \tau^2 \boldsymbol{\Psi}\big((\boldsymbol{\Phi}\widetilde{\mathbf{u}}^n) * (\boldsymbol{\Phi}\widetilde{\mathbf{u}}^n)\big) + \tau^2 \boldsymbol{\Psi}\mathbf{e}^n. \tag{39}$$

By construction, this defect is given by

$$e_j^n = \sum_{\mathbf{k}\in\mathcal{K}_j} d_j^{\mathbf{k}} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n} - \sum_{\mathbf{k}\in\mathbb{Z}^{M+1}\setminus\mathcal{K}_j} \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{j_1+j_2\equiv j} \phi(\tau\omega_{j_1}) z_{j_1}^{\mathbf{k}^1} \phi(\tau\omega_{j_2}) z_{j_2}^{\mathbf{k}^2} \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n},$$

where the defect $d_j^{\mathbf{k}}$ and the modulation functions are evaluated at $\delta^\nu t_n$.

Lemma 11. *The defect $\mathbf{e}^n$ in (39) satisfies, for $0 \leq t_n = n\tau \leq 1$,*

$$\|\boldsymbol{\Omega}\mathbf{e}^n\| \lesssim \delta^{2K(1-2\nu)}$$

with the norm $\|\cdot\|$ of Section 3.2.

Proof. We use that the set $\mathcal{K}_j$ was constructed in such a way that $\mathbf{k}$ belongs to $\mathcal{K}_j$ if $\mathbf{k} = \mathbf{k}^1 + \mathbf{k}^2$, $j \equiv j_1 + j_2$ and $m_1 + m_2 \leq 2K - 1$ for $\mathbf{k}^1, \mathbf{k}^2, j_1, j_2, m_1, m_2$ with $z_{j_1,m_1}^{\mathbf{k}^1} \neq 0$ and $z_{j_2,m_2}^{\mathbf{k}^2} \neq 0$. Using (38) to define $d_j^{\mathbf{k}}$ also for $\mathbf{k} \notin \mathcal{K}_j$, this yields the compact form

$$e_j^n = \sum_{\mathbf{k}\in\mathbb{Z}^{M+1}} d_j^{\mathbf{k}}(\delta^\nu t_n) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})t_n}$$

for the defect in (39). The statement of the lemma thus follows from Lemma 10. □

4.6 Bounds of the remainder

With the constructed modulation functions $z_j^{\mathbf{k}}$, we get the approximation $\widetilde{u}_j^n$ of (13) to the numerical solution u_j^n. From (16), we also get an approximation $\dot{\widetilde{u}}_j^n$ to $\dot{u}_j^n$. In the following lemma, we establish a bound for the approximation error.

Lemma 12. *The remainders $u_j^n - \widetilde{u}_j^n$ and $\dot{u}_j^n - \dot{\widetilde{u}}_j^n$ satisfy, for $0 \leq t_n \leq 1$,*

$$\big\|\boldsymbol{\Omega}\big(\mathbf{u}^n - \widetilde{\mathbf{u}}^n\big)\big\| + \big\|\dot{\mathbf{u}}^n - \dot{\widetilde{\mathbf{u}}}^n\big\| \lesssim \delta^{2K(1-2\nu)}.$$

Proof. (a) In a first step, we write the numerical scheme (6) and its approximation by a modulated Fourier expansion in one-step form. The method in one-step form reads

$$\begin{pmatrix} \mathbf{u}^{n+1} \\ \dot{\mathbf{u}}^{n+1} \end{pmatrix} = \begin{pmatrix} \cos(\tau\boldsymbol{\Omega}) & \boldsymbol{\Omega}^{-1}\sin(\tau\boldsymbol{\Omega}) \\ -\boldsymbol{\Omega}\sin(\tau\boldsymbol{\Omega}) & \cos(\tau\boldsymbol{\Omega}) \end{pmatrix} \begin{pmatrix} \mathbf{u}^n \\ \dot{\mathbf{u}}^n \end{pmatrix} + \frac{\tau}{2} \begin{pmatrix} \tau\boldsymbol{\Psi}\mathbf{g}^n \\ \cos(\tau\boldsymbol{\Omega})\boldsymbol{\Phi}\mathbf{g}^n + \boldsymbol{\Phi}\mathbf{g}^{n+1} \end{pmatrix} \tag{40}$$

with $\mathbf{g}^n = (\boldsymbol{\Phi}\mathbf{u}^n) * (\boldsymbol{\Phi}\mathbf{u}^n)$, see [15, Section XIII.2.2]. The first line is obtained by adding (6a) and (6c), and the second line is obtained by subtracting these equations with $n+1$ instead of n and using the first line to replace $\mathbf{u}^{n+1}$ as well as the symplecticity (7). In the same way, we derive for the approximations $\widetilde{\mathbf{u}}^n$ and $\dot{\widetilde{\mathbf{u}}}^n$, which satisfy (6c) exactly and (6a) up to a small defect given by (39),

$$\begin{pmatrix} \widetilde{\mathbf{u}}^{n+1} \\ \dot{\widetilde{\mathbf{u}}}^{n+1} \end{pmatrix} = \begin{pmatrix} \cos(\tau\boldsymbol{\Omega}) & \boldsymbol{\Omega}^{-1}\sin(\tau\boldsymbol{\Omega}) \\ -\boldsymbol{\Omega}\sin(\tau\boldsymbol{\Omega}) & \cos(\tau\boldsymbol{\Omega}) \end{pmatrix} \begin{pmatrix} \widetilde{\mathbf{u}}^{n} \\ \dot{\widetilde{\mathbf{u}}}^{n} \end{pmatrix} + \frac{\tau}{2} \begin{pmatrix} \tau\boldsymbol{\Psi}\widetilde{\mathbf{g}}^n \\ \cos(\tau\boldsymbol{\Omega})\boldsymbol{\Phi}\widetilde{\mathbf{g}}^n + \boldsymbol{\Phi}\widetilde{\mathbf{g}}^{n+1} \end{pmatrix} \quad (41)$$

with $\widetilde{\mathbf{g}}^n = \mathbf{g}^n + \mathbf{e}^n$.

(b) In the following, we write

$$\|(\mathbf{v}, \dot{\mathbf{v}})\| = \left(\|\boldsymbol{\Omega}\mathbf{v}\|^2 + \|\dot{\mathbf{v}}\|^2\right)^{1/2} \quad (42)$$

for a norm that is equivalent to the norm considered in the statement of the lemma. By induction on n, we prove that the numerical solutions stays small in this norm,

$$\|(\mathbf{u}^n, \dot{\mathbf{u}}^n)\| \leq \left(\sqrt{2C_0} + t_n\right)\delta^{1-2\nu} \qquad \text{for} \qquad 0 \leq t_n \leq 1$$

with C_0 from (20), provided that $\delta^{1-2\nu}$ is sufficiently small. This is true for $n = 0$ by the choice of the initial value (20). For $n > 0$, we first note that $\|\boldsymbol{\Omega}\mathbf{u}^n\| \leq C\delta^{1-2\nu}$ by induction and by the algebra property (19). We then observe that the matrix appearing in the one-step formulation (40) of the method preserves the norm (42). Moreover, by the algebra properties (18) and (19), the second term in the one-step formulation (40) can be estimated in the norm (42) by $\tau\delta^{2(1-2\nu)}$ up to a constant. This implies the stated bound of $\|(\mathbf{u}^n, \dot{\mathbf{u}}^n)\|$.

Using in addition Lemma 11 on the defect $\mathbf{e}^n$, we get for the modulated Fourier expansion with the help of (41) in the same way the bound

$$\|(\widetilde{\mathbf{u}}^n, \dot{\widetilde{\mathbf{u}}}^n)\| \leq \left(\sqrt{2C_0} + t_n\right)\delta^{1-2\nu} \qquad \text{for} \qquad 0 \leq t_n \leq 1.$$

(c) For the difference $(\mathbf{u}^n - \widetilde{\mathbf{u}}^n, \dot{\mathbf{u}}^n - \dot{\widetilde{\mathbf{u}}}^n)$, we subtract the above one-step formulations, and then proceed as in the proof of the smallness of $(\mathbf{u}^n, \dot{\mathbf{u}}^n)$ and $(\widetilde{\mathbf{u}}^n, \dot{\widetilde{\mathbf{u}}}^n)$ in (b). We use the smallness of $(\mathbf{u}^n, \dot{\mathbf{u}}^n)$ and $(\widetilde{\mathbf{u}}^n, \dot{\widetilde{\mathbf{u}}}^n)$ together with the algebra property (18) and Lemma 11 to control the difference $\mathbf{g}^n - \widetilde{\mathbf{g}}^n = -\mathbf{e}^n$. The mentioned lemma on the defect introduces the small parameter $\delta^{2K(1-2\nu)}$. □

With the results proven in this section, the proofs of Theorems 3 and 4 are complete.

5 Almost-invariant energies: Proofs of Theorems 5–7

5.1 Almost-invariant energies: Proof of Theorem 5

For the proof of Theorem 5, we sum the equality (22) to get

$$\mathcal{E}_l(t_n) - \mathcal{E}_l(0) = -\frac{\mathrm{i}}{2} \sum_{j=-M}^{M-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \sum_{\tilde{n}=1}^{n} \tau k_l \omega_l \phi(\tau\omega_j) z_{-j}^{-\mathbf{k}}(\delta^\nu t_{\tilde{n}}) d_j^{\mathbf{k}}(\delta^\nu t_{\tilde{n}}).$$

Note that, for $k_l \neq 0$, $z_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)})$ by Lemma 9 and $d_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)+(2K-e(l))(1-2\nu)})$ by Lemma 10 (since then $m(j,\mathbf{k}) \geq e(l)$). This shows that the variation in the almost-invariant energy $\mathcal{E}_l$ is indeed of the claimed order, and it shows the additional statement for $l = 1$.

To get the precise estimate of Theorem 5, we multiply the above equation with $\sigma_l \delta^{-2e(l)}$ and sum over l, we use (37) and $n\tau = t_n \leq 1$, and we apply the Cauchy-Schwarz inequality and the estimates of Lemma 9 and Lemma 10.

5.2 Transitions in the almost-invariant energies: Proof of Theorem 6

We consider the situation of Theorem 6, with a modulated Fourier expansion with coefficients $z_j^{\mathbf{k}}(\delta^\nu t)$, $0 \leq t \leq 1$, constructed from $(\mathbf{u}^0, \dot{\mathbf{u}}^0)$ and a modulated Fourier expansion with coefficients $\tilde{z}_j^{\mathbf{k}}(\delta^\nu t)$, $0 \leq t \leq 1$, constructed from $(\mathbf{u}^N, \dot{\mathbf{u}}^N)$, where N is such that $t_N = N\tau = 1$. For studying the difference of the corresponding almost-invariant energies, we first consider the difference of the modulation functions themselves.

Lemma 13. *Under the assumptions of Theorem 6, we have*

$$\sum_{j=-M}^{M-1} \sigma_j \Big(\sum_{\mathbf{k}\in\mathcal{K}_j} \gamma_j^{\mathbf{k}} \delta^{-2m(j,\mathbf{k})\nu} \big|\big|\big| z_j^{\mathbf{k}}(\cdot + \delta^\nu) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} - \tilde{z}_j^{\mathbf{k}} \big|\big|\big|_0 \Big)^2 \lesssim \delta^{4K(1-2\nu)}.$$

Proof. Recall that the functions $z_j^{\mathbf{k}}$ and $\tilde{z}_j^{\mathbf{k}}$ are constructed with the expansion (26). We consider here the truncated expansions

$$\big[z_j^{\mathbf{k}}\big]^\ell(\delta^\nu t) = \delta^{m(j,\mathbf{k})} \sum_{m=m(j,\mathbf{k})}^{\ell} \delta^{(m-m(j,\mathbf{k}))(1-2\nu)} z_{j,m}^{\mathbf{k}}(\delta^\nu t),$$

$$\big[\tilde{z}_j^{\mathbf{k}}\big]^\ell(\delta^\nu t) = \delta^{m(j,\mathbf{k})} \sum_{m=m(j,\mathbf{k})}^{\ell} \delta^{(m-m(j,\mathbf{k}))(1-2\nu)} \tilde{z}_{j,m}^{\mathbf{k}}(\delta^\nu t)$$

for $\ell = 1, \ldots, 2K-1$. Note that these truncations coincide for $\ell = 2K-1$ with $z_j^{\mathbf{k}}$ and $\tilde{z}_j^{\mathbf{k}}$, respectively. We therefore study the differences

$$\big[f_j^{\mathbf{k}}\big]^\ell(\delta^\nu t) = \delta^{-2m(j,\mathbf{k})\nu} \Big(\big[z_j^{\mathbf{k}}\big]^\ell(\delta^\nu t + \delta^\nu) \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} - \big[\tilde{z}_j^{\mathbf{k}}\big]^\ell(\delta^\nu t) \Big).$$

(a) We first derive equations for these differences from the defining equations (27) and (32). In order to derive an equation from (27) we use $a_1 a_2 - \tilde{a}_1 \tilde{a}_2 = (a_1 - \tilde{a}_1) a_2 + \tilde{a}_1 (a_2 - \tilde{a}_2)$, the symmetry of the sums in j_l and $\mathbf{k}^l$ and $\sum_{m=m(j,\mathbf{k})}^{\ell} \sum_{m_1+m_2=m} \% = \sum_{m_2=m(j_2,\mathbf{k}^2)}^{\ell-1} \sum_{m_1=m(j_1,\mathbf{k}^1)}^{\ell-m_2} \%$. This yields

$$\begin{aligned} &4 s_{\langle j\rangle-\mathbf{k}} s_{\langle j\rangle+\mathbf{k}} \big[f_j^{\mathbf{k}}\big]^\ell + 2\mathrm{i}\tau\delta^\nu s_{2\mathbf{k}} \big[\dot{f}_j^{\mathbf{k}}\big]^\ell + \tau^2 \delta^{2\nu} c_{2\mathbf{k}} \big[\ddot{f}_j^{\mathbf{k}}\big]^\ell + \ldots \\ &\quad = \tau^2 \psi(\tau\omega_j) \sum_{\mathbf{k}^1+\mathbf{k}^2=\mathbf{k}} \sum_{j_1+j_2\equiv j} \delta^{2(m(j_1,\mathbf{k}^1)+m(j_2,\mathbf{k}^2)-m(j,\mathbf{k}))\nu} \\ &\qquad \sum_{m_2=m(j_2,\mathbf{k}^2)}^{\ell-1} \delta^{m_2(1-2\nu)} \phi(\tau\omega_{j_1}) \big[f_{j_1}^{\mathbf{k}^1}\big]^{\ell-m_2} \phi(\tau\omega_{j_2}) \Big(\tilde{z}_{j_2,m_2}^{\mathbf{k}^2} + z_{j_2,m_2}^{\mathbf{k}^2} \mathrm{e}^{\mathrm{i}(\mathbf{k}^2\cdot\boldsymbol{\omega})} \Big), \end{aligned}$$

where $\widetilde{z}_j^{\mathbf{k}}$ and $f_j^{\mathbf{k}}$ are evaluated at $\delta^\nu t$ and $z_j^{\mathbf{k}}$ is evaluated at $\delta^\nu t + \delta^\nu$. Another equation can be derived from (32) for $\widetilde{z}$ (with u_j^N and $\dot{u}_j^N$ instead of u_j^0 and $\dot{u}_j^0$) and from (see (16))

$$u_j^N = \sum_{\mathbf{k}\in\mathcal{K}_j} \big[z_j^{\mathbf{k}}\big]^\ell(\delta^\nu)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} + \big[r_j^N\big]^\ell$$

$$\dot{u}_j^N = \big(\tau\,\mathrm{sinc}(\tau\omega_j)\big)^{-1} \sum_{\mathbf{k}\in\mathcal{K}_j} \Big(\mathrm{i}s_{2\mathbf{k}}\big[z_j^{\mathbf{k}}\big]^\ell(\delta^\nu)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} + \tau\delta^\nu c_{2\mathbf{k}}\big[\dot{z}_j^{\mathbf{k}}\big]^\ell(\delta^\nu)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} + \tfrac{\mathrm{i}}{2}\tau^2\delta^{2\nu}s_{2\mathbf{k}}\big[\ddot{z}_j^{\mathbf{k}}\big]^\ell(\delta^\nu)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} + \dots\Big) + \big[\dot{r}_j^N\big]^\ell$$

with the remainders $[r_j^N]^\ell$ and $[\dot{r}_j^N]^\ell$ of the truncated expansion. This yields, for $\ell \geq e(j)$,

$$\begin{aligned} 2\mathrm{i}\omega_j\big[f_j^{\pm\langle j\rangle}\big]^\ell = &-\mathrm{i}\sum_{\substack{\mathbf{k}\in\mathcal{K}_j\\ \mathbf{k}\neq\langle j\rangle,\,\mathbf{k}\neq-\langle j\rangle}} \delta^{2(m(j,\mathbf{k})-e(j))\nu}\Big(\omega_j \pm (\mathbf{k}\cdot\boldsymbol{\omega})\frac{\mathrm{sinc}(\tau(\mathbf{k}\cdot\boldsymbol{\omega}))}{\mathrm{sinc}(\tau\omega_j)}\Big)\big[f_j^{\mathbf{k}}\big]^\ell \\ &\mp \sum_{\mathbf{k}\in\mathcal{K}_j} \frac{\delta^{2(m(j,\mathbf{k})-e(j))\nu}}{\mathrm{sinc}(\tau\omega_j)}\Big(\delta^\nu c_{2\mathbf{k}}\big[\dot{f}_j^{\mathbf{k}}\big]^\ell + \tfrac{\mathrm{i}}{2}\tau\delta^{2\nu}s_{2\mathbf{k}}\big[\ddot{f}_j^{\mathbf{k}}\big]^\ell + \dots\Big) \\ &+ \delta^{-2e(j)\nu}\Big(\mathrm{i}\omega_j\big[r_j^N\big]^\ell \pm \big[\dot{r}_j^N\big]^\ell\Big), \end{aligned}$$

where $f_j^{\mathbf{k}}$ is evaluated at 0.

(b) As in Sections 4.5 and 4.6, we obtain for the remainders $[r_j^N]^\ell$ and $[\dot{r}_j^N]^\ell$ of the truncated expansions the bound

$$\|\boldsymbol{\Omega}[\mathbf{r}^N]^\ell\| + \|[\dot{\mathbf{r}}^N]^\ell\| \lesssim \delta^{(\ell+1)(1-2\nu)}.$$

Based on the equations derived in (a), we can then show as in the proof Lemma 9 by induction on ℓ that

$$\sum_{j=-M}^{M-1}\sigma_j\Big(\sum_{\mathbf{k}\in\mathcal{K}_j}\gamma_j^{\mathbf{k}}\big|\big|\big|\big[f_j^{\mathbf{k}}\big]^\ell\big|\big|\big|_0\Big)^2 \lesssim \delta^{2(\ell+1)(1-2\nu)}.$$

For $\ell = 2K-1$ this is the bound as stated in the lemma. □

Now we deduce Theorem 6 from this lemma. We consider the difference $\mathcal{E}_l(1) - \widetilde{\mathcal{E}}_l(0)$ of almost-invariants in the form (24). This difference consists of differences of the form

$$z_{-j}^{-\mathbf{k}}(\delta^\nu)\big(z_j^{\mathbf{k}}\big)^{(p)}(\delta^\nu) - \widetilde{z}_{-j}^{-\mathbf{k}}(0)\big(\widetilde{z}_j^{\mathbf{k}}\big)^{(p)}(0)$$

with $p = 0,1,2,\dots$ denoting derivatives. We rewrite these differences using

$$a_1a_2 - \widetilde{a}_1\widetilde{a}_2 = (a_1\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} - \widetilde{a}_1)a_2\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} + \widetilde{a}_1(a_2\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} - \widetilde{a}_2).$$

By the previous Lemma 13 we have $z_j^{\mathbf{k}}(\cdot+\delta^\nu)\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega})} - \widetilde{z}_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)+(2K-e(l))(1-2\nu)})$ if $k_l \neq 0$, and by Lemma 9 we have $z_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)})$ and $\widetilde{z}_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)})$ if $k_l \neq 0$. This yields the claimed order $\delta^{2e(l)+K(1-2\nu)}$ of the transition in the lth almost-invariant energy, and it yields the additional statement for $l = 1$. The precise statement of Theorem 6 is then obtained from Lemmas 9 and 13 together with (37).

5.3 Controlling mode energies by almost-invariant energies: Proof of Theorem 7

The proof of Theorem 7 is done in three steps. We first show in Lemma 14 below that the goal of controlling mode energies can be achieved by controlling certain dominant terms in the modulated Fourier expansion. Then we show, in Lemma 15 below, that these dominant terms can be controlled if only the diagonal ones among them are under control. Finally, we show in Lemma 16 below that these diagonal dominant terms can be described essentially by the almost-invariant energies of the modulated Fourier expansion. Throughout, we assume the conditions of Theorem 7 to be fulfilled.

Lemma 14. *For* $0 \le t \le 1$ *assume that*

$$|\mathrm{sinc}(\tau\omega_j)|^{-1}\gamma_j^{\mathbf{k}}\big|z_{j,e(j)}^{\mathbf{k}}(\delta^\nu t)\big| \le \mathcal{C}_0 \qquad \text{for} \qquad \mathbf{k} \ne \pm\langle j\rangle,\ |j| \le K,\ \mu(\mathbf{k}) \le K,$$
$$\sum_{j=-M}^{M-1} \sigma_j \omega_j^2 \big|z_{j,e(j)}^{\pm\langle j\rangle}(\delta^\nu t)\big|^2 \le \mathcal{C}_0.$$

Then, for $0 \le t_n = n\tau \le 1$,

$$\sum_{l=0}^{M} \sigma_l \delta^{-2e(l)} E_l^n \le \mathcal{C}$$

and

$$\sigma_1 \delta^{-2e(1)} \Big| E_1^n - \omega_1^2 \Big(\big|z_1^{\langle 1\rangle}(\delta^\nu t_n)\big|^2 + \big|z_1^{-\langle 1\rangle}(\delta^\nu t_n)\big|^2 \Big) \Big| \le \mathcal{C}\delta^{2(1-2\nu)},$$

where $\mathcal{C}$ *depends on* $\mathcal{C}_0$*, but is independent of* C_0 *of* (20) *if* $\delta^{1-2\nu}$ *is sufficiently small.*

Proof. We consider the two terms $\frac{1}{2}\omega_l^2|u_l^n|^2$ and $\frac{1}{2}|\dot{u}_l^n|^2$ in E_l^n separately.

(a) For the first term, we note that we have, with the notation $\widetilde{u}$ of (13),

$$|u_l^n| \le |u_l^n - \widetilde{u}_l^n| + \sum_{\mathbf{k}\in\mathcal{K}_l} \big|z_l^{\mathbf{k}}(\delta^\nu t_n)\big|.$$

We then insert the expansion (26) for $z_l^{\mathbf{k}}$, whose dominant term is $\delta^{m(l,\mathbf{k})} z_{l,m(l,\mathbf{k})}^{\mathbf{k}}$. Note that $m(l,\mathbf{k}) \ge e(l)$. This yields

$$\frac{1}{2}\sum_{l=0}^{M} \sigma_l \omega_l^2 \delta^{-2e(l)} |u_l^n|^2 \le \mathcal{C} + C\delta^{2(1-2\nu)},$$

with constants $\mathcal{C}$ and C. The constant $\mathcal{C}$ is the result of estimating the terms $z_{l,m(l,\mathbf{k})}^{\mathbf{k}}$ with $m(l,\mathbf{k}) = e(l)$ using the assumption on these terms (note that $z_{l,m(l,\mathbf{k})}^{\mathbf{k}} = 0$ for $|l| > K$ and $\mathbf{k} \ne \pm\langle l\rangle$ by (34) and (35) since $m(l,\mathbf{k}) = e(l) = K$ in this case). The constant $\mathcal{C}$ thus depends only on $\mathcal{C}_0$ and not on C_0 of (20). The second term $C\delta^{2(1-2\nu)}$ is the result of estimating the terms $z_{l,m(l,\mathbf{k})}^{\mathbf{k}}$ with $m(l,\mathbf{k}) > e(l)$ using Lemma 9 and the remainder term using Lemma 12. The constant C thus depends on C_0 of (20).

(b) For the second term, we proceed similarly. We start from

$$|\dot{u}_l^n| \le \big|\dot{u}_l^n - \dot{\widetilde{u}}_l^n\big| + \sum_{\mathbf{k}\ne\pm\langle l\rangle} \gamma_l^{\mathbf{k}} \frac{|\mathrm{sinc}(\tau(\mathbf{k}\cdot\boldsymbol{\omega}))|}{|\mathrm{sinc}(\tau\omega_l)|} \big|z_l^{\mathbf{k}}(\delta^\nu t_n)\big| + |\mathrm{sinc}(\tau\omega_l)|^{-1} \sum_{\mathbf{k}\in\mathcal{K}_l} \big|\big|\big|\dot{z}_l^{\mathbf{k}}\big|\big|\big|_{t_n},$$

which is obtained using the velocity approximation $\dot{\widetilde{u}}$ given by (16) and a Taylor expansion of $z_l^{\mathbf{k}}(\epsilon^{\nu/2} t_{n\pm1})$. We then insert the expansion (26) of $z_l^{\mathbf{k}}$ and isolate the dominant terms $\delta^{m(l,\mathbf{k})} z^{\mathbf{k}}_{l,m(l,\mathbf{k})}$. Using the assumption on the terms $z^{\mathbf{k}}_{l,e(l)}$ and noting that $\dot z^{\mathbf{k}}_{l,e(l)} = 0$ by (33), this yields similarly as in (a)

$$\frac{1}{2}\sum_{l=0}^{M} \sigma_l \omega_l^2 \delta^{-2e(l)} |\dot u_l^n|^2 \le \mathcal{C} + C\delta^{2(1-2\nu)}.$$

Combining this estimate with the corresponding estimate in (a), we get the estimate of the lemma, provided that $\delta^{1-2\nu}$ is sufficiently small.

(c) The additional estimate of the difference $E_1^n - \omega_1^2(|z_1^{\langle 1\rangle}(\delta^\nu t_n)|^2 + |z_1^{-\langle 1\rangle}(\delta^\nu t_n)|^2)$ is obtained as follows. We first note that, as in [12],

$$\begin{aligned}
2E_1^n &= |\omega_1 u_1^n|^2 + |\dot u_1^n|^2 \\
&= \omega_1^2 \big| z_1^{\langle 1\rangle}(\delta^\nu t_n) \mathrm{e}^{\mathrm{i}\omega_1 t_n} + z_1^{-\langle 1\rangle}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n}\big|^2 \\
&\quad + 2\omega_1^2 \mathrm{Re}\Big(\overline{\eta^n}\big(z_1^{\langle 1\rangle}(\delta^\nu t_n)\mathrm{e}^{\mathrm{i}\omega_1 t_n} + z_1^{-\langle 1\rangle}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n}\big)\Big) + \omega_1^2 |\eta^n|^2 \\
&\quad + \omega_1^2 \big| \mathrm{i} z_1^{\langle 1\rangle}(\delta^\nu t_n) \mathrm{e}^{\mathrm{i}\omega_1 t_n} - \mathrm{i} z_1^{-\langle 1\rangle}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n}\big|^2 \\
&\quad + 2\omega_1 \mathrm{Re}\Big(\overline{\vartheta^n}\big(\mathrm{i}z_1^{\langle 1\rangle}(\delta^\nu t_n)\mathrm{e}^{\mathrm{i}\omega_1 t_n} - \mathrm{i}z_1^{-\langle 1\rangle}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n}\big)\Big) + |\vartheta^n|^2,
\end{aligned}$$

but now with

$$\begin{aligned}
\eta^n &= \sum_{\pm\langle 1\rangle \ne \mathbf{k}\in\mathcal{K}_1} z_1^{\mathbf{k}}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n} + \big(u_1^n - \widetilde u_1^n\big) \\
\vartheta^n &= \sum_{\pm\langle 1\rangle \ne \mathbf{k}\in\mathcal{K}_1} \mathrm{i}(\mathbf{k}\cdot\boldsymbol{\omega}) \frac{\mathrm{sinc}(\tau(\mathbf{k}\cdot\boldsymbol{\omega}))}{\mathrm{sinc}(\tau\omega_1)} z_1^{\mathbf{k}}(\delta^\nu t_n) e^{-\mathrm{i}\omega_1 t_n} + \big(\dot u_1^n - \dot{\widetilde u}_1^n\big) \\
&\qquad + (\mathrm{sinc}(\tau\omega_1))^{-1} \sum_{\mathbf{k}\in\mathcal{K}_1} \big(c_{2\mathbf{k}} \delta^\nu \dot z_1^{\mathbf{k}}(\delta^\nu t_n) + \dots\big) e^{-\mathrm{i}\omega_1 t_n}.
\end{aligned}$$

By Lemma 9 and Lemma 12 we have $|\eta^n| \lesssim \delta^{1+2(1-2\nu)}$ and $|\vartheta^n| \lesssim \delta^{1+2(1-2\nu)}$ since $z^{\mathbf{k}}_{1,1} = z^{\mathbf{k}}_{1,2} = 0$ for $\mathbf{k} \ne \pm\langle 1\rangle$ and $\dot z^{\mathbf{k}}_{1,1} = \dot z^{\mathbf{k}}_{1,2} = 0$ for all $\mathbf{k}$. This yields the claimed estimate. □

Lemma 15. *For $0 \le t \le 1$ assume that*

$$\big|z^{\pm\langle j\rangle}_{j,e(j)}(\delta^\nu t)\big| \le \mathcal{C}_0 \qquad \textit{for} \qquad |j| \le K.$$

Then,

$$|\mathrm{sinc}(\tau\omega_j)|^{-1} \gamma_j^{\mathbf{k}} \big|z^{\mathbf{k}}_{j,e(j)}(\delta^\nu t)\big| \le \mathcal{C}_0 \qquad \textit{for} \qquad \mathbf{k} \ne \pm\langle j\rangle,\ |j| \le K,\ \mu(\mathbf{k}) \le K$$

where $\mathcal{C}$ depends on $\mathcal{C}_0$ but is independent of C_0 of (20).

Proof. This follows from the construction of off-diagonal modulation functions as in the case of the exact solution in [12, Lemma 4], see also the proof of Lemma 9. □

Lemma 16. *For $0 \le t \le 1$ we have*

$$\begin{aligned}
\delta^{-2e(0)} \Big|\mathcal{E}_0(t) - \tfrac{1}{2}\omega_0^2\Big(\big|z_0^{\langle 0\rangle}(\delta^\nu t)\big|^2 + \big|z_0^{-\langle 0\rangle}(\delta^\nu t)\big|^2\Big)\Big| &\lesssim \delta^{1-2\nu} \\
\delta^{-2e(1)} \Big|\mathcal{E}_1(t) - \omega_1^2\Big(\big|z_1^{\langle 1\rangle}(\delta^\nu t)\big|^2 + \big|z_1^{-\langle 1\rangle}(\delta^\nu t)\big|^2\Big)\Big| &\lesssim \delta^{2(1-2\nu)} \\
\sum_{l=1}^{M} \sigma_l \delta^{-2e(l)} \Big|\mathcal{E}_l(t) - \omega_l^2\Big(\big|z_l^{\langle l\rangle}(\delta^\nu t)\big|^2 + \big|z_l^{-\langle l\rangle}(\delta^\nu t)\big|^2\Big)\Big| &\lesssim \delta^{1-2\nu}.
\end{aligned}$$

Proof. We subtract $\omega_l^2(|z_l^{\langle l\rangle}|^2 + |z_l^{-\langle l\rangle}|^2)$ for $l \geq 1$ and $\frac{1}{2}\omega_0^2(|z_0^{\langle 0\rangle}|^2 + |z_0^{-\langle 0\rangle}|^2)$ for $l = 0$ from the almost-invariant energy $\mathcal{E}_l(t)$ in the form (24). For the modulation functions that appear in this difference we have by the expansion (26) and Lemma 9

- $z_j^{\mathbf{k}} = \mathcal{O}(\delta^{e(l)+1-2\nu})$ for $\mathbf{k} \neq \pm\langle j\rangle$ and $k_l \neq 0$ (for $\mathbf{k} \neq \pm\langle l\rangle$ this follows from (26) and (33) since then $m(j,\mathbf{k}) \geq e(l)$ and $\mu(\mathbf{k}) > e(l)$, and for $\mathbf{k} = \pm\langle l\rangle$ this follows from (26) and (36) since then $m(j,\mathbf{k}) \geq e(l)$),
- $z_l^{\pm\langle l\rangle} = \mathcal{O}(\delta^{e(l)})$ and $\dot{z}_l^{\pm\langle l\rangle} = \mathcal{O}(\delta^{e(l)+1-2\nu})$ (by (26) and (33)).

This shows that this difference is in fact of order $\mathcal{O}(\delta^{2e(l)+1-2\nu})$. The improved order for $l = 1$ follows from the fact that in this case $\dot{z}_{1,2}^{\pm\langle 1\rangle} = 0$, and hence $\dot{z}_1^{\pm\langle 1\rangle} = \mathcal{O}(\delta^{e(1)+2(1-2\nu)})$. To get the summed estimate, we multiply the difference with $\sigma_l\delta^{-2e(l)}$ and sum over l, we use (37), and we apply the Cauchy-Schwarz inequality and the estimates of Lemma 9. □

Combining Lemmas 14–16 yields the statement of Theorem 7, provided that $\delta^{1-2\nu}$ is sufficiently small.

Acknowledgement

We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) through CRC 1173 and project GA 2073/2-1.

References

[1] W. Bao, X. Dong, Analysis and comparison of numerical methods for the Klein-Gordon equation in the nonrelativistic limit regime, *Numer. Math.* **120** (2012), 189–229.

[2] B. Cano, Conservation of invariants by symmetric multistep cosine methods for second-order partial differential equations, *BIT* **53** (2013), 29–56.

[3] B. Cano, Conserved quantities of some Hamiltonian wave equations after full discretization, *Numer. Math.* **103** (2006), 197–223.

[4] D. Cohen, L. Gauckler, E. Hairer, C. Lubich, Long-term analysis of numerical integrators for oscillatory Hamiltonian systems under minimal non-resonance conditions, *BIT* **55** (2015), 705–732.

[5] D. Cohen, E. Hairer, C. Lubich, Conservation of energy, momentum and actions in numerical discretizations of non-linear wave equations, *Numer. Math.* **110** (2008), 113–143.

[6] P. Deuflhard, A study of extrapolation methods based on multistep schemes without parasitic solutions, *Z. Angew. Math. Phys.* **30** (1979), 177–189.

[7] X. Dong, Stability and convergence of trigonometric integrator pseudospectral discretization for N-coupled nonlinear Klein–Gordon equations, *Appl. Math. Comput.* **232** (2014), 752–765.

[8] E. Faou, B. Grébert, Hamiltonian interpolation of splitting approximations for nonlinear PDEs, *Found. Comput. Math.* **11** (2011), 381–415.

[9] E. Faou, B. Grébert, E. Paturel, Birkhoff normal form for splitting methods applied to semilinear Hamiltonian PDEs. I. Finite-dimensional discretization, *Numer. Math.* **114** (2010), 429–458.

[10] B. García-Archilla, J. M. Sanz-Serna, R. D. Skeel, Long-time-step methods for oscillatory differential equations, *SIAM J. Sci. Comput.* **20** (1999), 930–963.

[11] L. Gauckler, Error analysis of trigonometric integrators for semilinear wave equations, *SIAM J. Numer. Anal.* **53** (2015), 1082–1106.

[12] L. Gauckler, E. Hairer, C. Lubich, D. Weiss, Metastable energy strata in weakly nonlinear wave equations, *Comm. Partial Differential Equations* **37** (2012), 1391–1413.

[13] H. Grubmüller, H. Heller, A. Windemuth, K. Schulten, Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions, *Mol. Sim.* **6** (1991), 121–142.

[14] E. Hairer, C. Lubich, Spectral semi-discretisations of weakly nonlinear wave equations over long times, *Found. Comput. Math.* **8** (2008), 319–334.

[15] E. Hairer, C. Lubich, G. Wanner, *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations*, vol. 31 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 2006.

[16] M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* **97** (1992), 1990–2001.