Karlsruhe Institute of Technology

High-frequency wave-propagation: error analysis for analytical and numerical approximations

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

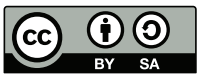
DISSERTATION

von

Julian Baumstark

Tag der mündlichen Prüfung: 06. Juli 2022

Referent: Prof. Dr. Tobias Jahnke
Korreferentin: Prof. Dr. Marlis Hochbruck



Acknowledgement

I gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – CRC 1173.

I continue in German.

Mein besonderer Dank gilt meinem Betreuer Prof. Dr. Tobias Jahnke. Nach meiner Masterarbeit bei ihm hat er mir das Vertrauen entgegengebracht, eine Promotion zu beginnen. Trotz überfülltem Terminkalender fand er immer Zeit für fachliche sowie moralische Unterstützung.
Außerdem bedanke ich mich bei Prof. Dr. Marlis Hochbruck für die Übernahme der Zweitbetreuung und ihre hilfreichen Anmerkungen. Vielen Dank für das stets offene Ohr bei allen auftretenden Belangen und Problemen der Promovierenden während meiner Arbeit als Sprecher der Doktoranden im SFB 1173. An dieser Stelle ein großes Dankeschön an Laurette Lauffer für die Unterstützung und die tolle Zusammenarbeit bei der Ausübung dieses Amtes.
Für die fachliche Unterstützung und konstruktive Kritik an dieser Dissertation bedanke ich mich speziell bei Benjamin Dörich. Ebenso Danke an Joëlle Kühn für ihre sprachlichen Korrekturvorschläge.
Danke an meine Kollegen der Arbeitsgruppe “Wissenschaftliches Rechnen” und der Arbeitsgruppe “Numerik” am Institut für Angewandte und Numerische Mathematik für das harmonische Arbeitsumfeld und die leckeren Kuchen. Meine wissenschaftliche Arbeit in Forschung und Lehre wäre ohne euch nicht so amüsant gewesen. Ich danke allen für die wunderschöne Zeit, die ich in bester Erinnerung behalte.
Zu guter Letzt möchte ich mich bei meiner Verlobten Joëlle und meiner Familie für ihre bedingungslose Liebe und Unterstützung bedanken. Ohne euch wäre diese Arbeit nicht möglich gewesen!

Contents

CHAPTER 1

Introduction and Motivation

1.1 High-frequency wave-propagation in physics

Hyperbolic partial differential equations and their applications are important in physics for a variety of reasons. The transmission of information by means of waves in a wide range of fields such as hearing, sight, television, and radio is modeled by this type of equations, see for example [37] and the references therein. The model equation for the first field is the acoustic wave equation and for the latter three the Maxwell equations. All of these equations have one thing in common: they often arise as descriptions of wave-propagation.

Wave phenomena. There are some general wave phenomena that all waves exhibit when they propagate. Since we are interested in the subject area of nonlinear optics, we briefly explain the following wave phenomena in connection with it. For more details see [15, 22, 35, 37, 38] and [41, Chapter 6].

- The main theory of geometric optics is that light takes the most efficient path between its source and the observer by traveling in straight lines along rays. With this physical theory, the *rectilinear propagation* of light and the laws of *reflection* and *refraction* are described. However, there are optical wave phenomena that cannot be described by geometrical optics. For example, light not only travels in straight lines, but also spreads out over long distances as it travels.

- Diffractive optics is the extension of geometric optics, where in addition to the usual rays of geometric optics diffracted rays are now included in the theory. Hence, *diffraction* is generally understood as the deviation of a wave propagation from the rectilinear path of rays, e.g. at the edge of an obstacle or a split. Diffraction is explained with the help of Huygens' wave principle [38, Subsection 4.4.1]. Roughly speaking, behind an obstacle the waves interact with each other and form "diffracted" wave fronts. The phenomenon of diffraction of light in everyday life can be observed, for example, in the fact that an opaque body does not cast a sharp shadow, but shows slightly

blurred shadow edges.
In addition to the size of the obstacle, the wavelength of the light is crucial for the extent of diffraction and the larger the wavelength, the stronger the diffraction.

Furthermore, nonlinear phenomena that may occur include the following:

- The first nonlinear phenomenon is the generation of new frequency components or, in other words, the generation of *higher harmonics*. For example, if we have a plane wave $a(t,x)e^{i\varphi(t,x)}$ with phase $\varphi(t,x)$ and amplitude $a(t,x)$, then nonlinear functions will produce waves with phases $j\varphi(t,x)$ for $j \in \mathbb{Z}$, where negative values of j come from the complex conjugate, by self-interaction. More details on higher harmonics are given in Subsection 3.2.2. Waves with these higher harmonics will then interact with each other. This generation and interaction of harmonics is one of the main characteristics of nonlinear problems.

- Another crucial wave phenomenon which occurs when considering wave-propagation is the interaction of waves with distinct phases: the phenomenon of *resonance*. For nonlinear problems there can be nontrivial interactions between the waves and, in particular, new phases may appear in the description of the solution. Suppose that waves coexist with phases $\varphi_i(t,x)$, $i = 1,2,3$. Then we call the three phases resonant if $\varphi_3(t,x) = \varphi_2(t,x) + \varphi_1(t,x)$. The analysis of all interactions is delicate, because in the case of near resonant interactions small divisors problems can occur. This causes the analyzed terms to become large. Details on resonances are presented in Section 3.7.

We end this part about wave phenomena with another physical effect that can occur in connection with waves. In the context of nonlinear optics, *dispersion* means that the speed of light depends on its wavelength. In dispersive media the waves pulse spreads out and changes its shape as it travels. Dispersion describes the interaction with the matter in which the wave propagates. For example, white light passing through a prism is decomposed into a spectrum of colors. Light with shorter wavelengths is bent more than light with longer wavelengths because it travels more slowly through glass.

In this thesis, we specifically investigate semilinear hyperbolic partial differential equations that have a special feature, namely that a small physical parameter occurs in these equations. Compared to the distance of propagation, or other physical scales of the solution, the wavelength of the solution is often short. Since the wavelength of a wave is inversely proportional to its frequency, the solution has a high frequency and is highly oscillatory. In optics, this small parameter corresponds to the wavelength of light. Throughout the thesis, we denote the small parameter by $\varepsilon \in \mathbb{R}$.

Numerical challenges. In the simulation of wave phenomena in general, many ordinary and partial differential equations which model the physical processes cannot be solved exactly. Therefore, numerical integrators are used to approximate the solution. For computing an approximation of the solution numerically, the underlying time interval and the domain in space are discretized into a finite number of points. At these chosen grid points approximations of the exact solution values are computed. However, even for linear highly-oscillatory equations like the harmonic oscillator, the explicit and implicit Euler methods, for example, fail if the frequency is large compared to the step-size. For more details on harmonic oscillators we refer to [41, Chapter 7].

Furthermore, implicit schemes which approximate the solution of linear highly-oscillatory differential equations very well become prohibitively costly if the differential equation is nonlinear. For this reason, numerical methods particularly suited for nonlinear differential equations such as exponential integrators, cf. [20], or splitting methods, cf. [33], were developed. However, it is well known that for standard splitting methods and standard exponential Runge–Kutta methods a very fine resolution has to be used to compute an approximation of a highly oscillatory solution numerically. In order to obtain a reasonable approximation, the step-size must be considerably smaller than the inverse of the highest frequency. Therefore, in the case of nonlinear optics, the number of grid-points in space and the number of time-steps must be inversely proportional to the parameter ε for times of length $\mathcal{O}(1)$. This reduces the efficiency significantly and leads to time-step restrictions which results in huge computational costs.
Moreover, based on geometric and diffractive optics, different time scales are distinguished. These time scales are relative to the wavelength and represent geometric and diffractive effects which lead to a different behaviour of the solution. With respect to the parameter ε, geometric optics describes the propagtion of light for times t of magnitude $\mathcal{O}(\varepsilon^0) = \mathcal{O}(1)$. The diffractive effects appear on long time scales, and therefore describe propagation for times t of order $\mathcal{O}(\varepsilon^{-1})$. In optical phenomena, long propagation times also increase the importance of nonlinear effects.
As a consequence, since we are interested in the regime of diffractive optics, the number of time-steps must be inversely proportional to ε^2 because of the long time interval. However, computing a lot of approximations leads to a long runtime and a lot of memory usage. In order to be efficient, numerical methods have to be tailor-made and thus one has to deal with the structure of the underlying problem and the occuring oscillations in a suitable way. One of our aims in this thesis is to avoid such costly numerical approximations and the associated effort. We shall see that this also requires analytical approximations that reduce the original problem to a simpler problem.

1.2 Semilinear hyperbolic systems and outline

A prominent example in which a small parameter occurs is the Maxwell–Lorentz system

$$\begin{aligned} \partial_t \mathbf{B} &= -\operatorname{curl} \mathbf{E}, \\ \partial_t \mathbf{E} &= \operatorname{curl} \mathbf{B} - \frac{1}{\varepsilon}\mathbf{Q}, \\ \partial_t \mathbf{Q} &= \frac{1}{\varepsilon}(\mathbf{E} - \mathbf{P}) + \varepsilon |\mathbf{P}|_2^2 \mathbf{P}, \\ \partial_t \mathbf{P} &= \frac{1}{\varepsilon}\mathbf{Q}, \\ \operatorname{div}(\mathbf{E} + \mathbf{P}) &= \operatorname{div} \mathbf{B} = 0, \end{aligned} \tag{1.1}$$

which models the propagation of a light beam in a Kerr medium. For further information on the Maxwell–Lorentz system see for example [11, 13, 15, 16, 24, 28, 29]. $\mathbf{E}$ describes the electric and $\mathbf{B}$ describes the magnetic field, respectively. Furthermore, the vector field $\mathbf{P}$ is the polarization and $\mathbf{Q}/\varepsilon$ its time derivative. Hence, Maxwell equations for $\mathbf{E}$ and $\mathbf{B}$ are coupled to two ordinary differential equations for $\mathbf{P}$ and $\mathbf{Q}$. The equations are normalized such that the speed of light is 1. In this example the small physical parameter $\varepsilon \in \mathbb{R}$ corresponds to the ratio between the wavelength of light and the next characteristic length of the problem, see for example [15].

In order to rewrite the Maxwell–Lorentz system (1.1) into a semilinear hyperbolic system, we introduce the vector $\mathbf{u}$ which consists of the four vector fields, the differential operator $A(\partial)$ and the matrix E given by

$$\mathbf{u} = \begin{pmatrix} \mathbf{B} \\ \mathbf{E} \\ \mathbf{Q} \\ \mathbf{P} \end{pmatrix}, \qquad A(\partial) = \begin{pmatrix} 0 & \nabla\times & 0 & 0 \\ -\nabla\times & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & -I & 0 & I \\ 0 & 0 & -I & 0 \end{pmatrix},$$

where in three-dimensional space ($d = 3$) the entries in the matrices are 3×3 matrices, which means $0 = 0_{3\times 3}$ and $I = I_{3\times 3}$. The identities in the matrix E correspond to the terms with $1/\varepsilon$ in the equations of the Maxwell–Lorentz system. Moreover, with a trilinearity T defined as

$$T(\mathbf{u}, \mathbf{u}, \mathbf{u}) = \begin{pmatrix} 0 \\ 0 \\ |\mathbf{P}|_2^2 \mathbf{P} \\ 0 \end{pmatrix},$$

the Maxwell–Lorentz system (1.1) is equivalent to

$$\partial_t \mathbf{u} + A(\partial)\mathbf{u} + \frac{1}{\varepsilon} E\mathbf{u} = \varepsilon T(\mathbf{u}, \mathbf{u}, \mathbf{u}). \tag{1.2}$$

The nonlinearity is a vector of size $s = 4d = 12$ and the matrices are of size 12×12.

Another semilinear hyperbolic model problem from physics is the Klein–Gordon system

$$\partial_t \mathbf{u} + \begin{pmatrix} 0 & \nabla \\ \nabla^\top & 0_{d\times d} \end{pmatrix} \mathbf{u} + \frac{1}{\varepsilon} \begin{pmatrix} 0 & -\upsilon^\top \\ \upsilon & 0_{d\times d} \end{pmatrix} \mathbf{u} = \varepsilon |\mathbf{u}|_2^2 \mathcal{M}\mathbf{u}. \tag{1.3}$$

Here, we have $\upsilon \in \mathbb{R}^d \backslash \{0\}$ and a skew-symmetric matrix $\mathcal{M} \in \mathbb{R}^{s\times s}$; cf. [11, 29]. The Klein–Gordon system is a nonlinear wave equation and the size of the vector $\mathbf{u}$ is given by $s = d + 1$. Similarly, as for the Maxwell–Lorentz system we can define

$$A(\partial) = \begin{pmatrix} 0 & \nabla \\ \nabla^\top & 0_{d\times d} \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & -\upsilon^\top \\ \upsilon & 0_{d\times d} \end{pmatrix}, \qquad T(\mathbf{u}, \mathbf{u}, \mathbf{u}) = |\mathbf{u}|_2^2 \mathcal{M}\mathbf{u},$$

so that the Klein–Gordon system (1.3) has the same form as the representation (1.2).

Problem setting. Next, we specify (1.2). With respect to the Maxwell–Lorentz system (1.1), the Klein–Gordon system (1.3), and the regime of diffractive geometric optics with dispersive effects (cf. Section 1.1), the specific type of semilinear hyperbolic systems which we focus on in this thesis is of the abstract form

$$\partial_t \mathbf{u} + A(\partial)\mathbf{u} + \frac{1}{\varepsilon} E\mathbf{u} = \varepsilon T(\mathbf{u}, \mathbf{u}, \mathbf{u}), \qquad t \in (0, t_{\text{end}}/\varepsilon],\ x \in \mathbb{R}^d, \tag{1.4a}$$

$$\mathbf{u}(0, x) = p(x) e^{\mathrm{i}(\kappa\cdot x)/\varepsilon} + c.c. \tag{1.4b}$$

with highly oscillatory initial data. The parameter $0 < \varepsilon \ll 1$ is considered to be small. For some finite time $t_{\text{end}} > 0$ and parameters $d, s \in \mathbb{N}$, $\mathbf{u}$ denotes a vector-valued solution which maps from $[0, t_{\text{end}}/\varepsilon] \times \mathbb{R}^d$

to $\mathbb{R}^s$. The differential operator $A(\partial)$ is defined as

$$A(\partial) = \sum_{\mu=1}^{d} A_\mu \partial_\mu, \tag{1.5}$$

where $A_1, \dots, A_d \in \mathbb{R}^{s\times s}$ are symmetric matrices. The matrix $E \in \mathbb{R}^{s\times s}$ is skew-symmetric, i.e. $E^\top = -E$, and induces dispersion of the different frequencies. For $E = 0_{s\times s}$ we are in the non-dispersive case. Since all the eigenvalues of the Hermitian matrix $\sum_{\mu=1}^{d} A_\mu - \mathrm{i}E$ are real, the system is hyperbolic. Such problems have been referred to as Friedrich systems, cf. [17, Chapter III] and references therein. On the right-hand side of (1.4a) the mapping $T : \mathbb{R}^s \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$ is a trilinear nonlinearity. The initial data (1.4b) are of the special form

$$\mathbf{u}(0, x) = p(x)e^{\mathrm{i}(\kappa\cdot x)/\varepsilon} + c.c.$$

and depend on a fixed and given wave vector $\kappa \in \mathbb{R}^d\backslash\{0\}$. This particular form of a rapidly oscillating exponential prefactor times a smooth envelope function $p : \mathbb{R}^d \to \mathbb{R}^s$ is called a wavetrain, cf. [2]. The dot is the Euclidean scalar product so that $\kappa \cdot x$ is a scalar. As usual, "c.c." means complex conjugation of the previous term, ensuring that the initial data is real.

Unfortunately, these specific semilinear hyperbolic systems are challenging. The small parameter ε makes it very delicate to treat the system (1.4a) numerically because the solution oscillates rapidly with a frequency of $\mathcal{O}(\varepsilon^{-1})$ in time *and* space. Therefore, the numerical challenges highlighted in Section 1.1 apply. The parameter ε occurs both in the PDE (1.4a) and in the initial data (1.4b). Furthermore, the problem is scaled in such a way that nonlinear and diffractive effects appear, which is only the case on a long time interval $[0, t_{\text{end}}/\varepsilon]$. Therefore, we cannot interpret the nonlinearity as a pertubation.
As a consequence, approximating the solution of (1.4) numerically with standard methods is prohibitively inefficient and even unfeasible. Therefore, special analytical and numerical approximations are needed.

We note that highly oscillatory problems have motivated many attempts to devise simpler models which are more suitable for numerical computations and at the same time provide a reasonable approximation to the corresponding solution. Asymptotic expansions of solutions to systems similar to (1.4) have been derived, e.g., in [16, 23, 26, 37] for geometric optics, i.e. for times of length $\mathcal{O}(1)$. The idea is to expand the approximate solution in an asymptotic series in ε. This means $\mathbf{u} \approx \sum_{j=1}^{\infty} \varepsilon^j U_j^\varepsilon$, where the higher order terms $\varepsilon^j U_j^\varepsilon$ serve as correctors. These correctors improve the approximation given by the leading order term U_0^ε. This asymptotic expansion is combined with a multiple scale ansatz which differs depending on which time scale is considered. For more information we refer for example to [37] where the author investigates the time scale of geometric optics. In contrast to [16, 23, 26, 37], as already mentioned, we seek approximations on long time intervals of length $\mathcal{O}(\varepsilon^{-1})$. In the diffractive regime approximations with the same asymptotic expansion but now with a multiple scale ansatz which includes an additional slow time variable exist. Approximations with infinitely small residual have been constructed in [14] for semilinear and quasilinear systems, but with εE rather than E/ε in (1.4a). More generalized nonlinear hyperbolic systems, but with $E = 0$ have been analyzed in [25]. Approximate solutions for quasilinear systems with dispersion have been analyzed in [28], and for dispersive problems with bilinear nonlinearity in [12], but without an explicit convergence rate.

The goal of this thesis. The goal of the thesis can be divided into two sub-goals. The first sub-goal is to derive a system of PDEs which is numerically more favorable than (1.4) but provides an analytical approximation to the solution $\mathbf{u}$ that is better than previous classical approximations in [11, 29]. This system is numerically more advantageous because there are no more ε-induced oscillations in space. However, the system still has oscillations in time. Therefore, secondly, we want to construct tailor-made time integrators that also allow large step-sizes and which are not limited by the smallness parameter ε. The thesis is organized as follows.

We present in Chapter 2 the basic notation which is used throughout this thesis. In Chapter 3 we investigate the specific form of the semilinear hyperbolic system (1.4) on which we focus in more detail and formulate a number of assumptions. Next, we state the slowly varying envelope approximation (SVEA) as a simplification of the problem (1.4). We extend the ansatz of the SVEA to a more accurate approximation to the solution of (1.4) by adding more terms, i.e.

$$\mathbf{u}(t,x) \approx \sum_j e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon} u_j(t,x),$$

where j is an odd integer, $\omega \in \mathbb{R}$ and the pair (ω,κ) satisfies the dispersion relation, for details see Chapter 3. The drawback of this analytical approximation is that we have more coefficients to calculate and not just one function. However, the bigger advantage is that the corresponding coefficients do no longer oscillate in space, since now, roughly speaking, the oscillations in space are in the prefactor. This analytical approximation raises questions about which PDEs solve the associated coefficients u_j and how accurate the approximation is. We prove well-posedness of this approximation in an adequate analytic setting. However, the associated system of PDEs still has ε-induced oscillations in time. Thus, we introduce a beneficial transformation. The resulting system appears to be more accessible for analytic investigation and constructing numerical schemes. The first main result is stated in Chapter 4. Here, we present an error bound for the SVEA that represents an improvement on previous results in the existing literature.

In addition to the analytical study of the introduced approximation in Chapter 4, we are also interested in computing this approximation via numerical methods in Chapter 5. Thus, in Chapter 5 we introduce a one-step time integrator. In particular, we state a rigorous error bound for this one-step method and show that the global error scales like $\mathcal{O}(\tau)$ with a constant independent of ε. Further, we extend this one-step method to a two-step method. Again, we prove that this method is a first-order method uniformly in ε. In addition, for step-sizes $\tau > \varepsilon$, the two-step method has the favourable property that its accuracy improves to $\mathcal{O}(\tau^2)$ with a constant independent of ε. This benefit is countered by the fact that each time-step requires the computation of nested multiple sums, which means higher computational costs. For this reason we reduce the workload by introducing an idea which we call "cherry picking". We validate these theoretical results with numerical experiments, focusing only on smaller problems such as the one-dimensional Klein–Gordon system (1.3).

In Chapter 6 we present another approach in which we attempt to address both problems, the ε-induced oscillations in space and time, simultaneously. We use an asymptotic expansion for the associated coefficients of this ansatz, which eventually leads to an analytical approximation with smooth coefficients. This approach is called modulated Fourier expansion and we present a corresponding error bound for this approximation as the main result in Chapter 6.

Finally, we close this thesis with a short summary of the main results combined with a brief outlook in Chapter 7.

CHAPTER 2

Preliminaries

Throughout this thesis, we use the following notation and abbreviations.

Miscellaneous. The one dimensional torus is denoted by $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ such that the d-dimensional torus $\mathbb{T}^d$ is given as the space $(\mathbb{R}\backslash 2\pi\mathbb{Z})^d$. The constant i denotes the imaginary unit, whereas i is used as an index in a few formulas. Moreover, the complex conjugate of a number or a vector a is denoted by $\overline{a}$. We use the abbreviation *c.c.* for "complex conjugate", so that for all $a \in \mathbb{C}^s$, $s \in \mathbb{N}$, the expression $a + c.c.$ is equivalent to $a + \overline{a}$.
Furthermore, $C > 0$ and $C(\cdot) > 0$ denote generic constants, which may have different values at different appearances. The notation $C(\cdot)$ means that the constant in front of the brackets depends only on the values specified in the brackets (This is not to be confused with the space of continuous function, which is also denoted by $C([0, t_{\text{end}}], X)$ for a function space X).
Let $j_{\max} \in \mathbb{N}$ be an odd integer. We define the sets

$$\begin{aligned} \mathcal{J} &= \big\{ j \in 2\mathbb{Z} - 1, \quad |j| \leqslant j_{\max} \big\}, \\ \mathcal{J}_+ &= \mathcal{J} \cap \mathbb{N}. \end{aligned}$$

Throughout the thesis, we often consider combinations of three $j_i \in \mathcal{J}$, $i = 1, 2, 3$, which we combine into one multi-index

$$J = (j_1, j_2, j_3) \quad \in \mathcal{J}^3,$$

and define the expression

$$\#J = j_1 + j_2 + j_3 \quad \in 2\mathbb{Z} - 1.$$

We list the acronyms which we will use frequently:

ODE	ordinary differential equation	PDE	partial differential equation
NLS	nonlinear Schrödinger	SVEA	slowly varying envelope approximation
KG	Klein–Gordon	ML	Maxwell–Lorentz
MFE	modulated Fourier expansion		

Vector algebra. For vectors $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^s$ or $\mathbb{C}^s$ with $\mathbf{a} = (a_1, ..., a_s)$ and $\mathbf{b} = (b_1, ..., b_s)$ we denote with

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^* \mathbf{b} = \sum_{m=1}^{s} \overline{a_m} b_m$$

the Hermitian scalar product.
Furthermore, $|\mathbf{a}|_q$ is the usual q-norm of the vector $\mathbf{a}$. For $q = 1, 2$, the norms are given by

$$|\mathbf{a}|_1 = \sum_{m=1}^{s} |a_m| \quad \text{and} \quad |\mathbf{a}|_2^2 = \sum_{m=1}^{s} |a_m|^2.$$

At this point we note that the estimates

$$|\mathbf{a}|_2 \leqslant |\mathbf{a}|_1 \leqslant \sqrt{s} |\mathbf{a}|_2 \tag{2.1}$$

hold. Moreover, we state some useful bounds. For vectors $\mathbf{a}$, $\mathbf{b} \in \mathbb{C}^s$ and a matrix $B \in \mathbb{C}^{s \times s}$, applying the Cauchy-Schwarz inequality results in

$$|\mathbf{a}^\top B \mathbf{b}| \leqslant |\mathbf{a}|_2 \, |B\mathbf{b}|_2 \leqslant |B|_2 \, |\mathbf{a}|_2 \, |\mathbf{b}|_2 \leqslant |B|_2 \, |\mathbf{a}|_1 \, |\mathbf{b}|_1, \tag{2.2}$$

where $|B|_2$ denotes the spectral norm. The second inequality in (2.2) follows since the Euclidean norm is submultiplicative which means

$$|B\mathbf{b}|_2 \leqslant |B|_2 \, |\mathbf{b}|_2.$$

Trilinear nonlinearity. Throughout the thesis the nonlinear operator T is trilinear in the sense that $T : \mathbb{R}^s \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$ is a function which is (real-)linear in each of the three variables. Its trilinear extension to $\mathbb{C}^s \times \mathbb{C}^s \times \mathbb{C}^s \to \mathbb{C}^s$ is also denoted by T. For all vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{C}^s$ and $\mathbf{a}_R$, $\mathbf{a}_I$, $\mathbf{b}_R$, $\mathbf{b}_I$, $\mathbf{c}_R$, $\mathbf{c}_I \in \mathbb{R}^s$ with $\mathbf{a} = \mathbf{a}_R + \mathrm{i}\mathbf{a}_I$, $\mathbf{b} = \mathbf{b}_R + \mathrm{i}\mathbf{b}_I$, $\mathbf{c} = \mathbf{c}_R + \mathrm{i}\mathbf{c}_I$ we can write by means of the linearity

$$\begin{aligned} T(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= T(\mathbf{a}_R + \mathrm{i}\mathbf{a}_I, \mathbf{b}_R + \mathrm{i}\mathbf{b}_I, \mathbf{c}_R + \mathrm{i}\mathbf{c}_I) \\ &= T(\mathbf{a}_R, \mathbf{b}_R, \mathbf{c}_R) + \mathrm{i}\Big(T(\mathbf{a}_I, \mathbf{b}_R, \mathbf{c}_R) + T(\mathbf{a}_R, \mathbf{b}_I, \mathbf{c}_R) + T(\mathbf{a}_R, \mathbf{b}_R, \mathbf{c}_I)\Big) \\ &\quad - \Big(T(\mathbf{a}_I, \mathbf{b}_I, \mathbf{c}_R) + T(\mathbf{a}_I, \mathbf{b}_R, \mathbf{c}_I) + T(\mathbf{a}_R, \mathbf{b}_I, \mathbf{c}_I)\Big) - \mathrm{i} T(\mathbf{a}_I, \mathbf{b}_I, \mathbf{c}_I). \end{aligned}$$

For this reason we obtain the fact that

$$\begin{aligned} \overline{T(\mathbf{a}, \mathbf{b}, \mathbf{c})} &= \overline{T(\mathbf{a}_R + \mathrm{i}\mathbf{a}_I, \mathbf{b}_R + \mathrm{i}\mathbf{b}_I, \mathbf{c}_R + \mathrm{i}\mathbf{c}_I)} \\ &= T(\mathbf{a}_R, \mathbf{b}_R, \mathbf{c}_R) - \mathrm{i}\Big(T(\mathbf{a}_I, \mathbf{b}_R, \mathbf{c}_R) + T(\mathbf{a}_R, \mathbf{b}_I, \mathbf{c}_R) + T(\mathbf{a}_R, \mathbf{b}_R, \mathbf{c}_I)\Big) \\ &\quad - \Big(T(\mathbf{a}_I, \mathbf{b}_I, \mathbf{c}_R) + T(\mathbf{a}_I, \mathbf{b}_R, \mathbf{c}_I) + T(\mathbf{a}_R, \mathbf{b}_I, \mathbf{c}_I)\Big) + \mathrm{i} T(\mathbf{a}_I, \mathbf{b}_I, \mathbf{c}_I) \\ &= T(\mathbf{a}_R - \mathrm{i}\mathbf{a}_I, \mathbf{b}_R - \mathrm{i}\mathbf{b}_I, \mathbf{c}_R - \mathrm{i}\mathbf{c}_I) = T(\overline{\mathbf{a}}, \overline{\mathbf{b}}, \overline{\mathbf{c}}) \end{aligned} \tag{2.3}$$

holds.

Another formulation which we often use later on concerns the difference of two trilinear nonlinearities. This difference can be rewritten by means of the trilinearity as

$$\begin{aligned} T(\mathbf{a},\mathbf{a},\mathbf{a}) - T(\mathbf{b},\mathbf{b},\mathbf{b}) &= T(\mathbf{a},\mathbf{a},\mathbf{a}) - T(\mathbf{b},\mathbf{a},\mathbf{a}) + T(\mathbf{b},\mathbf{a},\mathbf{a}) \\ &\quad - T(\mathbf{b},\mathbf{b},\mathbf{a}) + T(\mathbf{b},\mathbf{b},\mathbf{a}) - T(\mathbf{b},\mathbf{b},\mathbf{b}) \\ &= T(\mathbf{a}-\mathbf{b},\mathbf{a},\mathbf{a}) + T(\mathbf{b},\mathbf{a}-\mathbf{b},\mathbf{a}) + T(\mathbf{b},\mathbf{b},\mathbf{a}-\mathbf{b}). \end{aligned} \tag{2.4}$$

Finally, we state a helpful bound of the nonlinearity in the Euclidean vector norm. Because of the trilinearity we obtain for arbitrary vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \{\mathbb{R}^s, \mathbb{C}^s\}$

$$T(\mathbf{a},\mathbf{b},\mathbf{c}) = \sum_{i=1}^{s}\sum_{j=1}^{s}\sum_{m=1}^{s} a_i b_j c_m T(e_i, e_j, e_m),$$

where e_m denotes the m-th unit vector. Furthermore, it is important to note that for the $\mathbb{R}^s$ or $\mathbb{C}^s$ every basis only has a finite number of elements. Thus, there exists a constant C such that the bound $|T(e_i, e_j, e_m)|_2 \leqslant C$ holds for all $i, j, m = 1, \ldots, s$, and with (2.1) it follows that

$$\begin{aligned} |T(\mathbf{a},\mathbf{b},\mathbf{c})|_2 &\leqslant \Big|\sum_{i=1}^{s}\sum_{j=1}^{s}\sum_{m=1}^{s} a_i b_j c_m T(e_i, e_j, e_m)\Big|_2 \\ &\leqslant C|\mathbf{a}|_1|\mathbf{b}|_1|\mathbf{c}|_1 \leqslant C_T|\mathbf{a}|_2|\mathbf{b}|_2|\mathbf{c}|_2, \end{aligned} \tag{2.5}$$

where $C_T := Cs^{\frac{3}{2}}$.

Differential operators. Let $f : [0, t_{end}] \times \mathbb{R}^d \to \mathbb{R}$ with $d \in \mathbb{N}$ and $t_{\text{end}} > 0$ be a sufficiently smooth function. In this section $t \in [0, t_{end}]$ denotes the time variable and $x \in \mathbb{R}^d$ the spatial variable. We denote the partial derivative of f with respect to time by $\partial_t f$. Concerning the spatial derivatives, we denote the gradient of f by ∇f with $\nabla = (\partial_1, ..., \partial_d)$, where $\partial_\mu = \partial_{x_\mu}$ is the partial derivative with respect to the μ-th spatial direction. Note that in the special case $d = 1$, we simply write $\partial_x f$ instead of ∇f.

For a multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ we set $\partial^\alpha f := \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} f$.

Let $\mathcal{F}f$ or $\widehat{f}$ be the Fourier transform of a distribution $f \in \mathcal{S}'(\mathbb{R}^d)$, cf. Appendix A.2, then we obtain

$$\left(\mathcal{F}\left(\partial^\alpha f\right)\right)(k) = \mathrm{i}^{|\alpha|_1} k^\alpha \left(\mathcal{F}f\right)(k), \tag{2.6}$$

cf. [34, Theorem 4.26 (b)], where

$$k^\alpha = k_1^{\alpha_1} \cdots k_d^{\alpha_d} \quad \text{and} \quad |\alpha|_1 = \sum_{\mu=1}^{d} \alpha_\mu. \tag{2.7}$$

For $\mu \in \{1, \ldots, d\}$ we denote by D_μ the Fourier multiplicator $(D_\mu \widehat{f})(k) = \mathrm{i}k_\mu \widehat{f}(k)$ such that by definition $D_\mu \widehat{f}$ is the Fourier transform of $\partial_\mu f$. Thus, for notational simplicity we define

$$D^\alpha \widehat{f}(k) := \mathrm{i}^{|\alpha|_1} k^\alpha \widehat{f}(k) \tag{2.8}$$

such that by (2.6) we have that $D^\alpha \widehat{f}$ is the Fourier transform of $\partial^\alpha f$.

In functional analysis, a function $f : x \mapsto f(x)$ is often regarded as a point in a function space X and the spatial variable x is omitted for notational simplicity. In this sense, the function $x \mapsto f(t, x)$ at a fixed time t is simply denoted by $f(t)$ instead of $f(t, x)$. In the same spirit, the second argument of the (spatial) Fourier transform $\widehat{f}(t, k)$ of such a function will most often be omitted.

CHAPTER 3

Problem setting and ansatz

In this chapter, we investigate the specific semilinear hyberbolic system (1.4) with trilinear nonlinearity which we introduced in Section 1.2. We begin this chapter by taking a closer look at the numerical and analytical challenges. In Section 3.2 we state some important definitions and assumptions that are relevant within this thesis. Furthermore, we give an overview of classical results, such as the slowly varying envelope approximation. The goal is to derive a system of PDEs which provides an analytical approximation to the solution of (1.4) that is more accurate than the classical approximations. We present our ansatz which is a natural extension of the SVEA in Section 3.3 and introduce in Section 3.4 a suitable analytic setting for the analysis. The system of PDEs resulting from this approach is the starting point for further investigation. After formulating the evolution equations in Fourier space in Section 3.5, we establish a local well-posedness result on long time intervals in Section 3.6. To conclude this chapter, we introduce an additional transformation in Section 3.7 which leads to a new system of equations. Investigating the transformed system instead of the system (3.16) turns out to be advantageous. For this reason, we use the new system as a starting point for later analysis in Chapter 4.

3.1 Numerical challenges

The system (1.4) is the main object of research in this thesis. The particular focus of this work lies on constructing and analyzing suitable problem-adapted/tailor-made time-integration schemes. First, we go into more detail about the numerical challenges we have. The small parameter ε is crucial because it causes (almost) all the difficulties. The problems which occur are the following.

Oscillations in time and space. Physically relevant solutions oscillate rapidly in time and space with frequency $\sim 1/\varepsilon$. For this reason, we explain where these oscillations come from. For this purpose, it is sufficient to consider the linear case, where $T(\cdot,\cdot,\cdot) = 0$.

The linear system: We consider the linear hyperbolic system

$$\begin{aligned} \partial_t \mathbf{u} + A(\partial)\mathbf{u} + \frac{1}{\varepsilon} E\mathbf{u} = 0, \qquad & t \in (0, t_{\text{end}}/\varepsilon],\ x \in \mathbb{R}^d, \\ \mathbf{u}(0,x) = p(x) e^{\mathrm{i}(\kappa\cdot x)/\varepsilon} + c.c., & \end{aligned} \tag{3.1}$$

where the initial data correspond to (1.4b). Firstly, the solution oscillates in space because of the special form of the initial data. Since ε is small and $\kappa \in \mathbb{R}^d \backslash \{0\}$ is given by the data, the exponential term $\exp\left(\mathrm{i}\frac{\kappa\cdot x}{\varepsilon}\right)$ causes fast oscillations no matter how smooth the envelope p is. Simultaneously, the system (3.1) has oscillations in time. Even if we considered the linear system (3.1) without derivatives in space, meaning

$$\partial_t \mathbf{u} + \frac{1}{\varepsilon} E\mathbf{u} = 0, \qquad x \in \mathbb{R}^d, \quad t \in [0, t_{\text{end}}/\varepsilon],$$

the term $\frac{1}{\varepsilon} E\mathbf{u}$ would induce oscillations in time. Formally the time derivative is large and, thus, even the solution of the ODE would oscillate. Furthermore, even if we set $E = 0$ in (3.1), which means

$$\partial_t \mathbf{u} + A(\partial)\mathbf{u} = 0, \qquad x \in \mathbb{R}^d, \quad t \in [0, t_{\text{end}}/\varepsilon],$$

the term $A(\partial)\mathbf{u}$ would cause difficulties, although it has no factor $1/\varepsilon$. The reason is again the special form (1.4b) of the initial data. If the solution $\mathbf{u}$ has a similar form, we obtain a factor $1/\varepsilon$ by the derivative in space.
In summary, it follows that both terms $\frac{1}{\varepsilon} E\mathbf{u}$ and $A(\partial)\mathbf{u}$ produce oscillations in time and in space in their own way. Therefore, the solution behaves highly oscillatory in time and in space.

The long time interval and the nonlinearity. Now, we again consider the original system (1.4). We observe that both the oscillations in space and the oscillations in time become faster if ε becomes smaller. However, concurrently the time interval where we compute the solution increases. This is a second challenge that we have to face.
For a standard time integrator numerically speaking this would mean that if we decrease ε we have to choose more time-steps because the number of time-steps has to be inversely proportional to the parameter ε. However, a larger number of time-steps results in a longer calculation time and roundoff issues. In total the number of time-steps has to be inversely proportional to ε^2 due to the additional long time interval. The conclusion is that the computational work would increase significantly.
Finally, we have a nonlinear problem because of the nonlinearity T. The factor ε in front of the nonlinearity is favorable at first glance. One could think that the nonlinearity is only a small nonlinear pertubation. However, we consider time intervals of length $\mathcal{O}(\varepsilon^{-1})$, cf. (1.4a). The scaling of the systems we consider is chosen so that the nonlinearity has an effect of order $\mathcal{O}(1)$ on such long time intervals. In other words, the factor $1/\varepsilon$ of the long time interval and the factor ε in front of the nonlinearity evens out. Therefore, unfortunately we cannot consider the nonlinearity as a small perturbation of the linear system.
Furthermore, we know from the introduction that the nonlinearity also creates oscillations which may resonate with the linear propagator.

Unbounded domain. In order to perform numerical simulations we have to truncate the full space $\mathbb{R}^d$. The long-time simulation of wave-propagation is a challenging task. The reason is that in numerical

simulations we have to regard a finite space domain. In other words, we have to truncate the domain $\mathbb{R}^d$ and apply artificial conditions at the boundaries. The solution of wave-propagation leaves every bounded domain after a certain time. Consequently, this truncated domain has to be large enough such that the solution does not leave the domain to avoid effects from the artificial boundary conditions on the calculated solution. However, if the boundaries are set far apart, this increase the computational cost.

Conclusion. As a consequence of these challenges, applying standard time-integration methods to the system (1.4a) is prohibitively inefficient or even infeasible. Infeasible in the sense that the runtime is extremely long and the accuracy of the approximation is very poor. For this reason the idea is roughly speaking to introduce an analytical approximation of the exact solution, e.g. by a series expansion with new coefficient functions. Thereby one obtains a system for these new coefficient functions, which hopefully can be solved better numerically.

3.2 Assumptions and classification in the state of art

Again from a numerical perspective the main problem is that typical solutions oscillate rapidly in time and space such that it is inappropriate to apply standard numerical methods to compute an approximation. Standard methods will fail and constructing tailor-made methods for this problem class is a considerable challenge.

This problem has motivated many attempts to create suitable analytical approximations. The focus here is the attempt to devise simpler models which are more suitable for numerical computations and simultaneously provide a reasonable approximation to $\mathbf{u}$. Among these models, the nonlinear Schrödinger approximation is particularly appealing; cf. [11, 12, 14, 25, 27, 29, 39]. However, for example in [11, 29] the authors do not present any numerical methods but they focus on approximations. Since we aim to construct more accurate approximations than in [11, 29], we will briefly discuss their work. The central goal for them is to derive a nonlinear Schrödinger equation with a non-oscillatory solution that still allows them to approximate the solution of the original problem (1.4). This is attractive because solving the nonlinear Schrödinger equation is numerically a lot easier than computing an approximation of the original highly oscillatory problem. The first step on the way to establishing the nonlinear Schrödinger equation is another approximation, which is called the slowly varying envelope approximation. We will discuss this in more detail in Subsection 3.2.2.

We start this section with a number of definitions and assumptions which are based on [4, 11, 29].

3.2.1 Definitions and assumptions

For a vector $\beta \in \mathbb{R}^d$ we define a matrix

$$A(\beta) = \sum_{\mu=1}^{d} \beta_\mu A_\mu \qquad \in \mathbb{R}^{s\times s}. \tag{3.2}$$

We have seen a similar notation before. The notation (3.2) is consistent with the definition of $A(\partial)$ in (1.5). If we compare (3.2) with the definition of the differential operator (1.5) we see that the partial derivatives have been replaced by entries of the vector β. Moreover, thinking about the Fourier transform (cf. (2.6)), where the space derivatives turn into multiplications with numbers, it is clear that the matrix

$A(\beta)$ has a relation to the operator $A(\partial)$. Another important observation is that the matrix $A(\beta)$ is symmetric according to the assumption of the single matrices A_μ, $\mu = 1, \dots, d$.

Next, we define a matrix which will appear quiet frequently throughout because many terms can be expressed in a convenient form in terms of that matrix.
For $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^d$ we define

$$\mathcal{L}(\alpha, \beta) = -\alpha I + A(\beta) - \mathrm{i}E \qquad \in \mathbb{C}^{s\times s}. \tag{3.3}$$

The matrix $\mathcal{L}(0, \beta) = A(\beta) - \mathrm{i}E$ has a special role which will be seen later. A crucial observation is that $\mathcal{L}(\alpha, \beta)$ is Hermitian for all $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^d$, because according to the problem setting, $A(\beta)$ is symmetric and E is skew-symmetric. This is important since we directly know that $\mathcal{L}(\alpha, \beta)$ is unitarily diagonalizable with real eigenvalues.

Now, we consider a special form of the matrix (3.3), where we set $\beta = \kappa$. We recall that the wave vector $\kappa \in \mathbb{R}^d \backslash \{0\}$ is a fixed vector given by the data. From now on we set $\omega \in \mathbb{R}$ equal to an eigenvalue of $\mathcal{L}(0, \kappa)$. In other words, the parameter $\omega = \omega(\kappa)$ depends on κ and has to be an eigenvalue of the matrix $A(\kappa) - \mathrm{i}E$. In the literature one says that for fixed κ the pair $(\omega(\kappa), \kappa)$ satisfies the *dispersion relation*

$$\det\left(\mathcal{L}(\omega, \kappa)\right) = 0. \tag{3.4}$$

Hence, the matrix $\mathcal{L}(\omega, \kappa)$ is singular and has a non-trivial kernel, which becomes important later.
Furthermore, we define a quantity which is called the *group velocity* $c_g(\kappa) \in \mathbb{R}^d$ as a parameter which depends on κ and is given by

$$c_g(\kappa) = \nabla\omega(\kappa). \tag{3.5}$$

Here, ∇ denotes the derivative with respect to the wave vector κ. In summary, the group velocity $c_g(\kappa)$, the parameter $\omega(\kappa)$, and the wave vector κ are explicitly given in terms of the data and, therefore, in the following are fixed.

Next, the following assumptions are made. The first assumption is the *polarization condition*. The polarization condition is that the smooth envelope p of the initial data lies in the kernel of the matrix $\mathcal{L}(\omega, \kappa)$. This assumption is crucial for the classical results of the slowly varying envelope approximation and the nonlinear Schrödinger approximation. However, it is also an important assumption for our own work.

Assumption 3.2.1 (Polarization condition)**.**

The initial data (1.4b) *are polarized, i.e.*

$$p(x) \in \mathit{ker}\big(\mathcal{L}(\omega, \kappa)\big) \text{ for all } x \in \mathbb{R}^d.$$

The dimension of $\ker\big(\mathcal{L}(\omega, \kappa)\big)$ depends on the algebraic multiplicity of the eigenvalue ω of the matrix $\mathcal{L}(0, \kappa)$. Since $\mathcal{L}(0, \kappa)$ is Hermitian we know that the geometric multiplicity and algebraic multiplicity are equal for every eigenvalue of this matrix.

We remark that instead of Assumption 3.2.1 it is actually sufficient to assume that $p(x) = p_0(x) + \varepsilon p_1(x)$ with $p_0(x) \in \ker\big(\mathcal{L}(\omega, \kappa)\big)$. This assumption has also been made, e.g., in [11, 29]. However, for the sake of simplicity with respect to the presentation, we assume that $p_1 = 0$ and thus $p(x) = p_0(x)$.

Of course, Assumption 3.2.1 constrains the choice of initial data, however, we illustrate the advantage of this assumption with a simple example. Consider the linear system

$$\partial_t u + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(\omega,\kappa)u = 0,$$

where $u : [0, t_{end}/\varepsilon] \times \mathbb{R}^d \to \mathbb{C}^s$ and initial data $u(0,x) = u^0(x) \in \mathbb{R}^s$. By definition (3.3) for $\mathcal{L}(\omega,\kappa)$ we know that there exists a diagonalization $\mathcal{L}(\omega,\kappa) = \Psi\Lambda\Psi^*$ with a unitary matrix Ψ and a real diagonal matrix $\Lambda = \mathrm{diag}(\omega - \omega_1(\kappa), \dots, \omega - \omega_s(\kappa))$, where $\omega_m(\kappa)$, $m = 1, \dots, s$, denote the eigenvalues of the matrix $\mathcal{L}(0,\kappa)$. We perform the change of variables $v = \Psi^* u$ to obtain the diagonalized system

$$\begin{aligned} \partial_t v + \frac{\mathrm{i}}{\varepsilon}\Lambda v &= 0, \\ v(0,x) &= \Psi^* u^0(x). \end{aligned} \tag{3.6}$$

We observe that the linear part of the equation creates in the m-th component fast oscillations with frequencies $\omega - \omega_m(\kappa)$. To avoid this effect we choose, as already mentioned, the parameter ω equal to one of these eigenvalues $\omega_m(\kappa)$. This means the pair (ω,κ) satisfies the dispersion relation (3.4). If we now assume that the initial data $u^0(x)$ is contained in the corresponding eigenspace for all $x \in \mathbb{R}^d$, completely or up to $\mathcal{O}(\varepsilon)$ terms, the initial data $v(0,x)$ given by (3.6) have the same property. Hence, the creation of oscillations by the linear propagator are prevented.
In other words, the dispersion relation (3.4) combined with the polarization condition (Assumption 3.2.1) can be understood as filtering out the high oscillations for the part of the solution which lies in the kernel of $\mathcal{L}(\omega,\kappa)$. Unfortunately, it is not so simple in the nonlinear case because the nonlinearity also creates other oscillations which may resonate with the linear propagator.

The next assumptions mainly concern the matrix (3.3) and provide information about its properties.

Assumption 3.2.2 (Smooth eigendecomposition)**.**

The matrix $\mathcal{L}(0,\beta) = A(\beta) - \mathrm{i}E$ has a smooth eigendecomposition for $\beta \in \mathbb{R}^d\backslash\{0\}$. This means if $\lambda(\beta)$ is an eigenvalue of $\mathcal{L}(0,\beta)$, then $\lambda \in C^\infty(\mathbb{R}^d\backslash\{0\}, \mathbb{R})$, and there is a corresponding eigenvector $\psi(\beta)$ such that $|\psi(\beta)|_2 = 1$ for all β and $\psi \in C^\infty(\mathbb{R}^d\backslash\{0\}, \mathbb{C}^s)$.

Assumption 3.2.2 corresponds to Assumption 2 in [11].

Two natural questions arise. Firstly, are these assumptions also fulfilled by the two systems we are interested in? Secondly, why do we have to exclude $\beta = 0$?
At this point, it is helpful to explicitly specify the eigenvalues of the Klein–Gordon system and the Maxwell-Lorentz system. For the Maxwell–Lorentz system and the Klein–Gordon system the eigenvalues are stated in [11, Example 3 and 4]. However, for example with the Laplace expansion of determinants, cf. [30, Corollary 7.22], one can recalculate them oneself. We consider both systems separately. Instead of κ we consider at first an arbitrary vector β to gain an insight of the general form of the eigenvalues.

Example 3.2.3 (The Klein–Gordon system (1.3))**.** *In this case, for $d \in \mathbb{N}$, solving the equation*

$$\det\left(\mathcal{L}(0,\beta) - \lambda I\right) = 0$$

yields the three different eigenvalues

$$\begin{aligned}\lambda_1(\beta) &= \sqrt{|\beta|_2^2 + |\upsilon|_2^2},\\ \lambda_2(\beta) &= -\sqrt{|\beta|_2^2 + |\upsilon|_2^2},\\ \lambda_3(\beta) &= 0.\end{aligned}$$

The eigenvalues $\lambda_1(\beta)$ and $\lambda_2(\beta)$ have algebraic multiplicity 1. The eigenvalue $\lambda_3(\beta) = 0$ is an eigenvalue with algebraic multiplicity $d-1$. It is obvious that for the one-dimensional case $d = 1$, we only have the two eigenvalues $\lambda_1(\beta)$ and $\lambda_2(\beta)$.
From these formulas we observe that if $\beta = \upsilon = 0$ holds, the three eigenvalues will not be distinct anymore.

Example 3.2.4 (The Maxwell–Lorentz system (1.1))**.** *Here, we consider the three-dimensional case $d = 3$. In this case, solving the equation*

$$\det\left(\mathcal{L}(0,\beta) - \lambda I\right) = 0$$

yields seven different eigenvalues

$$\begin{aligned}\lambda_1(\beta) &= \frac{1}{2}\left(\sqrt{2\left(1+|\beta|_1\right)+|\beta|_2^2} + \sqrt{2\left(1-|\beta|_1\right)+|\beta|_2^2}\right),\\ \lambda_2(\beta) &= \sqrt{2},\\ \lambda_3(\beta) &= \frac{1}{2}\left(\sqrt{2\left(1+|\beta|_1\right)+|\beta|_2^2} - \sqrt{2\left(1-|\beta|_1\right)+|\beta|_2^2}\right),\\ \lambda_4(\beta) &= 0,\\ \lambda_5(\beta) &= -\lambda_3(\beta), \quad \lambda_6(\beta) = -\lambda_2(\beta), \quad \lambda_7(\beta) = -\lambda_1(\beta).\end{aligned}$$

The eigenvalues $\lambda_2(\beta)$ and $\lambda_6(\beta)$ have algebraic multiplicity 1. All the other eigenvalues are eigenvalues with algebraic multiplicity 2.
We observe that if $\beta = 0$, we have $\lambda_1(0) = \lambda_2(0)$, $\lambda_6(0) = \lambda_7(0)$, and $\lambda_3(0) = \lambda_5(0) = \lambda_4(0) = 0$.

For both systems it is crucial that the eigenvalues $\lambda_i(\beta)$ have constant multiplicities in a neighborhood of β, so they are distinct and their algebraic and geometric multiplicities are equal. Otherwise Assumption 3.2.2 is not necessarily fulfilled. In the literature, cf. [37, Chapter 3.I], such eigenvalues whose algebraic and geometric multiplicities are equal are called semisimple. Furthermore, there exists a result for semisimple eigenvalues, cf. [37, Theorem 3.I.1], which ensures that Assumption 3.2.2 is valid. From the observations which we made for the Klein–Gordon system and the Maxwell–Lorentz system we obtain the following. For every $\beta \in \mathbb{R}^d\backslash\{0\}$ the eigenvalues $\lambda_i(\beta)$ have constant multiplicities in a neighborhood of β.

Since we are interested in more general matrices $\mathcal{L}(\alpha,\beta)$ later, we make the following remark.

Remark 3.2.5. *We note that $\mathcal{L}(\alpha,\beta) = -\alpha I + \mathcal{L}(0,\beta)$ has the same eigenvectors as $\mathcal{L}(0,\beta)$ and that the eigenvalues are shifted by $-\alpha$. Hence, if Assumption 3.2.2 is fulfilled, then for every $\alpha \in \mathbb{R}$ the eigenvalues*

and eigenvectors of $\mathcal{L}(\alpha,\beta)$ have the smoothness specified in Assumption 3.2.2, too. This inheritance is a useful feature of the matrices $\mathcal{L}(\alpha,\beta)$.

The last assumption in this section concerns another special form of matrix $\mathcal{L}(\alpha,\beta)$.

Assumption 3.2.6. *For $j \in \{3,\dots,j_{max}+2\}$ the matrix $\mathcal{L}(j\omega, j\kappa)$ is invertible.*

We remark that for $j = 3$ Assumption 3.2.6 was also made in [11, Assumption 3]. We have expanded the assumption a bit.
It is important to point out that this assumption induces a restriction on the choice of ω. Recall that we set ω equal to an eigenvalue of the matrix $\mathcal{L}(0,\kappa)$, cf. (3.4). Now, suppose that for all $\beta \in \mathbb{R}^d\backslash\{0\}$, the matrix $\mathcal{L}(0,\beta)$ has one eigenvalue $\lambda^0(\beta)$ which is constantly equal to zero. For comparison see $\lambda_3(\beta)$ in Example 3.2.3 with $d > 1$ or $\lambda_4(\beta)$ in Example 3.2.4. Then, for $\kappa \neq 0$, λ^0 is also an eigenvalue of the matrix $\mathcal{L}(0, j\kappa)$ for $j \in \{3,\dots,j_{\max}+2\}$. However, with $\omega = \lambda^0 = 0$ the matrix $\mathcal{L}(j\omega, j\kappa) = \mathcal{L}(0, j\kappa)$ for $j \in \{3,\dots,j_{\max}+2\}$, and $\mathcal{L}(j\omega, j\kappa)$ is no longer invertible. Hence, this contradicts Assumption 3.2.6 and we outline that we should not choose $\omega = 0$.

In summary, the Assumptions 3.2.1, 3.2.2 and 3.2.6 are all technical assumptions that we will need later. At first glance it is not obvious that these assumptions are indeed true for the two systems, the Klein–Gordon system and the Maxwell–Lorentz system. However, if one computes explicitly the eigenvalues of the two systems, one can check that all these assumptions are valid.

3.2.2 Review of previous results in the literature

All the previously introduced assumptions and some further assumptions are needed to derive the slowly varying envelope approximation and the nonlinear Schrödinger approximation.
In [11] the classical nonlinear Schrödinger approximation is derived in two steps. The first step is known as the slowly varying envelope approximation.

The slowly varying envelope approximation. The idea of the slowly varying envelope approximation (SVEA) from [11] is that the exact solution is approximated by

$$\mathbf{u}(t,x) \approx \mathbf{u}_{\text{SVEA}}(t,x) = U_{\text{SVEA}}(t,x)e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c., \tag{3.7}$$

where $U_{\text{SVEA}} : [0, t_{\text{end}}/\varepsilon] \times \mathbb{R}^d \to \mathbb{C}^s$ is now complex-valued. Furthermore, the pair (ω,κ) satisfies the dispersion relation (3.4). If $t = 0$, then the ω term vanishs and we obtain a function in the form of (1.4b). By just comparing, it is clear that $U_{\text{SVEA}}(0,x)$ should exactly be $p(x)$.
In order to come up with a differential equation for the new variable U_{SVEA}, we formally substitute the ansatz (3.7) into the equation (1.4a). Since (3.7) is an ansatz with complex conjugate terms, every term and its associated complex conjugate appear. At this point we will go into more detail about the individual steps, since we will proceed similarly in our approach later on.
Using the product rule, the time derivative of $\mathbf{u}_{\text{SVEA}}$ yields with $U = U_{\text{SVEA}}(t,x)$

$$\partial_t \mathbf{u}_{\text{SVEA}}(t,x) = \left(\partial_t U - \frac{\mathrm{i}\omega}{\varepsilon}U\right) e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c. \tag{3.8}$$

For the spatial derivatives which are hidden in the differential operator $A(\partial)$ we obtain

$$A(\partial)\mathbf{u}_{\text{SVEA}}(t,x) = \left(\sum_{\mu=1}^{d} A_\mu \partial_\mu U + \sum_{\mu=1}^{d} \frac{\mathrm{i}\kappa_\mu}{\varepsilon} A_\mu U\right) e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c. \tag{3.9}$$

Therefore, plugging the ansatz (3.7) into the left-hand side of (1.4a) yields an expression which looks quiet complicated. However, by definitions of the differential operator $A(\partial)$ and the matrix $\mathcal{L}$, see (1.5) and (3.3), respectively, we are able to write the left-hand side into a more compact form. With $U = U_{\text{SVEA}}(t,x)$ it follows that

$$\begin{aligned}&\left(\partial_t U - \frac{\mathrm{i}\omega}{\varepsilon} U + \sum_{\mu=1}^{d} A_\mu \partial_\mu U + \sum_{\mu=1}^{d} \frac{\mathrm{i}\kappa_\mu}{\varepsilon} A_\mu U + \frac{1}{\varepsilon} E U\right) e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c.\\ &\quad = \left(\partial_t U + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(\omega,\kappa)U + A(\partial)U\right) e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c.\end{aligned} \tag{3.10}$$

Next, we consider the right-hand side of (1.4a). Substituting $\mathbf{u}_{\text{SVEA}}$ into the nonlinearity T yields with $U = U_{\text{SVEA}}(t,x)$

$$\begin{aligned}&T(\mathbf{u}_{\text{SVEA}}, \mathbf{u}_{\text{SVEA}}, \mathbf{u}_{\text{SVEA}})(t,x)\\ &\quad = e^{3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} T(U,U,U) + e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}\Big(T(U,U,\overline{U}) + T(U,\overline{U},U) + T(\overline{U},U,U)\Big)\\ &\qquad + e^{-\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}\Big(T(U,\overline{U},\overline{U}) + T(\overline{U},U,\overline{U},U) + T(\overline{U},\overline{U},U)\Big) + e^{-3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} T(\overline{U},\overline{U},\overline{U})\\ &\quad = e^{3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} T(U,U,U) + e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}\Big(T(U,U,\overline{U}) + T(U,\overline{U},U) + T(\overline{U},U,U)\Big) + c.c.,\end{aligned} \tag{3.11}$$

where we use the trilinearity and sort by exponentials.

Higher harmonics are precisely terms which involve exponential functions like $e^{\pm 3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$. In the next step, we aim to discard higher harmonics. The reason why we want to get rid of these terms is that in (3.10) we only have exponential terms of the form $e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$ or $e^{-\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$, where the minus one is hidden in the c.c. term. Therefore, we do not have a counterpart for $e^{\pm 3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$. This is precisely what causes the approximation error in the SVEA ansatz.

We compare in terms of $e^{\pm\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$ and ignore the higher harmonics. Because of is this we obtain two differential equations for U and $\overline{U}$ without any c.c. terms. Thus, with $U = U_{\text{SVEA}}(t,x)$ the envelope equation of U_{SVEA} is given by the PDE

$$\partial_t U + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(\omega,\kappa)U + A(\partial)U = \varepsilon\left(T(U,U,\overline{U}) + T(U,\overline{U},U) + T(\overline{U},U,U)\right). \tag{3.12}$$

The main advantage of (3.12) over (1.4) is that (3.12) has no ε-induced oscillations in *space* anymore. The reason is that the initial data $U_{\text{SVEA}}(0,x)$ is exactly $p(x)$ which is smooth in contrast to (1.4b). However, the fast oscillations in time of the nonpolarized modes must still be taken into account, which means that for this part of the solution the discretization step in numerical computations must still be small.

The accuracy of the SVEA is $\mathcal{O}(\varepsilon)$ on long time intervals $[0, t_{\text{end}}/\varepsilon]$ in suitable norms. We refer to [11, Theorem 1], where the error bound

$$\sup_{t\in[0,t_{\text{end}}/\varepsilon]} \|\mathbf{u}(t) - \mathbf{u}_{\text{SVEA}}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon \tag{3.13}$$

was shown under a lot of assumptions such as Assumptions 3.2.1, 3.2.2 and 3.2.6.

The fact that the discretization step in numerical computations must still be small for the nonpolarized modes motivates to simplify the problem further and to look for approximations that are independent of ε. In the second step, it is shown in [11] that the envelope equation can be replaced by the nonlinear Schrödinger equation without spoiling the accuracy. Thus, the NLS approximation is an approximation of the same precision as the slowly varying envelope approximation for long times $t \in [0, t_{\text{end}}/\varepsilon]$. We explain this in more detail.

Nonlinear Schrödinger approximation. The ansatz is the same as for the SVEA, see (3.7). In this model the idea is to approximate

$$\mathbf{u}(t,x) \approx \mathbf{u}_{\text{NLS}}(t,x) = U_{\text{NLS}}(t,x)e^{\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon} + c.c.,$$

where again (ω, κ) satisfy the dispersion relation (3.4) and $U_{\text{NLS}} : [0, t_{\text{end}}/\varepsilon] \times \mathbb{R}^d \to \mathbb{C}^s$ is now the solution of a nonlinear Schrödinger equation. For the initial data again the $U_{\text{NLS}}(0,x)$ coincide with the $p(x)$ of (1.4b).

At this point, we will not go into the derivation but will rather explain what makes this approximation so special.

If a co-moving coordinate system and a transformation in time are used, then the corresponding PDE of the NLS approximation does *not* depend on ε and only has to be solved on a time interval which is also independent of ε. The time interval has only a length of t_{end} instead of length $t_{\text{end}}/\varepsilon$; cf. Remark 8.v. in [11]. For the numerics this is very attractive because instead of solving a highly oscillatory problem with difficulties in space and in time, we simply have to solve an NLS equation on a time interval of $\mathcal{O}(1)$. In summary, the problem to be solved reduces to an NLS equation which does not depend on ε anymore and the parameter ε also disappears from the time interval. Thus, two numerical challenges which we stated in Section 3.1 disappear. As the name says, the NLS equation has still a nonlinearity but as long as there are no oscillations this is not a big problem. There are standard methods, like splitting methods, which can solve this efficiently.

Furthermore, the error bound

$$\sup_{t\in[0,t_{\text{end}}/\varepsilon]} \|\mathbf{u}(t) - \mathbf{u}_{\text{NLS}}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon$$

was shown under a number of assumptions in [11, Corollary 2]. In summary, the nonlinear Schrödinger approximation offers a possibility to approximate the solution of (1.4) up to $\mathcal{O}(\varepsilon)$ and the difficulties which are caused by oscillations or a long time interval no longer occur.

In some situations, however, a more accurate approximation to $\mathbf{u}$ is desirable. The approximation of PDEs by nonlinear Schrödinger equations and other modulation equations is extensively discussed in [39] and references therein.

Moreover, it is well-known that the accuracy of the nonlinear Schrödinger approximation deteriorates in the case of short or chirped pulses. A distinction between the two types of pulses can be found in [11]. In the case of short pulses, where the initial profile $p(x)$ in (1.4b) is given of the form

$$p(x) = f\left(\frac{x-x_0}{\varrho}\right) \quad \text{with } 0 < \varrho \ll 1 \text{ and } f \text{ smooth}, \tag{3.14}$$

there is a rule of thumb in [2] which suggests that the amplitude should not change more than 10% per wavelength. For much shorter pulses, the SVEA is not a reliable approximation and therefore inappro-

priate. This applies in consequence to the NLS approximation as well. For more information we refer for example to [2]. However, for such situations many improved models have been proposed and analyzed, e.g., in [1–3, 9, 11, 13, 29]. In contrast to these references, we are not interested in this situation. Instead of short or chirped pulses we restrict ourselves to wavetrains given by (3.14) with $\varrho = 1$. This is a restriction of the setting, however, we strive for higher accuracy. An illustrative example of a wavetrain and a short pulse is given in Figure 3.1. Here, the envelope p varies on a scale which is much larger than the wavelength of the oscillations and for $\varrho \to 0$ the width of the support of the envelope gets smaller. We note that we choose different scales for the x-axis for the two plots in Figure 3.1 because the envelope for the short pulse is narrower in the right plot. As a consequence, the number of optical cycles for such a short pulse is $\mathcal{O}(\varepsilon/\varrho)$.

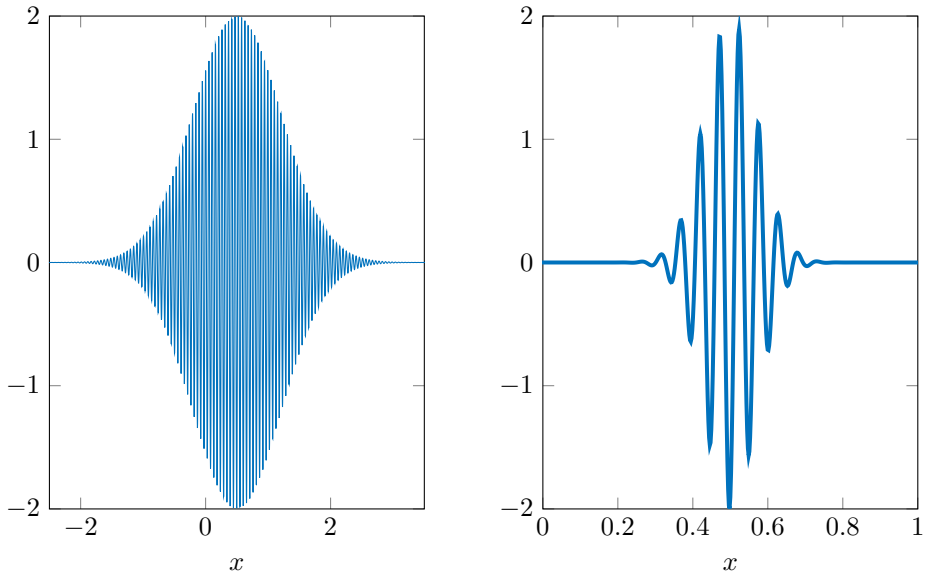


Figure 3.1: Initial data (1.4b) with $\varepsilon = 0.01$, $\kappa = 1.2$ and $p(x) = \exp\left(-\left(\frac{x-0.5}{\varrho}\right)^2\right)$. Wavetrain with $\varrho = 1$ on the left and short pulse with $\varrho = \sqrt{\varepsilon} = 0.1$ on the right.

Our goal is to derive a system of PDEs which is numerically more favorable than (1.4) but provides an approximation to the solution $\mathbf{u}$ up to an error of $\mathcal{O}(\varepsilon^2)$. More accurate approximations play an important role if the classical accuracy $\mathcal{O}(\varepsilon)$ is not enough. For numerical methods, the procedure is to refine the discretization and decrease the parameters, such as mesh width and time-step-size, to increase the accuracy. However, since ε is fixed and given by the model problem, we cannot decrease ε to obtain a better accuracy. For this reason, the accuracy must be improved by considering a higher-order extension of the classical slowly varying envelope approximation. This yields to an increase of the power of ε in the error approximation.

We recall that substituting $\mathbf{u}_{\text{SVEA}}$ into the nonlinearity T yields (3.11). The approximation error in the SVEA ansatz is caused by the higher harmonics terms with prefactor $e^{\pm 3\mathrm{i}(\kappa\cdot x-\omega t)/\varepsilon}$ which are ignored in the envelope equation.
This is the motivation to make the following ansatz for our work.

3.3 Ansatz

We start this section by defining our ansatz. In the following $j_{\max}$ is a positive odd integer. In our ansatz the exact solution of (1.4) is approximated by

$$\mathbf{u}(t,x) \approx \tilde{\mathbf{u}}^{(j_{\max})}(t,x) = \sum_{j\in\mathcal{J}} e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon} u_j(t,x), \tag{3.15}$$

for $\mathcal{J} = \{\pm 1, \pm 3, \dots, \pm j_{\max}\}$. As before, the pair (ω,κ) satisfies the dispersion relation (3.4). Furthermore, we assume that $u_{-j} = \overline{u_j}$ holds.

First, we explain in more detail why our ansatz (3.15) has this exact form. In contrast, for example, to the ansatz of the SVEA (3.7), we omit the expression "$+c.c.$" because the complex conjugate terms are hidden in the sum $\sum_{j\in\mathcal{J}}$. More precisely, with $\mathcal{J}_+ = \mathcal{J}\cap\mathbb{N}$ and $u_{-j} = \overline{u_j}$ we can equivalently write

$$\sum_{j\in\mathcal{J}} e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon} u_j(t,x) = \sum_{j\in\mathcal{J}_+} e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon} u_j(t,x) + c.c.,$$

where we also take into account higher harmonics. For $j_{\max} = 1$ our ansatz (3.15) looks very similar to (3.7). Later, after deriving the evolution equations for the coefficients u_j, we indeed obtain for $j_{\max} = 1$ the SVEA again. This is the reason why we interpret this ansatz as a natural extension of the SVEA.

Next, we know that the nonlinearity T is odd, or more precisely trilinear. Therefore, only odd harmonics are created by the nonlinearity because we are interested in initial data of the form (1.4b), where no even harmonic is given.

Since the procedure for deriving the evolution equations for the coefficients u_j is similar to that of the SVEA, this time we will not go into detail. The calculation of the time derivative and the spatial derivatives of $\tilde{\mathbf{u}}^{(j_{\max})}$ with the help of the product rule follows similarly to (3.8) and (3.9), respectively. However, the small difference is that a factor j occurs in the matrix $\mathcal{L}(j\omega, j\kappa)$. In comparison to (3.10), the matrix $\mathcal{L}(\omega,\kappa)$ is generalized by $\mathcal{L}(j\omega, j\kappa)$. Therefore, plugging the ansatz (3.15) into the left-hand side of (1.4a) yields

$$\sum_{j\in\mathcal{J}} \left(\partial_t u_j(t,x) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa)u_j(t,x) + A(\partial)u_j(t,x)\right) e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon}.$$

Substituting $\tilde{\mathbf{u}}^{(j_{\max})}$ into the nonlinearity T has in comparison to the SVEA a more complicated form. Again we use the trilinearity and sort by exponentials, however, the number of possible higher harmonics is larger. The number of possible multi-indices J is larger as well. Thus, we have to sum over all possible multi-indices $(j_1,j_2,j_3)\in\mathcal{J}^3$, where the sum $j_1+j_2+j_3 \in \{\pm 1, \pm 3, \dots, \pm 3j_{\max}\}$. We obtain

$$\begin{aligned} T\left(\tilde{\mathbf{u}}^{(j_{\max})}(t,x), \tilde{\mathbf{u}}^{(j_{\max})}(t,x), \tilde{\mathbf{u}}^{(j_{\max})}(t,x)\right) &= \sum_{(j_1,j_2,j_3)\in\mathcal{J}^3} e^{\mathrm{i}(j_1+j_2+j_3)(\kappa\cdot x-\omega t)/\varepsilon} T(u_{j_1},u_{j_2},u_{j_3})(t,x) \\ &= \sum_{\substack{j \text{ odd} \\ |j|\leqslant 3j_{\max}}} \sum_{j_1+j_2+j_3=j} e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon} T(u_{j_1},u_{j_2},u_{j_3})(t,x). \end{aligned}$$

Therefore, substituting the ansatz (3.15) into (1.4) and comparing the left- and right-hand side leads to the observation that we do not have a counterpart for the higher harmonics $e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon}$ with $j \in \{\pm j_{\max}\pm 2, \dots, \pm 3j_{\max}\}$. Similarly to the SVEA, this is the reason for discarding terms with prefactor

$e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon}$ if $|j|>j_{\max}$. All in all comparing in terms of $e^{\mathrm{i}j(\kappa\cdot x-\omega t)/\varepsilon}$ with $|j|\leqslant j_{\max}$ yields the system

$$\partial_t u_j+\frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega,j\kappa)u_j+A(\partial)u_j=\varepsilon\sum_{j_1+j_2+j_3=j}T(u_{j_1},u_{j_2},u_{j_3}),\tag{3.16}$$
$$\text{for } j\in\mathcal{J}_+=\mathcal{J}\cap\mathbb{N},\quad t\in(0,t_{\mathrm{end}}/\varepsilon],\quad x\in\mathbb{R}^d.$$

The sum on the right-hand side is taken over the set

$$\Big\{J=(j_1,j_2,j_3)\in\mathcal{J}^3:\#J:=j_1+j_2+j_3=j\Big\}.\tag{3.17}$$

This set has only finitely many elements. For example for $j_{\max}=3$, the set contains 12 multi-indices for $j=1$ and 10 multi-indices for $j=3$. We note that

$$|j_1|+|j_2|+|j_3|=|J|_1\geqslant|\#J|=|j_1+j_2+j_3|=|j|.$$

Since the condition $u_{-j}=\overline{u_j}$ holds, the PDEs for $u_{-1},\dots,u_{-j_{\max}}$ are redundant but compatible. This can be verified as follows. For the matrix $\mathcal{L}(j\omega,j\kappa)$ we deduce by definition (3.3) that

$$\begin{aligned}\overline{\mathcal{L}(j\omega,j\kappa)}&=-j\omega I+A(j\kappa)+\mathrm{i}E\\&=-\left(j\omega I+A(-j\kappa)-\mathrm{i}E\right)=-\mathcal{L}(-j\omega,-j\kappa).\end{aligned}$$

Thus, for the linear part of (3.16) that contains the matrix $\mathcal{L}(j\omega,j\kappa)$ we obtain together with $u_{-j}=\overline{u_j}$

$$\overline{\frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega,j\kappa)u_j}=\frac{\mathrm{i}}{\varepsilon}\mathcal{L}(-j\omega,-j\kappa)u_{-j}.$$

For the nonlinear part, the relation (2.3) yields

$$\sum_{j_1+j_2+j_3=j}\overline{T(u_{j_1},u_{j_2},u_{j_3})}=\sum_{j_1+j_2+j_3=-j}T(u_{j_1},u_{j_2},u_{j_3}).$$

Therefore, the PDE (3.16) is compatible with the condition that $u_{-j}=\overline{u_j}$. We end this section with a statement for the initial data and a remark. The coupled system (3.16) is endowed with initial data

$$u_{\pm1}(0,\cdot)=u_{\pm1}^{\varepsilon,0}:=p,\qquad u_{\pm j}(0,\cdot)=u_{\pm j}^0=0\quad\text{for }|j|>1.\tag{3.18}$$

Remark 3.3.1. *Similarly as for the SVEA, the main advantage of* (3.16) *over* (1.4) *is that the solution of* (3.16) *does not oscillate in* space, *because the initial data* (3.18) *are smooth in comparison to* (1.4b). *In contrast to the SVEA, the price to pay is that the total number of unknowns in* (3.16) *is* $(j_{max}+1)/2$ *times as large than in* (1.4). *If we compute only the first two coefficients, which means* $j_{max}=3$, *the number of unknowns is only twice as large, which is still a very cheap price for the numerical advantage. Since we have only removed the oscillations in space, typical solutions of* (3.16) *still oscillate in* time *due to the term* $\frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega,j\kappa)u_j$. *However, it turns out later that this situation is now more favourable. We present this in the next chapter in Remark 4.1.3.*

Of course investigating approximations of (1.4) is not new. Similar approximations have been considered in many other works. In nonlinear geometric or diffractive optics a well-known approach is to look for an approximation of the form $\mathbf{u}(t,x)\approx\mathcal{U}(t,x,(\kappa\cdot x-\omega t)/\varepsilon)$. Here $\mathcal{U}=\mathcal{U}(t,x,\theta)$ is a profile which is periodic with respect to the additional variable θ; cf. [2, 13, 14, 16, 23, 25, 28, 37]. In [8], the authors construct uniformly accurate numerical methods for highly oscillatory problems by means

of this approach. However, introducing an additional variable has a drawback. The additional variable increases the number of unknowns of the numerical discretization by a factor N_θ, where N_θ is the number of grid points in the θ-direction. This is the reason why we do not use a profile $\mathcal{U}(t,x,\theta)$ explicitly in our approach. However, we remark that the ansatz (3.15) can be interpreted as a truncated Fourier series of $\theta \mapsto \mathcal{U}(t,x,\theta)$, where all Fourier modes with index $j \notin \mathcal{J}$ are discarded.

The next goal is to define the proper analytical setting in which we aim to consider the problem class.

3.4 Analytical setting

With regard to analyze the system (3.16) and to investigate the approximation behavior of the ansatz (3.15) we establish a suitable analytical setting in this section.
Let $\mathcal{F}f = \widehat{f}$ be the Fourier transform of f, see (A.1), and let $\mathcal{S}'(\mathbb{R}^d)$ be the dual of the Schwartz space defined in Appendix A.2.

As in [11] we will work in the Wiener algebra, defined by

$$W\left(\mathbb{R}^d\right) = \left\{f \in \mathcal{S}'(\mathbb{R}^d) : \widehat{f} \in L^1(\mathbb{R}^d)\right\}$$

with the norm

$$\|f\|_{W(\mathbb{R}^d)} = \|\widehat{f}\|_{L^1(\mathbb{R}^d)} = \int_{\mathbb{R}^d} |\widehat{f}(k)| \, \mathrm{d}k.$$

For $r \in \mathbb{N}$ the Wiener algebra of order r is defined as

$$W^r\left(\mathbb{R}^d\right) = \left\{f \in W(\mathbb{R}^d) : \partial^\alpha f \in W(\mathbb{R}^d) \text{ for all } \alpha \in \mathbb{N}_0^d, |\alpha|_1 \leqslant r\right\},$$

cf. (16) in [11], where $|\alpha|_1$ is defined in (2.7). It is endowed with the norm

$$\|f\|_{W^r(\mathbb{R}^d)} = \sum_{|\alpha|_1 \leqslant r} \|\partial^\alpha f\|_{W(\mathbb{R}^d)}.$$

When $r = 0$, we write for simplicity $W(\mathbb{R}^d)$ instead of $W^0(\mathbb{R}^d)$.
For vector-valued versions of the function spaces W^r for $r \in \mathbb{N}_0$ and also L^1, we define the norm

$$\|\widehat{f}\|_{L^1(\mathbb{R}^d)^s} := \left\||\widehat{f}|_2\right\|_{L^1(\mathbb{R}^d)} = \int_{\mathbb{R}^d} |\widehat{f}(k)|_2 \, \mathrm{d}k, \qquad \widehat{f} \in L^1(\mathbb{R}^d)^s, \tag{3.19}$$

where $|\cdot|_2$ describes the Euclidean vector norm. Throughout the thesis we use the abbreviation $\|\cdot\|_{W^r} = \|\cdot\|_{W^r(\mathbb{R}^d)^s}$ for $r \in \mathbb{N}_0$.

In fact, many of the results that are presented in [11] can also be obtained in Sobolev spaces. However, in order to use Sobolev embedding to handle the nonlinearity we would need the additional assumption $r > \frac{d}{2}$. Furthermore, the Wiener algebras have favorable properties which are summarized later on. Another reason why people are interested in the Wiener algebra instead of Sobolev spaces is the following. The Wiener norm of the initial envelopes of short pulses (3.14) remain bounded when $\varrho \to 0$, whereas any Sobolev norm tends to infinity. For more information we refer the interested reader to [2, 11, 29].

Next, we state the classical properties of the Wiener algebra which are used in this thesis (cf. [29, Proposition 3.2], [11, Proposition 1] or [3]).

Properties of the Wiener algebra:

- The space $\left(W^r(\mathbb{R}^d), \|\cdot\|_W\right)$ is a Banach algebra which means for all $f, g \in W^r(\mathbb{R}^d)$ we have

$$\|fg\|_{W^r(\mathbb{R}^d)} \leqslant C\|f\|_{W^r(\mathbb{R}^d)}\|g\|_{W^r(\mathbb{R}^d)}.$$

- The Wiener algebra $W(\mathbb{R}^d)$ is continuously embedded in $L^\infty\left(\mathbb{R}^d\right)$.

Remark 3.4.1. *The first property in particular means that we have bilinear estimates. This property is very useful in order to handle the nonlinearity. The resulting estimates, such as Lemma 3.5.1, (3.30) and (3.31) are used quite often. A similar bilinear estimate is also known for Sobolev spaces. However, for Sobolev spaces this bilinear estimate is only true if $r > d/2$. For Wiener algebras there is no restriction on the parameter r. Thus, this bilinear estimate is even true for the space W, where r is equal to zero. The second property plays a crucial role in our error bounds later on.*

For estimates in the Wiener algebra it is convenient to consider the evolution equation (3.16) in Fourier space. The following sections of this chapter are based on [4]. Compared to [4], where $j_{\max} = 3$, we treat an arbitrary odd number $j_{\max}$.

3.5 Evolution equations in Fourier space

Formally the solutions u_j of the system (3.16) can be Fourier transformed by (A.2) with coefficients (A.1). If u_j is sufficiently smooth, then the derivatives in space for $\mu \in \{1, \dots, d\}$ are given by

$$\partial_\mu^r u_j(t,x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{u}_j(t,k)\partial_\mu^r e^{\mathrm{i}k\cdot x}\,\mathrm{d}k = (2\pi)^{-d/2} \int_{\mathbb{R}^d} (\mathrm{i}k_\mu)^r \widehat{u}_j(t,k)e^{\mathrm{i}k\cdot x}\,\mathrm{d}k$$

such that space derivatives correspond to multiplications of the Fourier coefficients. Differentiating (A.2) with respect to t formally gives

$$\partial_t u_j(t,x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \partial_t \widehat{u}_j(t,k)e^{\mathrm{i}k\cdot x}\,\mathrm{d}k \tag{3.20}$$

and

$$A(\partial)u_j(t,x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(\sum_{\mu=1}^d \mathrm{i}k_\mu A_\mu\right) \widehat{u}_j(t,k)e^{\mathrm{i}k\cdot x}\,\mathrm{d}k = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathrm{i}A(k)\widehat{u}_j(t,k)e^{\mathrm{i}k\cdot x}\,\mathrm{d}k. \tag{3.21}$$

Therefore, we obtain by inserting (3.20) and (3.21) into the left-hand side of the system (3.16)

$$\begin{aligned}
&\partial_t u_j(t,x) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa)u_j(t,x) + A(\partial)u_j(t,x) \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(\partial_t \widehat{u}_j(t,k) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa)\widehat{u}_j(t,k) + \mathrm{i}A(k)\widehat{u}_j(t,k)\right) e^{\mathrm{i}k\cdot x}\,\mathrm{d}k \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left(\partial_t \widehat{u}_j(t,k) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa + \varepsilon k)\widehat{u}_j(t,k)\right) e^{\mathrm{i}k\cdot x}\,\mathrm{d}k.
\end{aligned}$$

Throughout the thesis we introduce the shorthand notation

$$\mathcal{L}_j(\theta) := \mathcal{L}(j\omega, j\kappa + \theta) = -j\omega I + A(j\kappa + \theta) - \mathrm{i}E, \qquad \text{for} \quad j \in \mathcal{J}_+, \tag{3.22}$$

where we omit ω and κ because they are fixed. Thus, it follows that applying the Fourier transform to the left-hand side of (3.16) gives

$$\mathcal{F}\Big(\partial_t u_j + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa)u_j + A(\partial)u_j\Big)(t,k) = \partial_t \widehat{u}_j(t,k) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}_j(\varepsilon k)\widehat{u}_j(t,k).$$

Furthermore, the trilinear nonlinearity of (3.16) is formally given by

$$\begin{aligned}
&T(u_{j_1}, u_{j_2}, u_{j_3})(x) \\
&= (2\pi)^{-3d/2} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{\mathrm{i}(k^{(1)}+k^{(2)}+k^{(3)})\cdot x} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k^{(3)}))\, \mathrm{d}k^{(3)}\, \mathrm{d}k^{(2)}\, \mathrm{d}k^{(1)} \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left((2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k - k^{(1)} - k^{(2)}))\, \mathrm{d}k^{(2)}\, \mathrm{d}k^{(1)} \right) e^{\mathrm{i}k\cdot x}\, \mathrm{d}k \\
&= \mathcal{F}^{-1}\left((2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k - k^{(1)} - k^{(2)}))\, \mathrm{d}k^{(2)}\, \mathrm{d}k^{(1)} \right)(x),
\end{aligned}$$

where we substitute $k = k^{(1)} + k^{(2)} + k^{(3)}$.
With the notation $K = \big(k^{(1)}, k^{(2)}, k^{(3)}\big) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ and $\#K = k^{(1)} + k^{(2)} + k^{(3)} \in \mathbb{R}^d$ we obtain that the Fourier transform of $T(u_{j_1}, u_{j_2}, u_{j_3})$ is given by

$$\mathcal{F}\Big(T(u_{j_1}, u_{j_2}, u_{j_3})\Big)(k) = (2\pi)^{-d} \int_{\#K=k} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k^{(3)}))\, \mathrm{d}K =: \mathcal{T}\left(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\right)(k), \tag{3.23}$$

where we use the notation

$$\int_{\#K=k} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k^{(3)}))\, \mathrm{d}K = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} T(\widehat{u}_{j_1}(k^{(1)}), \widehat{u}_{j_2}(k^{(2)}), \widehat{u}_{j_3}(k - k^{(1)} - k^{(2)}))\, \mathrm{d}k^{(2)}\, \mathrm{d}k^{(1)}.$$

Hence, $u = (u_1, \ldots, u_{j_{\max}})$ solves the system (3.16) if and only if $\widehat{u} = (\widehat{u}_1, \ldots, \widehat{u}_{j_{\max}})$ solves the system

$$\partial_t \widehat{u}_j(t,k) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}_j(\varepsilon k)\widehat{u}_j(t,k) = \varepsilon \sum_{\#J=j} \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(t,k), \quad j \in \mathcal{J}_+, \quad t \in (0, t_{\mathrm{end}}/\varepsilon], \quad k \in \mathbb{R}^d, \tag{3.24}$$

where $\#J = j_1 + j_2 + j_3$, and with initial data

$$\widehat{u}_1(0,\cdot) = \widehat{p}, \qquad \widehat{u}_j(0,\cdot) = 0, \quad \text{for } j > 1, \tag{3.25}$$

where $\widehat{p}$ is the Fourier transform of p from (1.4b).
At this point, we note that the convention $u_{-j} = \overline{u_j}$ implies that $\widehat{u}_{-j}(t,k) = \overline{\widehat{u}_j(t,-k)}$. The reason is that we obtain

$$\begin{aligned}
u_j(x) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{u}_j(k) e^{\mathrm{i}k\cdot x}\, \mathrm{d}k, \\
u_{-j}(x) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \overline{\widehat{u}_j(k) e^{\mathrm{i}k\cdot x}}\, \mathrm{d}k = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \overline{\widehat{u}_j(k)} e^{-\mathrm{i}k\cdot x}\, \mathrm{d}k \\
&= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \overline{\widehat{u}_j(-k)} e^{\mathrm{i}k\cdot x}\, \mathrm{d}k.
\end{aligned}$$

Note that in general $\widehat{u}_{-j}(t,k) \neq \overline{\widehat{u}_j(t,k)}$, which we have to keep in mind in the implementation of numerical methods. For the analysis this implies

$$\|\widehat{u}_{-j}(t)\|_{L^1} = \int_{\mathbb{R}^d} |\widehat{u}_{-j}(t,k)|_2 \mathrm{d}k = \int_{\mathbb{R}^d} |\overline{\widehat{u}_j(t,-k)}|_2 \mathrm{d}k = \int_{\mathbb{R}^d} |\widehat{u}_j(t,-k)|_2 \mathrm{d}k = \|\widehat{u}_j(t)\|_{L^1}. \tag{3.26}$$

Furthermore, for negative indices it follows by means of the skew symmetry of E and the definition (3.22)

$$\mathcal{L}_{-j}(\theta) = j\omega I - A(j\kappa - \theta) - \mathrm{i}E = -\left(-j\omega I + A(j\kappa - \theta) - \overline{\mathrm{i}E}\right) = -\overline{\mathcal{L}_j(-\theta)} \qquad \text{for } j \in \mathcal{J}_+, \tag{3.27}$$

such that (3.24) holds also for $j \in \{-1, \dots, -j_{\max}\}$.

Now, we state some helpful bounds for the trilinear nonlinearity $\mathcal{T}$, defined in (3.23), which are frequently used later on.

Lemma 3.5.1. *For $f_1, f_2, f_3 \in W(\mathbb{R}^d)$ we have*

$$\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3)\right\|_{L^1} \leqslant C_{\mathcal{T}} \prod_{i=1}^{3} \|\widehat{f}_i\|_{L^1}, \tag{3.28}$$

where $C_{\mathcal{T}} := C_T(2\pi)^{-d}$ and C_T is the constant defined in (2.5).

If additionally $g_1, g_2, g_3 \in W(\mathbb{R}^d)$ and if $\|f_i\|_W$, $\|g_i\|_W \leqslant C$ for some $C > 0$, it follows that

$$\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3) - \mathcal{T}(\widehat{g}_1, \widehat{g}_2, \widehat{g}_3)\right\|_{L^1} \leqslant C_{\mathcal{T}} C^2 \sum_{i=1}^{3} \left\|\widehat{f}_i - \widehat{g}_i\right\|_{L^1}. \tag{3.29}$$

Proof. For $f_1, f_2, f_3 \in W(\mathbb{R}^d)$ we use the definition of L^1, (3.23), and obtain with (2.5)

$$\begin{aligned}
\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3)\right\|_{L^1} &\leqslant (2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\#K=k} \left|T(\widehat{f}_1(k^{(1)}), \widehat{f}_2(k^{(2)}), \widehat{f}_3(k^{(3)}))\right|_2 \mathrm{d}K \, \mathrm{d}k \\
&\leqslant C_T(2\pi)^{-d} \int_{\mathbb{R}^d} \int_{\#K=k} |\widehat{f}_1(k^{(1)})|_2 |\widehat{f}_2(k^{(2)})|_2 |\widehat{f}_3(k^{(3)})|_2 \, \mathrm{d}K \, \mathrm{d}k \\
&= C_T(2\pi)^{-d} \Big(\int_{\mathbb{R}^d} |\widehat{f}_1(k^{(1)})|_2 \, \mathrm{d}k^{(1)} \Big) \Big(\int_{\mathbb{R}^d} |\widehat{f}_2(k^{(2)})|_2 \, \mathrm{d}k^{(2)} \Big) \Big(\int_{\mathbb{R}^d} |\widehat{f}_3(k^{(3)})|_2 \, \mathrm{d}k^{(3)} \Big) \\
&= C_T(2\pi)^{-d} \|\widehat{f}_1\|_{L^1} \|\widehat{f}_2\|_{L^1} \|\widehat{f}_3\|_{L^1}.
\end{aligned}$$

For $f_1, f_2, f_3, g_1, g_2, g_3 \in W(\mathbb{R}^d)$ with $\|f_i\|_W \leqslant C$ and $\|g_i\|_W \leqslant C$ we have with (2.4)

$$\begin{aligned}
&\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3) - \mathcal{T}(\widehat{g}_1, \widehat{g}_2, \widehat{g}_3)\right\|_{L^1} \\
&\leqslant \left\|\mathcal{T}(\widehat{f}_1 - \widehat{g}_1, \widehat{f}_2, \widehat{f}_3)\right\|_{L^1} + \left\|\mathcal{T}(\widehat{g}_1, \widehat{f}_2 - \widehat{g}_2, \widehat{f}_3)\right\|_{L^1} + \left\|\mathcal{T}(\widehat{g}_1, \widehat{g}_2, \widehat{f}_3 - \widehat{g}_3)\right\|_{L^1} \\
&\leqslant C_T(2\pi)^{-d} \left(\left\|\widehat{f}_1 - \widehat{g}_1\right\|_{L^1} \|\widehat{f}_2\|_{L^1} \|\widehat{f}_3\|_{L^1} + \left\|\widehat{f}_2 - \widehat{g}_2\right\|_{L^1} \|\widehat{g}_1\|_{L^1} \|\widehat{f}_3\|_{L^1} + \left\|\widehat{f}_3 - \widehat{g}_3\right\|_{L^1} \|\widehat{g}_1\|_{L^1} \|\widehat{g}_2\|_{L^1} \right) \\
&\leqslant C_T(2\pi)^{-d} C^2 \sum_{i=1}^{3} \left\|\widehat{f}_i - \widehat{g}_i\right\|_{L^1},
\end{aligned}$$

where we use (3.28) for each of the three terms. ∎

Remark 3.5.2. *By definition of the Wiener algebra the estimates (3.28) and (3.29) are equivalent to*

$$\left\|T(f_1,f_2,f_3)\right\|_W \leqslant C_{\mathcal{T}} \prod_{i=1}^{3} \|f_i\|_W \tag{3.30}$$

and

$$\left\|T(f_1,f_2,f_3)-T(g_1,g_2,g_3)\right\|_W \leqslant C_{\mathcal{T}} C^2 \sum_{i=1}^{3} \|f_i-g_i\|_W. \tag{3.31}$$

After introducing the analytical setting, we establish a local well-posedness result in the next section.

3.6 Local well-posedness

In this section we show well-posedness of the system (3.16) on long time intervals, more precisely on a time interval $[0, t^\star_{\text{end}}/\varepsilon)$, where $t^\star_{\text{end}} > 0$ is independent of ε. The following results hold for arbitrary odd $j_{\max} > 0$. We start this section with the definition of a suitable norm and some helpful estimates.
For $v = (v_1, \dots, v_{j_{\max}}) \in W^s \times \dots \times W^s$ we define the norm

$$\|v\|_{W^s} = 2\|v_1\|_{W^s} + 2\|v_3\|_{W^s} + \dots + 2\|v_{j_{\max}}\|_{W^s}.$$

The factor 2 is introduced in order to account for the terms with negative indices which appear on the right-hand side of (3.24), hidden in the sum $\sum_{\#J=j}$. If we use the convention that $\widehat{v}_{-j}(k) = \overline{\widehat{v}_j(-k)}$ holds for $j \in \mathcal{J}_+$ as before, then we have with (3.26) that

$$\|v\|_{W^s} = \sum_{j\in\mathcal{J}} \|v_j\|_{W^s}.$$

For $v = (v_1, \dots, v_{j_{\max}}) \in W \times \dots \times W$ the inequalities

$$\begin{aligned}\sum_{j\in\mathcal{J}} \Big\| \sum_{\#J=j} \mathcal{T}(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3})\Big\|_{L^1} &\leqslant \sum_{j\in\mathcal{J}} \sum_{\#J=j} \left\|\mathcal{T}(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3})\right\|_{L^1} \leqslant \sum_{J\in\mathcal{J}^3} \left\|\mathcal{T}(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3})\right\|_{L^1} \\ &\leqslant C_{\mathcal{T}} \sum_{J\in\mathcal{J}^3} \|\widehat{v}_{j_1}\|_{L^1}\|\widehat{v}_{j_2}\|_{L^1}\|\widehat{v}_{j_3}\|_{L^1} \leqslant C_{\mathcal{T}}\|v\|_W^3 \end{aligned} \tag{3.32}$$

follow from (3.28) and the fact that the set of indices over which is summed in $\sum_{J\in\mathcal{J}^3}$ includes more elements than the index set of the sum $\sum_{j\in\mathcal{J}}\sum_{\#J=j}$. Let $u = (u_1, \dots, u_{j_{\max}})$ be another element in $W \times \dots \times W$. Similarly to the proof of Lemma 3.5.1 for the estimate (3.29), the trilinearity of T yields

$$\begin{aligned}&\sum_{j\in\mathcal{J}} \sum_{\#J=j} \left\|\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}) - \mathcal{T}(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3})\right\|_{L^1} \\ &\leqslant \sum_{j\in\mathcal{J}} \sum_{\#J=j} C_{\mathcal{T}}\Big(\|\widehat{u}_{j_1}-\widehat{v}_{j_1}\|_{L^1}\|\widehat{u}_{j_2}\|_{L^1}\|\widehat{u}_{j_3}\|_{L^1} + \|\widehat{u}_{j_2}-\widehat{v}_{j_2}\|_{L^1}\|\widehat{v}_{j_1}\|_{L^1}\|\widehat{u}_{j_3}\|_{L^1} + \|\widehat{u}_{j_3}-\widehat{v}_{j_3}\|_{L^1}\|\widehat{v}_{j_1}\|_{L^1}\|\widehat{v}_{j_2}\|_{L^1}\Big) \\ &\leqslant C_{\mathcal{T}}\Big(\|v\|_W^2 + \|u\|_W\|v\|_W + \|u\|_W^2\Big)\|u-v\|_W \end{aligned} \tag{3.33}$$

with the same reasoning as in (3.32).
After these preparations, we show local well-posedness of (3.16). We remark that the polarization of the initial data, cf. Assumption 3.2.1, is not required for the following result.

Lemma 3.6.1 (Local well-posedness)**.**

(i) If $p \in W$, then there is a existence time $t^\star_{end} > 0$ such that for every $\varepsilon \in (0,1]$ the system (3.16) with initial data (3.18) has a unique mild solution

$$u = (u_1, \dots, u_{j_{max}}), \qquad u_j \in C([0, t^\star_{end}/\varepsilon), W).$$

(ii) If $p \in W^1$ and $t_{end} < t^\star_{end}$ for some $t_{end} > 0$ independent of ε, then the mild solution on $[0, t_{end}/\varepsilon]$ is a classical solution $u = (u_1, \dots, u_{j_{max}})$ with

$$u_j \in C^1([0, t_{end}/\varepsilon], W) \cap C([0, t_{end}/\varepsilon], W^1).$$

(iii) If $p \in W^2$ and $t_{end} < t^\star_{end}$, then

$$u_j \in C^2([0, t_{end}/\varepsilon], W) \cap C^1([0, t_{end}/\varepsilon], W^1) \cap C([0, t_{end}/\varepsilon], W^2).$$

We immediately conclude by continuity the existence of constants $C_{u,1}$ and $C_{u,2}$ such that

$$\sup_{t \in [0, t_{\mathrm{end}}/\varepsilon]} \|u_j(t)\|_{W^1} \leqslant C_{u,1}, \qquad j \in \mathcal{J}_+ \tag{3.34}$$

in case (ii), and

$$\sup_{t \in [0, t_{\mathrm{end}}/\varepsilon]} \|u_j(t)\|_{W^2} \leqslant C_{u,2}, \qquad j \in \mathcal{J}_+ \tag{3.35}$$

in case (iii). In both cases the constant $C_{u,i}$ for $i = 1, 2$, depends only on t_{end}, $C_{\mathcal{T}}$ and on $\|p\|_{W^i}$, but not on ε.

Proof. The proof is based on classical arguments such as the varaition of constants formula and Banach's fixed point theorem, cf. [36, Chapter 6], but nevertheless we outline the main steps.

Step 1. First, we choose $\varepsilon \in (0,1]$ fixed. We define the operator

$$\mathcal{A} \begin{pmatrix} v_1 \\ \vdots \\ v_{j_{\max}} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_1 v_1 \\ \vdots \\ \mathcal{A}_{j_{\max}} v_{j_{\max}} \end{pmatrix}, \qquad \mathcal{A}_j = \frac{\mathrm{i}}{\varepsilon} \mathcal{L}(j\omega, j\kappa) + A(\partial) \tag{3.36}$$

with domain $D(\mathcal{A}) = W^1 \times \ldots \times W^1$. This operator generates a strongly continuous group $(e^{t\mathcal{A}})_{t \in \mathbb{R}}$ on $W \times \ldots \times W$. Moreover, we note that for $j \in \mathcal{J}_+$ and every $t \in \mathbb{R}$ the group operator $e^{t\mathcal{A}_j}$ is an isometry, because

$$\left\| e^{t\mathcal{A}_j} v_j \right\|_W = \left\| \mathcal{F}\left(e^{t\mathcal{A}_j} v_j\right) \right\|_{L^1} = \int_{\mathbb{R}^d} \left| e^{\mathrm{i}t\mathcal{L}_j(\varepsilon k)/\varepsilon} \widehat{v}_j(k) \right|_2 \mathrm{d}k = \int_{\mathbb{R}^d} \left| \widehat{v}_j(k) \right|_2 \mathrm{d}k = \|v_j\|_W$$

for all $v_j \in W$. Here, we use the fact that $\mathcal{L}_j(\varepsilon k)$ is Hermitian and thus the matrix $e^{\mathrm{i}t\mathcal{L}_j(\varepsilon k)/\varepsilon}$ is unitary. With the operator $\mathcal{A}_j$ from (3.36) the system (3.16) can be reformulated as

$$\partial_t u_j + \mathcal{A}_j u_j = \varepsilon \sum_{\#J = j} T(u_{j_1}, u_{j_2}, u_{j_3}) \qquad \text{for } j \in \mathcal{J}_+. \tag{3.37}$$

Step 2. We show existence in a closed ball via Banach's fixed point theorem. We introduce for a number $\tau > 0$ determined below the space

$$\mathcal{X} = C([0,\tau/\varepsilon], W) \times \ldots \times C([0,\tau/\varepsilon], W).$$

The corresponding norm is given by

$$\|v\|_{\mathcal{X}} = \sup_{t\in[0,\tau/\varepsilon]} \|v(t)\|_W = \sup_{t\in[0,\tau/\varepsilon]} \sum_{j\in\mathcal{J}} \|v_j(t)\|_W.$$

Now, we fix $u_j^0 = v_j(0) \in W$ and define the map

$$\Phi\colon \mathcal{X} \longrightarrow \mathcal{X}, \qquad \Phi(v) = v^{new} = \begin{pmatrix} v_1^{new} \\ \vdots \\ v_{j_{\max}}^{new} \end{pmatrix},$$

where for $t \in [0,\tau/\varepsilon]$

$$v_j^{new}(t) = e^{-t\mathcal{A}_j} u_j^0 + \varepsilon \sum_{\#J=j} \int_0^t e^{(\sigma-t)\mathcal{A}_j} T(v_{j_1}, v_{j_2}, v_{j_3})(\sigma)\, \mathrm{d}\sigma.$$

By definition of the Wiener algebra, the isometry property of the group operator $e^{t\mathcal{A}_j}$ and (3.23) we obtain

$$\|v_j^{new}(t)\|_W \leqslant \|u_j^0\|_W + \varepsilon \sum_{\#J=j} \int_0^t \big\|T(v_{j_1}, v_{j_2}, v_{j_3})(\sigma)\big\|_W \,\mathrm{d}\sigma = \|u_j^0\|_W + \varepsilon \sum_{\#J=j} \int_0^t \big\|\mathcal{T}(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3})(\sigma)\big\|_{L^1} \,\mathrm{d}\sigma.$$

With the estimate (3.32) it follows that

$$\begin{aligned}
\|\Phi(v)\|_{\mathcal{X}} &= \sup_{t\in[0,\tau/\varepsilon]} \sum_{j\in\mathcal{J}} \|v_j^{new}(t)\|_W \leqslant 2\|p\|_W + \varepsilon \sup_{t\in[0,\tau/\varepsilon]} \sum_{j\in\mathcal{J}} \sum_{\#J=j} \int_0^t \big\|\mathcal{T}(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3})(\sigma)\big\|_{L^1} \,\mathrm{d}\sigma \\
&\leqslant 2\|p\|_W + C_{\mathcal{T}}\varepsilon \sup_{t\in[0,\tau/\varepsilon]} \int_0^t \|v(\sigma)\|_W^3 \,\mathrm{d}\sigma \leqslant 2\|p\|_W + C_{\mathcal{T}}\tau \sup_{\sigma\in[0,\tau/\varepsilon]} \|v(\sigma)\|_W^3 \\
&= 2\|p\|_W + C_{\mathcal{T}}\tau \|v\|_{\mathcal{X}}^3.
\end{aligned}$$

Now we choose a parameter $\rho > 0$ and fix $r = 1 + \rho$. We define the closed ball with radius r as

$$B(r) = \big\{v \in \mathcal{X}\colon \|v\|_{\mathcal{X}} \leqslant r\big\}.$$

Then, for every $p \in W$ with $\|p\|_W \leqslant \frac{\rho}{2}$ we estimate with $v \in B(r)$

$$\|\Phi(v)\|_{\mathcal{X}} \leqslant \rho + C_{\mathcal{T}}\tau r^3.$$

Under the condition that $\tau \leqslant 1/(C_{\mathcal{T}} r^3)$ we obtain

$$\|\Phi(v)\|_{\mathcal{X}} \leqslant \rho + 1 = r,$$

which means that Φ maps the closed ball onto itself. Next, we show the contraction property of Φ for any $v, w \in B(r)$. It follows from (3.33) that

$$\begin{aligned}\|\Phi(v) - \Phi(w)\|_{\mathcal{X}} &= \varepsilon \sup_{t\in[0,\tau/\varepsilon]} \sum_{j\in\mathcal{J}} \sum_{\#J=j} \int_0^t \big\|T(v_{j_1}, v_{j_2}, v_{j_3})(\sigma) - T(w_{j_1}, w_{j_2}, w_{j_3})(\sigma)\big\|_W \,\mathrm{d}\sigma \\ &= 3C_{\mathcal{T}} r^2 \varepsilon \sup_{t\in[0,\tau/\varepsilon]} \int_0^t \|v(\sigma) - w(\sigma)\|_W \,\mathrm{d}\sigma \leqslant 3C_{\mathcal{T}} r^2 \tau \|v - w\|_{\mathcal{X}}.\end{aligned}$$

Under the condition that $\tau \leqslant 1/(6C_{\mathcal{T}} r^2)$ we have that Φ is Lipschitz on the closed ball $B(r)$ with a Lipschitz constant smaller than or equal to $\frac{1}{2}$. In total, if we choose

$$\tau = \min\big\{1/(C_{\mathcal{T}} r^3), 1/(6C_{\mathcal{T}} r^2)\big\}, \tag{3.38}$$

then $\Phi : B(r) \to B(r)$ is a contraction, and by Banach's fixed point theorem, there is a unique fixed point $u \in B(r)$ of Φ. We emphasize that by the choice (3.38) the number τ is independent of ε. By construction, this fixed point u is a mild solution of the system (3.16) with initial data (3.18). In other words, for fixed $\varepsilon \in (0, 1]$ and for $t \in [0, \tau/\varepsilon]$ we have

$$u_j(t) = e^{-t\mathcal{A}_j} p_j + \varepsilon \sum_{\#J=j} \int_0^t e^{(\sigma-t)\mathcal{A}_j} T(u_{j_1}, u_{j_2}, u_{j_3})(\sigma) \,\mathrm{d}\sigma.$$

We remark that at this point the proof only gives a conditional uniqueness result among functions belonging to a certain ball.

Step 3. In order to derive unconditional uniqueness we have to use the standard argument that we can glue and shift solutions. Thus, we glue together each local solution for short time intervals to cover any time interval which is not larger than the maximal existence time which we denote by $\tau^+(\varepsilon)/\varepsilon$. We refer to Theorem 1.4 and its proof in [36, Chapter 6] for more information. Therefore, the solution can be extended to the maximal time interval $[0, \tau^+(\varepsilon)/\varepsilon)$, and we define

$$t^\star_{\mathrm{end}} := \inf_{\varepsilon\in(0,1]} \tau^+(\varepsilon) \geqslant \tau > 0.$$

With this construction $t^\star_{\mathrm{end}}$ is uniformly bounded from below and, thus, independent of ε. This proves part (i).

Next, we prove part (ii).

Step 4. We aim to apply [36, Chapter 6, Theorem 1.5]. For this purpose we have to show that the nonlinearity $T : W \times W \times W$ is continuously differentiable and locally Lipschitz continuous. First, we note that for $(w_1, w_2, w_3), (h_1, h_2, h_3) \in W \times W \times W$ the trilinearity leads to

$$\begin{aligned}\Big(T(w_1 + h_1, w_2 + h_2, w_3 + h_3) - T(w_1, w_2, w_3)\Big) - \big(T(w_1, w_2, h_3) + T(w_1, h_2, w_3) + T(h_1, w_2, w_3)\big) \\ = T(w_1, h_2, h_3) + T(h_1, w_2, h_3) + T(h_1, h_2, w_3) + T(h_1, h_2, h_3).\end{aligned}$$

Furthermore, we have with $h = (h_1, h_2, h_3)$

$$\|h_i\|_W^2 \leqslant \|h_1\|_W^2 + \|h_2\|_W^2 + \|h_3\|_W^2 = \|h\|_{W\times W\times W}^2, \quad i = 1, 2, 3.$$

Together with the fact that W is an algebra, we obtain

$$\begin{aligned}
&\big\|T(w_1,h_2,h_3)+T(h_1,w_2,h_3)+T(h_1,h_2,w_3)+T(h_1,h_2,h_3)\big\|_W\\
&\quad\leqslant \big\|T(w_1,h_2,h_3)\big\|_W+\big\|T(h_1,w_2,h_3)\big\|_W+\big\|T(h_1,h_2,w_3)\big\|_W+\big\|T(h_1,h_2,h_3)\big\|_W\\
&\quad\leqslant C_T\big(\|w_1\|_W\|h_2\|_W\|h_3\|_W+\|h_1\|_W\|w_2\|_W\|h_3\|_W\\
&\qquad\quad+\|h_1\|_W\|h_2\|_W\|w_3\|_W+\|h_1\|_W\|h_2\|_W\|h_3\|_W\big)\\
&\quad\leqslant C_T\|h\|^2_{W\times W\times W}\big(\|w_1\|_W+\|w_2\|_W+\|w_3\|_W+\|h\|_{W\times W\times W}\big)\\
&\quad= \mathcal{O}(\|h\|^2_{W\times W\times W}).
\end{aligned}$$

Therefore, the Fréchet derivative is given by the linear map

$$(h_1,h_2,h_3)\mapsto T(w_1,w_2,h_3)+T(w_1,h_2,w_3)+T(h_1,w_2,w_3),$$

which is bounded because of (3.28) and

$$\begin{aligned}
\sup_{\|h\|_{W\times W\times W}=1}&\big\|T(w_1,w_2,h_3)+T(w_1,h_2,w_3)+T(h_1,w_2,w_3)\big\|_W\\
&\leqslant \|w_1\|_W\|w_2\|_W+\|w_1\|_W\|w_3\|_W+\|w_2\|_W\|w_3\|_W\leqslant C.
\end{aligned}$$

In summary the nonlinearity $T: W\times W\times W$ is Fréchet differentiable on $W\times W\times W$ and the continuity of the Fréchet derivative in every $(w_1,w_2,w_3)\in W\times W\times W$ is shown analogously. In addition, local Lipschitz continuity of T follows by (3.33).
Next, if $p\in W^1$, we then have by definition $(u_1^0,\dots,u_{j_{\max}}^0)=(p,0,\dots,0)\in D(\mathcal{A})$. Now, all the requirements are fulfilled to prove part (ii). With [36, Chapter 6, Theorem 1.5] it follows that for every $t_{\text{end}}<t_{\text{end}}^\star$, where t_{end} is independent of ε, the mild solution is in fact a classical solution on $[0,t_{\text{end}}/\varepsilon]$. This proves part (ii).

Step 5. For the part (iii) of the lemma, we set $u'=(u_1',\dots,u_{j_{\max}}')$ with $u_j'=\partial_t u_j$ and formally differentiate both sides of (3.37) with respect to t. This yields

$$\partial_t u_j'+\mathcal{A}_j u_j'=\varepsilon\sum_{\#J=j}\Big(T(u_{j_1}',u_{j_2},u_{j_3})+T(u_{j_1},u_{j_2}',u_{j_3})+T(u_{j_1},u_{j_2},u_{j_3}')\Big)$$

with initial data

$$u_j'(0)=-\mathcal{A}_j u_j^0+\varepsilon\sum_{\#J=j}T(u_{j_1}^0,u_{j_2}^0,u_{j_3}^0). \tag{3.39a}$$

Next, let $u=(u_1,\dots,u_{j_{\max}})$ be the classical solution constructed in part (ii) of this lemma. Then, we consider the linear problem

$$\begin{aligned}
&\partial_t u_j'+\mathcal{A}_j u_j'=\varepsilon\mathcal{B}_j(t,u'), &\text{(3.40a)}\\
&\mathcal{B}_j(t,u')=\sum_{\#J=j}\Big(T(u_{j_1}'(t),u_{j_2}(t),u_{j_3}(t))+T(u_{j_1}(t),u_{j_2}'(t),u_{j_3}(t))+T(u_{j_1}(t),u_{j_2}(t),u_{j_3}'(t))\Big)
\end{aligned}$$

with initial data (3.39a). Since we know by part (ii) that $u_j\in C^1([0,t_{\text{end}}/\varepsilon],W)\cap C([0,t_{\text{end}}/\varepsilon],W^1)$, the mapping

$$(t,u')\mapsto\mathcal{B}_j(t,u'),\qquad \mathcal{B}_j\colon[0,t_{\text{end}}/\varepsilon]\times(W\times\ldots\times W)\to W$$

is continuously differentiable and locally Lipschitz continuous. This follows again by (3.28). If $p \in W^2$, then $(u_1'(0), \dots, u_{j_{\max}}'(0)) \in D(\mathcal{A})$ due to (3.28). With the same justification as in part (ii) of the proof, the mild solution $u' = \partial_t u$ of (3.40a) with initial data (3.39a) is in fact a classical solution according to [36, Chapter 6, Theorem 1.5]. Thus, part (iii) is shown. ∎

We conclude this section with a brief remark on the local well-posedness of (1.4) in the Wiener algebra on long time intervals $[0, t_{\text{end}}/\varepsilon]$ for some $t_{\text{end}} > 0$ independent of ε. By adapting the proof of Lemma 3.6.1 we are also able to show local well-posedness of the original problem (1.4) in the Wiener algebra. The reason is that the operator $A(\partial) + \frac{1}{\varepsilon}E$ has the same group properties as the operator $\mathcal{A}$ defined in the proof of Lemma 3.6.1. By definition of the Wiener algebra and the isometry property of the group operator $\exp\left(t\left[A(\partial) + \frac{1}{\varepsilon}E\right]\right)$ we can prove local well-posedness of (1.4) via Banach's fixed point argument.

3.7 Transformation to smoother variables

In order to analyze the accuracy of the approximation (3.15), the estimates (3.34) and (3.35) have to be refined. In Section 4.4.1 with $j_{\max} = 3$ we will show that if the system (3.16) is considered with initial data (3.18), then for example the function u_3 stays small on long time intervals, i.e.

$$\sup_{t \in [0, t_{\text{end}}/\varepsilon]} \|u_3(t)\|_{W^1} \leqslant C\varepsilon^2.$$

A similar refined estimate of order $\mathcal{O}(\varepsilon)$ will be shown for a certain "part" of u_1 to be specified later in (4.10).

We first demonstrate that standard proof techniques do not help us to show this improved estimation. For a better explanation, we illustrate the problems that occur with an example, and consider (3.24) with $j_{\max} = 3$. In the following all calculations are formal. Applying Duhamels formula to (3.24) with $j = 3$ yields

$$\begin{aligned}\widehat{u}_3(t,k) &= \exp\left(-\frac{\mathrm{i}t}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right)\widehat{u}_3(0,k) + \varepsilon \sum_{\#J=3} \int_0^t \exp\left(\frac{\mathrm{i}(\sigma - t)}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right)\mathcal{T}(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3})(\sigma)\,\mathrm{d}\sigma \\ &= \varepsilon \int_0^t \exp\left(\frac{\mathrm{i}(\sigma - t)}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right)\mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)\,\mathrm{d}\sigma + \text{other terms},\end{aligned}$$

since $\widehat{u}_3(0,k) = 0$. However, $\mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)$ is formally $\mathcal{O}(1)$ and the factor ε in front of the integral counterbalances the long time interval $t \in [0, t_{\text{end}}/\varepsilon]$. Therefore, this integral term is $\mathcal{O}(1)$. However, we aim to show $\sup_{t \in [0, t_{\text{end}}/\varepsilon]} \|\widehat{u}_3(t)\|_{L^1} \leqslant C\varepsilon^2$. In order to gain one factor of ε from the oscillatory behavior of the integrand, we want to apply integration by parts, which leads to two problems. First, we have to ensure that $\Lambda_3(\varepsilon k)$ and thus $\mathcal{L}_3(\varepsilon k)$ is invertible for every $k \in \mathbb{R}^d$. However, this statement is not valid for general $k \in \mathbb{R}^d$. As an illustrative example, consider the eigenvalues of the matrix $\mathcal{L}_3(\theta)$, cf. (3.22), for $\theta \in \mathbb{R}$ in the one-dimensional case of the Klein-Gordon system (1.3) in Figure 3.2. First, we observe that the eigenvalue λ_{32} is bounded away from zero for all $\theta \in \mathbb{R}$. The eigenvalue λ_{31}, however, has two

intersections with the x-axis. Hence, the matrix $\mathcal{L}_3(\theta)$ is not invertible for all $\theta \in \mathbb{R}$. We only know that $\mathcal{L}_3(\theta)$ is invertible for $\theta = \varepsilon k = 0$ because of Assumption 3.2.6 and $\mathcal{L}_3(0) = \mathcal{L}(3\omega, 3\kappa)$.

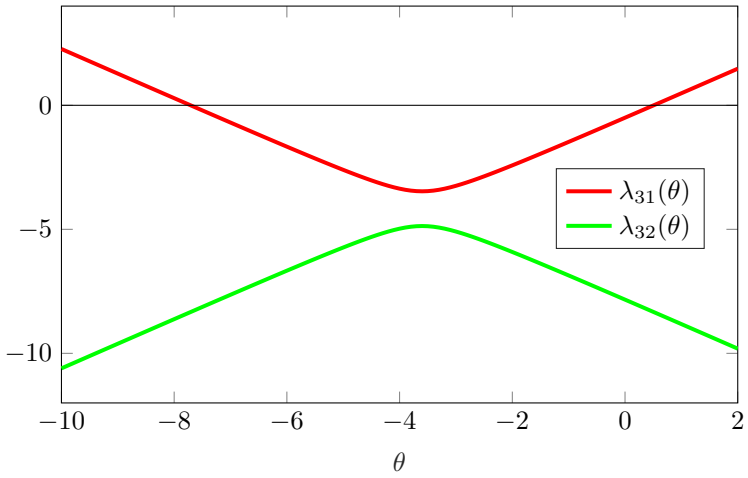


Figure 3.2: Eigenvalues $\lambda_{3m}(\theta)$ of the matrix $\mathcal{L}_3(\theta)$ of the one-dimensional Klein-Gordon system with $v = 0.7$, $\kappa = 1.2$ and $\omega = \max\{\omega_1(\kappa), \omega_2(\kappa)\}$, where ω_m is the m-th eigenvalue of $\mathcal{L}(0, \kappa)$.

In the resulting term, after integration by parts the next problem occurs. Under the assumption that $\mathcal{L}_3(\varepsilon k)$ is invertible for every $k \in \mathbb{R}^d$, we obtain

$$\begin{aligned}
&\varepsilon \int_0^t \exp\left(\frac{\mathrm{i}(\sigma - t)}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right) \mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)\, \mathrm{d}\sigma \\
&= \varepsilon^2 \left(\mathrm{i}\mathcal{L}_3(\varepsilon k)\right)^{-1} \left[\exp\left(\frac{\mathrm{i}(\sigma - t)}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right) \mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)\right]_{\sigma=0}^{t} \\
&\quad - \varepsilon^2 \left(\mathrm{i}\mathcal{L}_3(\varepsilon k)\right)^{-1} \int_0^t \exp\left(\frac{\mathrm{i}(\sigma - t)}{\varepsilon}\mathcal{L}_3(\varepsilon k)\right) \partial_t \mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)\, \mathrm{d}\sigma.
\end{aligned}$$

The first term is formally $\mathcal{O}(\varepsilon^2)$, however, the integral term is again only $\mathcal{O}(1)$, since $\partial_t \mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_1)(\sigma)$, cf. (3.24) with $j = 1$, and the length of the time interval are both $\mathcal{O}(\varepsilon^{-1})$. Thus, with this approach, the improved estimation cannot be shown.

In order to formulate and prove these refined estimates, it is very useful to consider a transformation of $\widehat{u}_j$ which we introduce in the following. This transformation remedies the problems encountered when using standard proof techniques.

For $j \in \mathcal{J}_+$ and every $\theta \in \mathbb{R}^d$ the Hermitian matrix $\mathcal{L}_j(\theta) = \mathcal{L}(j\omega, j\kappa + \theta)$ defined in (3.22) has an eigendecomposition

$$\mathcal{L}_j(\theta) = \Psi_j(\theta)\Lambda_j(\theta)\Psi_j^*(\theta) \tag{3.41}$$

with a unitary matrix $\Psi_j(\theta) \in \mathbb{C}^{s\times s}$ and a real diagonal matrix $\Lambda_j(\theta) \in \mathbb{R}^{s\times s}$ containing the eigenvalues of $\mathcal{L}_j(\theta)$.

We have the following relation of the unitary matrix $\Psi_j(\theta)$ and the diagonal matrix $\Lambda_j(\theta)$ for positive and negative numbers j. It follows from (3.27) that

$$\begin{aligned} \Psi_{-j}(\theta) &= -\overline{\Psi_j(-\theta)} \\ \text{and} \quad \Lambda_{-j}(\theta) &= -\overline{\Lambda_j(-\theta)} = -\Lambda_j(-\theta), \end{aligned} \tag{3.42}$$

since $\Lambda_j(\theta)$ is real for every $\theta \in \mathbb{R}^d$. Throughout this section we denote by $\psi_{jm}(\theta) \in \mathbb{C}^s$ the m-th column of $\Psi_j(\theta)$, and by $\lambda_{jm}(\theta) \in \mathbb{R}$ the m-th eigenvalue. Thus, for $m, \ell = 1, \dots, s$ the relations

$$\mathcal{L}_j(\theta)\psi_{jm}(\theta) = \lambda_{jm}(\theta)\psi_{jm}(\theta), \qquad \psi_{jm}(\theta) \cdot \psi_{j\ell}(\theta) = \begin{cases} 1 & \text{if } m = \ell, \\ 0 & \text{else} \end{cases} \tag{3.43}$$

hold.

We know that the eigenvalues of $\mathcal{L}(\alpha, \beta)$ are the eigenvalues of $\mathcal{L}(0, \beta)$ shifted by $-\alpha$. Thus, the general form of the eigenvalues $\lambda_{jm}(\theta)$ is given by

$$\lambda_{jm}(\theta) = -j\omega + \omega_m(j\kappa + \theta), \tag{3.44}$$

where $\omega_m(\beta)$ is the m-th eigenvalue of $\mathcal{L}(0, \beta)$ with $\beta \in \mathbb{R}^d \backslash \{0\}$. Furthermore, (3.42) implies

$$\lambda_{-jm}(\theta) = -\lambda_{jm}(-\theta) = j\omega - \omega_m(j\kappa - \theta).$$

By Remark 3.2.5 the eigenvalues have the smoothness specified in Assumption 3.2.2. We know that for the Klein-Gordon system the matrix $\mathcal{L}_1(0) = \mathcal{L}(\omega, \kappa)$ has a one-dimensional kernel since the eigenvalues, which are not constantly equal to zero, have algebraic multiplicity one. For the Maxwell-Lorentz system the matrix $\mathcal{L}_1(0)$ can have a one-dimensional or a two-dimensional kernel depending on the choice of ω. Hence, we introduce the parameter $m_\star \in \{1, 2\}$ such that $\mathcal{L}_1(0)$ has an $m_\star$-dimensional kernel. Then, the enumeration of the eigenvalues is chosen in such a way that $\lambda_{11}(0) = \lambda_{1m_\star}(0) = 0$. It follows that the kernel of $\mathcal{L}_1(0)$ is spanned by $\psi_{11}(0)$ and $\psi_{1m_\star}(0)$. This fact is important in the context of Assumption 3.2.1 as we will see below when we consider the initial data. Furthermore, the evaluation of the eigenvalues at $\theta = 0$ plays an important role later on.

Next, we define for every $j \in \mathcal{J}_+$, $\varepsilon > 0$, $t \geqslant 0$ and $k \in \mathbb{R}^d$ the matrix

$$S_{j,\varepsilon}(t, k) = \exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\Lambda_j(\varepsilon k)\right)\Psi_j^*(\varepsilon k) = \Psi_j^*(\varepsilon k)\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\mathcal{L}_j(\varepsilon k)\right) \qquad \in \mathbb{C}^{s\times s}. \tag{3.45}$$

For the last equality we use the relation

$$\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\mathcal{L}_j(\varepsilon k)\right) = \Psi_j(\varepsilon k)\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\Lambda_j(\varepsilon k)\right)\Psi_j^*(\varepsilon k),$$

which we obtain with the diagonalization (3.41).

The new variables $z_j : \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}^s$ are obtained by the transformation

$$z_j(t, k) = S_{j,\varepsilon}(t, k)\widehat{u}_j(t, k), \qquad j \in \mathcal{J}_+, \tag{3.46}$$

where $\widehat{u}_1(t, k), \dots, \widehat{u}_{j_{\max}}(t, k)$ is the solution of (3.24). The matrix (3.45) is unitary for every $j \in \mathcal{J}_+$, $\varepsilon > 0$, $t \geqslant 0$ and $k \in \mathbb{R}^d$ because by definition

$$\begin{aligned} S_{j,\varepsilon}^*(t, k)S_{j,\varepsilon}(t, k) &= \Psi_j(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\Lambda_j(\varepsilon k)\right)\exp\left(\tfrac{\mathrm{i}t}{\varepsilon^2}\Lambda_j(\varepsilon k)\right)\Psi_j^*(\varepsilon k) \\ &= \Psi_j(\varepsilon k)\Psi_j^*(\varepsilon k) = I = S_{j,\varepsilon}(t, k)S_{j,\varepsilon}^*(t, k). \end{aligned}$$

Thus, the transformation (3.46) is unitary, and we have

$$|\widehat{u}_j(t,k)|_2 = |z_j(t,k)|_2, \quad \text{which implies by definition (3.19)} \quad \|\widehat{u}_j(t)\|_{L^1} = \|z_j(t)\|_{L^1}. \tag{3.47}$$

Furthermore, for negative indices we set

$$S_{-j,\varepsilon}(t,k) := \overline{S_{j,\varepsilon}(t,-k)}, \qquad z_{-j}(t,k) := \overline{z_j(t,-k)},$$

which is in accordance with the convention $\widehat{u}_{-j}(t,k) = \overline{\widehat{u}_j(t,-k)}$.

We note that the linear part of the system (3.24) could be solved exactly in Fourier space. Therefore, the introduced transformation (3.46) can be interpreted as a transformation with the solution of this linear part. The advantage of this transformation is shown by deriving the equations of motion for the new variables.

We obtain for the time derivative of the matrix (3.45) for fixed $k \in \mathbb{R}^d$ and $j \in \mathcal{J}_+$

$$\partial_t S_{j,\varepsilon}(t,k) = \tfrac{\mathrm{i}}{\varepsilon}\Lambda_j(\varepsilon k) S_{j,\varepsilon}(t,k) = \tfrac{\mathrm{i}}{\varepsilon} S_{j,\varepsilon}(t,k)\mathcal{L}_j(\varepsilon k).$$

For the last equality we use the relation

$$\begin{aligned}\tfrac{\mathrm{i}}{\varepsilon}\Lambda_j(\varepsilon k) S_{j,\varepsilon}(t,k) &= \Psi_j^*(\varepsilon k)\Psi_j(\varepsilon k)\Lambda_j(\varepsilon k)\Psi_j^*(\varepsilon k)\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\mathcal{L}_j(\varepsilon k)\right) = \Psi_j^*(\varepsilon k)\mathcal{L}_j(\varepsilon k)\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\mathcal{L}_j(\varepsilon k)\right)\\ &= S_{j,\varepsilon}(t,k)\mathcal{L}_j(\varepsilon k),\end{aligned}$$

which we obtain with (3.45) and because $\Lambda_j(\varepsilon k)$ commutes with $\exp\left(\frac{\mathrm{i}t}{\varepsilon}\Lambda_j(\varepsilon k)\right)$. Hence, taking the time derivative of (3.46) and substituting (3.24) yields

$$\begin{aligned}\partial_t z_j(t,k) &= \left(\partial_t S_{j,\varepsilon}(t,k)\right)\widehat{u}_j(t,k) + S_{j,\varepsilon}(t,k)\partial_t\widehat{u}_j(t,k)\\ &= \tfrac{\mathrm{i}}{\varepsilon} S_{j,\varepsilon}(t,k)\mathcal{L}_j(\varepsilon k)\widehat{u}_j(t,k) - \tfrac{\mathrm{i}}{\varepsilon} S_{j,\varepsilon}(t,k)\mathcal{L}_j(\varepsilon k)\widehat{u}_j(t,k) + \varepsilon S_{j,\varepsilon}(t,k)\sum_{\#J=j}\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t,k)\\ &= \varepsilon S_{j,\varepsilon}(t,k)\sum_{\#J=j}\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t,k).\end{aligned}$$

To obtain a more compact notation, we write

$$\partial_t z_j(t) = \varepsilon\sum_{\#J=j}\mathbf{F}_\varepsilon(t,\widehat{u},J) \tag{3.48}$$

with $\widehat{u} = (\widehat{u}_1,\dots,\widehat{u}_{j_{\max}})$ and

$$\mathbf{F}_\varepsilon(t,\widehat{u},J) = S_{j,\varepsilon}(t)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t), \qquad j = \#J. \tag{3.49}$$

Comparing (3.24) with (3.48) shows that the dominating linear term

$$\frac{\mathrm{i}}{\varepsilon}\mathcal{L}_j(\varepsilon k)\widehat{u}_j(t,k)$$

in (3.24) is cancelled by the transformation. If we had no nonlinearity, which means $\mathcal{T}(\cdot,\cdot,\cdot) = 0$, it would follow from (3.48) and (3.49) that $\partial_t z_j(t) = 0$. Hence, the coefficient function $z_j(t) = z_j(0)$ would be constant in time. Therefore, in the linear case the exact solution of (3.24) is then simply

$$\widehat{u}_j(t,k) = S_{j,\varepsilon}^*(t,k)S_{j,\varepsilon}(0,k)\widehat{u}_j(0,k) = S_{j,\varepsilon}^*(t,k)z_j(0,k).$$

This is a favourable property which does not cure the oscillatory behaviour completely in the general (nonlinear) case. The right-hand side of the evolution equation (3.48) of z_j is formally $\mathcal{O}(\varepsilon)$, whereas the right-hand side of (3.24) is $\mathcal{O}(1/\varepsilon)$. Hence, the entries of z_j oscillate with a much smaller amplitude than the entries of $\widehat{u}_j$. This observation is our main motivation for considering transformed variables in the proofs of our main results.

A closed system of evolution equations for z_j with $j \in \mathcal{J}_+$ can be obtained by expressing the right-hand side of (3.49) by z_j via the inverse transformation

$$\widehat{u}_j(t,k) = S^*_{j,\varepsilon}(t,k) z_j(t,k). \tag{3.50}$$

In order to do this, the nonlinearity $\mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)$ has to be expressed in terms of the new variables. This leads to

$$\partial_t z_j(t) = \varepsilon \sum_{\#J=j} S_{j,\varepsilon}(t) \mathcal{T}\big(S^*_{j_1,\varepsilon} z_{j_1}, S^*_{j_2,\varepsilon} z_{j_2}, S^*_{j_3,\varepsilon} z_{j_3}\big)(t).$$

In order to gain insight into the interaction between the linear propagator $S_{j,\varepsilon}(t)$ and the nonlinearity, we have to consider single entries of the vectors $\widehat{u}_j = \big(\widehat{u}_{jm}\big)_{m=1}^s$ and $z_j = \big(z_{jm}\big)_{m=1}^s$, respectively. The transformation (3.50) is equivalent to

$$\begin{aligned} \widehat{u}_j(t,k) = S^*_{j,\varepsilon}(t,k) z_j(t,k) &= \Psi_j(\varepsilon k) \exp\big(-\tfrac{\mathrm{i}t}{\varepsilon} \Lambda_j(\varepsilon k)\big) z_j(t,k) \\ &= \sum_{m=1}^s \exp\big(-\tfrac{\mathrm{i}t}{\varepsilon} \lambda_{jm}(\varepsilon k)\big) z_{jm}(t,k) \psi_{jm}(\varepsilon k), \end{aligned} \tag{3.51}$$

where $\lambda_{jm}(\varepsilon k) \in \mathbb{R}$, $z_{jm}(t,k) \in \mathbb{C}$ and $\psi_{jm}(\varepsilon k) \in \mathbb{C}^s$. Together with (3.23) and the trilinearity we obtain

$$\begin{aligned} \mathcal{T}(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3})(t,k) &= (2\pi)^{-d} \int_{\#K=k} T(\widehat{u}(t,k^{(1)}), \widehat{u}(t,k^{(2)}), \widehat{u}(t,k^{(3)}))\, \mathrm{d}K \\ &= (2\pi)^{-d} \sum_M \int_{\#K=k} \exp\Big(-\tfrac{\mathrm{i}}{\varepsilon} t \lambda_{JM}(\varepsilon K)\Big) Z_{JM}(t,K) T(\psi_{JM}(\varepsilon K))\, \mathrm{d}K \end{aligned}$$

with summation over all $M = (m_1, m_2, m_3) \in \{1, \dots, s\}^3$ and the notation

$$\begin{aligned} K &= \big(k^{(1)}, k^{(2)}, k^{(3)}\big) && \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \\ \lambda_{JM}(\varepsilon K) &= \lambda_{j_1 m_1}\big(\varepsilon k^{(1)}\big) + \lambda_{j_2 m_2}\big(\varepsilon k^{(2)}\big) + \lambda_{j_3 m_3}\big(\varepsilon k^{(3)}\big) = \sum_{i=1}^3 \lambda_{j_i m_i}\big(\varepsilon k^{(i)}\big) && \in \mathbb{R} \\ Z_{JM}(t,K) &= z_{j_1 m_1}\big(t, k^{(1)}\big) z_{j_2 m_2}\big(t, k^{(2)}\big) z_{j_3 m_3}\big(t, k^{(3)}\big) = \prod_{i=1}^3 z_{j_i m_i}\big(t, k^{(i)}\big) && \in \mathbb{C} \\ T(\psi_{JM}(\varepsilon K)) &= T\Big(\psi_{j_1 m_1}\big(\varepsilon k^{(1)}\big), \psi_{j_2 m_2}\big(\varepsilon k^{(2)}\big), \psi_{j_3 m_3}\big(\varepsilon k^{(3)}\big)\Big) && \in \mathbb{C}^s. \end{aligned} \tag{3.52}$$

This yields the system of non-autonomous evolution equations

$$\partial_t z_j(t) = \varepsilon \sum_{\#J=j} \mathbf{F}_\varepsilon(t, z, J), \tag{3.53}$$

where $\mathbf{F}_\varepsilon$ and its m-th entry are given by

$$\begin{aligned} \mathbf{F}_\varepsilon(t,z,J) &= \Big(F_{\varepsilon,m}(t,z,J)\Big)_{m=1}^s, \\ F_{\varepsilon,m}(t,z,J)(k) &= \sum_M \int_{\#K=k} \exp\Big(\tfrac{\mathrm{i}t}{\varepsilon}[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)]\Big) Z_{JM}(t,K) \frac{\psi^*_{jm}(\varepsilon k) T\big(\psi_{JM}(\varepsilon K)\big)}{(2\pi)^d}\, \mathrm{d}K. \end{aligned} \tag{3.54}$$

Since these are rather complicated formulas, we will avoid these whenever possible, for example in the following lemma, which is helpful for the analysis in Chapter 4.

Lemma 3.7.1. *If $\widehat{v} = (\widehat{v}_1, \dots, \widehat{v}_{j_{max}})$ with $\widehat{v}_j \in L^1$ and $\widehat{v}_{-j}(k) = \overline{\widehat{v}_j(-k)}$ for $j \in \mathcal{J}_+$, then for every $J = (j_1, j_2, j_3) \in \mathcal{J}^3$ the inequality*

$$\left\|\mathbf{F}_\varepsilon(t, \widehat{v}, J)\right\|_{L^1} \leqslant C_{\mathcal{T}} \prod_{i=1}^{3} \|S_{j_i,\varepsilon}(t)\widehat{v}_{j_i}\|_{L^1},$$

holds for all $t \geqslant 0$.

Proof. The definition (3.49), the inequality (3.28) of Lemma 3.5.1 and the fact that $S_{j,\varepsilon}(t)$ is unitary for all $t \geqslant 0$ imply that

$$\left\|\mathbf{F}_\varepsilon(t, \widehat{v}, J)\right\|_{L^1} \leqslant \|\mathcal{T}(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3})\|_{L^1} \leqslant C_{\mathcal{T}} \|\widehat{v}_{j_1}\|_{L^1} \|\widehat{v}_{j_2}\|_{L^1} \|\widehat{v}_{j_3}\|_{L^1}.$$

Finally, the assertion follows from $\|\widehat{v}_{j_i}\|_{L^1} = \|S_{j_i,\varepsilon}(t)\widehat{v}_{j_i}\|_{L^1}$. ∎

For the rest of the chapter, we consider three issues related to the transformation. Furthermore, they serve as preparation for the next chapter.

Resonances. For the analysis the crucial term in (3.54) is

$$\exp\left(\tfrac{\mathrm{i}t}{\varepsilon}\left[\lambda_{jm}(\varepsilon k) - \sum_{i=1}^{3} \lambda_{j_i m_i}(\varepsilon k^{(i)})\right]\right) = \exp\left(\tfrac{\mathrm{i}t}{\varepsilon}[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)]\right). \tag{3.55}$$

depending on interaction of the eigenvalues. The eigenvalue $\lambda_{jm}(\varepsilon k)$ comes from the transformation (3.46), whereas the eigenvalues $\lambda_{j_i m_i}(\varepsilon k^{(i)})$ come from the inverse transformation (3.50). The term (3.55) leads to two problems, namely oscillations and resonances. To see the impact of the two effects, we consider the highly oscillatory integral

$$\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma, \tag{3.56}$$

for fixed k, K, m, M, j and J, where

$$\Delta\lambda_{jmJM}(\varepsilon, k, K) = \lambda_{jm}(\varepsilon k) - \sum_{i=1}^{3} \lambda_{j_i m_i}(\varepsilon k^{(i)}). \tag{3.57}$$

The integral (3.56) can be calculated explicitly. We distinguish two cases, where we obtain for $t \in [0, t_{\mathrm{end}}/\varepsilon]$

$$\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma = \begin{cases} \int_0^t 1\,\mathrm{d}\sigma = t, & \text{if } \Delta\lambda_{jmJM}(\varepsilon, k, K) = 0, \\ \frac{\varepsilon}{\mathrm{i}}\left(\Delta\lambda_{jmJM}(\varepsilon, k, K)\right)^{-1}\left[e^{\frac{\mathrm{i}t}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)} - 1\right], & \text{else.} \end{cases} \tag{3.58}$$

The difference $\Delta\lambda_{jmJM}(\varepsilon, k, K)/\varepsilon$ plays a key role in the analysis. If this term is large, then the integrand of (3.56) will be highly oscillatory. In this case, the interaction between the eigenvalues $\lambda_{jm}(\varepsilon k)$,

$\lambda_{j_1m_1}(\varepsilon k^{(1)})$, $\lambda_{j_2m_2}(\varepsilon k^{(2)})$ and $\lambda_{j_3m_3}(\varepsilon k^{(3)})$ is *non-resonant*. The value of the integral will be rather small. For example if $\Delta\lambda_{jmJM}(\varepsilon,k,K)=\mathcal{O}(1)$, we obtain with (3.58)

$$\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma=\mathcal{O}(\varepsilon)\qquad\text{for } t\in[0,t_{\mathrm{end}}/\varepsilon].$$

However, if $\left|\Delta\lambda_{jmJM}(\varepsilon,k,K)/\varepsilon\right|$ is equal or close to zero, the value of the integral will be large. The case that $\Delta\lambda_{jmJM}(\varepsilon,k,K)/\varepsilon$ is small means in our case that $|\Delta\lambda_{jmJM}(\varepsilon,k,K)|\ll\varepsilon$ and, hence, the term $\varepsilon\left(\Delta\lambda\right)^{-1}$ in (3.58) is large. For the special case $\Delta\lambda_{jmJM}(\varepsilon,k,K)=0$ the integral (3.58) grows with t, and thus for $t\in[0,t_{\mathrm{end}}/\varepsilon]$ the value is $\mathcal{O}(\varepsilon^{-1})$ in the worst case. In this case, there are *resonant* interactions of the eigenvalues. It needs some effort to control these resonant interactions. An important tool for controlling this resonant interactions is to impose that these interactions are fairly rare by means of non-resonance conditions. However, in general those non-resonance conditions are not fulfilled for *every* k, $k^{(i)}\in\mathbb{R}^d$ and *every* combination $j\in\mathcal{J}$, $m\in\{1,\dots,s\}$, $J\in\mathcal{J}^3$ and $M\in\{1,\dots,s\}^3$. These conditions mainly depend on the structure of the eigenvalues and this in turn depends on the data of the underlying problem. Therefore, our idea is to use non-resonance conditions which are fulfilled for a *special* value $\theta=\varepsilon k\in\mathbb{R}^d$ to control the whole difference $\Delta\lambda_{jmJM}(\varepsilon,k,K)$, or parts of it, for fixed k, $k^{(i)}\in\mathbb{R}^d$. We illustrate this idea with an example. It turns out that for the problem settings we consider the choice $\theta=0$ is beneficial. One reason is that for $\theta=0$ all eigenvalues, except $\lambda_{11}(0)$ and $\lambda_{1m_\star}(0)$, are not equal to zero. Consequently, the inverses of all these nonzero eigenvalues exist and we also know that $\Lambda_j(0)$ is invertible for all $j>1$ because of Assumption 3.2.6. With the abbreviation $\Delta\lambda_{jmJM}(0)=\Delta\lambda_{jmJM}\big(\varepsilon,0,(0,0,0)\big)$ we assume that there exists a constant $c_{\mathrm{res}}>0$ such that

$$|\Delta\lambda_{jmJM}(0)|=|\lambda_{jm}(0)-\lambda_{j_1m_1}(0)-\lambda_{j_2m_2}(0)-\lambda_{j_3m_3}(0)|\geqslant c_{\mathrm{res}}$$

for fixed $j\in\mathcal{J}$, $m\in\{1,\dots,s\}$, $J\in\mathcal{J}^3$ and $M\in\{1,\dots,s\}^3$. Then, we rewrite (3.56) as

$$\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma=\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(0)}e^{\frac{\mathrm{i}\sigma}{\varepsilon}(\Delta\lambda_{jmJM}(\varepsilon,k,K)-\Delta\lambda_{jmJM}(0))}\,\mathrm{d}\sigma$$

and use integration by parts. This yields

$$\begin{aligned}\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma&=\varepsilon\left(\mathrm{i}\Delta\lambda_{jmJM}(0)\right)^{-1}\left[e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\right]_{\sigma=0}^{t}\\&\quad-\left(\Delta\lambda_{jmJM}(0)\right)^{-1}\left(\Delta\lambda_{jmJM}(\varepsilon,k,K)-\Delta\lambda_{jmJM}(0)\right)\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma\end{aligned}$$

and for the absolute value we conclude

$$\left|\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma\right|\leqslant 2\varepsilon c_{\mathrm{res}}^{-1}+tc_{\mathrm{res}}^{-1}\left|\Delta\lambda_{jmJM}(\varepsilon,k,K)-\Delta\lambda_{jmJM}(0)\right|.$$

At first glance, this does not look like an improvement, since the second term still grows with t. However, if we are able to gain an additional factor ε by the difference $|\Delta\lambda_{jmJM}(\varepsilon,k,K)-\Delta\lambda_{jmJM}(0)|$, we achieve $\mathcal{O}(1)$ instead of $\mathcal{O}(\varepsilon^{-1})$ on long time intervals.
In order to be able to handle the difference $|\Delta\lambda_{jmJM}(\varepsilon,k,K)-\Delta\lambda_{jmJM}(0)|$ appropriately later, we make an additional assumption.

Assumption 3.7.2. *The map $\beta\mapsto\omega_m(\beta)$, where $\omega_m(\beta)$ is an eigenvalue of $\mathcal{L}(0,\beta)$, is globally Lipschitz continuous, i.e. there is a constant C such that*

$$|\omega_m(\widetilde{\beta})-\omega_m(\beta)|\leqslant C|\widetilde{\beta}-\beta|_1\qquad\text{for all }\ \widetilde{\beta},\beta\in\mathbb{R}^d.\tag{3.59}$$

Remark 3.7.3. *Because of the special form of the eigenvalues $\lambda_{jm}(\theta)$ given in (3.44), Assumption 3.7.2 implies (3.59) also for every $\lambda_{jm}(\theta)$ with $j \in \mathcal{J}_+$, $m \in \{1, \dots, s\}$ and $\theta \in \mathbb{R}^d$.*

Hence, with the triangle inequality applied on the difference $|\Delta\lambda_{jmJM}(\varepsilon, k, K) - \Delta\lambda_{jmJM}(0)|$ and by the Lipschitz continuity of every single eigenvalue, we obtain

$$\begin{aligned}\left|\int_0^t e^{\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)}\,\mathrm{d}\sigma\right| &\leqslant 2\varepsilon c_{\mathrm{res}}^{-1} + t c_{\mathrm{res}}^{-1}\left|\Delta\lambda_{jmJM}(\varepsilon,k,K) - \Delta\lambda_{jmJM}(0)\right| \\ &\leqslant 2\varepsilon c_{\mathrm{res}}^{-1} + \varepsilon t c_{\mathrm{res}}^{-1} C\left[|k|_1 + \sum_{i=1}^{3}|k^{(i)}|_1\right].\end{aligned}$$

Unfortunately, k and $k^{(i)}$ for $i = 1, 2, 3$ are not elements of a compact set later, but this problem is compensated by the fact that we consider integrals of the form (3.56) where the integrand is additionally multiplied by a function that decays faster than k and $k^{(i)}$ grow.

After deriving the equations of motion for the new variables z_j and highlighting the challenge of resonances, we now turn to the initial data for the new variables.

Initial data. The initial data for z_j with $j \in \mathcal{J}_+$ are obtained from (3.25) and (3.46). More precisely, the relation

$$z_j(0,k) = S_{j,\varepsilon}(0,k)\widehat{u}_j(0,k) = \begin{cases}\Psi_1^*(\varepsilon k)\widehat{p}(k) & \text{if } j = 1,\\ 0 & \text{if } j > 1\end{cases} \tag{3.60}$$

is valid. Since we know that $\ker\big(\mathcal{L}(\omega,\kappa)\big) = \ker\big(\mathcal{L}_1(0)\big) = \operatorname{span}\{\psi_{11}(0), \psi_{1m_\star}(0)\}$ and the initial data $\widehat{p}(k)$ can be written as a linear combination of the vectors $\psi_{11}(0)$ and $\psi_{1m_\star}(0)$, it follows from Assumption 3.2.1 that $\psi_{1m}^*(0)\widehat{p}(k) = 0$ for all k and all $m \in \{m_\star + 1, \dots, s\}$. Hence, the result follows by the orthogonality of the vectors $\psi_{1m}(0)$. However, this relation is in general not true for $\psi_{1m}^*(\varepsilon k)\widehat{p}(k)$. Indeed another helpful result can be proven. It has been shown in [11, proof of Lemma 3] under Assumption 3.7.2 that for initial data $p \in W^1$ the estimate

$$\|\psi_{1m}(\varepsilon\cdot)\psi_{1m}^*(\varepsilon\cdot)\widehat{p}\|_{L^1} \leqslant C\varepsilon\|\nabla p\|_W \tag{3.61}$$

holds for every $m \in \{m_\star + 1, \dots, s\}$. Since the proof in [11, proof of Lemma 3] is very concise, we elaborate it at this point.

Proof of (3.61): We know by Assumption 3.2.2 and the following remark that $\mathcal{L}_1(\theta)$ has the same eigenvectors as $\mathcal{L}(0, \kappa + \theta)$. Therefore, we also have the smoothness specified in Assumption 3.2.2. This means that the eigenvectors have the properties

- $|\psi_{1m}(\theta)|_2 = 1$ for all $\theta \in \mathbb{R}^d$ and all m,
- and $\psi_{1m} \in C^\infty(\mathbb{R}^d \backslash \{-\kappa\}, \mathbb{C}^n)$, since in this case $\beta := \kappa + \theta = 0$ if $\theta = -\kappa$.

In particular, we know that ψ_{1m} is C^∞ on the ball with center 0 and radius $\frac{|\kappa|_1}{2}$ which will be helpful later.

The first step is to rewrite for $m \in \{m_\star + 1, \dots, s\}$

$$\begin{aligned}\psi_{1m}(\varepsilon k)\psi_{1m}^*(\varepsilon k)\widehat{p}(k) &= \psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k) + \psi_{1m}(\varepsilon k)\psi_{1m}^*(0)\widehat{p}(k)\\ &= \psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k),\end{aligned}$$

since according to Assumption 3.2.1 we have $\psi_{1m}^*(0)\widehat{p}(k) = 0$ for all k and all $m \in \{m_\star + 1, \dots, s\}$.
In general the derivatives of ψ_{1m} are not bounded near the vector $-\kappa$, since $\psi_{1m} \in C^\infty(\mathbb{R}^d\backslash\{-\kappa\}, \mathbb{C}^n)$, and therefore, the term $\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)$ cannot be treated directly by means of a Taylor expansion. Hence, we decompose the difference as

$$\begin{aligned}\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right] &= \psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\chi_{\{\varepsilon|k|_1 \leqslant \frac{|\kappa|_1}{2}\}}\\ &\quad + \psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\chi_{\{\varepsilon|k|_1 > \frac{|\kappa|_1}{2}\}},\end{aligned}$$

where χ is the characteristic function.
Let $\overline{B}(0, \frac{|\kappa|_1}{2\varepsilon}) \subset \mathbb{R}^d$ be the closed ball with center 0 and radius $\frac{|\kappa|_1}{2\varepsilon}$. Thus, we divide

$$\begin{aligned}\|\psi_{1m}(\varepsilon\cdot)\psi_{1m}^*(\varepsilon\cdot)\widehat{p}\|_{L^1} &= \int_{\mathbb{R}^d} |\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k)|_2 \,\mathrm{d}k\\ &= \int_{\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} |\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k)|_2 \,\mathrm{d}k\\ &\quad + \int_{\mathbb{R}^d\backslash\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} |\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k)|_2 \,\mathrm{d}k.\end{aligned}$$

For the first term we are now in the position to use Taylor expansion, since ψ_{1m} is C^∞ on the ball with the origin as the center and radius $\frac{|\kappa|_1}{2\varepsilon}$. Therefore, we obtain

$$\int_{\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} \left|\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1\ell}^*(0)\right]\widehat{p}(k)\right|_2 \mathrm{d}k \leqslant C\varepsilon \int_{\overline{B}_0(\frac{|\kappa|_1}{2\varepsilon})} |k|_1 \left|\widehat{p}(k)\right|_2 \mathrm{d}k \leqslant C\varepsilon\|\widehat{\nabla p}\|_{L^1}.$$

For the second term we use the relation that for all $k \in \mathbb{R}^d\backslash\overline{B}(0, \frac{|\kappa|_1}{2\varepsilon})$ we have $\varepsilon|k|_1 \geqslant \frac{|\kappa|_1}{2}$, which can be written as $1 \leqslant \frac{2\varepsilon|k|_1}{|\kappa|_1}$. Hence, with $\psi_{1m}^*(0)\widehat{p}(k) = 0$ it follows that

$$\begin{aligned}\int_{\mathbb{R}^d\backslash\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} \left|\psi_{1m}(\varepsilon k)\left[\psi_{1m}^*(\varepsilon k) - \psi_{1m}^*(0)\right]\widehat{p}(k)\right|_2 \mathrm{d}k &= \int_{\mathbb{R}^d\backslash\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} \left|\psi_{1m}(\varepsilon k)\psi_{1m}^*(\varepsilon k)\right|_2 \left|\widehat{p}(k)\right|_2 \mathrm{d}k\\ &= \int_{\mathbb{R}^d\backslash\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} \left|\widehat{p}(k)\right|_2 \mathrm{d}k\\ &\leqslant \frac{2\varepsilon}{|\kappa|_1}\int_{\mathbb{R}^d\backslash\overline{B}(0,\frac{|\kappa|_1}{2\varepsilon})} |k|_1 \left|\widehat{p}(k)\right|_2 \mathrm{d}k \leqslant C\varepsilon\|\widehat{\nabla p}\|_{L^1},\end{aligned}$$

since $|\psi_{1m}(\theta)|_2 = 1$ for all $\theta \in \mathbb{R}^d$. This yields (3.61). ∎

This estimate (3.61) has an impact on the initial data of the new variable. With (3.51) we obtain

$$z_{1m}(0,k) = \psi_{1m}^*(\varepsilon k)\widehat{u}_1(0,k) = \psi_{1m}^*(\varepsilon k)\widehat{p}(k).$$

Expressed with the new variable z_1, since $|\psi_{1m}(\varepsilon k)|_2 = 1$ the estimate (3.61) and the definition of the Wiener algebra lead to

$$\|z_{1m}(0)\|_{L^1} \leqslant C\varepsilon\|\nabla p\|_W \qquad \text{for all } m \in \{m_\star + 1, \dots, s\}. \tag{3.62}$$

Consequently, we have a refined bound for certain parts of the initial data $z_1(0)$. In Chapter 4 we will show that a similar refined estimate of $\mathcal{O}(\varepsilon)$ holds for $z_{1m}(t)$, $m \in \{m_\star + 1, \dots, s\}$, for all times $t \in [0, t_{\text{end}}/\varepsilon]$, where in order to prove this estimate we need (3.62).

In view of Assumption 3.2.1 and its previously shown consequence (3.62) it is helpful to define the following projections.

Projections. The special role of the first entry of z_1 for $m_\star = 1$ or the first two entries for $m_\star = 2$ can be expressed with a slight abuse of notation for $m_\star = 1$ by means of the projection

$$P : \mathbb{C}^s \to \mathbb{C}^s, \quad (w_1, \dots, w_s)^\top \mapsto (w_1, w_{m_\star}, 0, \dots, 0)^\top. \tag{3.63}$$

Furthermore, we define the projection $P^\perp = (I - P)$ which sets the first $m_\star$ entries of a vector to zero. Then, Assumption 3.2.1 implies that $\|P^\perp z_1(0)\|_{L^1} = \mathcal{O}(\varepsilon)$, because

$$\begin{aligned}\|P^\perp z_1(0)\|_{L^1} &= \int_{\mathbb{R}^d} |P^\perp z_1(0,k)|_2 \,\mathrm{d}k \leqslant \int_{\mathbb{R}^d} |P^\perp z_1(0,k)|_1 \,\mathrm{d}k = \sum_{m=m_\star+1}^{s} \int_{\mathbb{R}^d} |z_{1m}(0,k)| \,\mathrm{d}k \\ &= \sum_{m=m_\star+1}^{s} \|z_{1m}(0)\|_{L^1} \leqslant C(s-m_\star)\varepsilon \|\widehat{\nabla p}\|_{L^1} \leqslant C\varepsilon \|p\|_{W^1}\end{aligned} \tag{3.64}$$

due to (3.62). We also define the projection

$$\widehat{w} \mapsto \mathcal{P}_\varepsilon \widehat{w}, \qquad \mathcal{P}_\varepsilon(k)\widehat{w}(k) = \sum_{m=1}^{m_\star} \psi_{1m}(\varepsilon k)\psi_{1m}^*(\varepsilon k)\widehat{w}(k) \tag{3.65}$$

which projects a vector-valued function $\widehat{w} : \mathbb{R}^d \to \mathbb{C}^s$ pointwise into the first eigenspace of $\mathcal{L}_1(\varepsilon k)$. In addition we define $\mathcal{P}_\varepsilon^\perp = I - \mathcal{P}_\varepsilon$ with

$$\mathcal{P}_\varepsilon^\perp(k)\widehat{w}(k) = \sum_{m=m_\star+1}^{s} \psi_{1m}(\varepsilon k)\psi_{1m}^*(\varepsilon k)\widehat{w}(k). \tag{3.66}$$

With these definitions it follows for the inverse transformation (3.50) with $j = 1$ that

$$\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k) = S_{1,\varepsilon}^*(t,k) P z_1(t,k). \tag{3.67}$$

Relation (3.67) holds because with the representation (3.51) for $j = 1$ and

$$\mathcal{P}_\varepsilon(k)\psi_{1m}(\varepsilon k) = \begin{cases} \psi_{1m}(\varepsilon k), & \text{for } m \in \{1, m_\star\}, \\ 0, & \text{else,} \end{cases}$$

we have

$$\begin{aligned}\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k) &= \mathcal{P}_\varepsilon(k) \sum_{m=1}^{s} \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m}(\varepsilon k)\right) z_{1m}(t,k)\psi_{1m}(\varepsilon k) \\ &= \sum_{m=1}^{m_\star} \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m}(\varepsilon k)\right) z_{1m}(t,k)\psi_{1m}(\varepsilon k) = \sum_{m=1}^{s} \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m}(\varepsilon k)\right) y_{1m}(t,k)\psi_{1m}(\varepsilon k) \\ &= S_{1,\varepsilon}^*(t,k) P z_1(t,k),\end{aligned}$$

where

$$y_1 = (z_{11}, z_{1m_\star}, 0, \dots, 0)^\top .$$

We recall that $S^*_{1,\varepsilon}(t,k)$ is unitary and thus, combining (3.47) with (3.67) and (3.63) yields

$$\|\mathcal{P}_\varepsilon(k)\widehat{u}_1(t)\|_{L^1} = \|Pz_1(t)\|_{L^1} \leqslant \|z_1(t)\|_{L^1} = \|\widehat{u}_1(t)\|_{L^1}. \tag{3.68}$$

Analogously we obtain

$$\mathcal{P}^\perp_\varepsilon(k)\widehat{u}_1(t,k) = S^*_{1,\varepsilon}(t,k)P^\perp z_1(t,k), \tag{3.69}$$

because

$$\begin{aligned}\mathcal{P}^\perp_\varepsilon(k)\widehat{u}_1(t,k) &= \mathcal{P}^\perp_\varepsilon(k)\sum_{m=1}^{s}\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m}(\varepsilon k)\right) z_{1m}(t,k)\psi_{1m}(\varepsilon k)\\ &= \sum_{m=m_\star+1}^{s}\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m}(\varepsilon k)\right) z_{1m}(t,k)\psi_{1m}(\varepsilon k) = S^*_{1,\varepsilon}(t,k)P^\perp z_1(t,k).\end{aligned}$$

The relation (3.69) is useful when we express the terms $\mathcal{P}^\perp_\varepsilon(k)\widehat{u}_1(t,k)$ by $P^\perp z_1(t,k)$ later in the analysis. Since $S^*_{1,\varepsilon}(t,k)$ is unitary we obtain the relation

$$\|\mathcal{P}^\perp_\varepsilon\widehat{u}_1(t)\|_{L^1} = \|P^\perp z_1(t)\|_{L^1}. \tag{3.70}$$

As mentioned previously, the refined bound (3.64) is needed in order to show a refined bound for $P^\perp z_1(t)$ for all $t \in [0, t_{\text{end}}/\varepsilon]$. With relation (3.70) this refined bound also holds for $\mathcal{P}^\perp_\varepsilon\widehat{u}_1(t)$.

We end the section with a brief summary. The bottom line is that the introduced transformation is the crucial point to show the higher accuracy of the SVEA in Chapter 4. It leads to the two important components of the error analysis. The first component is that we need suitable non-resonance conditions in order to apply integration by parts. In addition, it will be helpful to split the coefficient $\widehat{u}_1(t) = \mathcal{P}_\varepsilon\widehat{u}_1(t) + \mathcal{P}^\perp_\varepsilon\widehat{u}_1(t)$ by means of the introduced projections (3.63) and (3.65). With (3.67) and (3.69) a connection between the two terms of the splitting and the new variable is established. In summary, using the definitions, assumptions, results, and formulas introduced in this chapter, the next step is to prove the higher accuracy of the SVEA in the next chapter.

CHAPTER 4

An improved error bound for the SVEA

In this chapter we present our first main results. We recall that for the SVEA the error bound (3.13) was shown in [11, Theorem 1] under a number of assumptions. Adapted to our notation this means that there is a $t_\star \in (0, t_{\text{end}}]$ independent of ε such that for all $\varepsilon \in (0, 1]$

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(1)}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon, \tag{4.1}$$

where $\widetilde{\mathbf{u}}^{(1)}$ denotes the approximation (3.15) with $j_{\max} = 1$. In [4] we set $j_{\max} = 3$ and prove under certain assumptions the refined bounds

$$\sup_{t\in[0,t_\star/\varepsilon]} \left(\|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} + \sum_{\mu=1}^{d} \|D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \right) \leqslant C\varepsilon, \tag{4.2}$$

$$\sup_{t\in[0,t_\star/\varepsilon]} \|u_3(t)\|_{W^1} \leqslant C\varepsilon, \tag{4.3}$$

cf. [4, Proposition 3.2 and Proposition 3.6]. The definitions of the Fourier multiplicator D_μ and the projection $\mathcal{P}_\varepsilon^\perp$ are given at the end of Chapter 2 and in (3.66), respectively. Furthermore, we show in [4, Theorem 4.2] the error bound

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon^2. \tag{4.4}$$

Now in this thesis, one of the main results is that the accuracy (4.1) of the SVEA can be improved by one power with respect to ε. This is not only an improvement on the result from [11], but also on the result from [4], because already with $\widetilde{\mathbf{u}}^{(1)}$ instead of $\widetilde{\mathbf{u}}^{(3)}$ an accuracy of $\mathcal{O}(\varepsilon^2)$ is obtained. The chapter is structured as follows. Following the main ideas in [4, Proposition 3.2 and Proposition 3.6], we prove refined bounds for $P^\perp z_1(t)$ and $\mathcal{P}_\varepsilon^\perp \widehat{u}_1$ in the L^1-norm, as in (4.2). Roughly speaking, this part of the constructed approximation $\widetilde{\mathbf{u}}^{(1)}(t)$ is of $\mathcal{O}(\varepsilon)$ on long time intervals of length $\mathcal{O}(1/\varepsilon)$. This statement is made precise in Propositions 4.1.4 and 4.2.2. The main result is Theorem 4.3.4 which provides the improved error bound for the approximation (3.15) with $j_{\max} = 1$. We recall that for $j_{\max} = 1$ we have $\mathcal{J} = \{\pm 1\}$ and

the sum of the right-hand side of (3.16) for $j = 1$ is taken over the set $\{(1,1,-1),(1,-1,1),(-1,1,1)\}$, cf. (3.17). Instead of the error bound (3.13), we are able to prove

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(1)}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon^2$$

under a number of assumptions. These assumptions are similar to the assumptions which are made in [11, Theorem 1], however, we assume slightly higher regularity of the initial data and additional non-resonance assumptions. The proof relies on the important results in Propositions 4.1.4 and 4.2.2. Section 4.3 is mostly devoted to the proof of this theorem. A possible extension to higher accuracy is discussed in the last section. Instead of the refined bound (4.3) and the error bound (4.4), we show in Subsection 4.4.1 for $j_{\max} = 3$ how to proceed to show an improved refined bound

$$\sup_{t\in[0,t_\star/\varepsilon]} \|u_3(t)\|_{W^1} \leqslant C\varepsilon^2$$

and under certain non-resonance conditions the improved estimate

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_{L^\infty(\mathbb{R}^d)} \leqslant C\varepsilon^3.$$

We illustrate the analytical results by numerical experiments at the end of Section 4.3 and 4.4.

4.1 Refined bound

We recall that we introduced in Section 3.7 a parameter $m_\star \in \{1,2\}$ such that $\mathcal{L}_1(0)$ has an $m_\star$-dimensional kernel. For the sake of simplicity we assume throughout the chapter that $m_\star = 1$. However, all the results and proofs also work for $m_\star = 2$ at the cost of a more complicated notation. Thus, we state the following assumption.

Assumption 4.1.1. *The kernel of $\mathcal{L}(\omega,\kappa)$ is one-dimensional.*

At the end of Chapter 3 we explained how to incorporate integration by parts in the error analysis. For this purpose, let $u_1 \in C^1([0,t_{\text{end}}/\varepsilon],W) \cap C([0,t_{\text{end}}/\varepsilon],W^1)$ be a classical solution of (3.16) for $j_{\max} = 1$ with initial data (3.18) for some $p \in W^1$.
We observe that for $f \in C^1([0,t_{\text{end}}/\varepsilon],W) \cap C([0,t_{\text{end}}/\varepsilon],W^1)$ and Λ invertible, we obtain

$$\int_0^t \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\right) f(\sigma)\,\mathrm{d}\sigma = \Lambda^{-1}\tfrac{\varepsilon}{\mathrm{i}}\Big[\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\right) f(\sigma)\Big]_{\sigma=0}^t - \Lambda^{-1}\tfrac{\varepsilon}{\mathrm{i}}\int_0^t \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\right)\partial_t f(\sigma)\,\mathrm{d}\sigma, \tag{4.5}$$

such that with $|\Lambda^{-1}|_2 \leqslant C$ and $t \leqslant \frac{t_{\text{end}}}{\varepsilon}$

$$\Big|\int_0^t \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\right) f(\sigma)\,\mathrm{d}\sigma\Big|_2 \leqslant \varepsilon C\left[|f(t)|_2 + |f(0)|_2\right] + t_{\text{end}} C \sup_{\sigma\in[0,t_{\text{end}}/\varepsilon]} |\partial_t f(\sigma)|_2. \tag{4.6}$$

In order to have a uniform bound in ε of the left-hand side of (4.6), we need that $f(t)$ and $\partial_t f(t)$ are uniformly bounded in ε for all $t \in [0,t_{\text{end}}/\varepsilon]$.

As mentioned in Section 3.7, the ODE system (3.24) suggests that formally $\partial_t \widehat{u}_1(t) = \mathcal{O}(1/\varepsilon)$. The following lemma shows, however, that $\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(t)$ can be bounded independently of ε on long time intervals, which is later useful for the proof of Proposition 4.1.4, where we bound terms of the form (4.6). The lemma corresponds to [11, Lemma 2].

Lemma 4.1.2. *Under the assumptions of Lemma 3.6.1 (ii) with $j_{max} = 1$, Assumptions 3.7.2 and 4.1.1, there is a constant C such that*

$$\sup_{t\in[0,t_{end}/\varepsilon]} \|\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C.$$

C depends on the constant $C_{u,1}$ from (3.34) and thus also on t_{end}, but not on ε.

Proof. The proof is based on the definition of the projection $\mathcal{P}_\varepsilon(k)$, the dispersion relation (3.4), and the Lipschitz continuity of the eigenvalues, cf. Assumption 3.7.2. We observe that

$$\frac{\mathrm{i}}{\varepsilon}\mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k) = \frac{\mathrm{i}}{\varepsilon}\psi_{11}(\varepsilon k)\psi_{11}^*(\varepsilon k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k) = \frac{\mathrm{i}}{\varepsilon}\lambda_{11}(\varepsilon k)\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k) \tag{4.7}$$

because of (3.65) and (3.43). By Assumption 3.7.2 the Lipschitz continuity of the eigenvalues and the fact that $\lambda_{11}(0) = 0$ yield

$$|\lambda_{11}(\varepsilon k)| = |\lambda_{11}(\varepsilon k) - \lambda_{11}(0)| \leqslant C\varepsilon|k|_1. \tag{4.8}$$

Hence, we obtain in the Euclidean norm

$$\left|\frac{\mathrm{i}}{\varepsilon}\mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k)\right|_2 = \frac{1}{\varepsilon}|\lambda_{11}(\varepsilon k) - \lambda_{11}(0)|\left|\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)\right|_2 \leqslant C|k|_1\left|\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)\right|_2. \tag{4.9}$$

Applying the projection $\mathcal{P}_\varepsilon(k)$ to (3.24) with $j_{\max} = 1$ yields

$$\mathcal{P}_\varepsilon(k)\partial_t\widehat{u}_1(t,k) = -\frac{\mathrm{i}}{\varepsilon}\mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k) + \varepsilon\sum_{\#J=1}\mathcal{P}_\varepsilon(k)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t,k).$$

Together with (4.9) and the inequality (3.28) of Lemma 3.5.1 this shows that $\mathcal{P}_\varepsilon\partial_t\widehat{u}_1(t,k) = \partial_t\mathcal{P}_\varepsilon\widehat{u}_1(t,k)$ is uniformly bounded by

$$\sup_{t\in[0,t_{\mathrm{end}}/\varepsilon]} \|\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C_{u,1} + \varepsilon C_{\mathcal{T}} C_{u,1}^3.$$

■

Remark 4.1.3. *Under the assumptions of Lemma 4.1.2 we have with (3.68) that*

$$\sup_{t\in[0,t_{end}/\varepsilon]} \|\mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant \sup_{t\in[0,t_{end}/\varepsilon]} \|\widehat{u}_1(t)\|_{L^1} \leqslant C$$

uniformly in ε. Lemma 4.1.2 implies that formally the part $\mathcal{P}_\varepsilon\widehat{u}_1(t)$ of the classical solution u_1 is essentially non-oscillatory on long time intervals of length $\mathcal{O}(1/\varepsilon)$. Furthermore, the refined bound (4.10) for $\mathcal{P}_\varepsilon^\perp\widehat{u}_1$ means that $\widehat{u}_1(t) = \mathcal{P}_\varepsilon\widehat{u}_1(t) + \mathcal{O}(\varepsilon)$. We interpret this in the sense that the "main part" of the solution of (3.24) with $j_{max} = 1$ is $\mathcal{P}_\varepsilon\widehat{u}_1(t)$. This favourable property will be useful later as we split $\widehat{u}_1(t)$ into these two parts: the part $\mathcal{P}_\varepsilon\widehat{u}_1(t)$ which is $\mathcal{O}(1)$ but non-oscillatory and the part $\mathcal{P}_\varepsilon^\perp\widehat{u}_1(t)$ which is oscillatory but $\mathcal{O}(\varepsilon)$.

The main goal in this section is to prove that under certain assumptions there is a $t_\star \in (0, t_{\mathrm{end}}]$ independent of ε such that for all $\varepsilon \in (0, 1]$

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon, \tag{4.10}$$

uniformly in ε. With the relation (3.69) and since $S^*_{1,\varepsilon}(t,k)$ is unitary, this bound is equivalent to

$$\sup_{t\in[0,t_\star/\varepsilon]} \|P^\perp z_1(t)\|_{L^1} \leqslant C\varepsilon, \tag{4.11}$$

for all $\varepsilon \in (0,1]$.

In order to prove (4.11) we define the scaled norm

$$|||y|||_\varepsilon = 2\|Py_1\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp y_1\|_{L^1} \tag{4.12}$$

for all $y = (y_1, \dots, y_{j_{\max}})$ with $y_j \in L^1(\mathbb{R}^d, \mathbb{C}^n)$. As before, we set $y_{-1} = \overline{y_1}$. The factor 2 is introduced to take into account the terms with negative indices which appear due to the nonlinearity. By definition of the L^1-norm and the projection (3.63) we estimate

$$\|y_1\|_{L^1} \leqslant \|Py_1\|_{L^1} + \|P^\perp y_1\|_{L^1} \leqslant \|Py_1\|_{L^1} + \varepsilon^{-1}\|P^\perp y_1\|_{L^1} \leqslant |||y|||_\varepsilon, \tag{4.13}$$

which holds for all $\varepsilon \in (0,1]$. This estimate will be used frequently. In summary the goal in the following Proposition 4.1.4 is to prove that there is a constant C independent of ε such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_\varepsilon \leqslant C,$$

for all $\varepsilon \in (0,1]$. This estimate implies the refined bound (4.11) and hence also (4.10).

Proposition 4.1.4. *Let u_1 be the classical solution of (3.16) with $j_{max} = 1$ and initial data (3.18) for some $p \in W^1$. Let z_1 be the transformed variable defined in (3.46). For every sufficiently large $r > 0$ there is a $t_\star \in (0, t_{end}]$ such that under the Assumptions 3.2.1, 4.1.1, 3.2.2, 3.7.2*

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_\varepsilon \leqslant r \qquad \text{for all } \varepsilon \in (0,1].$$

The constant $t_\star$ depends on t_{end}, r, $C_{u,1}$, $C_{\mathcal{T}}$, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

"Sufficiently large" means that r must be larger than the constant $C_\bullet$ which occurs in the proof. This condition is required to ensure that $t_\star$ defined in (4.21) is positive.

Proof. The proof of Proposition 4.1.4 is subdivided into two steps. In the first step we show the general procedure to prove this proposition and in the second step we show that the estimates used in Step 1 are actually fulfilled.

Step 1. Integrating the system of PDEs (3.48) with $j = 1$ from 0 to t and applying the scaled norm (4.12) leads to

$$\begin{aligned} |||z(t)|||_\varepsilon &\leqslant |||z(0)|||_\varepsilon + |||\int_0^t \partial_t z(\sigma)\,\mathrm{d}\sigma|||_\varepsilon \\ &\leqslant |||z(0)|||_\varepsilon + 2\sum_{\#J=1}\left(\varepsilon\Big\|\int_0^t P\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} + \Big\|\int_0^t P^\perp\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}\right) \end{aligned} \tag{4.14}$$

with $\mathbf{F}_\varepsilon$ defined in (3.49) and P defined in (3.63). The first term of (4.14) is given by (4.12) as

$$|||z(0)|||_\varepsilon = 2\|Pz_1(0)\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp z_1(0)\|_{L^1}.$$

With (3.60) this term is uniformly bounded by

$$|||z(0)|||_\varepsilon \leqslant C\big(\|p\|_{W^1}\big), \tag{4.15}$$

because of $\|P^\perp z_1(0)\|_{L^1} \leqslant C\varepsilon\|p\|_{W^1}$ due to the estimate (3.64). Next, we define

$$a(t) := \|Pz_1(t)\|_{L^1} + \varepsilon^{-1}\|P^\perp z_1(t)\|_{L^1}. \tag{4.16}$$

According to (4.12) it follows that

$$2a(t) = |||z(t)|||_\varepsilon\,. \tag{4.17}$$

The goal is to prove that there are constants $C_\star$ and $\widehat{C}$ such that

$$\varepsilon\Big\|\int_0^t P\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} + \Big\|\int_0^t P^\perp\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} \leqslant C_\star + \widehat{C}\varepsilon\int_0^t a^3(\sigma)\,\mathrm{d}\sigma \tag{4.18}$$

holds for all $t\in[0,t_{\mathrm{end}}/\varepsilon]$ and for every $J=(j_1,j_2,j_3)\in\mathcal{J}^3$ with $\#J=1$. If the estimate (4.18) is true, then substituting into the initial estimate (4.14) yields

$$|||z(t)|||_\varepsilon \leqslant C_\bullet + 2\widehat{C}\varepsilon\sum_{\#J=1}\int_0^t a^3(\sigma)\,\mathrm{d}\sigma \leqslant C_\bullet + \widehat{C}\varepsilon\int_0^t\Big(2a(\sigma)\Big)^3\,\mathrm{d}\sigma = C_\bullet + \widehat{C}\varepsilon\int_0^t|||z(\sigma)|||_\varepsilon^3\,\mathrm{d}\sigma.$$

For the last equality we use (4.17). The constant $C_\bullet$ depends on $\|p\|_{W^1}$ from the estimate (4.15), a constant $C_\star$, which is determined below, and the (finite) number of multi-indices $J\in\mathcal{J}^3$ with $\#J=1$ which are in this case $(1,1,-1)$, $(1,-1,1)$ and $(-1,1,1)$. Next, we set

$$\rho_\varepsilon(t) := \sup_{\sigma\in[0,t]}|||z(\sigma)|||_\varepsilon = 2\sup_{\sigma\in[0,t]}a(\sigma),$$

which is a monotonically increasing function in t. Furthermore, we obtain

$$\rho_\varepsilon(t) \leqslant C_\bullet + \widehat{C}\varepsilon\int_0^t|||z(\sigma)|||_\varepsilon^3\,\mathrm{d}\sigma \leqslant C_\bullet + \widehat{C}\varepsilon t\rho_\varepsilon^3(t). \tag{4.19}$$

We set $r > C_\bullet$. If we now choose $t_\star$ in such a way that

$$C_\bullet + \widehat{C}t_\star\rho_\varepsilon^3(t_\star) \leqslant r, \tag{4.20}$$

then (4.19) and the fact that ρ_ε is monotonically increasing implies that $\rho_\varepsilon(t)\leqslant r$ for all $t\in[0,t_\star/\varepsilon]$. We choose

$$t_\star = \frac{r-C_\bullet}{\widehat{C}r^3}, \tag{4.21}$$

so that the condition (4.20) holds. The choice (4.21) is a worst-case estimate and in most cases too pessimistic. The important aspect is that $t_\star$ depends on $C_\bullet$, r, and $\widehat{C}$, but not on ε.

Step 2. The remaining main part of the proof is to show the inequality (4.18). We consider multi-indices $J \in \mathcal{J}^3$ with $\#J = 1$. By means of Lemma 3.7.1 the first term on the left-hand side of (4.18) is estimated in a straightforward way. We obtain

$$\varepsilon \Big\| \int_0^t P\mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)\, \mathrm{d}\sigma \Big\|_{L^1} \leqslant \varepsilon \int_0^t \big\| \mathbf{F}_\varepsilon(\sigma, \widehat{u}, J) \big\|_{L^1}\, \mathrm{d}\sigma \leqslant C_{\mathcal{T}}\, \varepsilon \int_0^t \prod_{i=1}^{3} \|z_{j_i}(\sigma)\|_{L^1}\, \mathrm{d}\sigma \leqslant C_{\mathcal{T}}\, \varepsilon \int_0^t a^3(\sigma)\, \mathrm{d}\sigma, \quad (4.22)$$

because by definition (4.16) and the estimate (4.13) we have that $\|z_{j_i}(\sigma)\|_{L^1} \leqslant a(\sigma)$. The inequality (4.22) is indeed an estimate of the form (4.18) with $C_\star = 0$. The main difficulty is to prove a bound for the term

$$\Big\| \int_0^t P^\perp \mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)\, \mathrm{d}\sigma \Big\|_{L^1}$$

for all $J \in \mathcal{J}^3$ with $\#J = 1$. We aim to gain one power of ε from the oscillatory behavior of $P^\perp \mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)$. There are three multi-indices $J \in \mathcal{J}^3$ with $\#J = 1$. As an example we consider $J = (1, 1, -1)$, because the other two permutations can be treated in the same way.

We use the definition (3.49) to obtain

$$\Big\| \int_0^t P^\perp \mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)\, \mathrm{d}\sigma \Big\|_{L^1} = \Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}(\widehat{u}_1, \widehat{u}_1, \widehat{u}_{-1})(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1}.$$

First, the nonlinearity is split into eight parts. With $\widehat{u}_1 = \mathcal{P}_\varepsilon \widehat{u}_1 + \mathcal{P}_\varepsilon^\perp \widehat{u}_1$ and $\widehat{u}_{-1} = \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1} + \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}$ we have

$$\begin{aligned} \mathcal{T}\big(\widehat{u}_1, \widehat{u}_1, \widehat{u}_{-1}\big) &= \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big) + \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}\big) + \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big) \\ &\quad + \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big) + \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}\big) \\ &\quad + \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}\big) + \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big) + \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}\big). \end{aligned} \quad (4.23)$$

All nonlinearities where the term $\mathcal{P}_\varepsilon^\perp \widehat{u}_{\pm 1}$ appears in at least one of the three arguments are easy to treat. Since $S_{1,\varepsilon}(\sigma)$ is unitary and with (3.28) and (3.47), we obtain for example

$$\begin{aligned} \Big\| \int_0^t S_{1,\varepsilon}(\sigma) \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} &\leqslant \int_0^t \big\| \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(\sigma) \big\|_{L^1}\, \mathrm{d}\sigma \\ &\leqslant C_{\mathcal{T}}\, \varepsilon \int_0^t \Big(\tfrac{1}{\varepsilon} \|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(\sigma)\|_{L^1} \|\mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\|_{L^1} \|\mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\|_{L^1} \Big)\, \mathrm{d}\sigma \\ &= C_{\mathcal{T}}\, \varepsilon \int_0^t \Big(\tfrac{1}{\varepsilon} \|P^\perp z_1(\sigma)\|_{L^1} \|P z_1(\sigma)\|_{L^1} \|P z_1(\sigma)\|_{L^1} \Big)\, \mathrm{d}\sigma \\ &\leqslant C_{\mathcal{T}}\, \varepsilon \int_0^t a^3(\sigma)\, \mathrm{d}\sigma. \end{aligned}$$

For the last inequality, we use the fact that the definition (4.16) implies that $\|P z_1(\sigma)\|_{L^1} \leqslant a(\sigma)$ and $\varepsilon^{-1} \|P^\perp z_1(\sigma)\|_{L^1} \leqslant a(\sigma)$. For all the other terms in (4.23), except $\mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)$, we obtain an

estimate of the form (4.18) with $C_\star = 0$ and $\widehat{C} = C_{\mathcal{T}}$.
The main difficulty is to prove that the only remaining term

$$\Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \tag{4.24}$$

is uniformly bounded in ε in spite of the integration over a possibly long time interval $[0, t_{\mathrm{end}}/\varepsilon]$. As mentioned previously, the idea is to gain one factor ε from the oscillatory behavior of the term under the integral. Since $\mathcal{P}_\varepsilon \widehat{u}_1(\sigma)$ is essentially non-oscillatory on long time intervals of length $\mathcal{O}\big(\varepsilon^{-1}\big)$, we consider the oscillatory transformation $S_{1,\varepsilon}(\sigma)$. The first idea is to rewrite the transformation in a suitable way. In the following we define $\Delta_1(\varepsilon k) := \Lambda_1(\varepsilon k) - \Lambda_1(0)$. By means of (3.45) we obtain

$$P^\perp S_{1,\varepsilon}(\sigma, k) = P^\perp \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Lambda_1(0)\big) \exp\big(-\tfrac{\mathrm{i}\sigma}{\varepsilon} \Lambda_1(0)\big) S_{1,\varepsilon}(\sigma, k) = \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Lambda_1(0)\big) P^\perp \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Delta_1(\varepsilon k)\big) \Psi_1^*(\varepsilon k),$$

because the projection $P^\perp$ commutes with every diagonal matrix. Hence, the remaining term (4.24) can be expressed as

$$\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Lambda_1(0)\big) P^\perp f_\varepsilon(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \quad \text{with} \quad f_\varepsilon(t,k) = \exp\big(\tfrac{\mathrm{i}t}{\varepsilon} \Delta_1(\varepsilon k)\big) \Psi_1^*(\varepsilon k) \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(t,k).$$

In order to gain the missing factor ε we want to integrate by parts. We recall that $\lambda_{11}(0) = 0$ (cf. Section 3.7), because of the dispersion relation (3.4) combined with Assumption 4.1.1. Therefore, the diagonal matrix $\Lambda_1(0) = \operatorname{diag}\big(\lambda_{11}(0), \dots, \lambda_{1n}(0)\big)$ is not invertible. Fortunately, this problem is compensated by the projection $P^\perp$ which sets by definition the first entry of a vector to zero. Hence, we simply replace the eigenvalue $\lambda_{11}(0)$ by 1 or any other nonzero number and consider a new diagonal matrix $\widetilde{\Lambda}_1(0) = \operatorname{diag}\big(1, \lambda_{12}(0), \dots, \lambda_{1s}(0)\big)$ instead of $\Lambda_1(0)$. Next, we estimate the new term which contains $\widetilde{\Lambda}_1(0)$. Therefore, the goal is to show

$$\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Lambda_1(0)\big) P^\perp f_\varepsilon(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} = \Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \widetilde{\Lambda}_1(0)\big) P^\perp f_\varepsilon(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \leqslant C,$$

since this is an estimate of the form (4.18) with $\widehat{C} = 0$. The advantage is that now the modified matrix $\widetilde{\Lambda}_1(0)$ is invertible because $\lambda_{1m}(0) \neq 0$ for $m > 1$ by the dispersion relation (3.4) and Assumption 4.1.1. Thus, we integrate by parts and obtain

$$\begin{aligned}
&\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \widetilde{\Lambda}_1(0)\big) P^\perp f_\varepsilon(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \\
&\quad \leqslant \Big\| \Big[\tfrac{\varepsilon}{\mathrm{i}} \widetilde{\Lambda}_1^{-1}(0) \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \widetilde{\Lambda}_1(0)\big) P^\perp f_\varepsilon(\sigma) \Big]_{\sigma=0}^t \Big\|_{L^1} + \Big\| \tfrac{\varepsilon}{\mathrm{i}} \widetilde{\Lambda}_1^{-1}(0) \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \widetilde{\Lambda}_1(0)\big) P^\perp \partial_t f_\varepsilon(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \\
&\quad \leqslant C\varepsilon \Big(\|f_\varepsilon(0)\|_{L^1} + \|f_\varepsilon(t)\|_{L^1} \Big) + C\varepsilon \int_0^t \|\partial_t f_\varepsilon(\sigma)\|_{L^1} \, \mathrm{d}\sigma
\end{aligned} \tag{4.25}$$

with a constant which depends on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$. With $|S_{1,\varepsilon}(\sigma, k)|_2 = 1$, (3.28), and (3.34) it follows that for all $t \in [0, t_{\mathrm{end}}/\varepsilon]$

$$\|f_\varepsilon(t)\|_{L^1} = \big\| \mathcal{T}\big(\mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(t) \big\|_{L^1} \leqslant C_{\mathcal{T}} \|\mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1}^3 \leqslant C,$$

where C depends on $C_{\mathcal{T}}$ and the constant $C_{u,1}$ from (3.34). Therefore, the first two terms of (4.25) are bounded and in fact the factor ε is not needed. In contrast, we need the gained factor ε for the integral term in (4.25) to compensate for the long time interval. Next, with the product rule we have

$$\begin{aligned}\partial_t f_\varepsilon(\sigma,k) = {}& \tfrac{\mathrm{i}}{\varepsilon}\Delta_1(\varepsilon k)\exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta_1(\varepsilon k)\big)\Psi_1^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\overline{\mathcal{P}_\varepsilon}\widehat{u}_{-1}\big)(\sigma,k)\\ &+\exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta_1(\varepsilon k)\big)\Psi_1^*(\varepsilon k)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\overline{\mathcal{P}_\varepsilon}\widehat{u}_{-1}\big)(\sigma,k).\end{aligned} \tag{4.26}$$

By Assumption 3.7.2 we know that Λ_1 is globally Lipschitz continuous. Thus, we estimate the difference $\Delta_1(\varepsilon k)$ in the Euclidean norm by

$$|\tfrac{\mathrm{i}}{\varepsilon}\Delta_1(\varepsilon k)|_2 = \tfrac{1}{\varepsilon}|\Lambda_1(\varepsilon k)-\Lambda_1(0)|_2 \leqslant C|k|_1, \tag{4.27}$$

where the constant C does not depend on ε and k. Substituting (4.26) and (4.27) into the remaining term of (4.25) yields for $t\in[0,t_{\mathrm{end}}/\varepsilon]$

$$\begin{aligned}\varepsilon\int_0^t \big\|\partial_t f_\varepsilon(\sigma)\big\|_{L^1}\,\mathrm{d}\sigma &\leqslant t_{\mathrm{end}}\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]}\big\|\partial_t f_\varepsilon(\sigma)\big\|_{L^1}\\ &\leqslant Ct_{\mathrm{end}}\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]}\int_{\mathbb{R}^d}|k|_1\big|\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\overline{\mathcal{P}_\varepsilon}\widehat{u}_{-1}\big)(\sigma,k)\big|_2\,\mathrm{d}k\\ &\quad+Ct_{\mathrm{end}}\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]}\int_{\mathbb{R}^d}\big|\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\overline{\mathcal{P}_\varepsilon}\widehat{u}_{-1}\big)(\sigma,k)\big|_2\,\mathrm{d}k\\ &\leqslant CC_{\mathcal{T}}t_{\mathrm{end}}\Big(C_{u,1}^3+C_{u,1}^2\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]}\|\partial_t\mathcal{P}_\varepsilon\widehat{u}_1(\sigma)\|_{L^1}\Big).\end{aligned}$$

For the last estimate we use (3.28) and (3.34). Furthermore, according to Lemma 4.1.2 the term $\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]}\|\partial_t\mathcal{P}_\varepsilon\widehat{u}_1(\sigma)\|_{L^1}$ is uniformly bounded. Thus, it follows that the remaining term of (4.25) is bounded by a constant which does not depend on ε. Overall, this shows that (4.24) is uniformly bounded of the form (4.18) with $\widehat{C}=0$. Together with the other considerations, this proves the inequality (4.18) in Step 1, and completes the proof of Proposition 4.1.4. ∎

In the next section, we show that under higher regularity assumptions, the refined bound for $P^\perp z_1(t)$ also holds in a stronger norm.

4.2 Extension to a stronger norm

As introduced in Chapter 2 we denote by D_μ the Fourier multiplicator $(D_\mu\widehat{w})(k) = \mathrm{i}k_\mu\widehat{w}(k)$ for $\mu\in\{1,\dots d\}$. If u_1 is the classical solution of (3.16) with $j_{\max}=1$ and initial data (3.18), then by multiplying (3.50) with D_μ it follows that

$$D_\mu\widehat{u}_1(t,k) = S_{1,\varepsilon}^*(t,k)D_\mu z_1(t,k)$$

is the Fourier transform of $\partial_\mu u_1(t,x)$, because the scalar multiplicator D_μ commutes with the matrix $S_{1,\varepsilon}^*(t,k)$ for every $t\in\mathbb{R}$ and $k\in\mathbb{R}^d$.

Next, we need a similar result as Lemma 4.1.2. However, now we prove that $\partial_t D_\mu\mathcal{P}_\varepsilon\widehat{u}_1(t)$ can be bounded independently of ε on long time intervals.

Lemma 4.2.1. *Under the assumptions of Lemma 3.6.1 (iii) with $j_{max} = 1$, Assumptions 3.7.2 and 4.1.1, there is a constant C such that*

$$\sup_{t\in[0,t_{end}/\varepsilon]} \|\partial_t D_\mu \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C.$$

The constant C depends on the constant $C_{u,2}$ from (3.35) and thus on t_{end}, but not on ε.

Proof. The proof is similar to the proof of Lemma 4.1.2. Applying the operator D_μ to both sides of (4.7) yields

$$\frac{\mathrm{i}}{\varepsilon} D_\mu \mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k) = \frac{\mathrm{i}}{\varepsilon}\lambda_{11}(\varepsilon k) D_\mu \mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k).$$

With (4.8) and $|D_\mu| = |k_\mu| \leqslant |k|_1$, we obtain

$$\begin{aligned}\Big|\frac{\mathrm{i}}{\varepsilon} D_\mu \mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k)\Big|_2 &\leqslant C|k|_1 \left|D_\mu \mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)\right|_2 = C|k|_1\,|k_\mu|\left|\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)\right|_2 \\ &\leqslant C|k|_1^2\left|\mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)\right|_2. \end{aligned} \tag{4.28}$$

Next, we consider D_μ applied to the nonlinear part. By definition (3.49) with $\#J = 1$ we have

$$D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J) = S_{1,\varepsilon}(t) D_\mu \mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t), \tag{4.29}$$

since the multiplicator D_μ commutes with the matrix $S_{1,\varepsilon}(t)$. Furthermore, the definition (3.23) of $\mathcal{T}$ implies that

$$D_\mu \mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big) = \mathcal{T}\big(D_\mu\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big) + \mathcal{T}\big(\widehat{u}_{j_1},D_\mu\widehat{u}_{j_2},\widehat{u}_{j_3}\big) + \mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},D_\mu\widehat{u}_{j_3}\big), \tag{4.30}$$

which corresponds to the product rule. It follows from (4.29) and (4.30) that

$$\begin{aligned}\big\|D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)\big\|_{L^1} &= \big\|D_\mu \mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t)\big\|_{L^1} \\ &\leqslant \|\mathcal{T}\big(D_\mu\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t)\|_{L^1} + \|\mathcal{T}\big(\widehat{u}_{j_1},D_\mu\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t)\|_{L^1} + \|\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},D_\mu\widehat{u}_{j_3}\big)(t)\|_{L^1},\end{aligned}$$

since $S_{1,\varepsilon}(t)$ is unitary. Next, we apply the inequality (3.28) from Lemma 3.5.1 to every term and obtain for all $t \in [0, t_{\mathrm{end}}/\varepsilon]$ that

$$\begin{aligned}\big\|D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)\big\|_{L^1} &\leqslant \|\mathcal{T}\big(D_\mu\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t)\|_{L^1} + \|\mathcal{T}\big(\widehat{u}_{j_1},D_\mu\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(t)\|_{L^1} + \|\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},D_\mu\widehat{u}_{j_3}\big)(t)\|_{L^1} \\ &\leqslant C_{\mathcal{T}}\Big(\|D_\mu\widehat{u}_{j_1}(t)\|_{L^1}\|\widehat{u}_{j_2}(t)\|_{L^1}\|\widehat{u}_{j_3}(t)\|_{L^1} + \|\widehat{u}_{j_1}(t)\|_{L^1}\|D_\mu\widehat{u}_{j_2}(t)\|_{L^1}\|\widehat{u}_{j_3}(t)\|_{L^1} \\ &\qquad + \|\widehat{u}_{j_1}(t)\|_{L^1}\|\widehat{u}_{j_2}(t)\|_{L^1}\|D_\mu\widehat{u}_{j_3}(t)\|_{L^1}\Big) \\ &\leqslant C_{\mathcal{T}} C_{u,1}^3. \end{aligned} \tag{4.31}$$

Together with (3.24) and the inequality (4.28) this shows that $D_\mu \mathcal{P}_\varepsilon(k)\partial_t\widehat{u}_1(t,k) = \partial_t D_\mu \mathcal{P}_\varepsilon(k)\widehat{u}_1(t,k)$ is uniformly bounded. ∎

For the approximation error of the SVEA which we consider in Section 4.3 we need the following version of Proposition 4.1.4, where $z(t)$ is replaced by $D_\mu z(t)$. The reason is that in the proof of Theorem 4.3.4 we need the refined bound of $P^\perp z_1$ in a stronger norm.

Therefore, the goal is to prove that there is a $t_\star \in (0, t_{\text{end}}]$ and a constant C such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||D_\mu z(t)|||_\varepsilon \leqslant C \qquad \text{for all } \varepsilon \in (0,1].$$

This bound implies by definition of the scaled norm

$$\sup_{t\in[0,t_\star/\varepsilon]} \|D_\mu P^\perp z_1(t)\|_{L^1} \leqslant C\varepsilon, \tag{4.32}$$

for all $\varepsilon \in (0,1]$. With the relation (3.69) and since $S^*_{1,\varepsilon}(t)$ is unitary and commutes with D_μ, the bound (4.32) is equivalent to

$$\sup_{t\in[0,t_\star/\varepsilon]} \|D_\mu \mathcal{P}^\perp_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon, \tag{4.33}$$

for all $\varepsilon \in (0,1]$.

Proposition 4.2.2. *Let u_1 be the classical solution of (3.16) with $j_{max} = 1$ and initial data (3.18) for some $p \in W^2$. Let z_1 be the transformed variables defined in (3.46), and let $\mu \in \{1,\dots,d\}$. Under the assumptions of Proposition 4.1.4 there is a constant C such that*

$$\sup_{t\in[0,t_\star/\varepsilon]} |||D_\mu z(t)|||_\varepsilon \leqslant C \qquad \textit{for all } \varepsilon \in (0,1]$$

with $t_\star$ from Proposition 4.1.4. The constant C depends on $\|p\|_{W^2}$, $C_{u,2}$ from (3.35), and on r from Proposition 4.1.4, but not on ε.

Proof. Proposition 4.1.4 is crucial for the proof of the theorem because it yields the refined bound (4.10). Therefore, all of the assumptions of Theorem 4.2.2 serve the purpose that we can apply this proposition with the same initial data $p \in W^2 \subset W^1$ and the same $t_\star$.
Similarly as the proof of Proposition 4.1.4, this proof is subdivided into two steps. In Step 1 the general procedure is shown which differs from the proof of Proposition 4.1.4. The required estimates used in Step 1 are proven in Step 2.

Step 1. In the following let $\mu \in \{1,\dots,d\}$ be fixed. Applying the operator D_μ to both sides of (3.48) gives

$$\partial_t D_\mu z_1(t) = \varepsilon \sum_{\#J=1} D_\mu \mathbf{F}_\varepsilon(t, \widehat{u}, J)(t). \tag{4.34}$$

According to (4.34) and the definition of the scaled norm (4.12) we have

$$|||D_\mu z(t)|||_\varepsilon \leqslant |||D_\mu z(0)\,|||_\varepsilon + |||\int_0^t \partial_t D_\mu z(\sigma)\,\mathrm{d}\sigma|||_\varepsilon \leqslant |||D_\mu z(0)\,|||_\varepsilon + 2\sum_{\#J=1} f_1(t,\varepsilon,\widehat{u},J)$$

with

$$f_1(t,\varepsilon,\widehat{u},J) = \varepsilon \Big\| \int_0^t P D_\mu \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma \Big\|_{L^1} + \Big\| \int_0^t P^\perp D_\mu \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma \Big\|_{L^1}.$$

The term

$$|||D_\mu z(0)|||_\varepsilon = 2\|D_\mu P z_1(0)\|_{L^1} + \frac{2}{\varepsilon}\|D_\mu P^\perp z_1(0)\|_{L^1}$$

is uniformly bounded with a constant which depends on $\|p\|_{W^2}$. For the first term it follows with (3.60) that

$$\|D_\mu P z_1(0)\|_{L^1} \leqslant \|D_\mu z_1(0)\|_{L^1} = \|D_\mu \widehat{p}\|_{L^1} \leqslant \|p\|_{W^1} \leqslant C$$

by assumption on the initial data. Secondly, we have a similar estimate as (3.64), however, now we obtain

$$\|P^\perp D_\mu z_1(0)\|_{L^1} \leqslant C\varepsilon\|p\|_{W^2}.$$

The overall goal is to prove that there are constants C_1 and C_2 such that the inequality

$$f_1(t,\varepsilon,\widehat{u},J) \leqslant C_1 + C_2\,\varepsilon \int_0^t |||D_\mu z(\sigma)|||_\varepsilon \;\mathrm{d}\sigma \tag{4.35}$$

holds for all $J \in \mathcal{J}^3$ with $\#J = 1$. In order to show (4.35) we investigate the term $D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)$, which appears in $f_1(t,\varepsilon,\widehat{u},J)$. Together with (4.29) we observe that if we consider $\widehat{u}_1$ as given, then (4.30) is *linear* with respect to $D_\mu \widehat{u}_1 = S^*_{1,\varepsilon} D_\mu z_1$. In comparison to the proof of Proposition 4.1.4, this is the reason why we can use Gronwall's lemma to prove boundedness of $|||D_\mu z(t)|||_\varepsilon$.
If (4.35) is true, then it follows with the uniform bound of $|||D_\mu z(0)|||_\varepsilon$ that

$$|||D_\mu z(t)|||_\varepsilon \leqslant c_1 + c_2\,\varepsilon \int_0^t |||D_\mu z(\sigma)|||_\varepsilon \;\mathrm{d}\sigma,$$

where the constant c_1 includes C_1 and a constant which depends on $\|p\|_{W^2}$. Therefore, applying Gronwall's lemma (cf. Lemma A.1.1) yields

$$\sup_{t\in[0,t_\star/\varepsilon]} |||D_\mu z(t)|||_\varepsilon \leqslant c_1 e^{c_2 t_\star},$$

which proves the assertion.
The main part of the proof is to show (4.35). This is done in the next step.

Step 2. We analyze the terms $PD_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)$ and $P^\perp D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)$, which appear in $f_1(t,\varepsilon,\widehat{u},J)$ separately. Since both terms contain the expression $D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)$ we consider it in more detail. From the proof of Lemma 4.2.1 we have (4.30).
Similarly to the proof of Lemma 4.2.1 we obtain (4.31) for all $t \in [0, t_\star/\varepsilon]$. Therefore, the first term of $f_1(t,\varepsilon,\widehat{u},J)$ is bounded with (4.31) and the definition of the scaled norm (4.12) by

$$\varepsilon\Big\| \int_0^t PD_\mu \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} \leqslant \varepsilon \int_0^t \big\|D_\mu \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\big\|_{L^1}\,\mathrm{d}\sigma \leqslant t_\star \sup_{t\in[0,t_\star/\varepsilon]} \big\|D_\mu \mathbf{F}_\varepsilon(t,\widehat{u},J)\big\|_{L^1} \leqslant t_\star C_{\mathcal{T}} C_{u,2}^3,$$

which is a bound of the type (4.35) with $C_2 = 0$.
The main difficulty is to show the boundedness of the second part of $f_1(t,\varepsilon,\widehat{u},J)$. Fortunately, this

boundedness can be shown by adapting the procedure from Step 2 of the proof of Proposition 4.1.4. With (4.30) we obtain

$$\begin{aligned}\Big\| \int_0^t P^\perp D_\mu \mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)\, \mathrm{d}\sigma \Big\|_{L^1} \leqslant & \Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(D_\mu \widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \\ & + \Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_{j_1}, D_\mu \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \\ & + \Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, D_\mu \widehat{u}_{j_3}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1}.\end{aligned}$$

As an example we consider for $J = (1, 1, -1)$ the term

$$\Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(D_\mu \widehat{u}_1, \widehat{u}_1, \widehat{u}_{-1}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1}.$$

Similarly to (4.23) we split the nonlinearity into eight parts. For all terms containing $\mathcal{P}_\varepsilon^\perp \widehat{u}_1$ or $\overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}$, but excluding $D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1$ or $D_\mu \overline{\mathcal{P}_\varepsilon^\perp} \widehat{u}_{-1}$, we obtain a bound of the type (4.35) with $C_1 = 0$, since Proposition 4.1.4 implies (4.10), which provides the required factor ε. For terms containing $D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1$ we also have a bound of the type (4.35) with $C_1 = 0$. Here, the required factor ε is obtained by the definition of the scaled norm (4.12). More precisely, with (3.69) and since $S_{1,\varepsilon}^*(\sigma)$ is unitary, we estimate

$$\frac{1}{\varepsilon} \|D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1(\sigma)\|_{L^1} = \frac{1}{\varepsilon} \|D_\mu P^\perp z_1(\sigma)\|_{L^1} \leqslant |||D_\mu z(\sigma)|||_\varepsilon .$$

The only remaining term

$$\Big\| \int_0^t P^\perp S_{1,\varepsilon}(\sigma) \mathcal{T}\big(D_\mu \mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \tag{4.36}$$

has to be treated in a similar way as (4.24). For this purpose, we express the term (4.36) as

$$\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \widetilde{\Lambda}_1(0)\big) P^\perp f_\varepsilon(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1}, \qquad f_\varepsilon(t,k) = \exp\big(\tfrac{\mathrm{i}t}{\varepsilon} \Delta_1(\varepsilon k)\big) \Psi_1^*(\varepsilon k) \mathcal{T}\big(D_\mu \mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \overline{\mathcal{P}_\varepsilon} \widehat{u}_{-1}\big)(t,k),$$

where $\widetilde{\Lambda}_1(0)$ and $\Delta_1(\varepsilon k)$ are defined as in the proof of Proposition 4.1.4. We integrate by parts and bound with the same reasoning as in the proof of Proposition 4.1.4 for all $t \in [0, t_\star/\varepsilon]$

$$\|f_\varepsilon(t)\|_{L^1} \leqslant C,$$

where C depends on $C_\mathcal{T}$ and the constant $C_{u,2}$ from (3.35). Furthermore, we bound for $t \in [0, t_\star/\varepsilon]$ with the product rule

$$\begin{aligned}\varepsilon \int_0^t \big\|\partial_t f_\varepsilon(\sigma)\big\|_{L^1}\, \mathrm{d}\sigma \leqslant & \, C C_\mathcal{T} t_\star \Big(C_{u,2}^3 + C_{u,2}^2 \sup_{\sigma \in [0, t_\star/\varepsilon]} \|\partial_t D_\mu \mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\|_{L^1} \Big) \\ & + \varepsilon 2 C C_\mathcal{T} \Big(C_{u,2}^2 + C_{u,2} \sup_{\sigma \in [0, t_\star/\varepsilon]} \|\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\|_{L^1} \Big) \int_0^t \big\|D_\mu \mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\big\|_{L^1}\, \mathrm{d}\sigma.\end{aligned}$$

Additionally to Lemma 4.1.2, we have to use Lemma 4.2.1 which yields that

$$\sup_{s\in[0,t_\star/\varepsilon]} \|\partial_t D_\mu \mathcal{P}_\varepsilon \widehat{u}_1(\sigma)\|_{L^1} \leqslant C$$

with C independent of ε. This leads to (4.35) with constants C_1 and C_2 which depend on $C_{u,2}$.
In a last step we adapt the procedure for the other two multi-indices J with $\#J = 1$ which completes the proof. ■

In the next section we consider the error bound for the approximation of $\widetilde{\mathbf{u}}^{(j_{\max})}$ defined in (3.15) with $j_{\max} = 1$.

4.3 Error bound for the approximation

In this section we state the main result of this chapter. Instead of the error bound (3.13) (cf. [11, Theorem 1]), we show an accuracy of the SVEA of $\mathcal{O}(\varepsilon^2)$. Here, the previously shown results of Section 4.1 and Section 4.2 play an important role for the proof of the error bound. We suppose in the following that u_1 is the solution of (3.16) for $j_{\max} = 1$ with initial data (3.18). Furthermore, we assume that a unique mild solution of (1.4) exists on the long time interval $[0, t_\star/\varepsilon]$ with the constant $t_\star$ from Proposition 4.1.4, and that there exists a constant $C_{\mathbf{u}}$ uniformly in ε, such that

$$C_{\mathbf{u}} := \max\Big\{ \sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t)\|_W, \sup_{t\in[0,t_\star/\varepsilon]} \|\widetilde{\mathbf{u}}^{(1)}(t)\|_W \Big\}. \tag{4.37}$$

In consequence of the approach (3.15), $\widetilde{\mathbf{u}}^{(1)}$ provides an approximation to the exact solution $\mathbf{u}$ of the original problem (1.4). The goal is to prove an error bound for this approximation of $\mathcal{O}(\varepsilon^2)$. This error bound requires some preparatory work. We start with an additional non-resonance assumption on the eigenvalues.

Assumption 4.3.1 (Non-resonance condition)**.** *The matrices $\mathcal{L}_3(0) = \mathcal{L}(3\omega, 3\kappa)$ and $\mathcal{L}_1(0) = \mathcal{L}(\omega, \kappa)$ have no common eigenvalues, i.e. $\lambda_{3m}(0) \neq \lambda_{1m_1}(0)$ for all $m, m_1 = 1, \dots, s$.*

Remark 4.3.2. *Since we know explicitly the eigenvalues of the matrix $\mathcal{L}(0,\kappa)$ for the Klein–Gordon and the Maxwell–Lorentz system, see Example 3.2.3 and 3.2.4, Assumption 4.3.1 can be verified with (3.44) if the corresponding eigenvalue $\omega_1(\beta)$ from the dispersion relation (3.4) is not constant in β.*

Similarly to Section 4.1 one component of the error analysis is to apply integration by parts. We observe that for $f \in C^2([0, t_{\text{end}}/\varepsilon], W) \cap C^1([0, t_{\text{end}}/\varepsilon], W^1) \cap C([0, t_{\text{end}}/\varepsilon], W^2)$ and Λ invertible, we obtain (4.5). If we again apply integration by parts, we obtain

$$\begin{aligned}\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\Big) f(\sigma)\,\mathrm{d}\sigma &= \Lambda^{-1}\tfrac{\varepsilon}{\mathrm{i}}\Big[\exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\Big) f(\sigma)\Big]_{\sigma=0}^t + \Lambda^{-2}\varepsilon^2\Big[\exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\Big)\partial_t f(\sigma)\Big]_{\sigma=0}^t \\ &\quad - \Lambda^{-2}\varepsilon^2 \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\Big)\partial_t^2 f(\sigma)\,\mathrm{d}\sigma,\end{aligned}$$

such that with $|\Lambda^{-1}|_2 \leqslant C$ and $t \leqslant \frac{t_{\text{end}}}{\varepsilon}$

$$\Big| \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda\Big) f(\sigma)\, \mathrm{d}\sigma \Big|_2 \leqslant \varepsilon C\big[|f(t)|_2 + |f(0)|_2\big] + \varepsilon^2 C\big[|\partial_t f(t)|_2 + |\partial_t f(0)|_2\big] + \varepsilon t_{\text{end}} C \sup_{\sigma\in[0,t_{\text{end}}/\varepsilon]} |\partial_t^2 f(\sigma)|_2. \tag{4.38}$$

In order to have a bound of $\mathcal{O}(\varepsilon)$ of the left-hand side of (4.38), we need that $f(t)$, $\partial_t f(t)$ and $\partial_t^2 f(t)$ are bounded with the right order in ε for all $t \in [0, t_{\text{end}}/\varepsilon]$. For this purpose, we state and prove the following lemma.

Lemma 4.3.3. *Under the assumptions of Lemma 3.6.1 (iii), Assumptions 3.7.2 and 4.1.1, there is a constant C such that*

$$\sup_{t\in[0,t_{end}/\varepsilon]} \|\partial_t^2 \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C.$$

The constant C depends on the constant $C_{u,2}$ from (3.35) and thus also on t_{end}, but not on ε.

Proof. The proof is similar to the proof of Lemma 4.1.2. We observe that equally to (4.30), we obtain for the time derivative that

$$\partial_t \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(t) = \mathcal{T}\big(\partial_t \widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(t) + \mathcal{T}\big(\widehat{u}_{j_1}, \partial_t \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(t) + \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \partial_t \widehat{u}_{j_3}\big)(t). \tag{4.39}$$

Analogously, we have

$$\Big|\frac{\mathrm{i}}{\varepsilon}\partial_t \mathcal{P}_\varepsilon \mathcal{L}_1(\varepsilon k)\widehat{u}_1(t,k)\Big|_2 \leqslant C|k|_1 \big|\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(t,k)\big|_2.$$

Together with (3.24), $\mathcal{P}_\varepsilon \partial_t^2 \widehat{u}_1(t,k) = \partial_t^2 \mathcal{P}_\varepsilon \widehat{u}_1(t,k)$, (4.39) and Lemma 4.2.1 this shows

$$\|\partial_t^2 \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C \int_{\mathbb{R}^d} |k|_1 \big|\partial_t \mathcal{P}_\varepsilon \widehat{u}_1(t,k)\big|_2 \,\mathrm{d}k + 9\varepsilon C_{\mathcal{T}} \|\partial_t \widehat{u}_1(t)\|_{L^1} \|\widehat{u}_1(t)\|_{L^1}^2 \leqslant C,$$

where the constant depends on $C_{u,2}$ and we use the fact that $\|\partial_t \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon^{-1}$. ∎

As mentioned in Section 3.7 we also take advantage of the fact that the entries of z_j oscillate with a much smaller amplitude than the entries of $\widehat{u}_j$. Lemma 3.7.1 together with (3.48) and (4.11) yield

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\partial_t z_1(t)\|_{L^1} \leqslant \varepsilon \sup_{t\in[0,t_\star/\varepsilon]} \sum_{\#J=1} \|\mathbf{F}_\varepsilon(t,\widehat{u},J)(t)\|_{L^1} \leqslant C\varepsilon. \tag{4.40}$$

Proposition 4.1.4 and 4.2.2 are crucial for the proof of the following theorem. Therefore, part of the assumptions of Theorem 4.3.4 serve the purpose that we can apply those two propositions. The crucial point is that Proposition 4.1.4 and 4.2.2 yield the bounds (4.10), (4.11), (4.32), (4.33) in addition to (3.35). We observe that combining (4.10) and (4.33) yields in particular that

$$\|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} + \sum_{\mu=1}^d \|D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon. \tag{4.41}$$

This bound will be helpful in the proof of Theorem 4.3.4. Now we state the first main result of this thesis.

Theorem 4.3.4. *Let $p \in W^2$ and let $\mathbf{u}$ be the solution of (1.4). Let u_1 be the classical solution of (3.16) with $j_{max} = 1$ established in part (iii) of Lemma 3.6.1, and let $\widetilde{\mathbf{u}}^{(1)}$ be the approximation defined in (3.15) with $j_{max} = 1$. Under Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2, and 4.3.1 there is a constant such that*

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(1)}(t)\|_W \leqslant C\varepsilon^2, \tag{4.42}$$

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(1)}(t)\|_{L^\infty} \leqslant C\varepsilon^2. \tag{4.43}$$

Proof. We only have to show (4.42). The second bound (4.43) of this theorem is an immediate consequence of the embedding $W(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$. Since the proof of Theorem 4.3.4 is rather lengthy, we subdivide it into several steps. In the first two steps we derive an error equation and reduce the problem to the crucial term for the error bound. The remaining steps deal with the elaborate part of the proof. Here, the required estimates are proven.

Step 1. In the following we denote the error between the exact solution $\mathbf{u}$ and the approximation by $\delta = \mathbf{u} - \widetilde{\mathbf{u}}^{(1)}$. The first goal is to derive an evolution equation for the error δ and its Fourier transform which will be used to apply Gronwall's lemma in Step 2. The approximation $\widetilde{\mathbf{u}}^{(1)}$, given in (3.15) with $j_{\max} = 1$, solves the semilinear hyperbolic system (1.4) up to the residual

$$R(t,x) = \varepsilon T(\widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)})(t,x) - \Big(\partial_t \widetilde{\mathbf{u}}^{(1)}(t,x) + A(\partial)\widetilde{\mathbf{u}}^{(1)}(t,x) + \frac{1}{\varepsilon} E\widetilde{\mathbf{u}}^{(1)}(t,x)\Big). \tag{4.44}$$

In order to derive a more useful and compact expression for R, we consider the two parts of (4.44) separately. First, we note that substituting the approximation $\widetilde{\mathbf{u}}^{(1)}$ into the left-hand side of (1.4) yields

$$\begin{aligned}
&\partial_t \widetilde{\mathbf{u}}^{(1)}(t,x) + A(\partial)\widetilde{\mathbf{u}}^{(1)}(t,x) + \frac{1}{\varepsilon} E\widetilde{\mathbf{u}}^{(1)}(t,x) \\
&\quad = \sum_{j\in\mathcal{J}} e^{\mathrm{i}j(\kappa\cdot x - \omega t)/\varepsilon}\Big(\partial_t u_j(t,x) + \tfrac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, j\kappa)u_j(t,x) + A(\partial)u_j(t,x)\Big) \\
&\quad = \varepsilon \sum_{j\in\mathcal{J}} \sum_{\#J=j} e^{\mathrm{i}j(\kappa\cdot x - \omega t)/\varepsilon} T(u_{j_1}, u_{j_2}, u_{j_3})(t,x),
\end{aligned} \tag{4.45}$$

where for the last equality we use (3.16) for $|j| = 1$. In contrast to (4.45), substituting the approximation $\widetilde{\mathbf{u}}^{(1)}$ into the right-hand side of (1.4) yields

$$\begin{aligned}
\varepsilon T(\widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)})(t,x) &= \varepsilon \sum_{J\in\mathcal{J}^3} e^{\mathrm{i}\#J(\kappa\cdot x - \omega t)/\varepsilon} T(u_{j_1}, u_{j_2}, u_{j_3})(t,x) \\
&= \varepsilon \sum_{\substack{j \text{ odd} \\ |j|\leqslant 3}} \sum_{\#J=j} e^{\mathrm{i}j(\kappa\cdot x - \omega t)/\varepsilon} T(u_{j_1}, u_{j_2}, u_{j_3})(t,x).
\end{aligned} \tag{4.46}$$

The difference is that (4.46) includes summands with $|j| = 3$, whereas $j \in \mathcal{J}$ with $j_{\max} = 1$ implies that $|j| = 1$ in (4.45). These additional summands are exactly the same higher harmonics which are omitted in the SVEA approach. Hence, inserting both expressions (4.45) and (4.46) into (4.44) yields for the residual

$$R(t,x) = \varepsilon \sum_{|j|=3} \sum_{\#J=j} e^{\mathrm{i}j(\kappa\cdot x - \omega t)/\varepsilon} T(u_{j_1}, u_{j_2}, u_{j_3})(t,x),$$

which is a more compact expression. Next, we write (4.44) as

$$\partial_t \widetilde{\mathbf{u}}^{(1)}(t,x) = -A(\partial)\widetilde{\mathbf{u}}^{(1)}(t,x) - \frac{1}{\varepsilon} E\widetilde{\mathbf{u}}^{(1)}(t,x) + \varepsilon T(\widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)}, \widetilde{\mathbf{u}}^{(1)})(t,x) - R(t,x).$$

We subtract this equation from (1.4). This shows that the error $\delta = \mathbf{u} - \widetilde{\mathbf{u}}^{(1)}$ solves the evolution equation

$$\partial_t \delta = -A(\partial)\delta - \frac{1}{\varepsilon} E\delta + \varepsilon \Big[T(\mathbf{u},\mathbf{u},\mathbf{u}) - T(\widetilde{\mathbf{u}}^{(1)},\widetilde{\mathbf{u}}^{(1)},\widetilde{\mathbf{u}}^{(1)}) \Big] + R \tag{4.47}$$

with initial data $\delta(0) = 0$. In order to derive a bound for the error δ in $\|\cdot\|_W$, we apply the Fourier transform to (4.47). Thus, we obtain for $\widehat{\delta} = \mathcal{F}\delta$ the evolution equation

$$\partial_t \widehat{\delta}(t,k) = -\big(\mathrm{i}A(k) + \tfrac{1}{\varepsilon}E\big)\widehat{\delta}(t,k) + \varepsilon \mathcal{G}\big(\mathcal{F}\mathbf{u}, \mathcal{F}\widetilde{\mathbf{u}}^{(1)}\big)(t,k) + \widehat{R}(t,k)$$

with

$$\begin{aligned} \mathcal{G}\big(\mathcal{F}\mathbf{u}, \mathcal{F}\widetilde{\mathbf{u}}^{(1)}\big) &= \mathcal{T}(\mathcal{F}\mathbf{u}, \mathcal{F}\mathbf{u}, \mathcal{F}\mathbf{u}) - \mathcal{T}(\mathcal{F}\widetilde{\mathbf{u}}^{(1)}, \mathcal{F}\widetilde{\mathbf{u}}^{(1)}, \mathcal{F}\widetilde{\mathbf{u}}^{(1)}) \\ \text{and} \qquad \widehat{R}(t,k) &= \varepsilon \sum_{|j|=3} \sum_{\#J=j} \mathcal{F}\left(T(u_{j_1}, u_{j_2}, u_{j_3}) e^{\mathrm{i}j\kappa\cdot x/\varepsilon} \right)(t,k) e^{-\mathrm{i}j\omega t/\varepsilon} \\ &= \varepsilon \sum_{|j|=3} \sum_{\#J=j} \mathcal{T}(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3})(t, k - \tfrac{j\kappa}{\varepsilon}) e^{-\mathrm{i}j\omega t/\varepsilon}, \end{aligned} \tag{4.48}$$

where the definition of $\mathcal{T}$ is given by (3.23). For the Fourier transform of the initial data we have $\widehat{\delta}(0) = 0$.

Step 2. In this step we aim for the estimate

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{\delta}(t)\|_{L^1} \leqslant C\varepsilon^2$$

by means of Gronwall's lemma.

Expressing the Fourier transform of the error $\widehat{\delta}(t,k)$ by the variation-of-constants formula yields with $\widehat{\delta}(0,k) = 0$

$$\begin{aligned} \widehat{\delta}(t,k) = \varepsilon \int_0^t \exp\big((\sigma - t)\big(\mathrm{i}A(k) + \tfrac{1}{\varepsilon}E\big)\big) \mathcal{G}\big(\mathcal{F}\mathbf{u}(\sigma), \mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big)(k)\, \mathrm{d}\sigma \\ + \int_0^t \exp\big((\sigma - t)\big(\mathrm{i}A(k) + \tfrac{1}{\varepsilon}E\big)\big) \widehat{R}(\sigma,k)\, \mathrm{d}\sigma. \end{aligned} \tag{4.49}$$

Next, we consider both integrals separately.

Since the matrix $A(k)$ is symmetric for every k and the matrix E is skew-symmetric, we observe that

$$\big(\mathrm{i}A(k) + \tfrac{1}{\varepsilon}E\big)^* = -\mathrm{i}A^\top(k) + \tfrac{1}{\varepsilon}E^\top = -\big(\mathrm{i}A(k) + \tfrac{1}{\varepsilon}E\big).$$

This means that the matrix $\mathrm{i}A(k) + \frac{1}{\varepsilon}E$ is skew-hermitian for every k which implies that the matrix exponential $\exp\big(t\big(\mathrm{i}A(k) + \frac{1}{\varepsilon}E\big)\big)$ is unitary for every $t \in \mathbb{R}$.

Furthermore, due to (3.33) we estimate

$$\begin{aligned}\big\|\mathcal{G}\big(\mathcal{F}\mathbf{u}(\sigma),\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big)\big\|_{L^1} &= \big\|\mathcal{T}(\mathcal{F}\mathbf{u},\mathcal{F}\mathbf{u},\mathcal{F}\mathbf{u})-\mathcal{T}(\mathcal{F}\widetilde{\mathbf{u}}^{(1)},\mathcal{F}\widetilde{\mathbf{u}}^{(1)},\mathcal{F}\widetilde{\mathbf{u}}^{(1)})(\sigma)\big\|_{L^1}\\ &\leqslant C_{\mathcal{T}}\big\|\widehat{\delta}(\sigma)\big\|_{L^1}\Big(\big\|\mathcal{F}\mathbf{u}(\sigma)\big\|_{L^1}^2+\big\|\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big\|_{L^1}\big\|\mathcal{F}\mathbf{u}(\sigma)\big\|_{L^1}+\big\|\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big\|_{L^1}^2\Big)\\ &= C_{\mathcal{T}}\big\|\widehat{\delta}(\sigma)\big\|_{L^1}\Big(\big\|\mathbf{u}(\sigma)\big\|_{W}^2+\big\|\widetilde{\mathbf{u}}^{(1)}(\sigma)\big\|_{W}\big\|\mathbf{u}(\sigma)\big\|_{L^1}+\big\|\widetilde{\mathbf{u}}^{(1)}(\sigma)\big\|_{W}^2\Big)\\ &\leqslant 3C_{\mathbf{u}}^2C_{\mathcal{T}}\big\|\widehat{\delta}(\sigma)\big\|_{L^1},\end{aligned}$$

where $C_{\mathbf{u}}$ is the constant defined in (4.37). For this reason the first term on the right-hand side of (4.49) can be bounded in L^1 by

$$\begin{aligned}&\varepsilon\int_0^t\int_{\mathbb{R}^d}\big|\exp\big((\sigma-t)\big(\mathrm{i}A(k)+\tfrac{1}{\varepsilon}E\big)\big)\big|_2\,\big|\mathcal{G}\big(\mathcal{F}\mathbf{u}(\sigma),\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big)(k)\big|_2\,\mathrm{d}k\,\mathrm{d}\sigma\\ &\quad=\varepsilon\int_0^t\int_{\mathbb{R}^d}\big|\mathcal{G}\big(\mathcal{F}\mathbf{u}(\sigma),\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big)(k)\big|_2\,\mathrm{d}k\,\mathrm{d}\sigma\\ &\quad=\varepsilon\int_0^t\big\|\mathcal{G}\big(\mathcal{F}\mathbf{u}(\sigma),\mathcal{F}\widetilde{\mathbf{u}}^{(1)}(\sigma)\big)\big\|_{L^1}\,\mathrm{d}\sigma\\ &\quad\leqslant 3C_{\mathcal{T}}C_{\mathbf{u}}^2\,\varepsilon\int_0^t\|\widehat{\delta}(\sigma)\|_{L^1}\,\mathrm{d}\sigma.\end{aligned}\tag{4.50}$$

The laborious part of the proof is to show that the remaining term of (4.49) can be estimated in L^1 by

$$\sup_{t\in[0,t_\star/\varepsilon]}\Big\|\int_0^t\exp\big((\sigma-t)\big(\mathrm{i}A(\cdot)+\tfrac{1}{\varepsilon}E\big)\big)\widehat{R}(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}\leqslant C\varepsilon^2\tag{4.51}$$

with a constant C which does not depend on ε. If we are able to show the estimate (4.51), then in combination with (4.50) we have all the required bounds. Together with (4.49) it follows that

$$\|\widehat{\delta}(t)\|_{L^1}\leqslant CC_{\mathbf{u}}^2\,\varepsilon\int_0^t\|\widehat{\delta}(\sigma)\|_{L^1}\,\mathrm{d}\sigma+C\varepsilon^2,$$

and applying Gronwall's lemma yields the desired bound

$$\sup_{t\in[0,t_\star/\varepsilon]}\|\mathbf{u}(t)-\widetilde{\mathbf{u}}^{(1)}(t)\|_W=\sup_{t\in[0,t_\star/\varepsilon]}\|\widehat{\delta}(t)\|_{L^1}\leqslant C\varepsilon^2e^{\gamma t_\star}$$

with $\gamma=CC_{\mathbf{u}}^2$, which proves (4.42). For the rest of the proof we aim to show (4.51).

Step 3. The goal of this step is to reformulate the integral term of (4.51) in an appropriate way. For this purpose we introduce the change of variables $k'=k-\frac{j\kappa}{\varepsilon}$, $\varepsilon k=j\kappa+\varepsilon k'$. Together with the definitions

(3.3), (3.22) and the representation (4.48), we obtain

$$\begin{aligned}
&\int_0^t \exp\big((\sigma-t)\big(\mathrm{i}A(k)+\tfrac{1}{\varepsilon}E\big)\big)\widehat{R}(\sigma,k)\,\mathrm{d}\sigma \\
&\quad= \varepsilon\int_0^t \exp\big((\sigma-t)\big(\mathrm{i}A(k)+\tfrac{1}{\varepsilon}E\big)\big)\sum_{j\in\{\pm3\}}\sum_{\#J=j}\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3})(\sigma,k-\tfrac{j\kappa}{\varepsilon})e^{-\mathrm{i}j\omega\sigma/\varepsilon}\,\mathrm{d}\sigma \\
&\quad= \varepsilon\sum_{j\in\{\pm3\}}\sum_{\#J=j}\int_0^t \exp\big(\tfrac{\mathrm{i}}{\varepsilon}(\sigma-t)\big(-j\omega+A(\varepsilon k)-\mathrm{i}E\big)\big)\,e^{-\mathrm{i}j\omega t/\varepsilon}\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3})(\sigma,k-\tfrac{j\kappa}{\varepsilon})\,\mathrm{d}\sigma \\
&\quad= \varepsilon\sum_{j\in\{\pm3\}}\sum_{\#J=j}\int_0^t \exp\big(\tfrac{\mathrm{i}}{\varepsilon}(\sigma-t)\mathcal{L}(j\omega,\varepsilon k)\big)\,e^{-\mathrm{i}j\omega t/\varepsilon}\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3})(\sigma,k-\tfrac{j\kappa}{\varepsilon})\,\mathrm{d}\sigma \\
&\quad= \varepsilon e^{-\mathrm{i}j\omega t/\varepsilon}\sum_{j\in\{\pm3\}}\sum_{\#J=j}\int_0^t \exp\big(\tfrac{\mathrm{i}}{\varepsilon}(\sigma-t)\mathcal{L}_j(\varepsilon k')\big)\,\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3})(\sigma,k')\,\mathrm{d}\sigma.
\end{aligned}$$

For the sake of simplicity we omit in the following the dash and write again k instead of k'. By considering the L^1-norm later, we integrate over k and, thus the difference does not matter. By means of the definitions (3.45) and (3.49) we conclude

$$\begin{aligned}
\exp\big(\tfrac{\mathrm{i}}{\varepsilon}(\sigma-t)\mathcal{L}_j(\varepsilon k)\big)\,\mathcal{T}(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3})(\sigma,k) &= \exp\big(\tfrac{\mathrm{i}}{\varepsilon}(\sigma-t)\mathcal{L}_j(\varepsilon k)\big)\,S^*_{j,\varepsilon}(\sigma,k)\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)(k) \\
&= \exp\big(-\tfrac{\mathrm{i}t}{\varepsilon}\mathcal{L}_j(\varepsilon k)\big)\,\Psi_j(\varepsilon k)\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)(k) \\
&= S^*_{j,\varepsilon}(t,k)\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)(k).
\end{aligned}$$

With $|e^{-\mathrm{i}j\omega t/\varepsilon}|=1$ and since $S^*_{j,\varepsilon}(t)$ is unitary, this yields the bound

$$\begin{aligned}
\Big\|\int_0^t \exp\big((\sigma-t)\big(\mathrm{i}A(\cdot)+\tfrac{1}{\varepsilon}E\big)\big)\widehat{R}(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} &\leqslant \varepsilon\sum_{j\in\{\pm3\}}\sum_{\#J=j}\Big\|S^*_{j,\varepsilon}(t)\int_0^t\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} \\
&= \varepsilon\sum_{j\in\{\pm3\}}\sum_{\#J=j}\Big\|\int_0^t\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}. \qquad (4.52)
\end{aligned}$$

This representation for the term (4.51) is more favourable, as we will see in the next steps.

Step 4. We now have reduced the required estimate (4.51) to

$$\sum_{j\in\{\pm3\}}\sum_{\#J=j}\Big\|\int_0^t\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}\leqslant C\varepsilon. \qquad (4.53)$$

The next goal in this and the following steps is to prove (4.53). If we combine this estimate with (4.52), we obtain the desired bound (4.51). We note that there is an extra factor ε on the right-hand side of (4.52) which together with the factor ε from the estimate (4.53) gives the required $\mathcal{O}(\varepsilon^2)$.
Now, we consider multi-indices $J\in\mathcal{J}^3$ with $\#J=j$, where $j\in\{\pm3\}$. The situation $|J|_1=|j|=3$ appears only if J is $(1,1,1)$ or $J=(-1,-1,-1)$. Hence, for $j=3$, we have

$$\mathbf{F}_\varepsilon(t,\widehat{u},J)=S_{3,\varepsilon}(t)\mathcal{T}\big(\widehat{u}_1,\widehat{u}_1,\widehat{u}_1\big)(t).$$

In order to take advantage of (4.10), we decompose

$$\begin{aligned}\mathcal{T}\big(\widehat{u}_1,\widehat{u}_1,\widehat{u}_1\big) &= \mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)+\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big)+\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)\\ &\quad+\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)+\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big)+\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big)\\ &\quad+\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)+\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big),\end{aligned}$$

similarly to the proof of Proposition 4.1.4. Now, all nonlinear terms $\mathcal{T}$ involving at least two terms $\mathcal{P}_\varepsilon^\perp\widehat{u}_1$ can be treated in a straightforward way because of Proposition 4.1.4 and the implied estimate (4.10). We obtain for example

$$\begin{aligned}\Big\|\int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} &\leqslant C_{\mathcal{T}}\,\varepsilon^2\int_0^t\Big(\tfrac{1}{\varepsilon}\|\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}\,\tfrac{1}{\varepsilon}\|\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}\,\|\mathcal{P}_\varepsilon\widehat{u}_1(\sigma)\|_{L^1}\Big)\,\mathrm{d}\sigma\\ &\leqslant \varepsilon C_{\mathcal{T}}\,t_\star C_{u,2}^3.\end{aligned}$$

The remaining nonlinearities contain one or zero terms $\mathcal{P}_\varepsilon^\perp\widehat{u}_1$. Therefore, the main difficulty is to prove

$$\Big\|\int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}\leqslant C\varepsilon \tag{4.54}$$

and

$$\Big\|\int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}\leqslant C\varepsilon. \tag{4.55}$$

The other two possible combinations $\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)$ and $\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big)$ of (4.54) can be treated in the same way. For this reason we consider (4.54) as an example.

To prove the bounds (4.54) and (4.55), we use that (3.67) and (3.69) in addition to (3.46) and (3.45), yield the representations

$$\begin{aligned}\mathcal{P}_\varepsilon\widehat{u}_1(t,k) &= S_{1,\varepsilon}^*(t,k)Pz_1(t,k)=\Psi_1(\varepsilon k)\exp\big(-\tfrac{\mathrm{i}t}{\varepsilon}\Lambda_1(\varepsilon k)\big)Pz_1(t,k)\\ &=\exp\big(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{11}(\varepsilon k)\big)z_{11}(t,k)\psi_{11}(\varepsilon k)\end{aligned} \tag{4.56}$$

and

$$\begin{aligned}\mathcal{P}_\varepsilon^\perp\widehat{u}_1(t,k) &= S_{1,\varepsilon}^*(t,k)P^\perp z_1(t,k)=\Psi_1(\varepsilon k)\exp\big(-\tfrac{\mathrm{i}t}{\varepsilon}\Lambda_1(\varepsilon k)\big)P^\perp z_1(t,k)\\ &=\sum_{m_1=2}^{s}\exp\big(-\tfrac{\mathrm{i}t}{\varepsilon}\lambda_{1m_1}(\varepsilon k)\big)z_{1m_1}(t,k)\psi_{1m_1}(\varepsilon k).\end{aligned} \tag{4.57}$$

The next two steps involve proving the bound (4.54). In the last step of this proof we show (4.55) and combine all the bounds to finish the proof.

Step 5. Combining the representations (4.56) and (4.57) with (3.23) and using

$$\Psi_3^*(\varepsilon k)=\sum_{m=1}^{s}e_m\psi_{3m}^*(\varepsilon k),$$

where e_m denotes the m-th unit vector, yields

$$S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)=F_\varepsilon(\sigma,z). \tag{4.58}$$

Here F_ε and its m-th entry are defined by

$$\begin{aligned} F_\varepsilon(t,z) &= \Big(F_{\varepsilon,m}(t,z)\Big)_{m=1}^{s}, \\ F_{\varepsilon,m}(t,z)(k) &= \sum_{m_1=2}^{s} \int_{\#K=k} \exp\Big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(t,K) c_{mm_1}(\varepsilon,k,K)\,\mathrm{d}K, \end{aligned}$$

where we have $\mathbb{1} = (1,1,1)$, $M = (m_1,1,1)$ and the notation (cf. (3.52) in Section 3.7)

$$\begin{aligned} K &= \big(k^{(1)},k^{(2)},k^{(3)}\big), \\ \Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K) &= \lambda_{3m}(\varepsilon k) - \lambda_{1m_1}\big(\varepsilon k^{(1)}\big) - \lambda_{11}\big(\varepsilon k^{(2)}\big) - \lambda_{11}\big(\varepsilon k^{(3)}\big), \\ Z_{\mathbb{1}M}(t,K) &= z_{1m_1}\big(t,k^{(1)}\big) z_{11}\big(t,k^{(2)}\big) z_{11}\big(t,k^{(3)}\big), \\ c_{mm_1}(\varepsilon,k,K) &= \frac{1}{(2\pi)^d} \psi^*_{3m}(\varepsilon k) T\Big(\psi_{1m_1}\big(\varepsilon k^{(1)}\big), \psi_{11}\big(\varepsilon k^{(2)}\big), \psi_{11}\big(\varepsilon k^{(3)}\big)\Big). \end{aligned}$$

We note that by definition $\psi_{jm}(\varepsilon k)$ is the m-th column of the unitary matrix $\Psi_j(\varepsilon k)$. Hence, it follows with (2.5) and $C_{\mathcal{T}} = C_T\,(2\pi)^{-d}$ that

$$|c_{mm_1}(\varepsilon,k,K)| \leqslant C_{\mathcal{T}} \qquad \text{for all } \varepsilon,k,K.$$

Together with the representation (4.58) and the definition of the L^1-norm we bound the left-hand side of (4.54) by

$$\begin{aligned} &\Big\| \int_0^t S_{3,\varepsilon}(\sigma) \mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma \Big\|_{L^1} \\ &\quad = \int_{\mathbb{R}^d} \Big| \int_0^t F_\varepsilon(\sigma,z)(k)\,\mathrm{d}\sigma \Big|_2 \,\mathrm{d}k \leqslant \sum_{m=1}^{s} \int_{\mathbb{R}^d} \Big| \int_0^t F_{\varepsilon,m}(\sigma,z)(k)\,\mathrm{d}\sigma \Big|\,\mathrm{d}k \\ &\quad \leqslant \sum_{m=1}^{s}\sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \Big| \int_{\#K=k} \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(\sigma,K)\,\mathrm{d}\sigma\; c_{mm_1}(\varepsilon,k,K)\,\mathrm{d}K \Big|\,\mathrm{d}k \\ &\quad \leqslant \sum_{m=1}^{s}\sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} \Big| \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(\sigma,K)\,\mathrm{d}\sigma \Big|\, |c_{mm_1}(\varepsilon,k,K)|\,\mathrm{d}K\,\mathrm{d}k \\ &\quad \leqslant C_{\mathcal{T}} \sum_{m=1}^{s}\sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} \Big| \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(\sigma,K)\,\mathrm{d}\sigma \Big|\,\mathrm{d}K\,\mathrm{d}k. \end{aligned} \tag{4.59}$$

The crucial term which we have to bound by $\mathcal{O}(\varepsilon)$ is the highly oscillatory integral in (4.59) for all m, m_1, K and $k = \#K$. We have already discussed the challenge of this type of terms in Section 3.7.

The next sub-goal in this step is to prove that we can bound the highly oscillatory integral by

$$\begin{aligned} &\Big| \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(\sigma,K)\,\mathrm{d}\sigma \Big| \\ &\qquad \leqslant C\varepsilon\Big(|Z_{\mathbb{1}M}(t,K)| + \sum_{i=1}^{3} |k^{(i)}|_1 \int_0^t |Z_{\mathbb{1}M}(\sigma,K)|\,\mathrm{d}\sigma + \int_0^t |\partial_t Z_{\mathbb{1}M}(\sigma,K)|\,\mathrm{d}\sigma\Big) \end{aligned} \tag{4.60}$$

for all m, m_1, K and $k = \#K$. For the rest of this step m, m_1, k, K are considered to be *fixed*. For this reason we simplify notation by setting

$$\begin{aligned}\Delta\lambda(\varepsilon) &= \Delta\lambda_{3m\mathbb{1}M}(\varepsilon, k, K) = \lambda_{3m}(\varepsilon k) - \lambda_{1m_1}\big(\varepsilon k^{(1)}\big) - \lambda_{11}\big(\varepsilon k^{(2)}\big) - \lambda_{11}\big(\varepsilon k^{(3)}\big),\\ Z(\sigma) &= Z_{\mathbb{1}M}(\sigma, K) = z_{1m_1}\big(\sigma, k^{(1)}\big) z_{11}\big(\sigma, k^{(2)}\big) z_{11}\big(\sigma, k^{(3)}\big).\end{aligned}$$

As mentioned in Section 3.7 the idea is to expand the highly oscillatory term $\exp\Big(\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\Big)$ in a suitable way. We underline that by Assumption 4.3.1 and since $\lambda_{11}(0) = 0$, we know

$$\Delta\lambda(0) = \lambda_{3m}(0) - \lambda_{1m_1}(0) \neq 0.$$

Furthermore, we define

$$Y(\sigma) = \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\big[\Delta\lambda(\varepsilon) - \Delta\lambda(0)\big]\Big) Z(\sigma).$$

With this extension we are able to integrate by parts in order to generate one additional power of ε. Since $|\exp\Big(\frac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)| = 1$ we obtain for the left-hand side of (4.60)

$$\Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\Big) Z(\sigma)\,\mathrm{d}\sigma\Big| = \Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big) Y(\sigma)\,\mathrm{d}\sigma\Big| \leqslant C\varepsilon\Big(|Z(t)| + |Z(0)|\Big) + C\varepsilon\int_0^t |\partial_t Y(\sigma)|\,\mathrm{d}\sigma.$$

The initial conditions (3.25) imply that $z_{1m_1}(0, k) = 0$ for all $m_1 \neq 1$ and hence that $Z(0) = Z_{\mathbb{1}M}(0, K) = 0$. By means of the product rule we have

$$\partial_t Y(\sigma) = \tfrac{\mathrm{i}}{\varepsilon}\big[\Delta\lambda(\varepsilon) - \Delta\lambda(0)\big] Y(\sigma) + \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\big[\Delta\lambda(\varepsilon) - \Delta\lambda(0)\big]\Big)\partial_t Z(\sigma).$$

Hence, for the integral term, we obtain

$$\int_0^t |\partial_t Y(\sigma)|\,\mathrm{d}\sigma \leqslant \frac{1}{\varepsilon}\int_0^t \Big|\big[\Delta\lambda(\varepsilon) - \Delta\lambda(0)\big]\Big|\,|Z(\sigma)|\,\mathrm{d}\sigma + \int_0^t |\partial_t Z(\sigma)|\,\mathrm{d}\sigma. \tag{4.61}$$

Next, we take advantage of the Lipschitz continuity of the eigenvalues. For the difference

$$\begin{aligned}\Delta\lambda(\varepsilon) - \Delta\lambda(0) &= \Delta\lambda_{3m\mathbb{1}M}(\varepsilon) - \Delta\lambda_{3m\mathbb{1}M}(0)\\ &= \big(\lambda_{3m}(\varepsilon k) - \lambda_{3m}(0)\big) - \big(\lambda_{1m_1}\big(\varepsilon k^{(1)}\big) - \lambda_{1m_1}(0)\big)\\ &\quad - \big(\lambda_{11}\big(\varepsilon k^{(2)}\big) - \lambda_{11}(0)\big) - \big(\lambda_{11}\big(\varepsilon k^{(3)}\big) - \lambda_{11}(0)\big)\end{aligned}$$

the Assumption 3.7.2 and the identity $k = \#K = k^{(1)} + k^{(2)} + k^{(3)}$ yield the inequality

$$|\Delta\lambda(\varepsilon) - \Delta\lambda(0)| \leqslant C|\varepsilon k|_1 + C\sum_{i=1}^3 |\varepsilon k^{(i)}|_1 \leqslant 2C\varepsilon\sum_{i=1}^3 |k^{(i)}|_1.$$

In this way, we gain an additional factor ε which counterbalances the factor $1/\varepsilon$ in front of the first term on the right-hand side of (4.61). All in all, this yields the bound

$$\Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\Big) Z(\sigma)\,\mathrm{d}\sigma\Big| \leqslant C\varepsilon\Big(|Z(t)| + \sum_{i=1}^3 |k^{(i)}|_1 \int_0^t |Z(\sigma)|\,\mathrm{d}\sigma + \int_0^t |\partial_t Z(\sigma)|\,\mathrm{d}\sigma\Big),$$

where the constant C depends on the inverse of $\Delta\lambda(0)$ and on the Lipschitz constant in Assumption 3.7.2, but not on ε. This proves the estimate (4.60).

Step 6. In the last step of proving the bound (4.54) we make our way back to the beginning. We substitute (4.60) into (4.59) and obtain

$$\begin{aligned}
&\Big\| \int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \\
&\quad \leqslant C_{\mathcal{T}} \sum_{m=1}^{s} \sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} \Big| \int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon} \Delta\lambda_{3m\mathbb{1}M}(\varepsilon,k,K)\Big) Z_{\mathbb{1}M}(\sigma,K)\, \mathrm{d}\sigma \Big|\, \mathrm{d}K\, \mathrm{d}k \\
&\quad \leqslant \varepsilon C_{\mathcal{T}} s\Big(X_1(t) + X_2(t) + X_3(t)\Big).
\end{aligned}$$

The factor s results from the fact that every term $X_i(t)$, $i = 1,2,3$, is independent of m, where we use the short-hand notation

$$\begin{aligned}
X_1(t) &= \sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} |Z_{\mathbb{1}M}(t,K)|\, \mathrm{d}K\mathrm{d}k, \\
X_2(t) &= \sum_{i=1}^{3} \sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} |k^{(i)}|_1 \int_0^t |Z_{\mathbb{1}M}(\sigma,K)|\, \mathrm{d}\sigma\, \mathrm{d}K\mathrm{d}k, \\
X_3(t) &= \sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} \int_0^t |\partial_t Z_{\mathbb{1}M}(\sigma,K)|\, \mathrm{d}\sigma\, \mathrm{d}K\mathrm{d}k.
\end{aligned}$$

To conclude the bound (4.54), it has to be shown that $X_i(t) \leqslant C$ for $i = 1,2,3$ and for all $t \in [0, t_\star/\varepsilon]$ with a constant C which is independent of ε.

Before bounding each term separately, we provide the estimate

$$\begin{aligned}
\sum_{m_1=2}^{s} |Z_{\mathbb{1}M}(t,K)| &= \Big(\sum_{m_1=2}^{s} \big|z_{1m_1}\big(t,k^{(1)}\big)\big| \Big) \big|z_{11}\big(t,k^{(2)}\big)\big|\, \big|z_{11}\big(t,k^{(3)}\big)\big| \\
&= \big|P^\perp z_1\big(t,k^{(1)}\big)\big|_1\, \big|z_{11}\big(t,k^{(2)}\big)\big|\, \big|z_{11}\big(t,k^{(3)}\big)\big| \\
&\leqslant \sqrt{s-1}\big|P^\perp z_1\big(t,k^{(1)}\big)\big|_2\, \big|Pz_1\big(t,k^{(2)}\big)\big|_2\, \big|Pz_1\big(t,k^{(3)}\big)\big|_2 \\
&\leqslant \sqrt{s-1}\big|\mathcal{P}_\varepsilon^\perp \widehat{u}_1\big(t,k^{(1)}\big)\big|_2\, \big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\, \big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2,
\end{aligned}$$

which follows from the definition of $Z_{\mathbb{1}M}$, (2.1) and (3.47).
Now we consider $X_1(t)$. Together with the previous estimate, this leads to

$$\begin{aligned}
X_1(t) &= \sum_{m_1=2}^{s} \int_{\mathbb{R}^d} \int_{\#K=k} |Z_{\mathbb{1}M}(t,K)|\, \mathrm{d}K\mathrm{d}k \\
&\leqslant \sqrt{s-1} \int_{\mathbb{R}^d} \int_{\#K=k} \big|\mathcal{P}_\varepsilon^\perp \widehat{u}_1\big(t,k^{(1)}\big)\big|_2\, \big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\, \big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2\, \mathrm{d}K\mathrm{d}k \\
&= \sqrt{s-1}\big\|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\big\|_{L^1}\, \|\widehat{u}_1(t)\|_{L^1}^2 \leqslant C.
\end{aligned}$$

We remark that in fact, it even follows that $X_1(t) \leqslant C\varepsilon$ according to (4.10). Next, in order to bound

$X_2(t)$ we write

$$\begin{aligned}
&\sum_{i=1}^{3}\sum_{m_1=2}^{s}|k^{(i)}|_1|Z_{\mathbb{1}M}(t,K)| \\
&\quad=\sqrt{s-1}\Big(|k^{(1)}|_1\,\big|\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big(t,k^{(1)}\big)\big|_2\,\big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\,\big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2 \\
&\qquad+\big|\mathcal{P}_\varepsilon^\perp\widehat{u}_1\big(t,k^{(1)}\big)\big|_2\Big[|k^{(2)}|_1\,\big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\,\big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2+\big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\,|k^{(3)}|_1\,\big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2\Big]\Big)
\end{aligned}$$

and by the definition of the Wiener algebra

$$\int_{\mathbb{R}^d}|k^{(i)}|_1\,\big|\widehat{u}_{j_i}\big(t,k^{(i)}\big)\big|_2\,\mathrm{d}k^{(i)}\leqslant\|\widehat{u}_{j_i}(t)\|_{L^1}+\int_{\mathbb{R}^d}|k^{(i)}|_1\,\big|\widehat{u}_{j_i}\big(t,k^{(i)}\big)\big|_2\,\mathrm{d}k^{(i)}=\|u_{j_i}(t)\|_{W^1}$$

for $i\in\{1,2,3\}$ and $J=(j_1,j_2,j_3)=(1,1,1)$. In a similar way as before it follows with these bounds that

$$\begin{aligned}
X_2(t)&=\sum_{i=1}^{3}\sum_{m_1=2}^{s}\int_{\mathbb{R}^d}\int_{\#K=k}|k^{(i)}|_1\int_0^t|Z_{\mathbb{1}M}(\sigma,K)|\,\mathrm{d}\sigma\,\mathrm{d}K\mathrm{d}k\\
&\leqslant C\int_0^t\left(\|\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}+\sum_{\mu=1}^{d}\|D_\mu\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}\right)\left(\|\widehat{u}_1(\sigma)\|_{L^1}+\sum_{\mu=1}^{d}\|D_\mu\widehat{u}_1(\sigma)\|_{L^1}\right)^2\mathrm{d}\sigma\\
&\leqslant Ct_\star\sup_{\sigma\in[0,t_\star/\varepsilon]}\left[\varepsilon^{-1}\left(\|\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}+\sum_{\mu=1}^{d}\|D_\mu\mathcal{P}_\varepsilon^\perp\widehat{u}_1(\sigma)\|_{L^1}\right)\|u_1(\sigma)\|_{W^1}^2\right]\\
&\leqslant C
\end{aligned}$$

due to (4.41) and (3.35). Finally, we consider $X_3(t)$. Again due to the definition of $Z_{\mathbb{1}M}$ and (2.1) it follows with the product rule that

$$\begin{aligned}
\sum_{m_1=2}^{s}|\partial_tZ_{\mathbb{1}M}(t,K)|\leqslant\sqrt{s-1}\Big(&\big|\partial_tP^\perp z_1\big(t,k^{(1)}\big)\big|_2\,\big|Pz_1\big(t,k^{(2)}\big)\big|_2\,\big|Pz_1\big(t,k^{(3)}\big)\big|_2\\
&+\big|P^\perp z_1\big(t,k^{(1)}\big)\big|_2\Big[\big|\partial_tPz_1\big(t,k^{(2)}\big)\big|_2\,\big|Pz_1\big(t,k^{(3)}\big)\big|_2+\big|Pz_1\big(t,k^{(2)}\big)\big|_2\,\big|\partial_tPz_1\big(t,k^{(3)}\big)\big|_2\Big]\Big).
\end{aligned}$$

Together with (4.11) and (4.40) we obtain

$$\begin{aligned}
X_3(t)&=\sum_{m=1}^{n}\int_{\mathbb{R}^d}\int_{\#K=k}\int_0^t|\partial_tZ_{\mathbb{1}M}(\sigma,K)|\,\mathrm{d}\sigma\;\;\mathrm{d}K\;\mathrm{d}k\\
&\leqslant C\int_0^t\Big(\|\partial_tP^\perp z_1(\sigma)\|_{L^1}\|z_1(\sigma)\|_{L^1}^2+2\|P^\perp z_1(\sigma)\|_{L^1}\|z_1(\sigma)\|_{L^1}\|\partial_tz_1(\sigma)\|_{L^1}\Big)\,\mathrm{d}\sigma\\
&\leqslant C\tfrac{t_\star}{\varepsilon}(\varepsilon+2\varepsilon^2)\leqslant C.
\end{aligned}$$

Thus, all terms $X_i(t)$ for $i=1,2,3$ and $t\in[0,t_\star/\varepsilon]$ are uniformly bounded. This implies the bound (4.54).

Step 7. The main goal in this step is to prove (4.55) uniformly in ε in spite of the integration over a possibly long time interval which finally leads to (4.51). The technique differs from the proof of (4.54), since we do not use the notation (3.52). The ideas are similar to those in the proof of Proposition 4.1.4

Step 2, where we bound (4.24). However, there is one crucial difference. We have to apply integration by parts twice to obtain the required factors of ε. Again the idea is to expand the highly oscillatory part of the integral

$$\int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma$$

in a suitable way. Since every term $\mathcal{P}_\varepsilon\widehat{u}_1(\sigma)$ is essentially non-oscillatory, the whole nonlinearity $\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)$ is non-oscillatory on the time intervals of length $\mathcal{O}\big(\varepsilon^{-1}\big)$. Thus, the term $S_{3,\varepsilon}(\sigma,k)$ is the only highly oscillatory part in (4.55). In the following we define $\Delta_3(\varepsilon k)=\Lambda_3(\varepsilon k)-\Lambda_3(0)$ and by definition (3.45) we have

$$S_{3,\varepsilon}(\sigma,k)=\exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)\exp\big(-\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)S_{3,\varepsilon}(\sigma,k)=\exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)\exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k).$$

Hence, the left-hand side of (4.54) can be expressed as

$$\Big\|\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)f_\varepsilon(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1},\qquad f_\varepsilon(t,k)=\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k).$$

In order to gain a factor ε we integrate by parts. This is possible since the matrix $\mathcal{L}_3(0)=\mathcal{L}(3\omega,3\kappa)$ is invertible by Assumption 3.2.6. Thus, we obtain

$$\Big\|\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)f_\varepsilon(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}\leqslant\varepsilon C\Big(\|f_\varepsilon(0)\|_{L^1}+\|f_\varepsilon(t)\|_{L^1}\Big)+\varepsilon C\Big\|\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)\partial_t f_\varepsilon(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}.$$

We note that the constant depends on the inverse of the eigenvalues of $\Lambda_3(0)$. Next, we consider each term separately. With $|S_{3,\varepsilon}(\sigma,k)|_2=1$, (3.28), and (3.34) it follows that for all $t\in[0,t_\star/\varepsilon]$

$$\|f_\varepsilon(t)\|_{L^1}=\big\|\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(t)\big\|_{L^1}\leqslant C_{\mathcal{T}}\|\mathcal{P}_\varepsilon\widehat{u}_1(t)\|_{L^1}^3\leqslant C \tag{4.62}$$

with a constant which depends on $C_{\mathcal{T}}$ and the constant $C_{u,2}$ from (3.35).
Furthermore, we write

$$\begin{aligned}\partial_t f_\varepsilon(t)&=\partial_t\Big(\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k)\Big)\\&=\tfrac{\mathrm{i}}{\varepsilon}\Delta_3(\varepsilon k)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k)\\&\quad+\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k)\\&=g_1(t)+g_2(t).\end{aligned}$$

At this point it is crucial to not take the norm under the integral and to bound $\partial_t f_\varepsilon$ straightforwardly. The reason is that we have

$$\|g_1(t)\|_{L^1}=\mathcal{O}(1)\quad\text{and}\quad\|g_2(t)\|_{L^1}=\mathcal{O}(1).$$

These statements will be shown later. The long time interval counterbalances the factor ε which we gained by integration by parts and the factor ε is missing on the right-hand side of (4.55). Therefore, the idea is to apply again integration by parts on each integral term of g_1 and g_2. Thus, we consider

$$\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)\partial_t f_\varepsilon(\sigma)\,\mathrm{d}\sigma=\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)g_1(\sigma)\,\mathrm{d}\sigma+\int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)g_2(\sigma)\,\mathrm{d}\sigma.$$

In a first step we treat these integral terms in general. Since the matrix $\mathcal{L}_3(0) = \mathcal{L}(3\omega, 3\kappa)$ is invertible by Assumption 3.2.6, we again integrate by parts and obtain for $i = 1, 2$

$$\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big) g_i(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \leqslant C\varepsilon\Big(\|g_i(0)\|_{L^1} + \|g_i(t)\|_{L^1}\Big) + C\varepsilon \Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big)\partial_t g_i(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1}.$$

Recall that Λ_3 is globally Lipschitz continuous by Assumption 3.7.2, which means

$$|\tfrac{\mathrm{i}}{\varepsilon}\Delta_3(\varepsilon k)|_2 = \tfrac{1}{\varepsilon}|\Lambda_3(\varepsilon k) - \Lambda_3(0)|_2 \leqslant C|k|_1$$

with a constant C which does not depend on ε and k. With $|S_{3,\varepsilon}(\sigma, k)|_2 = 1$, (3.28), and (3.34) it follows for $t \in [0, t_{\mathrm{end}}/\varepsilon]$ that

$$\begin{aligned}\|g_1(t)\|_{L^1} &= \varepsilon^{-1}\big\|\Delta_3(\varepsilon\cdot)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon\cdot)\big)\Psi_3^*(\varepsilon\cdot)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t)\big\|_{L^1} \\ &\leqslant C\int_{\mathbb{R}^d} |k|_1 \big|\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k)\big|_2\, \mathrm{d}k \leqslant CC_{\mathcal{T}}C_{u,1}^3\end{aligned}$$

and with the product rule (4.39) that

$$\begin{aligned}\|g_2(t)\|_{L^1} &= \big\|\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon\cdot)\big)\Psi_3^*(\varepsilon\cdot)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t)\big\|_{L^1} = \big\|\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t)\big\|_{L^1} \\ &\leqslant 3C_{\mathcal{T}}\big\|\partial_t\mathcal{P}_\varepsilon\widehat{u}_1(t)\big\|_{L^1}\|\mathcal{P}_\varepsilon\widehat{u}_1\|_{L^1}^2 \leqslant C,\end{aligned}$$

where the constant depends on $C_{u,1}$ because of Lemma 4.1.2. Furthermore, we write

$$\begin{aligned}\partial_t g_1(t) &= \partial_t\big(\tfrac{\mathrm{i}}{\varepsilon}\Delta_3(\varepsilon k)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k)\big) \\ &= -\tfrac{1}{\varepsilon^2}\Delta_3^2(\varepsilon k)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k) \\ &\quad + \tfrac{\mathrm{i}}{\varepsilon}\Delta_3(\varepsilon k)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t,k) \\ &= h_1(t) + h_0(t)\end{aligned}$$

and

$$\begin{aligned}\partial_t g_2(t) &= \partial_t\big(\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon\cdot)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t)\big) \\ &= \tfrac{\mathrm{i}}{\varepsilon}\Delta_3(\varepsilon k)\exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t) \\ &\quad + \exp\big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta_3(\varepsilon k)\big)\Psi_3^*(\varepsilon k)\partial_t^2\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(t) \\ &= h_0(t) + h_2(t).\end{aligned}$$

If we are able to show that

$$\|h_0(t)\|_{L^1} + \|h_1(t)\|_{L^1} + \|h_2(t)\|_{L^1} = \mathcal{O}(1),$$

then the integral terms can be estimated in a straightforward way. We note that by

$$\varepsilon\Big\| \int_0^t \exp\big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Lambda_3(0)\big) h_i(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} \leqslant t_{\mathrm{end}} \sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \big\|h_i(\sigma)\big\|_{L^1},$$

it remains to bound $\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \big\|h_i(\sigma)\big\|_{L^1}$ uniformly for $i = 0, 1, 2$. The long time interval counterbalances the factor ε which we gained by the second integration by parts. However, the factor ε which we

gained by the first integration by parts provides the required factor ε on the right-hand side of (4.55). Again with the Lipschitz continuity of Λ_3, $|S_{3,\varepsilon}(\sigma,k)|_2 = 1$, (3.28), and (3.35) it follows that

$$\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \|h_1(\sigma)\|_{L^1} \leqslant \sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \int_{\mathbb{R}^d} |k|_1^2 \left|\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma,k)\right|_2 \mathrm{d}k \leqslant CC_{\mathcal{T}}C_{u,2}^3$$

and

$$\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \|h_0(\sigma)\|_{L^1} \leqslant \sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \int_{\mathbb{R}^d} |k|_1 \left|\partial_t\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma,k)\right|_2 \mathrm{d}k \leqslant CC_{\mathcal{T}}C_{u,2}^3,$$

where again we use Lemma 4.1.2 and Lemma 4.2.1. For the remaining term we estimate

$$\sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \|h_2(\sigma)\|_{L^1} \leqslant C \sup_{\sigma\in[0,t_{\mathrm{end}}/\varepsilon]} \int_{\mathbb{R}^d} \left|\partial_t^2\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma,k)\right|_2 \mathrm{d}k \leqslant CC_{\mathcal{T}}C_{u,2}^3,$$

where we use Lemma 4.3.3.

All in all this proves the inequality (4.55). Combining all the results implies (4.51) and completes the proof of Theorem 4.3.4. ∎

Remark 4.3.5. *The natural question is why the accuracy of the SVEA cannot be even better than $\mathcal{O}(\varepsilon^2)$ for the SVEA. The limiting term in the analysis is (4.62), which is of order $\mathcal{O}(1)$. There is no possibilty to gain a factor ε for this term and we cannot improve the estimate (4.55). Thus, we cannot be better than*

$$\left\| \int_0^t S_{3,\varepsilon}(\sigma)\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma \right\|_{L^1} \leqslant C\varepsilon.$$

Our theoretical considerations are confirmed by the following numerical example.

Numerical experiment

We conclude this section with a numerical experiment to illustrate Theorem 4.3.4. In the following we consider the one-dimensional Klein–Gordon system (1.3) with $\kappa = 1.2$, $\upsilon = 0.7$, $\mathcal{M} = E$, $\omega = \max\{\omega_1(\kappa), \omega_2(\kappa)\}$, where ω_m is the m-th eigenvalue of $\mathcal{L}(0,\kappa)$, and with initial data $p(x) = \psi_{11}(0)\exp\big(-(x-0.5)^2\big)$. We set $t_{\mathrm{end}} = 1$ and consider 2^{14} equidistant grid points in the interval $[-64, 64]$ with periodic boundary conditions. The reference solution is computed by the approximation (3.15) with $j_{\max} = 5$. At the moment, it is not clear why this choice is a reasonable reference solution. The justification why we can use $\widetilde{\mathbf{u}}^{(5)}$ as the reference solution in this numerical experiment is given in the following section in Remark 4.4.1. The code to reproduce the plots in this and the next section is available on https://www.doi.org/10.5445/IR/1000149721.

Figure 4.1 shows the accuracy of the SVEA compared to the reference solution considered with different values of ε. We observe that the accuracy improves quadratically in accordance with Theorem 4.3.4. The solutions of (3.16) with $j_{\max} = 1$ and $j_{\max} = 5$ are approximated by the Strang splitting method with $N = 10^5$ time-steps. We remark that the number of time-steps is chosen large enough in comparison to the choice of ε, which we consider, so that $\varepsilon^2 N > 1$ holds. The dashed red line is a reference line for order two.

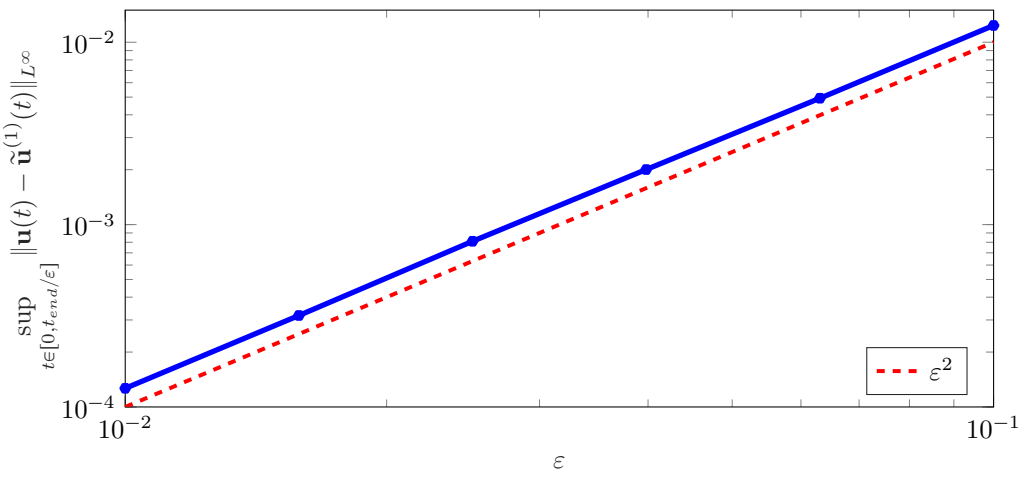


Figure 4.1: Accuracy of the SVEA for different values of ε in blue. The dashed red line is a reference line for order two.

Strang splitting. For the Strang splitting method we split the PDE (3.16) with $j_{\max} \in \{1, 5\}$ into the linear and the nonlinear part. Thus, we have the linear subproblem

$$\partial_t u_j^\bullet(t) = -\mathcal{A}_j u_j^\bullet(t) \qquad \text{with given } u_j^\bullet(0), \qquad \text{for } j \in \mathcal{J}_+, \tag{4.63}$$

where $\mathcal{A}_j$ is defined in (3.36) and the nonlinear subproblem

$$\partial_t u_j^{\bullet\bullet} = \varepsilon \sum_{\#J=j} T(u_{j_1}^{\bullet\bullet}, u_{j_2}^{\bullet\bullet}, u_{j_3}^{\bullet\bullet}) \qquad \text{with given } u_j^{\bullet\bullet}(0) \qquad \text{for } j \in \mathcal{J}_+. \tag{4.64}$$

The operator $\mathcal{A}_j$ generates a strongly continuous group on $W(\mathbb{R}^d)$. Thus, for $t \geqslant 0$ we obtain a solution of the linear subproblem via

$$u_j^\bullet(t) = e^{-t\mathcal{A}_j} u_j^\bullet(0) \tag{4.65}$$

and, hence, (4.63) can be solved exactly in Fourier space. Since we cannot solve the nonlinear subproblem (4.64) exactly, we approximate the solution with Heun's method which is a Runge-Kutta method of order two.
In total we obtain for $j \in \mathcal{J}_+$ an approximation $u_j^n \approx u_j(t_n)$ recursively, meaning by solving the subproblems (4.63) and (4.64) in alternating fashion. In order to calculate u_j^{n+1} we first approximate the solution of (4.64) via Heun's method with one half time-step $\frac{\tau}{2}$ and initial data u_j^n which yields $u_j^{n+1,-}$. Next, we compute $u_j^{n+1,+}$ by taking a full time-step τ of the exact solution (4.65), where now $u_j^{n+1,-}$ is the initial data. Finally, we approximate the solution of (4.64) via Heun's method again with one half time-step and initial data $u_j^{n+1,+}$ which yields the approximation u_j^{n+1}.

After proving this improved error bound for the SVEA and observing the accuracy numerically, we discuss an extension for higher accuracies in the last section of this chapter. The natural question that arises is what happens for higher $j_{\max} > 1$?

4.4 Extension to approximations with higher accuracy

We start this section by considering the same one-dimensional setting for the KG system (1.3) as above numerically. However, now we set $j_{\max} = 3$ in the ansatz (3.15).

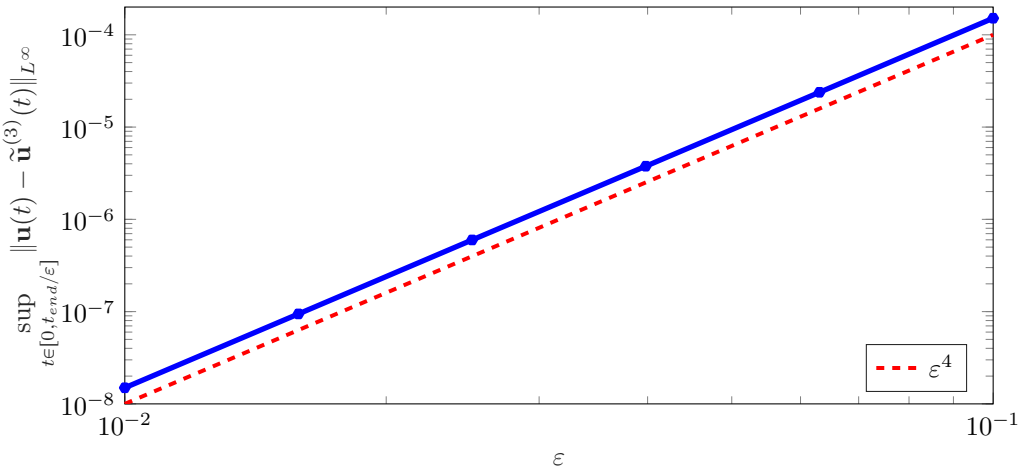


Figure 4.2: Accuracy of $\widetilde{\mathbf{u}}^{(3)}$ for different values of ε in blue. The dashed red line is a reference line for order four.

Figure 4.2 shows the accuracy of $\widetilde{\mathbf{u}}^{(3)}$ compared to the reference solution considered with different values of ε. The dashed red line is a reference line for order four. As before, we take as a reference solution $\widetilde{\mathbf{u}}^{(5)}$. For justification see Remark 4.4.1 below. The solutions of (3.16) with $j_{\max} = 3$ and $j_{\max} = 5$ are approximated by the Strang splitting method with $N = 10^5$ time steps. We observe that the accuracy improves quartically. Therefore, we conjecture that if we include more coefficients in the ansatz (3.15), we will also obtain better accuracy for the associated approximation.
Next, we set $j_{\max} = 5$ in (3.15) and consider the corresponding coefficients u_j for $j \in \mathcal{J}_+$.

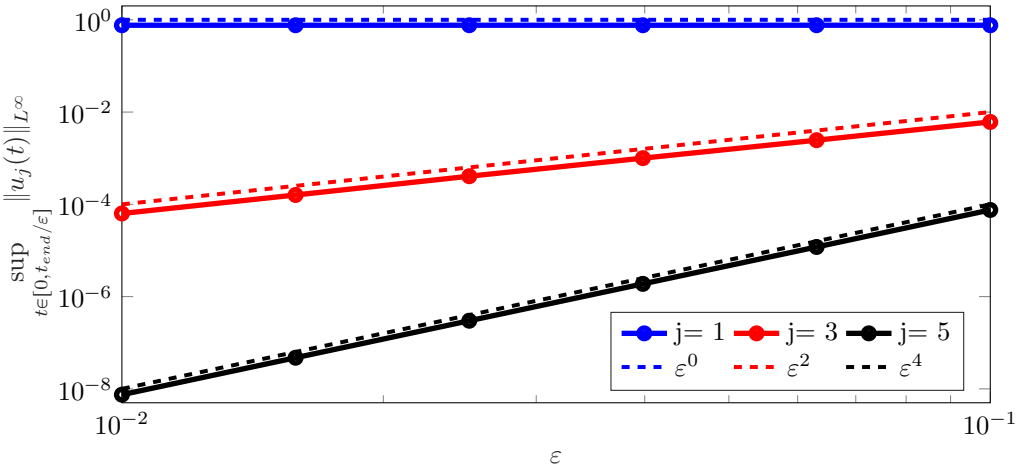


Figure 4.3: Illustration that the coefficients $u_3(t)$ and $u_5(t)$ remain small with respect to ε on long time intervals in red and black. The dashed red line is a reference line for order two, whereas the dashed black line is a reference line for order four. The coefficient $u_1(t)$ and a reference line for order zero is given in blue and dashed blue, respectively.

Figure 4.3 illustrates an important property of the coefficients $u_3(t)$ and $u_5(t)$ for different values of ε on long time intervals. We observe that the coefficients $u_j(t)$ for $j > 1$ stay small on long time intervals. We suspect that this behavior of the coefficients is crucial in order to prove error bounds for $\widetilde{\mathbf{u}}^{(j_{\max})}$ with $j_{\max} > 1$. Based on the numerical experiments, the following question arises.

Question. Under which conditions does a constant $C > 0$ exist such that

$$\sup_{t\in[0,t_\star/\varepsilon]} \|u_j(t)\|_W \leqslant C\varepsilon^{|j|-1}, \qquad \text{for } j \in \mathcal{J}, \tag{4.66}$$

and

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(j_{\max})}(t)\|_W \leqslant C\varepsilon^{j_{\max}+1} \tag{4.67}$$

hold?

General procedure. In order to prove (4.66) for $j_{\max} > 1$ we define the scaled norm

$$|||y|||_\varepsilon = 2\|Py_1\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp y_1\|_{L^1} + \sum_{j\in\mathcal{J}_+\setminus\{1\}} \frac{2}{\varepsilon^{j-1}}\|y_j\|_{L^1} \tag{4.68}$$

for all $y = (y_1, \dots, y_{j_{\max}})$ with $y_j \in L^1(\mathbb{R}^d, \mathbb{C}^n)$. For comparison the definitions (4.68) and (4.12) are equal for $j_{\max} = 1$. As before, we set $y_{-j} = \overline{y_j}$. Analogously to Section 4.1, the first goal is to prove (4.66) or equivalently that there is a constant C such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_\varepsilon \leqslant C,$$

for all $\varepsilon \in (0, 1]$. To this end, we investigate every term on the right-hand side of

$$\begin{aligned}
|||z(t)|||_\varepsilon &\leqslant |||z(0)\,|||_\varepsilon + |||\int_0^t \partial_t z(\sigma)\,\mathrm{d}\sigma|||_\varepsilon \\
&\leqslant |||z(0)\,|||_\varepsilon + 2\sum_{\#J=1}\left(\varepsilon\Big\|\int_0^t P\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} + \Big\|\int_0^t P^\perp\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}\right) \\
&\quad + 2\sum_{j\in\mathcal{J}_+\setminus\{1\}}\sum_{\#J=j}\frac{1}{\varepsilon^{j-2}}\Big\|\int_0^t \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}.
\end{aligned}$$

By definition of $\mathbf{F}_\varepsilon$ we have to estimate terms of the form

$$\Big\|\int_0^t S_{j,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1}$$

for $1 < j \leqslant j_{\max}$. At this point we note that in the proof of the approximation error in Section 4.3 we have already handled terms of this form, see for example (4.54) and (4.55). However, for the approximation error we consider those terms with $j > j_{\max}$. We express the procedure that we have seen for the case $j_{\max} = 1$ in Section 4.1-4.3 in generalized terms:
The goal is to establish

$$\Big\|\int_0^t S_{j,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} \leqslant \begin{cases} C_\star\varepsilon^{j-2} + \widehat{C}\varepsilon^{j-1}\int_0^t \prod_{i=1}^3 a_{j_i}(\sigma)\,\mathrm{d}\sigma & \text{for } 1 < j \leqslant j_{\max}, \\ C\varepsilon^{j-2} & \text{for } j > j_{\max} \end{cases} \tag{4.69}$$

for all $t \in [0, t_\star/\varepsilon]$ and for every $J = (j_1, j_2, j_3) \in \mathcal{J}^3$ with $\#J = j$. In contrast to definition (4.16), we now define

$$a_j(t) = \begin{cases} \|Pz_1(t)\|_{L^1} + \varepsilon^{-1}\|P^\perp z_1(t)\|_{L^1} & \text{if } j = \pm 1, \\ \varepsilon^{1-|j|}\|z_j(t)\|_{L^1} & \text{if } |j| > 1, \end{cases} \tag{4.70}$$

which according to (4.68) means that

$$\sum_{j\in\mathcal{J}} a_j(t) = 2\sum_{j\in\mathcal{J}_+} a_j(t) = |||z(t)|||_\varepsilon .$$

In order to show (4.69), we distinguish two cases, which have to be treated separately.

Case 1: $|J|_1 > j$. In this case we have $|J|_1 \geqslant j+2$ because $|J|_1$ is odd. Lemma 3.7.1 and the fact that

$$\varepsilon^{|J|_1-3}\prod_{i=1}^{3}\varepsilon^{1-|j_i|} = \varepsilon^0 = 1$$

yields

$$\begin{aligned}\Big\| \int_0^t S_{j,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} &\leqslant C_\mathcal{T}\int_0^t \prod_{i=1}^{3}\|z_{j_i}(\sigma)\|_{L^1}\,\mathrm{d}\sigma = C_\mathcal{T}\varepsilon^{|J|_1-3}\int_0^t\prod_{i=1}^{3}\Big(\varepsilon^{1-|j_i|}\|z_{j_i}(\sigma)\|_{L^1}\Big)\,\mathrm{d}\sigma \\ &\leqslant C_\mathcal{T}\varepsilon^{j-1}\int_0^t\prod_{i=1}^{3}a_{j_i}(\sigma)\,\mathrm{d}\sigma,\end{aligned}$$

because $\varepsilon^{|J|_1-3} \leqslant \varepsilon^{j-1}$ and $\varepsilon^{1-|j_i|}\|z_{j_i}(\sigma)\|_{L^1} \leqslant a_{j_i}(\sigma)$ by definition (4.70). This yields an estimate of the form (4.69) with $C_\star = 0$ and $\widehat{C} = C_\mathcal{T}$.

Case 2: $|J|_1 = j$. In this situation, the simple argument from Case 1 is not enough to prove the desired bound, because now $\varepsilon^{|J|_1-3} = \varepsilon^{j-3}$. As an example, this case appears for $j = 3$ only if $J = (1,1,1)$.
As in the proofs of Section 4.1-4.3 we treat the coefficent $\widehat{u}_1$ in a special way. If one of the coefficients $\widehat{u}_{j_i}$ in the nonlinearity has an index $j_i = 1$, we split this term into $\widehat{u}_1 = \mathcal{P}_\varepsilon\widehat{u}_1 + \mathcal{P}_\varepsilon^\perp\widehat{u}_1$.
This decomposition helps us because all terms where $\mathcal{P}_\varepsilon^\perp\widehat{u}_1$ appears in at least two of the three arguments can be estimated as in Case 1, but now with $\varepsilon^{-1}\|P^\perp z_{\pm1}(\sigma)\|_{L^1} \leqslant a_{\pm1}(\sigma)$. For the remaining terms we aim to gain additional factors of ε from the oscillatory behavior of $\mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)$ by means of integration by parts. These remaining terms, exemplary for $j \in \{3, 5\}$, are

- for $j = 3$: the term on the left-hand side of (4.54) plus its permutations, and (4.55),
- for $j = 5$:

$$\Big\| \int_0^t S_{5,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_3, \mathcal{P}_\varepsilon\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} + \quad \text{permutations} \tag{4.71}$$

$$\text{and} \qquad \Big\| \int_0^t S_{5,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_3, \mathcal{P}_\varepsilon^\perp\widehat{u}_1, \mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} + \quad \text{permutations.} \tag{4.72}$$

Using Step 5 in the proof of Theorem 4.3.4 as an example, the idea is to rewrite the integrand by means of the representations (4.56), (4.57), whereas in the case where $j_i \neq 1$ we use (3.51) for $i = 1, 2, 3$. Together with $\Psi_j^*(\varepsilon k) = \sum_{m=1}^s e_m \psi_{jm}^*(\varepsilon k)$ this leads to a function $F_\varepsilon(\sigma, z)$ given by

$$\begin{aligned} F_\varepsilon(t,z) &= \Big(F_{\varepsilon,m}(t,z)\Big)_{m=1}^s, \\ F_{\varepsilon,m}(t,z)(k) &= \sum_M \int_{\#K=k} \exp\Big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)\Big) Z_{JM}(t,K) c_{jmJM}(\varepsilon,k,K)\,\mathrm{d}K, \end{aligned} \tag{4.73}$$

where we use the notations (3.52), (3.57) and

$$c_{jmJM}(\varepsilon,k,K) = \frac{1}{(2\pi)^d}\psi_{jm}^*(\varepsilon k) T\Big(\psi_{JM}(\varepsilon K)\Big).$$

In the following we will see that there are restrictions and that, in general, the bound (4.66) is only true for $j_{\max} \in \{1,3\}$, whereas the error bound (4.67) is only true for $j_{\max} = 1$.

Restrictions. The crucial term in (4.73) which leads to restrictions is

$$\exp\Big(\tfrac{\mathrm{i}t}{\varepsilon}\Delta\lambda_{jmJM}(\varepsilon,k,K)\Big), \tag{4.74}$$

where the definition of $\Delta\lambda_{jmJM}(\varepsilon,k,K)$ is given by (3.57). We have to assume $\Delta\lambda_{jmJM}(\varepsilon,0,0) \neq 0$ in order to apply integration by parts. However, if one eigenvalue of the matrix $\mathcal{L}(0,\beta)$ is constantly equal to 0 for all $\beta \in \mathbb{R}^d\backslash\{0\}$, i.e. $\omega_m(\beta) = 0$ with $m \neq 1$, and there exists an eigenvalue $\omega_{m_2}(\beta) = -\omega_1(\beta)$ with $1 \neq m_2 \neq m$ such that $\omega_{m_2}(\kappa) = -\omega_1(\kappa) = -\omega$, then the assumption $\Delta\lambda_{jmJM}(\varepsilon,0,0) \neq 0$ is not fulfilled anymore for every j, m, J and M. As an example for the reader we consider the term (4.72). In this case we have with (3.44) and fixed $m = m_1 \neq 1$ with $\omega_m(\beta) = 0$ and fixed $m_2 \neq 1$ with $\omega_{m_2}(\beta) = -\omega_1(\beta)$ for example

$$\begin{aligned} \Delta\lambda_{5mJM}(\varepsilon,0,0) &= \lambda_{5m}(0) - \lambda_{3m}(0) - \lambda_{1m_2}(0) \\ &= -5\omega + \omega_m(5\kappa) + 3\omega - \omega_m(3\kappa) + \omega - \omega_{m_2}(\kappa) = -\omega + \omega_1(\kappa) = 0. \end{aligned}$$

There are more of such combinations also for $j > 5$. If there are eigenvalues of the matrix $\mathcal{L}(0,\beta)$ which are constant in β, there is always the possibility to have resonances. For this example we cannot apply integration by parts for every j, m, J and M. Thus, if $\sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{u}_3(t)\|_{L^1} \leqslant C\varepsilon^2$ and $\sup_{t\in[0,t_\star/\varepsilon]} \|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon$ hold, we can only show with the straightforward argument of Case 1

$$\Big\| \int_0^t S_{5,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_3, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma \Big\|_{L^1} \leqslant C\varepsilon^2$$

and not $\mathcal{O}(\varepsilon^3)$ as required in (4.69).

In summary, the required non-resonance conditions are a limitation for this technique of proof. For $j > 3$, these are generally no longer satisfied for the Klein–Gordon system with $d > 1$ and the Maxwell–Lorentz system. This can be verified since we know explicitly the eigenvalues of the matrix $\mathcal{L}(0,\kappa)$ for the Klein–Gordon and the Maxwell–Lorentz system, see Example 3.2.3 and 3.2.4. Hence, in general the bound (4.66) is only true for $j_{\max} \in \{1,3\}$ and (4.67) is only true for $j_{\max} = 1$, because for $j_{\max} = 3$ we would need for the error approximation

$$\Big\| \int_0^t S_{5,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_3, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma \Big\|_{L^1} \leqslant C\varepsilon^3.$$

Remark 4.4.1. *The Klein–Gordon system with $d = 1$ is a special case. Here, these non-resonance conditions are satisfied, since there is no constant eigenvalue in $\beta \in \mathbb{R}\backslash\{0\}$ (see Example 3.2.3). Therefore, we suspect that (4.66) and (4.67) can also be shown for higher j_{max}. This conjecture is reinforced by the numerical experiments above, illustrated in Figure 4.2 and 4.3. This is also the reason why we have used $\widetilde{\mathbf{u}}^{(5)}$ as the reference solution in all the numerical experiments in this chapter.*

Therefore, for $d > 1$ we show a weaker error bound for the approximation (3.15) with $j_{\max} = 3$ than (4.67) in the next subsection.

4.4.1 The case $j_{\max} = 3$ with $d > 1$

The first part of this subsection is based on [4], where we prove the error bound (cf. [4, Theorem 4.2])

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_W \leqslant C\varepsilon^2.$$

However, under slightly stronger assumptions we can prove an improved error approximation of the form

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_W \leqslant C\varepsilon^3. \tag{4.75}$$

We list the following points, which must be shown one after the other in order to be able to establish a theorem for the error approximation for (3.15) with $j_{\max} = 3$ which fulfills (4.75). The first two points are adopted from [4].

- In [4, Proposition 3.2] we prove under Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2 and with initial data $p \in W^1$ that there exists a $t_\star \in (0, t_{\text{end}}]$ and a sufficiently large $r > 0$ such that

  $$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_{\varepsilon,1} \leqslant r \qquad \text{for all } \varepsilon \in (0,1],$$

  where

  $$|||y|||_{\varepsilon,1} = 2\|Py_1\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp y_1\|_{L^1} + \frac{2}{\varepsilon}\|y_3\|_{L^1}$$

  for all $y = (y_1, y_3)$ with $y_j \in L^1(\mathbb{R}^d, \mathbb{C}^n)$. We note that $|||\cdot|||_{\varepsilon,1}$ denotes a weaker scaled norm which at first implies for $j = 3$ only the weaker bound

  $$\sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{u}_3(t)\|_{L^1} \leqslant C\varepsilon.$$

- In [4, Proposition 3.6] we show under the same assumptions but with initial data $p \in W^2$ that there is a constant C which does not depend on ε such that

  $$\sup_{t\in[0,t_\star/\varepsilon]} |||D_\mu z(t)|||_{\varepsilon,1} \leqslant C \qquad \text{for all } \varepsilon \in (0,1].$$

The following parts are not included in [4].

- Under the same assumptions as in [4, Proposition 3.6] and Assumption 4.3.1 we can proceed as in the proof of Theorem 4.3.4 Steps 4-7 to prove that there exists a $t_\star \in (0, t_{\text{end}}]$ and a sufficiently large $r > 0$ such that

  $$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_{\varepsilon} \leqslant r \qquad \text{for all } \varepsilon \in (0,1].$$

By definition (4.68) this boundedness implies the refined bounds

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon \qquad \text{and} \qquad \sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{u}_3(t)\|_{L^1} \leqslant C\varepsilon^2. \tag{4.76}$$

We omit the detailed steps, as these are very similar to [4, Proof of Proposition 3.6] combined with the techniques of the Steps 4-7 of Theorem 4.3.4.

- In order to prove the approximation error (4.75) we need the refined bounds (4.76) in a stronger norm, meaning

$$\sup_{t\in[0,t_\star/\varepsilon]} \|D_\mu \mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \leqslant C\varepsilon \qquad \text{and} \qquad \sup_{t\in[0,t_\star/\varepsilon]} \|D_\mu \widehat{u}_3(t)\|_{L^1} \leqslant C\varepsilon^2. \tag{4.77}$$

To show (4.77), it is sufficient to prove

$$\sup_{t\in[0,t_\star/\varepsilon]} |||D_\mu z(t)|||_\varepsilon \leqslant C \qquad \text{for all } \varepsilon \in (0,1]$$

under the additional assumption $p \in W^3$. To this end, we need a similar lemma as Lemma 4.3.3 which ensures that there exists a constant C such that

$$\sup_{t\in[0,t_{\text{end}}/\varepsilon]} \|D_\mu \partial_t^2 \mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1} \leqslant C.$$

- Next, we investigate for the error bound the terms (4.71) and (4.72). Since, we already know that we cannot apply integration by parts to (4.72), we obtain by means of the refined bounds (4.76)

$$\begin{aligned}\Big\| \int_0^t S_{5,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_3, \mathcal{P}_\varepsilon^\perp \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma)\, \mathrm{d}\sigma \Big\|_{L^1} &\leqslant C_{\mathcal{T}} \tfrac{t_\star}{\varepsilon} \sup_{t\in[0,t_\star/\varepsilon]} \big(\|\widehat{u}_3(t)\|_{L^1} \|\mathcal{P}_\varepsilon^\perp \widehat{u}_1(t)\|_{L^1} \|\mathcal{P}_\varepsilon \widehat{u}_1(t)\|_{L^1}\big) \\ &\leqslant C\varepsilon^2. \end{aligned} \tag{4.78}$$

In order to bound the terms in (4.71) we need an additional non-resonance assumption which can be verified with the same reasoning as in Remark 4.3.2.

Assumption 4.4.2. *The matrices $\mathcal{L}_5(0) = \mathcal{L}(5\omega, 5\kappa)$ and $\mathcal{L}_3(0) = \mathcal{L}(3\omega, 3\kappa)$ have no common eigenvalues, i.e. $\lambda_{5m}(0) \neq \lambda_{3m_1}(0)$ for all $m, m_1 = 1, \dots, s$.*

Now the preliminary work with all the important requirements and estimates is done and we state the following theorem:

Theorem 4.4.3. *Let $p \in W^3$ and let $\mathbf{u}$ be the solution of (1.4). Let $u = (u_1, u_3)$ with $u_j \in C^2([0, t_{end}/\varepsilon], W^1) \cap C^1([0, t_{end}/\varepsilon], W^2) \cap C([0, t_{end}/\varepsilon], W^3)$ be the classical solution of (3.16), and let $\widetilde{\mathbf{u}}^{(3)}$ be the approximation defined in (3.15) with $j_{max} = 3$. Under Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2, 4.3.1 and 4.4.2, there is a constant such that*

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_W \leqslant C\varepsilon^3,$$
$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \widetilde{\mathbf{u}}^{(3)}(t)\|_{L^\infty} \leqslant C\varepsilon^3.$$

Proof. The proof is similar to [4, Proof of Theorem 4.2]. Step 1 to Step 4 can be adopted with the slightly difference that the goal is to show

$$\sum_{|j|\in\{5,7,9\}} \sum_{\#J=j} \Big\| \int_0^t S_{j,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(\sigma)\,\mathrm{d}\sigma \Big\|_{L^1} \leqslant C\varepsilon^2$$

with a constant C which does not depend on ε, whereas in [4] we have on the right-hand side ε. Analogous to [4], in Step 4 we make a distinction between the possible combinations of multi-indices J in two cases. At the beginning of Section 4.4 we already outline the cases $|J|_1 > j$ and $|J|_1 = j$, which can be directly transferred to the setting $j = 5$. Additionally, by means of the refined bounds (4.76) we already showed (4.78). The main difficulty is to prove that

$$\Big\| \int_0^t S_{5,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_3,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma)\,\mathrm{d}\sigma \Big\|_{L^1} \leqslant C\varepsilon^2. \tag{4.79}$$

At this point we note that formally the order of the integrand is $\mathcal{O}\big(\varepsilon^2\big)$. At first we might think we only have to apply integration by parts once, which provides an additional factor ε, to show (4.71), but unfortunately this is not correct. This becomes clearer after adopting Step 5 of the proof for (4.71) and integrating by parts. For $J = (3,1,1)$, $M = (m_1,1,1)$ and fixed m, m_1, k and K in (4.73) we set

$$\begin{aligned}
\Delta\lambda(\varepsilon) &= \Delta\lambda_{5mJM}(\varepsilon,k,K),\\
Z(\sigma) &= Z_{JM}(\sigma,K) = z_{3,m_1}(\sigma,k^{(1)})z_{1,1}(\sigma,k^{(2)})z_{1,1}(\sigma,k^{(3)}),\\
Y(\sigma) &= \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\big[\Delta\lambda(\varepsilon)-\Delta\lambda(0)\big]\Big)Z(\sigma).
\end{aligned} \tag{4.80}$$

As mentioned in Section 3.7 the idea is to expand the highly oscillatory term (4.74) in a suitable way. By Assumption 4.4.2, (4.80) and integration by parts we obtain

$$\begin{aligned}
\Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\Big)Z(\sigma)\,\mathrm{d}\sigma\Big| &= \Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)Y(\sigma)\,\mathrm{d}\sigma\Big|\\
&\leqslant \frac{\varepsilon}{|\Delta\lambda(0)|}\Big|\Big[\exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)Y(\sigma)\Big]_{\sigma=0}^t\Big| + \frac{\varepsilon}{|\Delta\lambda(0)|}\Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)\partial_t Y(\sigma)\,\mathrm{d}\sigma\Big|\\
&\leqslant C\varepsilon\Big(|Z(t)|+|Z(0)|\Big) + C\varepsilon\Big|\int_0^t \frac{\big[\Delta\lambda(\varepsilon)-\Delta\lambda(0)\big]}{\varepsilon}\exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)Y(\sigma)\,\mathrm{d}\sigma\Big|\\
&\quad + C\varepsilon\Big|\int_0^t \exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\Big)\partial_t Z(\sigma)\,\mathrm{d}\sigma\Big|.
\end{aligned} \tag{4.81}$$

For $t \in [0, t_\star/\varepsilon]$ and $Z_m(t) = Z(t)$, we estimate

$$\begin{aligned}
\sum_{m=1}^s \int_{\mathbb{R}^d}\int_{\#K=k} |Z_m(t,K)|\,\mathrm{d}K\,\mathrm{d}k &\leqslant \sqrt{s}\int_{\mathbb{R}^d}\int_{\#K=k} \big|\widehat{u}_3\big(t,k^{(1)}\big)\big|_2\,\big|\widehat{u}_1\big(t,k^{(2)}\big)\big|_2\,\big|\widehat{u}_1\big(t,k^{(3)}\big)\big|_2\,\mathrm{d}K\,\mathrm{d}k\\
&= \sqrt{s}\|u_3(t)\|_W\,\|u_1(t)\|_W^2 \leqslant C\varepsilon^2
\end{aligned}$$

by means of the refined bound (4.76) for $\widehat{u}_3$. Furthermore, with

$$\Big|\int_0^t \frac{\big[\Delta\lambda(\varepsilon)-\Delta\lambda(0)\big]}{\varepsilon}\exp\Big(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(0)\Big)Y(\sigma)\,\mathrm{d}\sigma\Big| \leqslant C\sum_{i=1}^3\int_0^t |k|_1\,|Z(\sigma)|\,\mathrm{d}\sigma \leqslant Ct_\star\sum_{i=1}^3 \sup_{\sigma\in[0,t]}\big(\varepsilon^{-1}|k|_1\,|Z(\sigma)|\big),$$

and the refined bound (4.76) for $\widehat{u}_3$ we obtain

$$\sum_{i=1}^{3}\sum_{m=1}^{s}\int_{\mathbb{R}^d}\int_{\#K=k}|k^{(i)}|_1\int_0^t|Z_m(\sigma,K)|\,\mathrm{d}\sigma\,\mathrm{d}K\,\mathrm{d}k\leqslant C\int_0^t\|u_3(\sigma)\|_{W^1}\,\|u_1(\sigma)\|_{W^1}^2\,\mathrm{d}\sigma$$
$$\leqslant C\frac{t_\star}{\varepsilon}\sup_{t\in[0,t_\star/\varepsilon]}\left(\|u_3(t)\|_{W^1}\,\|u_1(t)\|_{W^1}^2\right)\leqslant C\varepsilon.$$

For the last term of (4.81) we distinguish two cases. With the product rule we have

$$\partial_t Z_m(t,K)=\partial_t z_{3m}(t,k^{(1)})z_{11}(t,k^{(2)})z_{11}(t,k^{(3)})$$
$$+z_{3m}(t,k^{(1)})\left(\partial_t z_{11}(t,k^{(2)})z_{11}(t,k^{(3)})+z_{11}(t,k^{(2)})\partial_t z_{11}(t,k^{(3)})\right).$$

Firstly, the refined bound (4.76) for $\widehat{u}_3$ combined with (4.40) yields

$$\sum_{m=1}^{s}\int_{\mathbb{R}^d}\int_{\#K=k}\int_0^t\left|z_{3m}(\sigma,k^{(1)})\left(\partial_t z_{11}(\sigma,k^{(2)})z_{11}(\sigma,k^{(3)})+z_{11}(\sigma,k^{(2)})\partial_t z_{11}(\sigma,k^{(3)})\right)\right|\mathrm{d}\sigma\,\mathrm{d}K\,\mathrm{d}k$$
$$\leqslant 2\sqrt{s}\int_0^t\|z_3(\sigma)\|_{L^1}\,\|z_1(\sigma)\|_{L^1}\,\|\partial_t z_1(\sigma)\|_{L^1}\,\mathrm{d}\sigma\leqslant C\frac{t_\star}{\varepsilon}\sup_{t\in[0,t_\star/\varepsilon]}\left(\|z_3(t)\|_{L^1}\,\|z_1(t)\|_{L^1}\,\|\partial_t z_1(t)\|_{L^1}\right)\leqslant C\varepsilon^2.$$

Next, since (4.40) holds also for $j=3$, we aim to gain again an additional factor ε by integration by parts. However, first we insert the evolution equation (3.53) combined with (3.54) for $\partial_t z_{3m_1}$. This yields

$$\begin{aligned}
&\varepsilon\left|\int_0^t\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\right)\partial_t z_{3m_1}(\sigma,k^{(1)})z_{11}(\sigma,k^{(2)})z_{11}(\sigma,k^{(3)})\,\mathrm{d}\sigma\right|\\
&\quad=\varepsilon^2\left|\int_0^t\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\lambda(\varepsilon)\right)\sum_{\#J_1=3}\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\lambda_{3m_1}(\varepsilon k^{(1)})\right)\psi^*_{3m_1}(\varepsilon k^{(1)})\right.\\
&\qquad\left.\times\,\mathcal{T}\big(\widehat{u}_{j_{11}},\widehat{u}_{j_{12}},\widehat{u}_{j_{13}}\big)(\sigma,k^{(1)})z_{11}(\sigma,k^{(2)})z_{11}(\sigma,k^{(3)})\,\mathrm{d}\sigma\right|\\
&\quad=\varepsilon^2\left|\int_0^t\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon}\Delta\widetilde{\lambda}(\varepsilon)\right)\psi^*_{3m_1}(\varepsilon k^{(1)})\mathcal{T}\big(\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1,\mathcal{P}_\varepsilon\widehat{u}_1\big)(\sigma,k^{(1)})z_{11}(\sigma,k^{(2)})z_{11}(\sigma,k^{(3)})\,\mathrm{d}\sigma\right| \qquad (4.82)\\
&\qquad+\mathcal{O}\big(\varepsilon^2\big),
\end{aligned}$$

where

$$\Delta\widetilde{\lambda}(\varepsilon)=\lambda_{5m}(\varepsilon k)-\lambda_{11}(\varepsilon k^{(2)})-\lambda_{11}(\varepsilon k^{(3)}).$$

In order to apply again integration by parts, it is sufficient to assume $\lambda_{5m}(0)\neq 0$ for all $m=1,\dots,s$. Fortunately, this condition is satisfied because of Assumption 3.2.6. Applying integration by parts to (4.82) and estimating every single term similarly to [4, Proof of Theorem 4.2, Step 7] yields (4.79) and completes the proof. ∎

4.4.2 The case $j_{\max} > 3$ with $d > 1$

Based on the insights of Subsection 4.4.1, we conjecture that under slightly higher regularity assumptions and under additional non-resonance assumptions there exists a constant $C > 0$ such that

$$\sup_{t\in[0,t_\star/\varepsilon]} \|u_j(t)\|_W \leqslant C\varepsilon^{(|j|+1)/2}, \qquad \text{for } j \in \mathcal{J}\backslash\{\pm 1\}, \tag{4.83}$$

and

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\mathbf{u}(t) - \tilde{\mathbf{u}}^{(j_{\max})}(t)\|_W \leqslant C\varepsilon^{(j_{\max}+3)/2}, \qquad \text{for } j_{\max} > 3, \tag{4.84}$$

hold. For $j_{\max} = 3$, the error bound (4.84) coincides with Theorem 4.4.3. We emphasize that we do not present a rigorous proof of the estimates (4.83) and (4.84) in this subsection. Rather, we only outline the main steps and ideas required to do so.

The general procedure in this subsection is similar to the procedure at the beginning of Section 4.4, except that the scaling changes.

General procedure. In order to prove (4.83) for $j_{\max} > 3$ and $d > 1$ we define the scaled norm

$$|||y|||_\varepsilon = 2\|Py_1\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp y_1\|_{L^1} + \sum_{j\in\mathcal{J}_+\backslash\{1\}} \frac{2}{\varepsilon^{(j+1)/2}}\|y_j\|_{L^1} \tag{4.85}$$

for all $y = (y_1, \dots, y_{j_{\max}})$ with $y_j \in L^1(\mathbb{R}^d, \mathbb{C}^n)$. In comparison to the definition (4.68) the scaling for $j > 3$ is different. As before, we set $y_{-j} = \overline{y_j}$ and the first goal is to prove (4.83) or equivalently that there is a constant C such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_\varepsilon \leqslant C,$$

for all $\varepsilon \in (0, 1]$. To this end, we investigate every term on the right-hand side of

$$\begin{aligned}|||z(t)|||_\varepsilon \leqslant |||z(0)\,|||_\varepsilon + 2\sum_{\#J=1}\left(\varepsilon\Big\|\int_0^t P\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1} + \Big\|\int_0^t P^\perp\mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}\right)\\ + 2\sum_{j\in\mathcal{J}_+\backslash\{1\}}\sum_{\#J=j}\frac{1}{\varepsilon^{(j-1)/2}}\Big\|\int_0^t \mathbf{F}_\varepsilon(\sigma,\widehat{u},J)\,\mathrm{d}\sigma\Big\|_{L^1}.\end{aligned}$$

In contrast to (4.69), the goal is to establish

$$\Big\|\int_0^t S_{j,\varepsilon}(\sigma)\mathcal{T}\big(\widehat{u}_{j_1},\widehat{u}_{j_2},\widehat{u}_{j_3}\big)(\sigma)\,\mathrm{d}\sigma\Big\|_{L^1} \leqslant \begin{cases} C_\star\varepsilon^{(j-1)/2} + \widehat{C}\varepsilon^{(j+1)/2}\int_0^t\prod_{i=1}^3 a_{j_i}(\sigma)\,\mathrm{d}\sigma & \text{for } 1 < j \leqslant j_{\max},\\ C\varepsilon^{(j-1)/2} & \text{for } j > j_{\max}\end{cases} \tag{4.86}$$

for all $t \in [0, t_\star/\varepsilon]$ and for every $J = (j_1, j_2, j_3) \in \mathcal{J}^3$ with $\#J = j$, and in comparison to definition (4.70), we now define

$$a_j(t) = \begin{cases}\|Pz_1(t)\|_{L^1} + \varepsilon^{-1}\|P^\perp z_1(t)\|_{L^1} & \text{if } j = \pm 1,\\ \varepsilon^{-(1+|j|)/2}\|z_j(t)\|_{L^1} & \text{if } |j| \geqslant 3.\end{cases} \tag{4.87}$$

As before, we distinguish two cases to show (4.86). However, the procedure in both cases is a little different than before because of the different scaling.

Case 1: $|J|_1 > j$. In this case we have $|J|_1 \geqslant j+2$ because $|J|_1$ is odd. Lemma 3.7.1 and the fact that

$$\varepsilon^{(|J|_1+3)/2} \prod_{i=1}^{3} \varepsilon^{-(1+|j_i|)/2} = \varepsilon^0 = 1$$

yields

$$\begin{aligned}
\Big\| \int_0^t S_{j,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} &\leqslant C_{\mathcal{T}} \int_0^t \prod_{i=1}^{3} \|z_{j_i}(\sigma)\|_{L^1} \, \mathrm{d}\sigma \\
&= C_{\mathcal{T}} \varepsilon^{(|J|_1+3)/2} \int_0^t \prod_{i=1}^{3} \Big(\varepsilon^{-(1+|j_i|)/2} \|z_{j_i}(\sigma)\|_{L^1} \Big) \, \mathrm{d}\sigma \\
&\leqslant C_{\mathcal{T}} \varepsilon^{(j+1)/2} \varepsilon^2 \int_0^t \prod_{i=1}^{3} \Big(\varepsilon^{-(1+|j_i|)/2} \|z_{j_i}(\sigma)\|_{L^1} \Big) \, \mathrm{d}\sigma \\
&\leqslant C_{\mathcal{T}} \varepsilon^{(j+1)/2} \int_0^t \prod_{i=1}^{3} a_{j_i}(\sigma) \, \mathrm{d}\sigma, \qquad (4.88)
\end{aligned}$$

since $\varepsilon^{(|J|_1+3)/2} \leqslant \varepsilon^{(j+5)/2} = \varepsilon^{(j+1)/2}\varepsilon^2$. In addition, the estimate (4.88) is valid due to $\varepsilon^{1-(1+|j_i|)/2} = \varepsilon^{(1-|j_i|)/2}$ and

$$\begin{aligned}
\varepsilon^{(1-|j_i|)/2} \|z_{j_i}(\sigma)\|_{L^1} = \|z_{j_i}(\sigma)\|_{L^1} \leqslant a_{j_i}(\sigma) \qquad &\text{for } |j_i| = 1, \\
\varepsilon^{-(1+|j_i|)/2} \|z_{j_i}(\sigma)\|_{L^1} \leqslant a_{j_i}(\sigma) \qquad &\text{for } |j_i| \geqslant 3
\end{aligned}$$

by definition (4.87). We remark that in this case at most two components of J can be $|j_i| = 1$, and we need one factor ε of $\varepsilon^{(j+1)/2}\varepsilon^2$ to obtain the correct scaling for $\varepsilon^{-1}\|z_{\pm 1}(\sigma)\|_{L^1}$. This yields an estimate of the form (4.86) with $C_\star = 0$ and $\widehat{C} = C_{\mathcal{T}}$.

Case 2: $|J|_1 = j$. In this situation, the simple argument from Case 1 is not enough to prove the desired bound for all possible multi-indices J, because now $\varepsilon^{(|J|_1+3)/2} = \varepsilon^{(j+3)/2} = \varepsilon^{(j+1)/2}\varepsilon$. In comparison to Case 1, the additional factor ε is only enough to estimate

$$\Big\| \int_0^t S_{j,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_{j_1}, \widehat{u}_{j_2}, \widehat{u}_{j_3}\big)(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} \leqslant C_{\mathcal{T}} \varepsilon^{(j+1)/2} \int_0^t \prod_{i=1}^{3} a_{j_i}(\sigma) \, \mathrm{d}\sigma,$$

if at most one component of J is equal to 1. For all multi-indices J with two or three components equal to 1, we proceed as follows. We note that this case appears for $j \geqslant 3$ only if $J = (j-2, 1, 1)$ and for its permutations.

We treat the coefficent $\widehat{u}_1$ in a special way and split this term into $\widehat{u}_1 = \mathcal{P}_\varepsilon \widehat{u}_1 + \mathcal{P}_\varepsilon^\perp \widehat{u}_1$.

All terms where $\mathcal{P}_\varepsilon^\perp \widehat{u}_1$ appears in at least once of the arguments can be estimated as in Case 1, but now with $\varepsilon^{-1}\|P^\perp z_{\pm 1}(\sigma)\|_{L^1} \leqslant a_{\pm 1}(\sigma)$. We aim to gain additional factors of ε from the oscillatory behavior of $\mathbf{F}_\varepsilon(\sigma, \widehat{u}, J)$ by means of integration by parts for the remaining terms, which, exemplary for $j \in \{5, 7\}$, are

- for $j = 5$: the term on the left-hand side of (4.79) plus its permutations,
- for $j = 7$:

$$\Big\| \int_0^t S_{7,\varepsilon}(\sigma) \mathcal{T}\big(\widehat{u}_5, \mathcal{P}_\varepsilon \widehat{u}_1, \mathcal{P}_\varepsilon \widehat{u}_1\big)(\sigma) \, \mathrm{d}\sigma \Big\|_{L^1} + \quad \text{permutations.}$$

To apply integration by parts, we require the following non-resonance assumption in accordance to Assumptions 4.3.1 and 4.4.2.

Assumption 4.4.4. *The matrices $\mathcal{L}_{j+2}(0)$ and $\mathcal{L}_j(0)$ have no common eigenvalues, i.e. $\lambda_{(j+2)\,m}(0) \neq \lambda_{jm_1}(0)$ for all $m, m_1 = 1, \dots, s$ and $j \in \mathcal{J}_+$.*

The difficulty which we have already seen in Subsection 4.4.1 is that we cannot show (4.86) for $1 < j \leqslant j_{\max}$ directly. The reason is that as soon as we apply integration by parts we obtain by the Lipschitz continuity factors $|k_i|_1$. Hence, we would need the bounds (4.83) in the stronger norm as well. However, this is not yet given at the beginning and has to be shown first. Similarly to Subsection 4.4.1, we list the following points, which must be shown one after the other in order to be able to establish (4.86) for $1 < j \leqslant j_{\max}$, which implies (4.83).

- We start with the weaker scaled norm

$$|||y|||_{\varepsilon,1} = 2\|Py_1\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp y_1\|_{L^1} + \frac{2}{\varepsilon}\sum_{j\in\mathcal{J}_+\setminus\{1\}} \|y_j\|_{L^1}$$

  for all $y = (y_1, \dots, y_{j_{\max}})$ with $y_j \in L^1(\mathbb{R}^d, \mathbb{C}^n)$. If we are able to show that there exists a $t_\star \in (0, t_{\text{end}}]$ and a sufficiently large $r > 0$ such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_{\varepsilon,1} \leqslant r \qquad \text{for all } \varepsilon \in (0,1],$$

  this result implies for $j \geqslant 3$ only the weaker bound

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{u}_j(t)\|_{L^1} \leqslant C\varepsilon.$$

- In the next step we show this weaker bounds in the stronger norm as in Proposition 4.2.2.

- With this result of weaker bounds in the stronger norm, we can proceed as in the proof of Theorem 4.3.4 Steps 4-7 to prove that there exists a $t_\star \in (0, t_{\text{end}}]$ and a sufficiently large $r > 0$ such that

$$\sup_{t\in[0,t_\star/\varepsilon]} |||z(t)|||_{\varepsilon,2} = 2\|Pz_1(t)\|_{L^1} + \frac{2}{\varepsilon}\|P^\perp z_1(t)\|_{L^1} + \frac{2}{\varepsilon^2}\sum_{j\in\mathcal{J}_+\setminus\{1\}} \|z_j(t)\|_{L^1} \leqslant r \qquad \text{for all } \varepsilon \in (0,1].$$

  For $j \geqslant 3$ this estimate yields

$$\sup_{t\in[0,t_\star/\varepsilon]} \|\widehat{u}_j(t)\|_{L^1} \leqslant C\varepsilon^2.$$

- Again this weaker refined bounds can be shown in the stronger norm and so on.

Remark 4.4.5.

(i) With this procedure, we gradually increase the power of ε on the right side of the weaker norms up to the required scaled norm (4.85). Similarly to the proof of Theorem 4.4.3, to increase the power in some cases we have to apply integration by parts more than once to gain the required factors of ε. Hence, we need the refined bounds in even stronger norms. Consequently, this increases the regularity assumption on the initial data.

(ii) We note that if we apply integration by parts to an integral term with $\partial_t z_{j_i m_i}(\sigma)$, we first substitute the evolution equation for $\partial_t z_{j_i m_i}(\sigma)$ in order to avoid terms of the form $\partial_t^2 z_{j_i m_i}(\sigma)$ or higher derivatives in time for $j_i \neq 1$. As an example we refer to the proof of Theorem 4.4.3.

Finally, to prove the error bound (4.84) we proceed similarly to the proof of Theorem 4.4.3. The difference is that we apply the Cases 1 and 2 described above to show (4.86) for $j > j_{\max}$.

CHAPTER 5

Numerical methods

In addition to the analytical study of the exact solution of the system (3.16) with $j_{\max} = 1$, we are also interested in approximations of the exact solution u_1 by numerical computations. From a numerical point of view the advantage of considering the evolution equation of u_1 compared to the full system (1.4) is that in system (3.16) with $j_{\max} = 1$ there are no more spatial oscillations caused by ε. However, the coefficient u_1 still oscillates in time which causes difficulties for standard numerical methods. The overall goal is to achieve an approximation for the semilinear hyperbolic system (1.4) by the analytical approximation (3.15) with $j_{\max} = 1$ in combination with numerical computations. Thus, the error of the overall approximation consists of two parts. The first approximation is given by the analytical approximation and, therefore, the first approximation error depends only on the smallness parameter ε. The second approximation error is made by solving the transformed system (3.53) for $j = 1$ with a numerical time integrator. This error mainly depends on the choice of the step-size τ. In general the power of τ depends on the numerical time integrator which is used. According to Theorem 4.3.4, the accuracy of the SVEA is of order $\mathcal{O}(\varepsilon^2)$ and, hence, for a method of order p the accuracy is in $\mathcal{O}\big(\max\{\varepsilon^2, \tau^p\}\big)$, if the error constants of both approximations are similar. Consequently, for $\tau < \varepsilon^{\frac{2}{p}}$ the accuracy is limited by the analytical approximation. This has the following implications for the choice of step-sizes of the numerical method. For a first-order method, it is reasonable to use step-sizes in the regime $\varepsilon^2 \leqslant \tau$, whereas for a second-order method, it is only useful to choose step-sizes with $\varepsilon \leqslant \tau$. Of course we can use step-sizes smaller than the restriction, although, computing a numerical approximation finer than $\mathcal{O}\big(\varepsilon^2\big)$ gives no advantage because we are limited by the analytical approximation error. The chapter is structured as follows.

After transforming the general system (3.16) in Section 5.1, we will first construct a one-step method and prove that this method converges with order 1 uniformly in ε. First-order methods as the one-step method in Section 5.2 are certainly not satisfactory to approximate solutions. However, studying this method permits valuable insight for the construction and the analysis of more elaborate methods.

After the investigation of the one-step method, we aim to construct a method with higher accuracy.

Instead of investigating a one-step method of order two, we extend our approach and consider a two-step method. We provide the rationale in Subsection 5.2.3. The goal is to prove that the two-step method has a higher accuracy. The majority of this chapter is devoted to error analysis. For this purpose we have to rewrite the two-step method into an equivalent one-step method as a first step in order to be able to apply the typical strategy that stability combined with the local error bound leads to the global error bound. Section 5.3 contains two different global error estimates. These are dependent on the associated assumptions we make. Additional assumptions improve the error estimate of the local error. Here, the bounds of the coefficients (shown in Chapter 4) and the associated assumptions play a major role again. They are crucial to improve the error estimates. It is important to emphasize that the bound of the coefficients alone is not sufficient to improve the error estimate. However, similar ideas and techniques are used as in Chapter 4.
Since the constructed numerical methods are costly due to the fact that each time-step requires the computation of nested multiple sums, we end this chapter by reducing the computational workload in a suitable way. We start this chapter by transforming the system (1.4).

5.1 Transformation of the system

5.1.1 Co-moving coordinate system and rescaling of time

We know that the solution of the system (1.4) propagates in time. Consequently, the computational domain in space has to be large enough such that the solution does not leave this domain. In order to avoid a large finite space domain we introduce a time dependent shift in the spatial coordinate to go into a co-moving coordinate system. In Remark 4.1.3 we interpret the non-oscillatory part of $\widehat{u}_1$ as the main part of the solution. This main part travels with the group velocity $c_g = c_g(\kappa)$, cf. (3.5). If we define $t' = \varepsilon t$, the time dependent spatial shift $x' = x - tc_g$ and the new variable

$$\mathbf{v}(t', x') := \mathbf{u}\left(\frac{t'}{\varepsilon}, x' + \frac{t'}{\varepsilon} c_g\right), \quad \text{equivalently} \quad \mathbf{v}(\varepsilon t, x - tc_g) = \mathbf{u}(t, x),$$

then we obtain via the chain rule

$$\begin{aligned} \partial_t \mathbf{u}(t, x) &= \partial_t \mathbf{v}(t', x') = \varepsilon \partial_{t'} \mathbf{v}(t', x') - c_g \cdot \nabla_{x'} \mathbf{v}(t', x'), \\ A(\partial_x)\mathbf{u}(t, x) &= A(\partial_{x'})\mathbf{v}(t', x'). \end{aligned}$$

Thus, the change of variables turns the original problem (1.4) into

$$\partial_{t'} \mathbf{v} + \frac{1}{\varepsilon} B(\partial_{x'})\mathbf{v} + \frac{1}{\varepsilon^2} E\mathbf{v} = T(\mathbf{v}, \mathbf{v}, \mathbf{v}), \qquad x' \in \mathbb{R}^d,\ t' \in (0, t_{\text{end}}], \tag{5.1}$$

where we define

$$B(\partial_{x'}) = \sum_{\mu=1}^{d} B_\mu \frac{\partial}{\partial x'_\mu}, \qquad B_\mu = A_\mu - (c_g)_\mu I. \tag{5.2}$$

Next, we set $v_j(t', x') = u_j(t, x)$ such that the change of variables turns the left-hand side of the system of PDEs (3.16) into

$$\begin{aligned}
&\partial_t u_j(t,x) + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, jk)u_j(t,x) + A(\partial_x)u_j(t,x) \\
&= \varepsilon\partial_{t'} v_j(t',x') - c_g \cdot \nabla_{x'} v_j(t',x') + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, jk)v_j(t',x') + A(\partial_{x'})v_j(t',x') \\
&= \varepsilon\partial_{t'} v_j(t',x') + \frac{\mathrm{i}}{\varepsilon}\mathcal{L}(j\omega, jk)v_j(t',x') + B(\partial_{x'})v_j(t',x').
\end{aligned}$$

Remark 5.1.1. *For the sake of simplicity we omit in the following the prime and write again t and x instead of t' and x', respectively. We explicitly note that the variables t and x in the rest of the chapter are different variables than in the previous chapters.*

With the abbreviations $v_J = (v_{j_1}, v_{j_2}, v_{j_3})$ and $B(\partial) = B(\partial_x)$ we obtain the system

$$\partial_t v_j + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega, jk)v_j + \frac{1}{\varepsilon}B(\partial)v_j = \sum_{\#J=j} T(v_J), \quad j \in \mathcal{J}_+, \quad t \in (0, t_{\mathrm{end}}], \quad x \in \mathbb{R}^d. \tag{5.3}$$

The sum on the right-hand side is taken over the set (3.17). Analogously to (3.16), the PDE (5.3) is compatible with the condition that $v_{-j} = \overline{v_j}$. In the new variables the approximation defined in (3.15) reads

$$\mathbf{v}(t,x) \approx \sum_{j\in\mathcal{J}} e^{\mathrm{i}jk\cdot x/\varepsilon} e^{\mathrm{i}j(k\cdot c_g - \omega)t/\varepsilon^2} v_j(t,x), \qquad v_{-j} = \overline{v_j}. \tag{5.4}$$

In principle, splitting methods could be used as numerical integrators for solving the system (5.3). However, numerical experiments illustrate that these methods are not suitable for highly-oscillatory problems because of step-size restrictions, which will be seen in Section 5.6. Furthermore, we do not directly apply exponential integrators to the system (5.3). The reason is that the techniques in the error analysis always use Taylor expansion of the exact solution. However, those Taylor expansions are no longer applicable since higher-order time derivatives of v_j will have large norms because of the factor $\frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega, jk)$. Therefore, we transform the system (5.3) similarly to Chapter 3 and construct new methods.

5.1.2 Evolution equations on $\mathbb{T}^d$

In order to do numerical simulations we have to truncate the full space $\mathbb{R}^d$. Therefore, we replace for simplicity $\mathbb{R}^d$ by $\mathbb{T}^d$ in (5.3), where $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ means $[-\pi, \pi]$ with periodic boundary conditions. Of course, periodic boundary conditions are a simplification, however, we will ignore this for the time being. First, we derive the evolution equations on the torus and then we end this subsection with an estimate important for the later analysis.

For the rest of the chapter we use the same letter v_j as in (5.3) but now for the variable on the torus. Formally the solutions v_j, after transformation in space and time, of the system

$$\partial_t v_j + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega, jk)v_j + \frac{1}{\varepsilon}B(\partial)v_j = \sum_{\#J=j} T(v_J), \quad x \in \mathbb{T}^d, \quad t \in (0, t_{\mathrm{end}}] \tag{5.5}$$

can be represented by the semidiscrete Fourier transform

$$v_j(t,x) = \sum_{k\in\mathbb{Z}^d} \widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x} \tag{5.6}$$

with Fourier coefficients

$$\widehat{v}_j(t,k) = (2\pi)^{-d} \int_{\mathbb{T}^d} v_j(t,x) e^{-\mathrm{i}k\cdot x}\,\mathrm{d}x, \quad k \in \mathbb{Z}^d. \tag{5.7}$$

If v is sufficiently smooth, then the derivatives in space for $\mu \in \{1,\dots,d\}$ are given by

$$\partial_{x_\mu}^r v(t,x) = \sum_{k\in\mathbb{Z}^d} (\mathrm{i}k_\mu)^r\, \widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x}$$

such that space derivatives correspond to multiplications of the Fourier coefficients. Differentiating (5.6) formally gives

$$\partial_t v_j(t,x) = \sum_{k\in\mathbb{Z}^d} \partial_t \widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x} \tag{5.8}$$

and

$$B(\partial)v_j(t,x) = \sum_{k\in\mathbb{Z}^d} \left(\sum_{\mu=1}^d B_\mu(\mathrm{i}k_\mu)\right) \widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x} = \sum_{k\in\mathbb{Z}^d} \mathrm{i}B(k)\widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x}. \tag{5.9}$$

Therefore, we obtain by inserting (5.8) and (5.9) into the left-hand side of the system (5.5)

$$\partial_t v_j + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega, j\kappa)v_j + \frac{1}{\varepsilon}B(\partial)v_j = \sum_{\widetilde{k}\in\mathbb{Z}^d} \left(\partial_t \widehat{v}_j(t,\widetilde{k}) + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega,j\kappa)\widehat{v}_j(t,\widetilde{k}) + \frac{\mathrm{i}}{\varepsilon}B(\widetilde{k})\widehat{v}_j(t,\widetilde{k})\right) e^{\mathrm{i}\widetilde{k}\cdot x}.$$

Since for $m \in \mathbb{Z}^d$

$$(2\pi)^{-d} \int_{\mathbb{T}^d} e^{\mathrm{i}(m-k)\cdot x}\,\mathrm{d}x = \begin{cases} 1, & \text{if } m = k, \\ 0, & \text{else}, \end{cases} \tag{5.10}$$

applying the Fourier transform gives

$$\begin{aligned}
&\mathcal{F}\Big(\partial_t v_j + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega,j\kappa)v_j + \frac{1}{\varepsilon}B(\partial)v_j\Big)(k) \\
&\quad = \sum_{\widetilde{k}\in\mathbb{Z}^d} (2\pi)^{-d} \int_{\mathbb{T}^d} e^{\mathrm{i}(\widetilde{k}-k)\cdot x}\,\mathrm{d}x \left(\partial_t \widehat{v}_j(t,\widetilde{k}) + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(j\omega,j\kappa)\widehat{v}_j(t,\widetilde{k}) + \frac{\mathrm{i}}{\varepsilon}B(\widetilde{k})\widehat{v}_j(t,\widetilde{k})\right) \\
&\quad = \partial_t \widehat{v}_j(t,k) + \frac{\mathrm{i}}{\varepsilon^2}\widetilde{\mathcal{L}}_j(\varepsilon k)\widehat{v}_j(t,k)
\end{aligned}$$

with the shorthand notation

$$\widetilde{\mathcal{L}}_j(\theta) := -j\omega I - c_g\cdot\theta + A(j\kappa+\theta) - \mathrm{i}E. \tag{5.11}$$

We note here the difference to the definition of $\mathcal{L}_j$ in Section 3.5, cf. (3.22). Because of the co-moving coordinate system we obtain the additional term $-c_g\cdot\theta$. We emphasize that by Remark 3.2.5 the smoothness property from Assumption 3.2.2 also holds for the matrix $\widetilde{\mathcal{L}}_j(\theta)$ with $\theta \in \mathbb{R}^d\backslash\{0\}$.
Next, we derive the Fourier transform of the nonlinearity T, where in comparison to Section 3.5 we have the notation $K = \big(k^{(1)},k^{(2)},k^{(3)}\big) \in \mathbb{Z}^d\times\mathbb{Z}^d\times\mathbb{Z}^d$, $\#K = k^{(1)}+k^{(2)}+k^{(3)} \in \mathbb{Z}^d$ and sums instead of integrals. The trilinear nonlinearity of (5.5) for functions (5.6) is given by

$$\begin{aligned}
T(v_{j_1},v_{j_2},v_{j_3})(x) &= \sum_{k^{(1)}\in\mathbb{Z}^d}\sum_{k^{(2)}\in\mathbb{Z}^d}\sum_{k^{(3)}\in\mathbb{Z}^d} e^{\mathrm{i}(k^{(1)}+k^{(2)}+k^{(3)})\cdot x}\, T(\widehat{v}_{j_1}(k^{(1)}),\widehat{v}_{j_2}(k^{(2)}),\widehat{v}_{j_3}(k^{(3)})) \\
&= \sum_{\#K\in\mathbb{Z}^d}\sum_{k^{(1)}+k^{(2)}+k^{(3)}=\#K} e^{\mathrm{i}\#K\cdot x}\, T(\widehat{v}_{j_1}(k^{(1)}),\widehat{v}_{j_2}(k^{(2)}),\widehat{v}_{j_3}(k^{(3)})).
\end{aligned}$$

With (5.10) we obtain

$$\begin{aligned}\mathcal{F}\Big(T(v_{j_1},v_{j_2},v_{j_3})\Big)(k) &= (2\pi)^{-d}\sum_{k^{(1)}\in\mathbb{Z}^d}\sum_{k^{(2)}\in\mathbb{Z}^d}\sum_{k^{(3)}\in\mathbb{Z}^d}\int_{\mathbb{T}^d} e^{\mathrm{i}(\#K-k)\cdot x}\,\mathrm{d}x\,T(\widehat{v}_{j_1}(k^{(1)}),\widehat{v}_{j_2}(k^{(2)}),\widehat{v}_{j_3}(k^{(3)}))\\ &= \sum_{\#K=k} T(\widehat{v}_{j_1}(k^{(1)}),\widehat{v}_{j_2}(k^{(2)}),\widehat{v}_{j_3}(k^{(3)}))\\ &=: \mathcal{T}\left(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3}\right)(k).\end{aligned} \tag{5.12}$$

Thus, we have for every $j\in\mathcal{J}_+$ a system with infinitely many ODEs

$$\partial_t\widehat{v}_j(t,k)+\frac{\mathrm{i}}{\varepsilon^2}\widetilde{\mathcal{L}}_j(\varepsilon k)\widehat{v}_j(t,k)=\sum_{\#J=j}\mathcal{T}\big(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3}\big)(t,k),\quad j\in\mathcal{J}_+,\quad t\in(0,t_{\mathrm{end}}],\quad k\in\mathbb{Z}^d \tag{5.13}$$

by equating coefficients for fixed k. We note that the convention $v_{-j}=\overline{v_j}$ implies that

$$\widehat{v}_{-j}(t,k)=\overline{\widehat{v}_j(t,-k)}. \tag{5.14}$$

Hence, a family of functions v_j solves the system (5.5) if and only if their Fourier transforms $\widehat{v}_j$ solve the system (5.13) with initial data

$$\widehat{p}:=\widehat{v}(0)=\left(\widehat{v}_j(0)\right)_{j\in\mathcal{J}_+}.$$

We end this section with an observation on the eigenvalues of the matrix $\widetilde{\mathcal{L}}_1(\varepsilon k)$. In comparison to (3.44), now the general structure of the eigenvalues $\lambda_{1\ell}$ is

$$\lambda_{1\ell}(\theta)=-\left(\omega+c_g(\kappa)\cdot\theta\right)+\omega_\ell\left(\kappa+\theta\right).$$

We already know that because of the dispersion relation (3.4) and Assumption 4.1.1 the eigenvalue $\lambda_{\pm 11}$ is a special case. The enumeration is chosen in such a way that $\omega=\omega_1(\kappa)$ and

$$\lambda_{11}(0)=-\omega+\omega_1(\kappa)=0=\omega-\omega_1(\kappa)=\lambda_{-11}(0).$$

Furthermore, the co-moving coordinate system results in the feature $\nabla\lambda_{\pm 11}(0)=0$. The fact that $\lambda_{\pm 11}(0)$ and $\nabla\lambda_{\pm 11}(0)$ are zero will be used in a Taylor expansion. For this purpose, we calculate the gradient of the eigenvalues $\lambda_{\pm 11}$ with respect to θ. With $\vartheta:=\kappa+\theta\in\mathbb{R}^d$ it follows that

$$\begin{aligned}\partial_{\theta_\mu}\lambda_{11}(\theta)&=-\{c_g(\kappa)\}_\mu+\partial_{\vartheta_\mu}\omega_1\left(\kappa+\theta\right)\underbrace{\frac{\partial\vartheta_\mu}{\partial\theta_\mu}}_{=1},\\ \partial_{\theta_\mu}\lambda_{-11}(\theta)&=-\{c_g(\kappa)\}_\mu-\partial_{\vartheta_\mu}\omega_1\left(\kappa-\theta\right)\underbrace{\frac{\partial\vartheta_\mu}{\partial\theta_\mu}}_{=-1}=-\{c_g(\kappa)\}_\mu+\partial_{\vartheta_\mu}\omega_1\left(\kappa-\theta\right),\end{aligned}$$

where $\{c_g(\kappa)\}_\mu$ denotes the μ-th entry of the vector $c_g(\kappa)$. Thus, we obtain

$$\nabla\lambda_{\pm 11}(\theta)=-c_g(\kappa)+\nabla\omega_1\left(\kappa\pm\theta\right).$$

Furthermore, by definition (3.5) we have $c_g(\kappa)=\nabla\omega(\kappa)=\nabla\omega_1(\kappa)$ and, thus, for $\theta=0$

$$\nabla\lambda_{11}(0)=-c_g(\kappa)+\nabla\omega_1\left(\kappa\right)=0=\nabla\lambda_{-11}(0).$$

Moreover, the second derivative yields

$$\partial_{\theta_\mu}\partial_{\theta_\nu}\lambda_{\pm 11}(\theta) = \pm\partial_{\vartheta\mu}\partial_{\vartheta_\nu}\omega_1\left(\kappa \pm \theta\right)$$

and we obtain

$$\nabla^2\lambda_{\pm 11}(\theta) = \pm\nabla^2\omega_1\left(\kappa \pm \theta\right).$$

One of the techniques which we use later on is a formal Taylor expansion for the eigenvalue $\lambda_{\pm 11}(\varepsilon k)$ for $\varepsilon k \in \mathbb{R}^d$ and, for this reason we write

$$\lambda_{\pm 11}(\varepsilon k) = \lambda_{\pm 11}(0) + \varepsilon k^T\nabla\lambda_{\pm 11}(0) + \varepsilon^2\int_0^1 (1-\vartheta)k^T\nabla^2\lambda_{\pm 11}(\vartheta\varepsilon k)k\,\mathrm{d}\vartheta.$$

Thus, with $\lambda_{\pm 11}(0) = 0$ and $\nabla\lambda_{\pm 11}(0) = 0$ we conclude

$$\lambda_{\pm 11}(\varepsilon k) = \varepsilon^2\int_0^1 (1-\vartheta)k^T\nabla^2\lambda_{\pm 11}(\vartheta\varepsilon k)k\,\mathrm{d}\vartheta. \tag{5.15}$$

With the Assumptions 3.2.2 and 3.7.2 we obtain for all $\varepsilon k \in \mathbb{R}^d$

$$\sum_{k\in\mathbb{Z}^d}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)| \leqslant C\sum_{k\in\mathbb{Z}^d}|k|_1^2. \tag{5.16}$$

Proof of (5.16). We note that because of Assumption 3.2.2, in general the derivatives of $\lambda_{\pm 11}$ are not bounded near the vector $\mp\kappa$. Hence, we decompose

$$\sum_{k\in\mathbb{Z}^d}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)| = \sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1<\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)| + \sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1\geqslant\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)|.$$

For the first term we now use Taylor expansion and obtain with (5.15)

$$\begin{aligned}\sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1<\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)| &= \sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1<\frac{|\kappa|_1}{2\varepsilon}}}\left|\int_0^1(1-\vartheta)k^T\nabla^2\lambda_{\pm 11}(\vartheta\varepsilon k)k\,\mathrm{d}\vartheta\right| \leqslant \sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1<\frac{|\kappa|_1}{2\varepsilon}}}\sup_{\vartheta\in(0,1)}\left|\nabla^2\omega_1(\kappa\pm\vartheta\varepsilon k)\right|_2|k|_1^2\\ &\leqslant C\sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1<\frac{|\kappa|_1}{2\varepsilon}}}|k|_1^2 \leqslant C\sum_{k\in\mathbb{Z}^d}|k|_1^2.\end{aligned}$$

For the second term it follows with $1 \leqslant \frac{2\varepsilon|k|_1}{|\kappa|_1}$, $\lambda_{\pm 11}(0) = 0$ and Assumption 3.7.2 that

$$\begin{aligned}\sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1\geqslant\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)| &= \sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1\geqslant\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon^2}|\lambda_{\pm 11}(\varepsilon k)-\lambda_{\pm 11}(0)| \leqslant C\sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1\geqslant\frac{|\kappa|_1}{2\varepsilon}}}\frac{1}{\varepsilon}|k|_1 \leqslant \frac{2}{|\kappa|_1}C\sum_{\substack{k\in\mathbb{Z}^d\\ |k|_1\geqslant\frac{|\kappa|_1}{2\varepsilon}}}|k|_1^2\\ &\leqslant C\sum_{k\in\mathbb{Z}^d}|k|_1^2.\end{aligned}$$

Combining the two estimates yield (5.16). ∎

5.1.3 Transformation to smoother variables on $\mathbb{T}^d$

Since the matrix (5.11) is Hermitian, a similar eigendecomposition as (3.41) exists where now $\Psi_j(\theta)$ denotes a unitary matrix and $\Lambda_j(\theta)$ is a real diagonal matrix containing the eigenvalues of $\widetilde{\mathcal{L}}_j(\theta)$. As in Chapter 3 the eigenvectors $\psi_{jm}(\theta)$ are orthonormal and the enumeration is chosen in such a way that $\lambda_{11}(0) = 0$. We emphasize that we use the same letter z for the transformed variable as in Section 3.7, however the definition is different. The variables $z_j : \mathbb{R} \times \mathbb{Z}^d \to \mathbb{C}^s$ are obtained by the transformation

$$z_j(t,k) = \widetilde{S}_{j,\varepsilon}(t,k)\widehat{v}_j(t,k),$$

where we define for every $\varepsilon > 0$, $t \geqslant 0$ and $k \in \mathbb{Z}^d$ the matrix

$$\widetilde{S}_{j,\varepsilon}(t,k) = \exp\left(\tfrac{\mathrm{i}t}{\varepsilon^2}\Lambda_j(\varepsilon k)\right)\Psi_j^*(\varepsilon k). \tag{5.17}$$

We note the difference to Chapter 3, where we have $\frac{\mathrm{i}t}{\varepsilon}$ instead of $\frac{\mathrm{i}t}{\varepsilon^2}$ in (3.45). In accordance with (3.45) the matrix $\widetilde{S}_{j,\varepsilon}(t,k) \in \mathbb{C}^{s\times s}$ is unitary for every $t \in \mathbb{R}$ and $k \in \mathbb{Z}^d$. Next, we derive equations of motion for the transformed variables. Analogously to Section 3.7 the dominating term vanishes and we obtain

$$\partial_t z_j(t,k) = \sum_{\#J=j} F_\varepsilon(t,z,J)(k), \quad j \in \mathcal{J}_+, \quad t \in (0,t_{\mathrm{end}}], \quad k \in \mathbb{Z}^d, \tag{5.18}$$

where

$$F_\varepsilon(t,z,J)(k) = \widetilde{S}_{j,\varepsilon}(t,k)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}z_{j_1}, \widetilde{S}^*_{j_2,\varepsilon}z_{j_2}, \widetilde{S}^*_{j_3,\varepsilon}z_{j_3}\right)(t,k) \tag{5.19}$$

$$= \widetilde{S}_{j,\varepsilon}(t,k)\mathcal{T}\left(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3}\right)(t,k). \tag{5.20}$$

As we will see later, we need both notations (5.19) and (5.20), depending on what we are looking at. The difference to Section 3.7 is that because of the transformation in time the ε in front of the nonlinear part vanishes. The entries of z_j still oscillate with a much smaller amplitude than the entries of $\widehat{v}_j$, since the right-hand side of (5.18) is formally $\mathcal{O}(1)$ instead of $\mathcal{O}\big(\varepsilon^{-2}\big)$ in (5.13). This fact is our main motivation for considering the transformed variables, and it will be very advantageous in our analysis below.

5.1.4 Analytical setting on $\mathbb{T}^d$

So far, we have performed several transformations to obtain the system (5.18). With regard to the analysis of the system (5.18) and to the investigation of the error behavior of the numerical methods, which we introduce in this thesis, we establish a suitable analytic setting in this section.
In Section 5.1.2, we truncated the full space $\mathbb{R}^d$ and replaced for simplicity $\mathbb{R}^d$ by $\mathbb{T}^d$. Hence, we adapt the Wiener algebra introduced in Section 3.4.

Let $W\left(\mathbb{T}^d\right)$ denote the Wiener algebra on $\mathbb{T}^d$, which is defined as the set of all functions whose Fourier series converge absolutely, i.e.

$$f \in W\left(\mathbb{T}^d\right) \quad \Leftrightarrow \quad \widehat{f} \in \ell^1(\mathbb{Z}^d) \quad \Leftrightarrow \quad \sum_{k\in\mathbb{Z}^d} |\widehat{f}(k)| < \infty$$

with norm

$$\|f\|_{W(\mathbb{T}^d)} = \|\widehat{f}\|_{\ell^1(\mathbb{Z}^d)} = \sum_{k\in\mathbb{Z}^d} |\widehat{f}(k)|,$$

where the k-th Fourier coefficient of f is given by (5.7). Then, for $r \in \mathbb{N}$ and with the definitions (2.7) and (2.8), we define the spaces

$$W^r(\mathbb{T}^d) = \left\{ f \in W(\mathbb{T}^d) : \partial^\alpha f \in W(\mathbb{T}^d) \text{ for all } \alpha \in \mathbb{N}_0^d, |\alpha|_1 \leqslant r \right\},$$
$$\|f\|_{W^r} = \sum_{|\alpha|_1 \leqslant r} \|\partial^\alpha f\|_{W(\mathbb{T}^d)} = \sum_{|\alpha|_1 \leqslant r} \|D^\alpha \widehat{f}\|_{\ell_1}$$

with

$$\partial^\alpha f(x) = \sum_{k \in \mathbb{Z}^d} (\mathrm{i}k_1)^{\alpha_1} \cdots (\mathrm{i}k_d)^{\alpha_d} \widehat{f}(k) e^{\mathrm{i}k\cdot x} = \sum_{k \in \mathbb{Z}^d} \mathrm{i}^{|\alpha|_1} k^\alpha \widehat{f}(k) e^{\mathrm{i}k\cdot x} = \sum_{k \in \mathbb{Z}^d} D^\alpha \widehat{f}(k) e^{\mathrm{i}k\cdot x},$$

such that f is in $W^r(\mathbb{T}^d)$ if and only if for all $|\alpha|_1 \leqslant r$

$$\left((\mathrm{i}k_1)^{\alpha_1} \cdots (\mathrm{i}k_d)^{\alpha_d} \widehat{f}(k) \right)_{k \in \mathbb{Z}^d} \in \ell^1\left(\mathbb{Z}^d\right).$$

Furthermore, we define the norm

$$\|\widehat{f}\|_{\ell^1_r} := \sum_{|\alpha|_1 \leqslant r} \|D^\alpha \widehat{f}\|_{\ell_1}$$

and the corresponding Banach space

$$\ell^1_r := \left\{ z_j \in \mathbb{C}^s \,:\, \|z_j\|_{\ell^1_r} < \infty \right\}.$$

In the following we write for simplicity $\|\cdot\|_W$ instead of $\|\cdot\|_{W(\mathbb{T}^d)}$. However, the same properties as for the Wiener algebra on the full space hold (cf. [7]). This means: the space $\left(W^r(\mathbb{T}^d), \|\cdot\|_W\right)$ is a Banach algebra and the Wiener algebra $W(\mathbb{T}^d)$ is continuously embedded in $L^\infty\left(\mathbb{T}^d\right)$.

Since we consider vectors of length s, we introduce vector-valued versions of the function spaces. Therefore, we replace in the definitions $W^r(\mathbb{T}^d)$ by $W^r(\mathbb{T}^d)^s$ for $r \in \mathbb{N}_0$ and $\ell^1\left(\mathbb{Z}^d\right)$ by $\ell^1\left(\mathbb{Z}^d\right)^s$. The corresponding norm is now given by

$$\|\widehat{f}\|_{\ell^1(\mathbb{Z}^d)^s} := \left\| |\widehat{f}|_2 \right\|_{\ell^1(\mathbb{Z}^d)} = \sum_{k \in \mathbb{Z}^d} |\widehat{f}(k)|_2, \qquad \widehat{f} \in \ell^1\left(\mathbb{Z}^d\right)^s,$$

where $|\cdot|_2$ describes the Euclidean vector norm.

Remark 5.1.2.

(i) For simplicity we write throughout the sections $W(\mathbb{T}^d)$ or W instead of $W(\mathbb{T}^d)^s$ and ℓ^1 instead of $\ell^1\left(\mathbb{Z}^d\right)^s$. Nevertheless, the quantities of the spaces are still $\mathbb{C}^s$-valued.

(ii) We have for the initial data

$$z(0) \in \ell^1_r \quad \Leftrightarrow \quad \widehat{p} \in \ell^1_r.$$

5.1.5 Preliminary considerations

Until the end of Chapter 5 we set $j_{\max} = 1$. We will derive error bounds for different step-sizes for a one-step and a two-step method as an integrator of our system

$$\partial_t z_1(k) = \sum_{\#J=1} F_\varepsilon(t, z, J)(k), \qquad t \in (0, t_{\mathrm{end}}], \quad k \in \mathbb{Z}^d.$$

In the case where one component of $J = (j_1, j_2, j_3)$ is negative we know that $z_{-1}(t,k) = \overline{z_1(t,-k)}$. In the following, we define the abbreviation $z(t,k) = z_1(t,k)$. The initial data is given by

$$z(0) = \Psi_1^*(\varepsilon k)\widehat{p}(k).$$

We recall that in Chapter 3 we showed local well-posedness of the coefficients u_j in Lemma 3.6.1. Careful inspections of the proofs in Chapter 3 show that all arguments are still true after a transformation in space and time and if we replace $\mathbb{R}^d$ by $\mathbb{T}^d$. Therefore, we assume on the torus:

Assumption 5.1.3. *If $p \in W^r(\mathbb{T}^d)$, $r = 1, 2$, then there exist a time $t_{end} > 0$ independent of ε such that the system (5.3) with $j_{max} = 1$ has a unique classical solution*

$$v_1 \in \bigcap_{i=0}^{r} C^i([0, t_{end}], W^{r-i}(\mathbb{T}^d)).$$

We conclude by continuity that there exist a constant C_r such that

$$\max_{t\in[0,t_{end}]} \|v_1(t)\|_{W^r} = \max_{t\in[0,t_{end}]} \|\widehat{v}_1(t)\|_{\ell^1_r} = \max_{t\in[0,t_{end}]} \|z_1(t)\|_{\ell^1_r} \leqslant C_r. \tag{5.21}$$

The constant C_r for $r = 1, 2$, depends only on t_{end}, C_T and on $\|p\|_{W^r}$, but not on ε.

Next, we state the helpful results needed for the error analysis later, without proving them again. In comparison to Chapter 3 the difference in the proofs is that instead of the L^1-norm we now consider the ℓ^1-norm and thus, roughly speaking, we have to replace the integrals in the proofs by sums.

As in Section 5.1.3 let $\widehat{v}_1$ be the function obtained by applying the inverse transformation $\widetilde{S}^*_{1,\varepsilon}$ to z_1, meaning

$$\widehat{v}_1(t,k) = \widetilde{S}^*_{1,\varepsilon}(t,k) z_1(t,k), \tag{5.22}$$

where $\widehat{v}_1$ solves the evolution equation

$$\partial_t \widehat{v}_1(t,k) + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}_1(\varepsilon k)\widehat{v}_1(t,k) = \sum_{\#J=1} \mathcal{T}(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3})(t,k), \quad t \in [0, t_{\mathrm{end}}], \quad k \in \mathbb{Z}^d. \tag{5.23}$$

The inverse transformation (5.22) is equivalent to

$$\widehat{v}_1(t,k) = \sum_{m=1}^{s} \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{1m}(\varepsilon k)\right) z_{1m}(t,k)\psi_{1m}(\varepsilon k). \tag{5.24}$$

Moreover, (5.24) together with the definition (5.12) of $\mathcal{T}$ yield

$$\mathcal{T}\left(\widehat{v}_{j_1}, \widehat{v}_{j_2}, \widehat{v}_{j_3}\right)(t,k) = \sum_{M} \sum_{\#K=k} \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{JM}(\varepsilon K)\right) Z_{JM}(t,K) T\left(\psi_{JM}(\varepsilon K)\right)$$

with summation over all $M = (m_1, m_2, m_3)$ and the notation (3.52). The difference is that now we have $K = \left(k^{(1)}, k^{(2)}, k^{(3)}\right) \in \mathbb{Z}^d \times \mathbb{Z}^d \times \mathbb{Z}^d$.

We start this section with some helpful (in)equalities.

Equivalences. Since $\widetilde{S}^*_{j,\varepsilon}$ is unitary the relation

$$\|\widehat{v}_j(t)\|_{\ell^1} = \sum_{k\in\mathbb{Z}^d} |\widetilde{S}^*_{j,\varepsilon}(t,k) z_j(t,k)|_2 = \sum_{k\in\mathbb{Z}^d} |z_j(t,k)|_2 = \|z_j(t)\|_{\ell^1} \tag{5.25}$$

is satisfied.

Estimates. Estimating products of infinite vector-sequences plays an important role in this section. In these estimates, we frequently employ the Banach algebra structure of ℓ^1_r.
Since the proofs of the following estimates are analogous to the proofs in Chapter 3, we give only the results. The minor difference is the norm in which we prove these results, where in comparison to Chapter 3, we have to replace integrals with sums. The following bounds for the trilinear nonlinearity $\mathcal{T}$ are analogous to Lemma 3.5.1.

Lemma 5.1.4. *For $f_1, f_2, f_3 \in W^r(\mathbb{T}^d)$ we have*

$$\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3)\right\|_{\ell^1_r} \leqslant C_T \prod_{i=1}^{3} \|\widehat{f}_i\|_{\ell^1_r}, \tag{5.26}$$

where C_T is the constant defined in (2.5).
If additionally $g_1, g_2, g_3 \in W^r(\mathbb{T}^d)$ and if $\|f_i\|_{W^r}$, $\|g_i\|_{W^r} \leqslant C$ for some $C > 0$, it follows that

$$\left\|\mathcal{T}(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3) - \mathcal{T}(\widehat{g}_1, \widehat{g}_2, \widehat{g}_3)\right\|_{\ell^1_r} \leqslant C_T C^2 \sum_{i=1}^{3} \left\|\widehat{f}_i - \widehat{g}_i\right\|_{\ell^1_r}. \tag{5.27}$$

Since $\widetilde{S}_{1,\varepsilon}$ is unitary, the definition (5.19) of F_ε and (5.26) imply the following lemma, which is analogous to Lemma 3.7.1.

Lemma 5.1.5. *Let be $z \in \ell^1_r$ and $J \in \mathcal{J}^3$, then we obtain for all $t \geqslant 0$*

$$\|F_\varepsilon(t, z, J)\|_{\ell^1_r} \leqslant C_T \prod_{i=1}^{3} \|z_{j_i}(t)\|_{\ell^1_r}.$$

With the estimate from Lemma 5.1.5 we conclude the following bound for the time-derivative of z_j.

Lemma 5.1.6. *Let z be the solution of (5.18) with initial data $z(0) \in \ell^1_{r_+}$, $r_+ = \max\{1, r\}$, $r \in \mathbb{N}_0$. Then, we have for all $t \in [0, t_{end}]$*

$$\|\partial_t z(t)\|_{\ell^1_r} \leqslant C, \tag{5.28}$$

where C depends on C_T and C_{r_+} but not on ε.

Proof. We immediately obtain with Lemma 5.1.5 the estimate

$$\|\partial_t z(t)\|_{\ell^1_r} = \Big\| \sum_{\#J=1} F_\varepsilon(t, z, J) \Big\|_{\ell^1_r} \leqslant \sum_{\#J=1} \|F_\varepsilon(t, z, J)\|_{\ell^1_r} \leqslant C_T \sum_{\#J=1} \prod_{i=1}^{3} \|z_{j_i}(t)\|_{\ell^1_r}.$$

Thus, with (5.21) the claim follows directly. ∎

Remark 5.1.7. *Lemma 5.1.6 shows the advantage in considering the smooth variables z_1 instead of $\widehat{v}_1$. By Assumption 5.1.3 adapted to the transformed system (5.5) it follows that $\widehat{v}_1(t) \in \ell^1_2$ and $\partial_t \widehat{v}_1(t) \in \ell^1_1$ for $t \in [0, t_{end}]$ under the condition that $\widehat{p} \in \ell^1_2$. Under the same condition, however, we even have that $z_1(t)$ and $\partial_t z_1(t) \in \ell^1_2$ for $t \in [0, t_{end}]$.*

In the following sections we consider two numerical methods. We start with an explicit one-step method and afterwards we extend the explicit one-step method into an explicit two-step method.

5.2 One-step method

In this section, we consider an explicit one-step method as a first integrator of the system (5.18). First, we explain how we construct the one-step method in Subsection 5.2.1. Then, in Subsection 5.2.2 we state the results regarding the error bounds of the one-step method. We prove in Theorem 5.2.1 that the one-step method is a first-order method. As previously mentioned first-order methods are not satisfactory, however, they often permit valuable insight. At the end of the chapter we illustrate the behavior of the method by numerical examples in Section 5.6.

5.2.1 Construction of the one-step method

The solution of the equation (5.18) at time $t_{n+1} = t_n + \tau$, $t_0 = 0$, $n = 0, 1, \dots$ with time-step-size τ is given by the fundamental theorem of calculus

$$z(t_{n+1}, k) = z(t_n, k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k) \, \mathrm{d}\sigma. \tag{5.29}$$

Since the solution is highly oscillatory in time it is not practicable to use a standard explicit method where one uses a quadrature formula on the whole integrand. Therefore, the key idea is to retain the integral in (5.29) over the highly oscillatory phases which are hidden in $\widetilde{S}_{1,\varepsilon}(\sigma)$ and $\widetilde{S}^*_{j_i,\varepsilon}(\sigma)$ for $i = 1, 2, 3$, cf. (5.19).

Hence, to obtain a first-order method we freeze $z(\sigma)$ in the nonlinearity (5.19) in (5.29) at $\sigma = t_n$, meaning

$$z(t_{n+1}, k) \approx z(t_n, k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma, k) \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma) z_{j_3}(t_n)\right)(k) \, \mathrm{d}\sigma.$$

This motivates the one-step method

$$z^{n+1}(k) = z^n(k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} F_\varepsilon(\sigma, z^n, J)(k) \, \mathrm{d}\sigma. \tag{5.30}$$

In comparison to Lawson methods, cf. [20], which are a variant of exponential integrators only the term $z_{j_i}(\sigma)$ is frozen at time $\sigma = t_n$ in the nonlinearity $\mathcal{T}$, and not the whole expression $\widetilde{S}^*_{j_i,\varepsilon}(\sigma) z_{j_i}(\sigma) = \widehat{v}_{j_i}(\sigma)$ for $i = 1, 2, 3$. The next goal is to state an error bound for the constructed method (5.30).

5.2.2 Error analysis for the one-step method

The global error of the first-order method applied to the our system satisfies the following bound.

Theorem 5.2.1. *Let $z \in C^1([0, t_{end}]; \ell^1) \cap C([0, t_{end}]; \ell^1_1)$ be the solution of (5.18), then for sufficiently small step-sizes τ the global error of the scheme (5.30) is bounded by*

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant C\tau, \qquad \tau n \leqslant t_{end},$$

where C depends on t_{end}, C_T and $\|z(0)\|_{\ell^1_1}$ but not on ε.

Remark 5.2.2. *In Theorem 5.2.1 and in all subsequent theorems, where we state the global errors, we require "sufficiently small step-sizes". The reason is that in order to ensure the boundedness of the numerical solutions in ℓ^1 (cf. Proposition 5.2.6), we have this restriction on the step-size τ.*

In order to prove Theorem 5.2.1, first, we state an error estimate for the local error. We consider the stability and the proof for the global error afterwards. However, before we consider the error estimate of the local error, we require some preliminary work. In comparison to $F_\varepsilon(\sigma, z(\sigma), J) = F_\varepsilon(\sigma, z, J)$ which is defined in (5.19), the nonlinearity evaluated at a constant vector $z(t_n)$ is given by

$$F_\varepsilon(t, z(t_n), J) = \widetilde{S}_{1,\varepsilon}(t)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(t)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(t)z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(t)z_{j_3}(t_n)\right). \tag{5.31}$$

As a final preparation, we investigate the difference

$$F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J).$$

By means of the formulations (5.19) and (5.31), we extend the difference by adding suitable terms similarly to (2.4) and use the fundamental theorem of calculus

$$z_{j_i}(\sigma) = z_{j_i}(t_n) + \int_{t_n}^{\sigma} \partial_t z_{j_i}(\mu)\,\mathrm{d}\mu, \qquad \text{for } i = 1, 2, 3. \tag{5.32}$$

Hence, we obtain

$$\begin{aligned}
F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J) &= \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}z_{j_1}, \widetilde{S}^*_{j_2,\varepsilon}z_{j_2}, \widetilde{S}^*_{j_3,\varepsilon}z_{j_3}\right)(\sigma) \\
&\quad - \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)z_{j_3}(t_n)\right) \\
&= \widetilde{S}_{1,\varepsilon}(\sigma)\Big(\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)\left(z_{j_1}(\sigma) - z_{j_1}(t_n)\right), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)z_{j_2}(\sigma), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)z_{j_3}(\sigma)\right) \\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\left(z_{j_2}(\sigma) - z_{j_2}(t_n)\right), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)z_{j_3}(\sigma)\right) \\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)\left(z_{j_3}(\sigma) - z_{j_3}(t_n)\right)\right)\Big) \\
&= \widetilde{S}_{1,\varepsilon}(\sigma)\Big(\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)\int_{t_n}^{\sigma}\partial_t z_{j_1}(\mu)\,\mathrm{d}\mu, \widehat{v}_{j_2}(\sigma), \widehat{v}_{j_3}(\sigma)\right) \\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma}\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \widehat{v}_{j_3}(\sigma)\right) \\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)\int_{t_n}^{\sigma}\partial_t z_{j_3}(\mu)\,\mathrm{d}\mu\right)\Big),
\end{aligned} \tag{5.33}$$

where for (5.33) we substitute the definition of $\widehat{v}_{j_i}(\sigma)$, cf. (5.22).

In order to prove first-order convergence uniformly in ε, we follow the classical concept of "stability and consistency yields convergence". Inserting the exact solution value $z(t_n, k)$ into the numerical scheme (5.30) and subtracting from the exact solution at time t_{n+1} (5.29) yields the local error (consistency)

$$\begin{aligned}
\delta^{n+1}(k) &:= z(t_n + \tau, k) - z(t_n, k) - \sum_{\#J=1}\int_{t_n}^{t_{n+1}} F_\varepsilon(\sigma, z(t_n), J)(k)\,\mathrm{d}\sigma \\
&= \sum_{\#J=1}\int_{t_n}^{t_{n+1}}\Big[F_\varepsilon(\sigma, z(\sigma), J)(k) - F_\varepsilon(\sigma, z(t_n), J)(k)\Big]\,\mathrm{d}\sigma.
\end{aligned}$$

Now, we state and prove the local error bound of the one-step method.

Proposition 5.2.3 (Local error). *If $z(0) \in \ell^1_1$, then the local error of the one-step method applied to (5.18) satisfies*

$$\|\delta^{n+1}\|_{\ell^1} \leqslant \tau^2 C, \qquad (n+1)\tau \leqslant t_{end},$$

where C depends on C_T, $\|z(0)\|_{\ell^1_1}$, but not on ε.

By definition of the local error we estimate

$$\|\delta^{n+1}\|_{\ell^1} \leqslant \sum_{\#J=1} \Big\| \int_{t_n}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \Big\|_{\ell^1}.$$

Hence, in order to prove the bound of the local error it is sufficient to show the following lemma.

Lemma 5.2.4. *Let be z the exact solution of (5.18) with initial data $z(0) \in \ell^1_1$ and $J \in \mathcal{J}^3$ with $\#J = 1$. Then, we have*

$$\Big\| \int_{t_n}^{t_{n+1}} (F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J))\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \tau^2 C, \qquad (n+1)\tau \leqslant t_{end},$$

where the constant C depends on C_T and C_1 but not on ε.

Proof. We define the short-hand notation $\Gamma_n := [t_n, t_{n+1}]$. Let $\widehat{v}_1$ be defined as in (5.22), whereas for $j_i = -1$ we use the convention (5.14). Since $\widetilde{S}_{1,\varepsilon}$ is unitary, it follows with (5.33) that

$$\begin{aligned}
&\Big\| \int_{t_n}^{t_{n+1}} (F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J))\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\quad \leqslant \int_{t_n}^{t_{n+1}} \Big\| \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t z_{j_1}(\mu)\,\mathrm{d}\mu, \widehat{v}_{j_2}(\sigma), \widehat{v}_{j_3}(\sigma) \right) \Big\|_{\ell^1} \\
&\qquad + \Big\| \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \widehat{v}_{j_3}(\sigma) \right) \Big\|_{\ell^1} \\
&\qquad + \Big\| \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t z_{j_3}(\mu)\,\mathrm{d}\mu \right) \Big\|_{\ell^1} \mathrm{d}\sigma.
\end{aligned}$$

Using the bound for the trilinearity (5.26) for every term, we obtain

$$\begin{aligned}
&\Big\| \int_{t_n}^{t_{n+1}} (F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J))\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\leqslant C_T \Bigg[\int_{t_n}^{t_{n+1}} \Big\| \int_{t_n}^{\sigma} \partial_t z_{j_1}(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \|\widehat{v}_{j_2}(\sigma)\|_{\ell^1} \|\widehat{v}_{j_3}(\sigma)\|_{\ell^1} + \|z_{j_1}(t_n)\|_{\ell^1} \Big\| \int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \|\widehat{v}_{j_3}(\sigma)\|_{\ell^1} \\
&\qquad + \|z_{j_1}(t_n)\|_{\ell^1} \|z_{j_2}(t_n)\|_{\ell^1} \Big\| \int_{t_n}^{\sigma} \partial_t z_{j_3}(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \mathrm{d}\sigma \Bigg] \\
&\leqslant C_T \Bigg[\int_{t_n}^{t_{n+1}} \int_{t_n}^{\sigma} \|\partial_t z_{j_1}(\mu)\|_{\ell^1}\,\mathrm{d}\mu \max_{\sigma\in\Gamma_n} \|\widehat{v}_{j_2}(\sigma)\|_{\ell^1} \max_{\sigma\in\Gamma_n} \|\widehat{v}_{j_3}(\sigma)\|_{\ell^1} \\
&\qquad + \|z_{j_1}(t_n)\|_{\ell^1} \int_{t_n}^{\sigma} \|\partial_t z_{j_2}(\mu)\|_{\ell^1}\,\mathrm{d}\mu \max_{\sigma\in\Gamma_n} \|\widehat{v}_{j_3}(\sigma)\|_{\ell^1} + \|z_{j_1}(t_n)\|_{\ell^1} \|z_{j_2}(t_n)\|_{\ell^1} \int_{t_n}^{\sigma} \|\partial_t z_{j_3}(\mu)\|_{\ell^1}\,\mathrm{d}\mu\,\mathrm{d}\sigma \Bigg] \\
&\leqslant C_T \int_{t_n}^{t_{n+1}} |\sigma - t_n|\,\mathrm{d}\sigma \sum_{m=1}^{3} \max_{\sigma\in\Gamma_n} \|\partial_t z_{j_m}(\sigma)\|_{\ell^1} \prod_{\substack{i=1 \\ i\neq m}}^{3} \max_{\sigma\in\Gamma_n} \|z_{j_i}(\sigma)\|_{\ell^1} \\
&= C_T \frac{\tau^2}{2} \sum_{m=1}^{3} \max_{\sigma\in\Gamma_n} \|\partial_t z_{j_m}(\sigma)\|_{\ell^1} \prod_{\substack{i=1 \\ i\neq m}}^{3} \max_{\sigma\in\Gamma_n} \|z_{j_i}(\sigma)\|_{\ell^1},
\end{aligned}$$

where we use the fact that

$$\int_{t_n}^{t_{n+1}} |\sigma - t_n| \,\mathrm{d}\sigma = \int_{t_n}^{t_{n+1}} (\sigma - t_n) \,\mathrm{d}\sigma = \left[\frac{1}{2}(\sigma - t_n)^2\right]_{t_n}^{t_{n+1}} = \frac{1}{2}\tau^2,$$

relation (5.25) and $\|z_{j_i}(t_n)\|_{\ell^1} \leqslant \max_{\sigma\in\Gamma_n} \|z_{j_i}(\sigma)\|_{\ell^1}$. As a last step, we only substitute the bounds $\max_{t\in[0,t_{\mathrm{end}}]} \|z(t)\|_{\ell^1} \leqslant C_0$ and (5.28), which means $\max_{t\in[0,t_{\mathrm{end}}]} \|\partial_t z(t)\|_{\ell^1} \leqslant \widetilde{C}$, where $\widetilde{C}$ is the constant from Lemma 5.1.6. Therefore, the estimate follows directly. ∎

With the result from Lemma 5.2.4, Proposition 5.2.3 follows directly. Before we state and prove the proposition concerning the stability of the one-step method, we make a few preparations. We denote by $\Phi^n_{\tau,t_\bullet}(f)$ the result of $n \in \mathbb{N}$ steps of the numerical method (5.30) with step-size τ starting at time $t_\bullet$ with initial data f. If $n = 1$, we simply write $\Phi_{\tau,t_\bullet}(f)$ instead of $\Phi^1_{\tau,t_\bullet}(f)$. Moreover, for $n_\star$ and n in $\mathbb{N}$ the relations

$$\Phi^0_{\tau,t_\bullet}(f) = f \quad \text{and} \quad \Phi^n_{\tau,t_{n_\star}}(f) = \Phi_{\tau,t_{n+n_\star-1}}\left(\Phi^{n-1}_{\tau,t_{n_\star}}(f)\right)$$

follow directly from this definition.

Proposition 5.2.5 (Stability)**.** *Let $n \in \mathbb{N}$ with $t_{n+1} = t_n + \tau \leqslant t_{end}$. For f and g in ℓ^1 with $\mathcal{C} := \max\{\|f\|_{\ell^1}, \|g\|_{\ell^1}\}$, the numerical method satisfies*

$$\|\Phi_{\tau,t_n}(f) - \Phi_{\tau,t_n}(g)\|_{\ell^1} \leqslant e^{\tau C}\|f - g\|_{\ell^1},$$

where C depends on C_T and $\mathcal{C}$.

Proof. Inserting f and g in the numerical method yields

$$\Phi_{\tau,t_n}(f) - \Phi_{\tau,t_n}(g) = f - g + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} \left[F_\varepsilon(\sigma, f, J) - F_\varepsilon(\sigma, g, J)\right] \mathrm{d}\sigma.$$

In the following we set $f_1 = f$ and $f_{-1} = \overline{f}$. The same notation holds for g. Let $\widehat{f}_1$ be the function obtained by applying the inverse transformation $\widetilde{S}_{1,\varepsilon}$ to f_1, i.e.

$$\widehat{f}_1(t,k) = \Psi_1(\varepsilon k) \exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\Lambda_1(\varepsilon k)\right) f_1(k)$$

with the relation $\|\widehat{f}_1(t)\|_{\ell^1} = \|f_1\|_{\ell^1}$ because $\widetilde{S}^*_{1,\varepsilon}$ is unitary. Then, it follows with (2.4) similarly to the proof of (5.27) that

$$\begin{aligned}
\left\| \int_{t_n}^{t_{n+1}} \left[F_\varepsilon(\sigma, f, J) - F_\varepsilon(\sigma, g, J)\right] \mathrm{d}\sigma \right\|_{\ell^1} &\leqslant \tau \max_{\sigma\in\Gamma_n} \|F_\varepsilon(\sigma, f, J) - F_\varepsilon(\sigma, g, J)\|_{\ell^1} \\
&\leqslant \tau C_T \Big[\|f_{j_1} - g_{j_1}\|_{\ell^1}\|f_{j_2}\|_{\ell^1}\|f_{j_3}\|_{\ell^1} + \|f_{j_2} - g_{j_2}\|_{\ell^1}\|g_{j_1}\|_{\ell^1}\|f_{j_3}\|_{\ell^1} \\
&\quad + \|f_{j_3} - g_{j_3}\|_{\ell^1}\|g_{j_1}\|_{\ell^1}\|g_{j_2}\|_{\ell^1} \Big].
\end{aligned}$$

In total we estimate

$$\begin{aligned}
\|\Phi_{\tau,t_n}(f) - \Phi_{\tau,t_n}(g)\|_{\ell^1} &\leqslant \|f - g\|_{\ell^1} + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} \|F_\varepsilon(\sigma, f, J) - F_\varepsilon(\sigma, g, J)\|_{\ell^1} \,\mathrm{d}\sigma \\
&\leqslant \|f - g\|_{\ell^1} + 3\tau C_T \mathcal{C}^2 \Big[\sum_{i=1}^{3} \|f_{j_i} - g_{j_i}\|_{\ell^1} \Big] \\
&\leqslant \|f - g\|_{\ell^1} + 9\tau C_T \mathcal{C}^2 \|f - g\|_{\ell^1}.
\end{aligned}$$

With the relation $1 + x \leqslant e^x$ we can rewrite this to

$$\|\Phi_{\tau,t_n}(f) - \Phi_{\tau,t_n}(g)\|_{\ell^1} \leqslant e^{9\tau C_T \mathcal{C}^2} \|f - g\|_{\ell^1},$$

and the claim follows. ∎

A crucial ingredient for the error analysis of the first-order numerical method is the boundedness of the numerical scheme in ℓ^1. We state and prove a proposition ensuring this boundedness under suitable conditions.

Proposition 5.2.6. *If $z(0) \in \ell^1_1$ and if z is a solution of (5.18) with initial data $z(0)$. Then, for sufficiently small step-sizes $\tau \in (0, \tau_0]$, where $\tau_0 \in (0, t_{end}]$, the numerical solution z^n stays bounded in ℓ^1 for $n \in \mathbb{N}$ with $\tau n \leqslant t_{end}$.*

Proof. We show this proposition by an induction argument. First, we choose a constant $\mathcal{C}_\star > \mathcal{C} := \max_{t\in[0,t_{\mathrm{end}}]} \|z(t)\|_{\ell^1}$. Clearly, the bound

$$\|\Phi^0_{\tau,t_0}(z(0))\|_{\ell^1} = \|z(0)\|_{\ell^1} \leqslant \mathcal{C}_\star$$

follows. Now, we assume that

$$\|\Phi^{n_\star}_{\tau,t_\ell}(z(t_\ell))\|_{\ell^1} \leqslant \mathcal{C}_\star \quad \text{for all } \ell \in \mathbb{N}_0, \quad n_\star = 0, ..., n-1, \quad \ell + n_\star \leqslant N.$$

For the induction step, we will prove that if the step-size τ is sufficiently small, then

$$\|\Phi^{n}_{\tau,t_\ell}(z(t_\ell))\|_{\ell^1} \leqslant \mathcal{C}_\star \quad \text{for all } \ell \in \mathbb{N}_0, \quad \ell + n \leqslant N. \tag{5.34}$$

Since the argument holds for arbitrary starting times t_ℓ, we assume that $\ell = 0$ with no loss of generality. Representing $\Phi^n_{\tau,t_0}(z(0))$ by the telescoping sum

$$\Phi^n_{\tau,t_0}(z(0)) = z(t_n) + \sum_{m=0}^{n-1} \left(\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1})) \right)$$

allows us to estimate

$$\|\Phi^n_{\tau,t_0}(z(0))\|_{\ell^1} \leqslant \|z(t_n)\|_{\ell^1} + \sum_{m=0}^{n-1} \|\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1}.$$

According to the assumption $\|\Phi^{n_\star}_{\tau,t_\ell}(z(t_\ell))\|_{\ell^1} \leqslant \mathcal{C}_\star$, we now apply the stability result, Proposition 5.2.5, of the first-order numerical method to each summand in the previous equation and obtain for $n-m-1 \geqslant 1$

$$\begin{aligned}\|\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1} &= \|\Phi_{\tau,t_m}\left(\Phi^{n-m-1}_{\tau,t_m}(z(t_m))\right) - \Phi_{\tau,t_m}\left(\Phi^{n-m-2}_{\tau,t_{m+1}}(z(t_{m+1}))\right)\|_{\ell^1} \\ &\leqslant e^{\tau C_T \mathcal{C}_\star^2} \|\Phi^{n-m-1}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-2}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1}.\end{aligned}$$

Applying this procedure recursively, we obtain

$$\begin{aligned}\|\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1} &\leqslant e^{(n-m-1)\tau C_T \mathcal{C}_\star^2} \|\Phi_{\tau,t_m}(z(t_m)) - z(t_{m+1})\|_{\ell^1} \\ &= e^{(n-m-1)\tau C_T \mathcal{C}_\star^2} \|\delta^{m+1}\|_{\ell^1} \leqslant e^{t_{\mathrm{end}} C_T \mathcal{C}_\star^2} \|\delta^{m+1}\|_{\ell^1},\end{aligned}$$

where we used $(n-m-1)\tau \leqslant t_{\text{end}}$ for all n, m.
Using the bound of the local error yields

$$\|\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1} \leqslant e^{t_{\text{end}}C_T\mathcal{C}_\star^2}\tau^2\tilde{C},$$

where $\tilde{C}$ is the constant in the local error bound from the Proposition 5.2.3.
So the estimate of the numerical method after n steps with step-size τ at time $t_0 = 0$ with initial data $z(0)$ can be written as

$$\|\Phi^n_{\tau,t_0}(z(0))\|_{\ell^1} \leqslant \|z(t_n)\|_{\ell^1} + ne^{t_{\text{end}}C_T\mathcal{C}_\star^2}\tau^2\tilde{C} \leqslant \mathcal{C} + e^{t_{\text{end}}C_T\mathcal{C}_\star^2}\tau t_{\text{end}}\tilde{C},$$

since $n\tau \leqslant t_{\text{end}}$.
Hence, if τ is so small that it satisfies the inequality

$$\tau \leqslant \frac{\mathcal{C}_\star - \mathcal{C}}{t_{\text{end}}\tilde{C}}e^{-t_{\text{end}}C_T\mathcal{C}_\star^2} =: \tau_0,$$

then we obtain $\|\Phi^n_{\tau,t_0}(z(0))\|_{\ell^1} \leqslant \mathcal{C}_\star$ as desired. ∎

With these preparations, we are now in a position to show the global error bound by combining the stability result and the local error bound with the classical telescoping sum argument of Lady Windermere's fan.

Proof of Theorem 5.2.1. The first-order bound for the global error in ℓ^1 follows with the telescoping sum

$$\|z^n - z(t_n)\|_{\ell^1} = \|\Phi^n_{\tau,t_0}(z(0)) - z(t_n)\|_{\ell^1} \leqslant \sum_{m=0}^{n-1}\|\Phi^{n-m}_{\tau,t_m}(z(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(z(t_{m+1}))\|_{\ell^1}.$$

Thanks to (5.34) the stability result (Proposition 5.2.5) can be applied repeatedly, which yields

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant e^{t_{\text{end}}C_T\mathcal{C}_\star^2}\sum_{m=0}^{n-1}\|\Phi_{\tau,t_m}(z(t_m)) - z(t_{m+1})\|_{\ell^1}.$$

Finally, we obtain with the local error result (Proposition 5.2.3)

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant ne^{t_{\text{end}}C_T\mathcal{C}_\star^2}\tau^2\tilde{C} \leqslant e^{t_{\text{end}}C_T\mathcal{C}_\star^2}\tau t_{\text{end}}\tilde{C},$$

where $\tilde{C}$ is the constant from the local error. ∎

5.2.3 A naive approach towards second-order methods

The next goal is to construct higher order methods. As mentioned before, to obtain for example a second-order method, it is crucial not to approximate the highly oscillatory integral in (5.29) naively by a quadrature formula like the explicit midpoint method.
If we extend our approach to higher-order methods, we have to expand the exact solution by applying the variation of constants formula or fundamental theorem of calculus recursively. The drawback of this procedure will be seen next.

To obtain a second-order one-step method, the first idea would be not to freeze the nonlinearity in (5.29) under the integral but to use again the fundamental theorem of calculus for each component of the nonlinearity. With the exact solution $z_{j_i}(\sigma)$ be written as (5.32) we would obtain for the exact solution at time t_{n+1} in (5.29)

$$\begin{aligned} z(t_{n+1},k) &= z(t_n,k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k)\,\mathrm{d}\sigma \\ &= z(t_n,k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} F_\varepsilon\left(\sigma, z(t_n) + \int_{t_n}^{\sigma} \partial_t z(\mu)\,\mathrm{d}\mu, J\right)(k)\,\mathrm{d}\sigma \\ &= z(t_n,k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} \Big[F_\varepsilon(\sigma, z(t_n), J)(k) + F_\varepsilon^1(\sigma, t_n, z, J)(k) + h.o.t.\Big]\,\mathrm{d}\sigma, \end{aligned}$$

where

$$\begin{aligned} F_\varepsilon^1(\sigma,t_n,z,J) = \widetilde{S}_{1,\varepsilon}(\sigma)\Big[&\mathcal{T}\Big(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) \sum_{\#J_1=j_1} \int_{t_n}^{\sigma} F_\varepsilon(\mu, z(\mu), J_1)\,\mathrm{d}\mu, \widetilde{S}^*_{j_2,\varepsilon}(\sigma) z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma) z_{j_3}(t_n)\Big) \\ &+ \mathcal{T}\Big(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \sum_{\#J_2=j_2} \int_{t_n}^{\sigma} F_\varepsilon(\mu, z(\mu), J_2)\,\mathrm{d}\mu, \widetilde{S}^*_{j_3,\varepsilon}(\sigma) z_{j_3}(t_n)\Big) \\ &+ \mathcal{T}\Big(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma) \sum_{\#J_3=j_3} \int_{t_n}^{\sigma} F_\varepsilon(\mu, z(\mu), J_1)\,\mathrm{d}\mu\Big)\Big]. \end{aligned}$$

Next, we would freeze $z(\mu)$ in the nonlinearity at $\mu = t_n$ in the term $F_\varepsilon^1(\sigma, t_n, z, J)$ and omit the higher order terms. This would yield

$$z^{n+1}(k) = z^n(k) + \sum_{\#J=1} \int_{t_n}^{t_{n+1}} \Big[F_\varepsilon(\sigma, z^n, J)(k) + F_\varepsilon^1(\sigma, z^n, z^n, J)(k)\Big]\,\mathrm{d}\sigma.$$

However, in order to compute the term $F_\varepsilon^1(\sigma, z^n, z^n, J)$, we would have to evaluate an additional nonlinearity in each step.

Because evaluating the right-hand side of the system (5.18) is rather expensive due to the multiple sum structure, we restrict ourselves to one evaluation in each time-step. This leads towards a two-step method. In the next section, we explain how we construct the two-step method. Following the construction, we state the results of the error bounds of this numerical method. It turns out that our approach does not give a "classical" second-order method. Instead, the error behaviour is special in the sense that we obtain various levels of accuracy for different ranges of the step-size. However, for step-sizes $\tau > \varepsilon$ we have the "classical" second-order convergence.

5.3 Two-step method

In this section, we consider an explicit two-step method as an extension to the one-step method investigated in Section 5.2. The two-step method is constructed in Subsection 5.3.1. Then, in the next section we rewrite the introduced two-step method into an equivalent one-step method. In Subsection 5.3.2 we state the results of the error analysis of the two-step method. We prove in Theorem 5.3.2 two different bounds depending on the assumptions. At the end of the chapter we illustrate the behavior of the method by numerical examples in Section 5.6.

5.3.1 Construction of the two-step method

Recall that the solution of the equation (5.18) at time t_{n+1} with step-size τ is given by (5.29). To obtain a two-step method, we use the same approach as for the one-step method, but now at time t_n. Thus, we have for the exact solution at time t_n

$$z(t_n,k) = z(t_{n-1},k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_n} F_\varepsilon(\sigma, z(\sigma), J)(k)\,\mathrm{d}\sigma. \tag{5.35}$$

Inserting equation (5.35) for $z(t_n)$ into (5.29) yields the following two-step equation for the exact solution at time t_{n+1}

$$z(t_{n+1},k) = z(t_{n-1},k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k)\,\mathrm{d}\sigma. \tag{5.36}$$

An approximation for the exact solution at time t_{n+1} can be obtained by

$$z(t_{n+1},k) \approx z(t_{n-1},k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(t_n), J)(k)\,\mathrm{d}\sigma,$$

where we freeze $z(\sigma)$ in the nonlinearity at the midpoint $\sigma = t_n$. This yields the following two-step method for $n \geqslant 1$

$$z^{n+1}(k) = z^{n-1}(k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z^n, J)(k)\,\mathrm{d}\sigma, \tag{5.37}$$

where for the starting step, i.e $n = 0$, we use the one-step method, see (5.30),

$$z^1(k) = z^0(k) + \sum_{\#J=1} \int_{t_0}^{t_1} F_\varepsilon(\sigma, z^0, J)(k)\,\mathrm{d}\sigma.$$

5.3.2 Equivalent one-step method

In order to state a rigorous error analysis we reformulate the two-step method (5.37) as a one-step method. This allows us to apply the typical strategy "stability and consistency yields convergence", which we already used for the one-step method in Section 5.2.2. Therefore, we obtain with the abbreviations

$$\mathbf{z}^n(k) = \begin{pmatrix} z^n(k) \\ z^{n-1}(k) \end{pmatrix}, \qquad \mathbf{F}(\mathbf{z}^n, t_n, J)(k) = \begin{pmatrix} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z^n, J)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix},$$

$$\mathbf{z}(t_n,k) = \begin{pmatrix} z(t_n,k) \\ z(t_{n-1},k) \end{pmatrix}, \qquad \mathbf{F}(\mathbf{z}, t_n, J)(k) = \begin{pmatrix} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix}, \qquad \mathcal{M} = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

the one-step formulation

$$\mathbf{z}^{n+1}(k) = \mathcal{M}\mathbf{z}^n(k) + \sum_{\#J=1} \mathbf{F}(\mathbf{z}^n, t_n, J)(k)$$

and the exact solution

$$\mathbf{z}(t_{n+1},k) = \mathcal{M}\mathbf{z}(t_n,k) + \sum_{\#J=1} \mathbf{F}(\mathbf{z}, t_n, J)(k).$$

We note that the first argument of $\mathbf{F}(\mathbf{z}, t_n, J)$ is a time-dependent function, whereas the first argument of $\mathbf{F}(\mathbf{z}^n, t_n, J)$ is a constant vector. We interpret $\mathbf{z}^n$ and later $\mathbf{z}(t_n)$ as functions which are constant in time. Inserting the exact solution $\mathbf{z}(t_n, k)$ into the numerical scheme and subtracting from the exact solution at time t_{n+1} yields

$$\begin{aligned}\mathbf{z}(t_{n+1},k) - \mathcal{M}\mathbf{z}(t_n,k) - \sum_{\#J=1} \mathbf{F}(\mathbf{z}(t_n),t_n,J)(k) &= \sum_{\#J=1} \big(\mathbf{F}(\mathbf{z},t_n,J) - \mathbf{F}(\mathbf{z}(t_n),t_n,J)\big)(k) \\ &= \sum_{\#J=1} \begin{pmatrix} \int_{t_{n-1}}^{t_{n+1}} \big(F_\varepsilon(\sigma,z(\sigma),J) - F_\varepsilon(\sigma,z(t_n),J)\big)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix}.\end{aligned}$$

Therefore, we define the local error for $n \geqslant 1$

$$\mathbf{d}^{n+1}(k) = \sum_{\#J=1} \big(\mathbf{F}(\mathbf{z},t_n,J) - \mathbf{F}(\mathbf{z}(t_n),t_n,J)\big)(k). \tag{5.38}$$

Remark 5.3.1. *For the starting step $n = 0$ we have*

$$\mathbf{F}(\mathbf{z}(t_0),t_0,J)(k) = \begin{pmatrix} \int_{t_0}^{t_1} F_\varepsilon(\sigma,z(t_0),J)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix}.$$

5.3.3 Error analysis for the two-step method

The goal of this section is to formulate a theorem for the global error of the numerical scheme. We consider two error estimates: the first error estimate requires the same assumptions as for the one-step method in Section 5.2, however, the order of the accuracy is the same as for the one-step method. For the second error estimate, we need additional assumptions, which lead to the fact that we can increase the accuracy for step-sizes $\tau > \varepsilon$. The price we have to pay for this improvement is a more complicated and more elaborate error analysis. This leads to two different error estimates for the local error, which we consider separately.
The global error of the two-step method applied to the system (5.18) satisfies the following bounds.

Theorem 5.3.2.

a) Let $z \in C^1([0,t_{end}];\ell^1) \cap C([0,t_{end}];\ell^1_1)$ be the solution of (5.18), then for sufficiently small step-sizes τ the global error of the scheme (5.37) is bounded by

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant \tau C, \qquad \tau n \leqslant t_{end}, \qquad \text{for } \tau > 0,$$

where C depends on t_{end}, C_T and $\|z(0)\|_{\ell^1_1}$, but not on ε.

b) Let $z \in C^2([0,t_{end}];\ell^1) \cap C^1([0,t_{end}];\ell^1_1) \cap C([0,t_{end}];\ell^1_2)$ be the solution of (5.18). If Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2, and 4.3.1 hold, then for sufficiently small step-sizes τ the global error of the scheme (5.37) is bounded by

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant \left(\tau^2 + \varepsilon^2\right) C,$$

where C depends on t_{end}, C_T, $\|z(0)\|_{\ell^1_2}$, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

First, we investigate part a) of the global error result. Part b) is elaborate and we postpone the investigation to Section 5.4.

Global error result part a)

We observe that we make no further assumptions for part a) of the global error result, in contrast to part b). The local error bound is of order τ^2. This is exactly the same order as the one-step method we analyzed in Section 5.2. The reason is that the proofs in this subsection are the same with the minor difference that now we have the integral over $[t_{n-1}, t_{n+1}]$ instead of $[t_n, t_{n+1}]$.
We define the following short-hand notation

$$\Gamma_n := \begin{cases} [t_0, t_1], & \text{for } n = 0, \\ [t_{n-1}, t_{n+1}], & \text{for } n \geqslant 1. \end{cases}$$

Proposition 5.3.3 (Local error of order 2)**.** *If $\mathbf{z}(0) \in \ell^1_1$, then the local error of the equivalent one-step method applied to (5.18) satisfies*

$$\|\mathbf{d}^{n+1}\|_{\ell^1} \leqslant \tau^2 C, \qquad (n+1)\tau \leqslant t_{end},$$

where C depends on C_T, $\|z(0)\|_{\ell^1_1}$, but not on ε.

Proof. We set $b = t_{n+1}$ and

$$a = \begin{cases} t_0, & \text{for } n = 0, \\ t_{n-1}, & \text{for } n \geqslant 1. \end{cases}$$

Then, we obtain with (5.38) and Remark 5.3.1 for $n \geqslant 0$

$$\|\mathbf{d}^{n+1}\|_{\ell^1} \leqslant \sum_{\#J=1} \left\|\mathbf{F}(\mathbf{z}, t_n, J) - \mathbf{F}(\mathbf{z}(t_n), t_n, J)\right\|_{\ell^1} = \sum_{\#J=1} \left\| \int_a^b F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \right\|_{\ell^1}.$$

Analogously to the proof of Lemma 5.2.4, with the minor difference that now we have the integral over $[t_{n-1}, t_{n+1}]$ instead of $[t_n, t_{n+1}]$ for $n \geqslant 1$, we bound

$$\begin{aligned} \sum_{\#J=1} \left\| \int_a^b F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \right\|_{\ell^1} &\leqslant C_T \tau^2 \sum_{\#J=1} \sum_{m=1}^{3} \max_{\sigma\in\Gamma_n} \|\partial_t z_{j_m}(\sigma)\|_{\ell^1} \prod_{\substack{i=1 \\ i\neq m}}^{3} \max_{\sigma\in\Gamma_n} \|z_{j_i}(\sigma)\|_{\ell^1} \\ &\leqslant \tau^2 C, \end{aligned}$$

where the constant C depends on C_T and C_1. ∎

From the error analysis of the one-step method in Section 5.2 we already know that this local error result of Proposition 5.3.3 combined with a stability result only leads to first-order convergence of the time integration scheme with a constant independent of ε.

We state two propositions concerning the stability of the two-step method and the boundedness of the numerical scheme in ℓ^1 without proving them in detail.

Proposition 5.3.4 (Stability)**.** *Let $n \in \mathbb{N}_0$ with $t_{n+1} = t_n + \tau \leqslant t_{end}$. For $\mathbf{f}$ and $\mathbf{g} \in \ell^1$ and with $\mathcal{C} := \max\{\|\mathbf{f}\|_{\ell^1}, \|\mathbf{g}\|_{\ell^1}\}$, the numerical method satisfies*

$$\|\Phi_{\tau,t_n}(\mathbf{f}) - \Phi_{\tau,t_n}(\mathbf{g})\|_{\ell^1} \leqslant e^{\tau C} \|\mathbf{f} - \mathbf{g}\|_{\ell^1},$$

where C depends on C_T and $\mathcal{C}$.

Proof. The proof of Proposition 5.2.5 can be adopted to prove stability of the two-step method written as an equivalent one-step method. We note that $|\mathcal{M}|_2 = 1$, such that $\|\mathcal{M}(\mathbf{f}-\mathbf{g})\|_{\ell^1} = \|\mathbf{f}-\mathbf{g}\|_{\ell^1}$ holds for $\mathbf{f}, \mathbf{g} \in \ell^1$. ∎

Proposition 5.3.5. *If $\mathbf{z}(0) \in \ell^1$ and if $\mathbf{z}$ is a solution of the system of envelope equations with initial data $\mathbf{z}(0)$, then for sufficiently small step-size τ the numerical solution $\mathbf{z}^n$ stays bounded in ℓ^1 for $n \in \mathbb{N}$ with $\tau n \leqslant t_{end}$.*

The proof for the boundedness of the numerical solution is analogous to the proof of Proposition 5.2.6 with the difference that we apply Proposition 5.3.3 instead of Proposition 5.2.3. Therefore, we omit the details at this point.

Equipped with the results for the local error, the stability and the boundedness of the two-step method, we prove Theorem 5.3.2 a).

Proof of Theorem 5.3.2 a).

The bound for the global error in ℓ^1 follows from the fact that we can play back the global error from the equivalent one-step method in the larger space to the global error of the two-step method. With the telescoping sum we obtain

$$\begin{aligned}\|z^n - z(t_n)\|_{\ell^1} \leqslant \|\mathbf{z}^n - \mathbf{z}(t_n)\|_{\ell^1} &= \|\Phi^n_{\tau,t_0}(\mathbf{z}(0)) - \mathbf{z}(t_n)\|_{\ell^1} \\ &\leqslant \sum_{m=0}^{n-1} \|\Phi^{n-m}_{\tau,t_m}(\mathbf{z}(t_m)) - \Phi^{n-m-1}_{\tau,t_{m+1}}(\mathbf{z}(t_{m+1}))\|_{\ell^1}.\end{aligned}$$

Thanks to the boundedness of the numerical solution

$$\|\Phi^n_{\tau,t_\ell}(\mathbf{z}(t_\ell))\|_{\ell^1} \leqslant \mathcal{C}_\star \quad \text{for all } \ell \in \mathbb{N}_0, \quad \ell + n \leqslant N,$$

the stability result (Proposition 5.3.4) can be applied repeatedly, which yields

$$\begin{aligned}\|\mathbf{z}^n - \mathbf{z}(t_n)\|_{\ell^1} &\leqslant e^{t_{\text{end}} C_T \mathcal{C}_\star^2} \sum_{m=0}^{n-1} \|\Phi_{\tau,t_m}(\mathbf{z}(t_m)) - \mathbf{z}(t_{m+1})\|_{\ell^1} \\ &= e^{t_{\text{end}} C_T \mathcal{C}_\star^2} \left(\sum_{m=1}^{n-1} \|\Phi_{\tau,t_m}(\mathbf{z}(t_m)) - \mathbf{z}(t_{m+1})\|_{\ell^1} + \|\mathbf{d}^1\|_{\ell^1} \right).\end{aligned}$$

Finally, we obtain with the local error result (Proposition 5.3.3) and the fact that $n\tau \leqslant t_{\text{end}}$

$$\begin{aligned}\|\mathbf{z}^n - \mathbf{z}(t_n)\|_{\ell^1} &\leqslant e^{t_{\text{end}} C_T \mathcal{C}_\star^2} \left((n-1)\tau^2 \tilde{C} + \tau^2 \tilde{C} \right) \\ &\leqslant e^{t_{\text{end}} C_T \mathcal{C}_\star^2} \tau t_{\text{end}} \tilde{C}.\end{aligned}$$

∎

5.4 Proof of part b) of Theorem 5.3.2

For part b) of the global error result we make additional assumptions which allow us to use the techniques of proof from Chapter 4.

Before we state and prove an improved local error result, we make some preconsiderations and recall the results from the analysis part. For the rest of this chapter, we suppose that the Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2, and 4.3.1 are valid. Under these assumptions we proved in Chapter 4 bounds for the coefficient u_1 in Proposition 4.1.4 and 4.2.2, and an error bound for the approximation $\widetilde{\mathbf{u}}^{(1)}$ in Theorem 4.3.4. Careful inspections of the proofs in Chapter 4 show that all arguments are still true after a transformation in space and time and if we replace $\mathbb{R}^d$ by $\mathbb{T}^d$. Therefore, we assume the following on the torus.

Assumption 5.4.1.

(i) The time interval $[0, t_\star]$, where the following bounds in (ii) hold, could, in principle, be smaller than the interval $[0, t_{end}]$, but for the sake of simplicity we assume that $t_\star = t_{end}$.

(ii) If $z(0) \in \ell^1_r$, then we have for $r = 1, 2$

$$\begin{aligned} \max_{t\in[0,t_{end}]} \|Pz_1(t)\|_{\ell^1_r} &\leqslant \max_{t\in[0,t_{end}]} \|z_1(t)\|_{\ell^1_r} \leqslant C_r, \\ \max_{t\in[0,t_{end}]} \|P^\perp z_1(t)\|_{\ell^1_{r-1}} &\leqslant \varepsilon C, \end{aligned} \tag{5.39}$$

where $C > 0$ is a constant which is independent of ε and C_r the constant from Assumption 5.1.3. The estimate (5.39) is equivalent to

$$\max_{t\in[0,t_{end}]} \|\mathcal{P}_\varepsilon^\perp \widehat{v}_1(t)\|_{\ell^1_{r-1}} \leqslant \varepsilon C. \tag{5.40}$$

Assumption 5.4.1 (ii) motivates the definition

$$C_r^\varepsilon := \max_{t\in[0,t_{\mathrm{end}}]} \left\{ \|z_1(t)\|_{\ell^1_r}, \varepsilon^{-1} \|P^\perp z_1(t)\|_{\ell^1} \right\}.$$

Analogously to (3.65) we define the projection which projects a vector-valued function pointwise into the first eigenspace of $\widetilde{\mathcal{L}}_1(\varepsilon k)$. For later use we also need the decompositions

$$\widehat{v}_{\pm1}(t,k) = \mathcal{P}_\varepsilon \widehat{v}_{\pm1}(t,k) + \mathcal{P}_\varepsilon^\perp \widehat{v}_{\pm1}(t,k) = \begin{cases} \mathcal{P}_\varepsilon(k)\widehat{v}_1(t,k) + \mathcal{P}_\varepsilon^\perp(k)\widehat{v}_1(t,k), \\ \overline{\mathcal{P}_\varepsilon(-k)}\widehat{v}_{-1}(t,k) + \overline{\mathcal{P}_\varepsilon^\perp(-k)}\widehat{v}_{-1}(t,k), \end{cases} \tag{5.41}$$

$$z_{\pm1}(t,k) = Pz_{\pm1}(t,k) + P^\perp z_{\pm1}(t,k), \tag{5.42}$$

where the definition of P is given by (3.63). Analogous to (3.67) and (3.69), we also write

$$\mathcal{P}_\varepsilon(k)\widehat{v}_1(t,k) = \widetilde{S}^*_{1,\varepsilon}(t,k) P z_1(t,k) \qquad \text{and} \qquad \mathcal{P}_\varepsilon^\perp(k)\widehat{v}_1(t,k) = \widetilde{S}^*_{1,\varepsilon}(t,k) P^\perp z_1(t,k).$$

Since $\widetilde{S}_{1,\varepsilon}$ is unitary, we obtain

$$\|\mathcal{P}_\varepsilon \widehat{v}_1(t)\|_{\ell^1_r} = \|Pz_1(t)\|_{\ell^1_r}, \tag{5.43}$$

$$\|\mathcal{P}_\varepsilon^\perp \widehat{v}_1(t)\|_{\ell^1_r} = \|P^\perp z_1(t)\|_{\ell^1_r}. \tag{5.44}$$

Furthermore, with the decomposition (5.41) we obtain for (5.20) in general by the trilinearity of the nonlinearity

$$\widetilde{S}_{1,\varepsilon}(t)\mathcal{T}\left(\widehat{v}_{j_1}(t),\widehat{v}_{j_2}(t),\widehat{v}_{j_3}(t)\right)=\widetilde{S}_{1,\varepsilon}(t)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t)\right)+\mathcal{N}_1^\perp(t,\widehat{v},J)+\mathcal{N}_2^\perp(t,\widehat{v},J), \tag{5.45}$$

where

$$\begin{aligned}\mathcal{N}_1^\perp(t,\widehat{v},J)=\widetilde{S}_{1,\varepsilon}(t)\Big[&\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_3}(t)\right)\\&+\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t)\right)+\mathcal{T}\left(\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t)\right)\Big],\\ \mathcal{N}_2^\perp(t,\widehat{v},J)=\widetilde{S}_{1,\varepsilon}(t)\Big[&\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_3}(t)\right)+\mathcal{T}\left(\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_3}(t)\right)\\&+\mathcal{T}\left(\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t)\right)+\mathcal{T}\left(\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_1}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_2}(t),\mathcal{P}_\varepsilon^\perp\widehat{v}_{j_3}(t)\right)\Big].\end{aligned}$$

A similar relation holds for (5.19) with the decomposition (5.42). These splittings of the nonlinearity will be helpful later on.

Since $\widetilde{S}_{\pm1,\varepsilon}$ is unitary, the estimate (5.26) combined with (5.43), (5.44) and the bounds (5.39), (5.40) directly yield

$$\|\mathcal{N}_1^\perp(t,\widehat{v},J)\|_{\ell^1}\leqslant 3C_T\|Pz_1(t)\|_{\ell^1}^2\|P^\perp z_1(t)\|_{\ell^1}\leqslant\varepsilon C(C_T,C_0^\varepsilon), \tag{5.46}$$

$$\|\mathcal{N}_2^\perp(t,\widehat{v},J)\|_{\ell^1}\leqslant C_T\left[3\|Pz_1(t)\|_{\ell^1}\|P^\perp z_1(t)\|_{\ell^1}^2+\|P^\perp z_1(t)\|_{\ell^1}^3\right]\leqslant\varepsilon^2C(C_T,C_0^\varepsilon). \tag{5.47}$$

Moreover, we state and prove the following useful result which corresponds to Lemma 4.1.2.

Lemma 5.4.2. *Under the Assumptions 3.2.2 and 3.7.2 we have the following two bounds.*

a) For $\widehat{p}\in\ell_1^1$ there is a constant C such that

$$\max_{t\in[0,t_{end}]}\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(t)\|_{\ell^1}\leqslant\varepsilon^{-1}C, \tag{5.48}$$

where C depends on C_1, C_T and t_{end}, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

b) Furthermore, for $\widehat{p}\in\ell_2^1$, there is a constant such that

$$\max_{t\in[0,t_{end}]}\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(t)\|_{\ell^1}\leqslant C,$$

where C depends on C_2, C_T and t_{end}, but not on ε.

Proof. The evolution equation (5.23) implies for $j=1$ that

$$\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(t,k)=-\tfrac{\mathrm{i}}{\varepsilon^2}\mathcal{P}_\varepsilon\mathcal{L}_1(\varepsilon k)\widehat{v}_1(t,k)+\sum_{\#J=1}\mathcal{P}_\varepsilon\mathcal{T}\left(\widehat{v}_{j_1},\widehat{v}_{j_2},\widehat{v}_{j_3}\right)(t,k)$$

and since $\mathcal{P}_\varepsilon\mathcal{L}_1(\varepsilon k)$ is the projection into the first eigenspace of $\mathcal{L}_1(\varepsilon k)$, we obtain with the definition

(5.22) of $\widehat{v}_1$

$$\begin{aligned}\mathcal{L}_1(\varepsilon k)\widehat{v}_1(t,k) &= \Psi_1(\varepsilon k)\Lambda_1(\varepsilon k)\Psi_1^*(\varepsilon k)\Psi_1(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\Lambda_1(\varepsilon k)\right) z_1(t,k)\\ &= \Psi_1(\varepsilon k)\begin{pmatrix}\lambda_{11}(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{11}(\varepsilon k)\right) z_{11}(t,k)\\ \vdots\\ \lambda_{1s}(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{1s}(\varepsilon k)\right) z_{1s}(t,k)\end{pmatrix}\\ &= \sum_{\ell=1}^{s}\lambda_{1\ell}(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{1\ell}(\varepsilon k)\right) z_{1\ell}(t,k)\psi_{1\ell}(\varepsilon k),\end{aligned}$$

such that

$$\begin{aligned}\tfrac{\mathrm{i}}{\varepsilon^2}\mathcal{P}_\varepsilon(k)\mathcal{L}_1(\varepsilon k)\widehat{v}_1(t,k) &= \tfrac{\mathrm{i}}{\varepsilon^2}\psi_{11}(\varepsilon k)\psi_{11}^*(\varepsilon k)\sum_{\ell=1}^{s}\lambda_{1\ell}(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{1\ell}(\varepsilon k)\right) z_{1\ell}(t,k)\psi_{1\ell}(\varepsilon k)\\ &= \tfrac{\mathrm{i}}{\varepsilon^2}\psi_{11}(\varepsilon k)\lambda_{11}(\varepsilon k)\exp\left(-\tfrac{\mathrm{i}t}{\varepsilon^2}\lambda_{11}(\varepsilon k)\right) z_{11}(t,k)\\ &= \tfrac{\mathrm{i}}{\varepsilon^2}\lambda_{11}(\varepsilon k)\mathcal{P}_\varepsilon(k)\widehat{v}_1(t,k).\end{aligned}$$

a) Since by definition $\lambda_{11}(0) = 0$, the Lipschitz continuity of the eigenvalues 3.7.2 implies that

$$\tfrac{1}{\varepsilon^2}|\lambda_{11}(\varepsilon k)| = \tfrac{1}{\varepsilon^2}|\lambda_{11}(\varepsilon k) - \lambda_{11}(0)| \leqslant \tfrac{1}{\varepsilon}C|k|_1. \tag{5.49}$$

Hence, one factor ε^{-1} is exchanged for a factor $|k|_1$. The nonlinear term is bounded by (5.26). Thus, we have with

$$\sum_{k\in\mathbb{Z}^d}|k|_1\left|\widehat{v}_1(k)\right|_2 = \sum_{k\in\mathbb{Z}^d}\sum_{i=1}^{d}|k_i|\left|\widehat{v}_1(k)\right|_2 = \sum_{i=1}^{d}\sum_{k\in\mathbb{Z}^d}\left|k_i\widehat{v}_1(k)\right|_2 \leqslant \|\widehat{v}_1\|_{\ell^1_1}$$

that

$$\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(t)\|_{\ell^1} \leqslant \frac{C}{\varepsilon}\|\mathcal{P}_\varepsilon\widehat{v}_1(t)\|_{\ell^1_1} + C_T\sum_{\#J=1}\prod_{i=1}^{3}\|\widehat{v}_{j_i}(t)\|_{\ell^1}$$

and the claimed estimate follows.

b) Assumptions 3.2.2 and 3.7.2 imply (5.16) instead of (5.49), and hence, two factors ε^{-1} are exchanged for two factors $|k|_1$. The nonlinear term is bounded by (5.26) as before and we use

$$\sum_{k\in\mathbb{Z}^d}|k|_1^2\left|\widehat{v}_1(k)\right|_2 \leqslant 2\sum_{k\in\mathbb{Z}^d}\sum_{|\alpha|=2}|k^\alpha|\left|\widehat{v}_1(k)\right|_2 = 2\sum_{|\alpha|=2}\sum_{k\in\mathbb{Z}^d}\left|k_1^{\alpha_1}\cdots k_d^{\alpha_d}\widehat{v}_1(k)\right|_2 \leqslant 2\|\widehat{v}_1\|_{\ell^1_2} \tag{5.50}$$

to prove the assertion. ∎

Remark 5.4.3. *As already indicated in the lemma, the difference of both estimates is the regularity assumptions that are made. At first glance, the estimate in a) looks unfavorable, due to the factor ε^{-1} on the right-hand side of (5.48). But we will see later that this factor is not critical in some cases and the advantage is the lower regularity assumption.*

Furthermore, we state a lemma where we bound the difference $P\left(\partial_t z_1(\mu) - \partial_t z_1(t_n)\right)$, $\mu \in \Gamma_n$, in a suitable way, where P is the projection (3.63) which sets the first entry of a vector to zero.

Lemma 5.4.4. *Let be z the solution of (5.18) with initial data $z(0) \in \ell^1_2$ and $J \in \mathcal{J}^3$. Under the Assumptions 3.2.1, 4.1.1, 3.2.2 and 3.7.2, we have for all $\mu \in \Gamma_n$*

$$\|P\big(\partial_t z_1(\mu) - \partial_t z_1(t_n)\big)\|_{\ell^1} \leqslant (|\mu - t_n| + \varepsilon)\, C,$$

where C depends on C_T and C_2^ε, but not on ε.

Proof. By (5.18) it follows that

$$\partial_t P z_1(t) = \sum_{\#J=1} PF_\varepsilon(t, z, J)$$

and, thus, by definition of the ℓ^1-norm we consider the formulation

$$\|P\big(\partial_t z_1(\mu) - \partial_t z_1(t_n)\big)\|_{\ell^1} \leqslant \sum_{\#J=1} \big\|PF_\varepsilon(\mu, z, J) - PF_\varepsilon(t_n, z(t_n), J)\big\|_{\ell^1}.$$

By definition (5.20) combined with the decomposition (5.41) we obtain (5.45). Furthermore, for $F_\varepsilon(t_n, z(t_n), J)$ we have by definition (5.31) and the decomposition $z_{j_i}(t_n) = Pz_{j_i}(t_n) + P^\perp z_{j_i}(t_n)$ a similar relation. This yields

$$\begin{aligned}
&\sum_{\#J=1} \big\|PF_\varepsilon(\mu, z, J) - PF_\varepsilon(t_n, z(t_n), J)\big\|_{\ell^1} \\
&\quad\leqslant \sum_{\#J=1} \Big[\big\|PF_\varepsilon(\mu, Pz, J) - PF_\varepsilon(t_n, Pz(t_n), J)\big\|_{\ell^1} + \big\|\mathcal{N}(\mu, \widehat{v}, t_n, z(t_n), J)\big\|_{\ell^1}\Big] \\
&\quad= \sum_{\#J=1} \Big[\Big\|\int_{t_n}^{\mu} \partial_t PF_\varepsilon(\nu, Pz, J)\,\mathrm{d}\nu\Big\|_{\ell^1} + \big\|\mathcal{N}(\mu, \widehat{v}, t_n, z(t_n), J)\big\|_{\ell^1}\Big],
\end{aligned}$$

where

$$\mathcal{N}(\mu, \widehat{v}, t_n, z(t_n), J) = P\mathcal{N}_1^\perp(\mu, \widehat{v}, J) + P\mathcal{N}_2^\perp(\mu, \widehat{v}, J) - P\mathcal{N}_1^\perp(t_n, z(t_n), J) - P\mathcal{N}_2^\perp(t_n, z(t_n), J).$$

Every nonlinear term $\mathcal{T}$ containing $\mathcal{P}_\varepsilon^\perp \widehat{v}_{j_i}(\nu)$ or $P^\perp z_{j_i}(t_n)$ is collected in the terms $\mathcal{N}_1^\perp$ and $\mathcal{N}_2^\perp$, cf. (5.43) and (5.44). These terms can be estimated straightforwardly by (5.46) and (5.47), such that

$$\begin{aligned}
\|\mathcal{N}(\mu, \widehat{v}, t_n, z(t_n), J)\|_{\ell^1} &\leqslant \|P\mathcal{N}_1^\perp(\mu, \widehat{v}, J)\|_{\ell^1} + \|P\mathcal{N}_2^\perp(\mu, \widehat{v}, J)\|_{\ell^1} \\
&\quad + \|P\mathcal{N}_1^\perp(t_n, z(t_n), J)\|_{\ell^1} + \|P\mathcal{N}_2^\perp(t_n, z(t_n), J)\|_{\ell^1} \\
&\leqslant \varepsilon C(C_T, C_2^\varepsilon).
\end{aligned}$$

Therefore, so far we have shown the bound

$$\|P\big(\partial_t z_1(\mu) - \partial_t z_1(t_n)\big)\|_{\ell^1} \leqslant \sum_{\#J=1} \Big\|\int_{t_n}^{\mu} \partial_t PF_\varepsilon(\nu, Pz, J)\,\mathrm{d}\nu\Big\|_{\ell^1} + \varepsilon C, \tag{5.51}$$

where C depends on C_2^ε.
Taking the time-derivative of (5.20) for $j = 1$ yields

$$\begin{aligned}
\partial_t F_\varepsilon(t, Pz(t), J) = \tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon k)\widetilde{S}_{1,\varepsilon}(t, k)\mathcal{T}\left(\mathcal{P}_\varepsilon \widehat{v}_{j_1}, \mathcal{P}_\varepsilon \widehat{v}_{j_2}, \mathcal{P}_\varepsilon \widehat{v}_{j_3}\right)(t, k) \\
+ \widetilde{S}_{1,\varepsilon}(t, k)\partial_t \mathcal{T}\left(\mathcal{P}_\varepsilon \widehat{v}_{j_1}, \mathcal{P}_\varepsilon \widehat{v}_{j_2}, \mathcal{P}_\varepsilon \widehat{v}_{j_3}\right)(t, k).
\end{aligned}$$

Thus, we have to bound

$$\left\| \int_{t_n}^{\mu} \partial_t P F_\varepsilon(\nu, Pz, J)\,\mathrm{d}\nu \right\|_{\ell^1} \leqslant \left\| \int_{t_n}^{\mu} \tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon\cdot) P\widetilde{S}_{1,\varepsilon}(\nu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu)\,\mathrm{d}\nu \right\|_{\ell^1} + \left\| \int_{t_n}^{\mu} P\widetilde{S}_{1,\varepsilon}(\nu)\partial_t\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu)\,\mathrm{d}\nu \right\|_{\ell^1}. \tag{5.52}$$

Assumptions 3.2.2 and 3.7.2 imply (5.16) and together with (2.2) we have

$$\sum_{k\in\mathbb{Z}^d} \left|\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon k)Pf\right|_2 = \sum_{k\in\mathbb{Z}^d} \left|\tfrac{\mathrm{i}}{\varepsilon^2}\lambda_{11}(\varepsilon k) f_1\right| \leqslant \sum_{k\in\mathbb{Z}^d} C|k|_1^2 \left|Pf\right|_2, \tag{5.53}$$

since $|f_1| = |Pf|_2$. With (5.50) it follows for the first term of (5.52) that

$$\begin{aligned}
&\left\| \int_{t_n}^{\mu} \tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon\cdot) P\widetilde{S}_{1,\varepsilon}(\nu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu)\,\mathrm{d}\nu \right\|_{\ell^1} \\
&\quad\leqslant \int_{t_n}^{\mu} \left\| \tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon\cdot) P\widetilde{S}_{1,\varepsilon}(\nu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu) \right\|_{\ell^1} \mathrm{d}\nu \\
&\quad\leqslant C\left(C_T\right)|(\mu - t_n)| \max_{\nu\in\Gamma_n} \prod_{i=1}^{3} \|\mathcal{P}_\varepsilon\widehat{v}_{j_i}(\nu)\|_{\ell^1_2} \\
&\quad\leqslant |\mu - t_n| C\left(C_T, C_2\right).
\end{aligned} \tag{5.54}$$

The second term of (5.52) is bounded with Lemma 5.4.2 b) by

$$\begin{aligned}
\left\| \int_{t_n}^{\mu} P\widetilde{S}_{1,\varepsilon}(\nu)\partial_t\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu)\,\mathrm{d}\nu \right\|_{\ell^1} &\leqslant \int_{t_n}^{\mu} \left\| P\widetilde{S}_{1,\varepsilon}(\nu)\partial_t\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_{j_1}, \mathcal{P}_\varepsilon\widehat{v}_{j_2}, \mathcal{P}_\varepsilon\widehat{v}_{j_3}\right)(\nu) \right\|_{\ell^1} \mathrm{d}\nu \\
&\leqslant C_T|\mu - t_n| \max_{\nu\in\Gamma_n} \sum_{i=1}^{3} \|\partial_t\mathcal{P}_\varepsilon\widehat{v}_{j_i}(\nu)\|_{\ell^1} \|\mathcal{P}_\varepsilon\widehat{v}_1(\nu)\|_{\ell^1}^2 \\
&\leqslant |\mu - t_n| C\left(C_T, C_2\right).
\end{aligned} \tag{5.55}$$

Combining the bounds (5.51), (5.54) and (5.55) yield the asserted estimate. ∎

The hope is that with the bounds (5.39), (5.40) and the techniques within their proofs, we can improve Proposition 5.3.3 for step-sizes $\tau > \varepsilon$. We aim for a local error of $\mathcal{O}(\tau\varepsilon^2 + \varepsilon\tau^2 + \tau^3)$.

Proposition 5.4.5 (Refined local error)**.** *If $\mathbf{z}(0) \in \ell^1_2$ and the Assumptions 3.2.1, 4.1.1, 3.2.2 and 3.7.2 hold, then the local error of the equivalent one-step method applied to (5.18) satisfies*

$$\|\mathbf{d}^{n+1}\|_{\ell^1} \leqslant \left(\varepsilon^2\tau + \tau^2\varepsilon + \tau^3\right)\tilde{C}, \qquad (n+1)\tau \leqslant t_{end}, \qquad n \in \mathbb{N},$$

where $\tilde{C}$ depends on C_T, $\|z(0)\|_{\ell^1_2}$, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.
For $n = 0$ we have

$$\|\mathbf{d}^1\|_{\ell^1} \leqslant \tau^2\tilde{C},$$

where $\tilde{C}$ is the constant from Proposition 5.3.3.

Remark 5.4.6. *In comparison to Proposition 5.3.3 the local error estimate is refined in the sense that now the estimate $\|\mathbf{d}^{n+1}\|_{\ell^1} \leqslant \tau^3 C$ holds for large step-sizes $\tau > \varepsilon$ and $n \in \mathbb{N}$.*

From the proof of the local error bound of order 2 (Proposition 5.3.3) we know that it is sufficient to show

$$\sum_{\#J=1} \left\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \right\|_{\ell^1} \leqslant \left(\tau^3 + \tau^2\varepsilon + \tau\varepsilon^2\right) C. \tag{5.56}$$

Since the proof of Proposition 5.4.5 is rather lengthy, we subdivide it into several lemmata. The idea is to decompose the difference in (5.56) into several subterms and to investigate them separately. Combining all the corresponding lemmata at the end proves the assertion.

First, we start to reformulate the difference in (5.56). Recall that we have for $J \in \mathcal{J}^3$ with $\#J = 1$ the formulation (5.33). We have that all components of the multi-index satisfy $|j_i| = 1$ for $i = 1, 2, 3$. We consider as an example the second term of (5.33). The idea is to apply the decompositions (5.41) and (5.42) to the first and third argument of the nonlinearity $\mathcal{T}$. The second argument containing the integral remains unchanged which yields

$$\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \widehat{v}_{j_3}(\sigma)\right) = \widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{G}_2(\sigma, t_n, z, J) + \mathcal{R}_2(\sigma, t_n, z, J)\big],$$

where for $\#J = 1$

$$\begin{aligned}
\mathcal{G}_2(\sigma, t_n, z, J) &= \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\right), \qquad (5.57)\\
\mathcal{R}_2(\sigma, t_n, z, J) &= \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P^\perp z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\right)\\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}^\perp_\varepsilon \widehat{v}_{j_3}(\sigma)\right) \qquad (5.58)\\
&\quad + \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P^\perp z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}^\perp_\varepsilon \widehat{v}_{j_3}(\sigma)\right).
\end{aligned}$$

The decomposition into $\mathcal{G}_2$ and $\mathcal{R}_2$ is convenient because in Lemma 5.4.7 we will see that $\mathcal{R}_2$ is a priori small with respect to ε. The other two terms of (5.33) can then be treated in the same way. This leads to

$$\begin{aligned}
F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J) &= \widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{G}_1(\sigma, t_n, z, J) + \mathcal{G}_2(\sigma, t_n, z, J) + \mathcal{G}_3(\sigma, t_n, z, J)\big] \qquad (5.59)\\
&\quad + \widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{R}_1(\sigma, t_n, z, J) + \mathcal{R}_2(\sigma, t_n, z, J) + \mathcal{R}_3(\sigma, t_n, z, J)\big],
\end{aligned}$$

where we have (5.57) and

$$\begin{aligned}
\mathcal{G}_1(\sigma, t_n, z, J) &= \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_1}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_2}(\sigma), \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\right),\\
\mathcal{G}_3(\sigma, t_n, z, J) &= \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) P z_{j_2}(t_n), \widetilde{S}^*_{j_3,\varepsilon}(\sigma)\int_{t_n}^{\sigma} \partial_t z_{j_3}(\mu)\,\mathrm{d}\mu\right).
\end{aligned}$$

The terms $\mathcal{R}_i(\sigma, t_n, z, J)$ for $i = 1, 3$ are similar to $\mathcal{R}_2(\sigma, t_n, z, J)$.
In the first lemma, we investigate the terms $\mathcal{R}_i(\sigma, t_n, z, J)$ for $i = 1, 2, 3$, because they are easy to treat. Using the a priori bounds (5.39) and (5.40), these terms already contain parts of $\mathcal{O}(\varepsilon)$, which directly leads to the required estimate of the form (5.56).

Lemma 5.4.7. *Under the assumptions of Proposition 5.4.5 we have for all $n \geqslant 1$*

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{R}_i(\sigma, t_n, z, J) \, \mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \varepsilon \tau^2 C, \qquad \text{for } i = 1,2,3 \quad \text{and} \quad \#J = 1,$$

where the constant C depend on C_T and C_2^ε, but not on ε.

Proof. We consider as an example $\mathcal{R}_2$, cf. (5.58). The other two terms $\mathcal{R}_i$ for $i = 1, 3$ can then be treated in the same way.
We observe that all terms $\mathcal{T}$ in (5.58) contain at least one term $P^\perp z_{j_i}$ or $\mathcal{P}_\varepsilon^\perp \widehat{v}_{j_i}$. Hence, they can be bounded in the following way. As an example, we bound the first term of (5.58) by

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{T} \left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P^\perp z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t P z_{j_2}(\mu) \, \mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma) \right) \mathrm{d}\sigma \Big\|_{\ell^1} \\
&\quad \leqslant \int_{t_{n-1}}^{t_{n+1}} \Big\| \mathcal{T} \left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P^\perp z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t P z_{j_2}(\mu) \, \mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma) \right) \Big\|_{\ell^1} \mathrm{d}\sigma \\
&\quad \leqslant C_T \int_{t_{n-1}}^{t_{n+1}} \|P^\perp z_{j_1}(t_n)\|_{\ell^1} \Big\| \int_{t_n}^{\sigma} \partial_t P z_{j_2}(\mu) \, \mathrm{d}\mu \Big\|_{\ell^1} \|\mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\|_{\ell^1} \, \mathrm{d}\sigma \\
&\quad \leqslant C_T \varepsilon \left(\varepsilon^{-1} \|P^\perp z_{j_1}(t_n)\|_{\ell^1} \right) \max_{\sigma \in \Gamma_n} \|P z_{j_3}(\sigma)\|_{\ell^1} \max_{\mu \in \Gamma_n} \|\partial_t P z_{j_2}(\mu)\|_{\ell^1} \int_{t_{n-1}}^{t_{n+1}} |\sigma - t_n| \, \mathrm{d}\sigma \\
&\quad \leqslant \tau^2 \varepsilon C \left(C_T, C_2^\varepsilon \right),
\end{aligned}$$

where we apply (5.26), (5.43), (5.39) and (5.28). Applying this procedure to the remaining terms of (5.58) yields the assertion. ∎

Using (5.59) and applying Lemma 5.4.7, so far we have shown

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J) \, \mathrm{d}\sigma \Big\|_{\ell^1} \qquad (5.60) \\
&\quad \leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma) \big[\mathcal{G}_1(\sigma, t_n, z, J) + \mathcal{G}_2(\sigma, t_n, z, J) + \mathcal{G}_3(\sigma, t_n, z, J) \big] \, \mathrm{d}\sigma \Big\|_{\ell^1} + \tau^2 \varepsilon C \qquad \text{for } \#J = 1.
\end{aligned}$$

The remaining goal is to show that for every single term $\mathcal{G}_i$ with $i = 1, 2, 3$ we can estimate

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_i(\sigma, t_n, z, J) \, \mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \left(\tau^3 + \tau^2 \varepsilon + \tau \varepsilon^2 \right) C \qquad \text{for } \#J = 1.$$

For the rest of the section we only investigate the term $\mathcal{G}_2$. The other two terms $\mathcal{G}_1$ and $\mathcal{G}_3$ can be treated in the same way. The next idea is to apply the decomposition (5.42) to $\partial_t z_{j_i}$ in (5.57). This yields

$$\mathcal{G}_2(\sigma, t_n, z, J) = \mathcal{G}_{2P}(\sigma, t_n, z, J) + \mathcal{G}_{2P^\perp}(\sigma, t_n, z, J),$$

where for $\square \in \{P, P^\perp\}$

$$\mathcal{G}_{2\,\square}(\sigma, t_n, z, J) = \mathcal{T} \left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} \partial_t \, \square \, z_{j_2}(\mu) \, \mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma) \right). \qquad (5.61)$$

Our goal is to prove that

$$\begin{aligned}\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_2(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1} &\leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1} \\ &\quad + \Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P^\perp}(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1} \\ &\leqslant \left(\tau^3+\tau^2\varepsilon+\tau\varepsilon^2\right) C \qquad \text{for } \#J=1.\end{aligned}$$

If we bound the terms

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1} \tag{5.62}$$

and

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P^\perp}(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1}. \tag{5.63}$$

with the same technique as in Lemma 5.2.4, we only obtain a bound of order $\mathcal{O}(\tau^2)$, since $\max_{\sigma\in\Gamma_n}\|Pz_{\pm1}(\sigma)\|_{\ell^1}$ and $\max_{\mu\in\Gamma_n}\|\partial_t z_{\pm1}(\mu)\|_{\ell^1}$ are both of order $\mathcal{O}(1)$ only. Therefore, we consider the terms (5.62) and (5.63) in more detail in order to bound these terms in a suitable way. First, we investigate (5.63). We note that in the second argument of the nonlinearity $\mathcal{T}$ in (5.61) with $\square = P^\perp$, we have the integral of $\partial_t P^\perp z_{j_2}(\mu)$ from $\mu\in[t_n,\sigma]$.

If z is a solution of the problem (5.18), we frequently use the estimate

$$\Big\| \int_{t_n}^{\sigma} \partial_t P^\perp z_{j_2}(\mu)\,\mathrm{d}\mu\Big\|_{\ell^1} = \Big\| \sum_{\#J_2=j_2} \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J_2)\,\mathrm{d}\mu\Big\|_{\ell^1} \leqslant \sum_{\#J_2=j_2} \Big\| \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J_2)\,\mathrm{d}\mu\Big\|_{\ell^1}. \tag{5.64}$$

In the previous procedure we directly took the norm under the integral and have used (5.28), cf. the proof of Lemma 5.4.7. Instead, now we aim to gain powers of ε from the oscillatory behavior of $P^\perp F_\varepsilon(\mu,z(\mu),J)$. In the next lemma, we state a useful bound for the right-hand side of (5.64).

Lemma 5.4.8. *Let be z the exact solution of (5.18) with initial data $z(0)\in\ell^1_1$. Under the Assumptions 3.2.1, 4.1.1, 3.2.2 and 3.7.2, we have for all $n\geqslant 0$ and $\sigma\in\Gamma_n$*

$$\sum_{\#J=1} \Big\| \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J)\,\mathrm{d}\mu\Big\|_{\ell^1} \leqslant C\left(\varepsilon|\sigma-t_n|)+\varepsilon^2\right),$$

where C depends on C_T, C_1^ε, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

Proof. The goal is to gain one power of ε from the oscillatory behavior of $P^\perp F_\varepsilon(\mu,z,J)$. We only consider the multi-index $J=(1,1,-1)$, because the other two permutations can be treated in the same way. For the proof, we use the relation (5.20), where we have $\widehat{v}_{\pm1}$ in the nonlinearity, and obtain

$$\Big\| \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J)\,\mathrm{d}\mu\Big\|_{\ell^1} = \Big\| \int_{t_n}^{\sigma} P^\perp \widetilde{S}_{1,\varepsilon}(\mu)\mathcal{T}\left(\widehat{v}_1,\widehat{v}_1,\widehat{v}_{-1}\right)(\mu)\,\mathrm{d}\mu\Big\|_{\ell^1}.$$

The proof is divided in two steps. First, we take advantage of the bound (5.40). In the second step, we apply integration by parts to obtain the asserted bound.

Step 1. With the decomposition (5.41), we split the nonlinearity under the integral into eight parts which we cluster into three terms, cf. (5.45). We bound every term in $\mathcal{N}_1^\perp(\mu,\widehat{v},J)$ and $\mathcal{N}_2^\perp(\mu,\widehat{v},J)$ containing $\mathcal{P}_\varepsilon^\perp\widehat{v}_{\pm1}$ in a straightforward way. Hence, using

$$\Big\| \int_{t_n}^{\sigma} P^\perp \left(\mathcal{N}_1^\perp(\mu,\widehat{v},J) + \mathcal{N}_2^\perp(\mu,\widehat{v},J)\right) \mathrm{d}\mu \Big\|_{\ell^1} \leqslant \varepsilon|\sigma - t_n|\, C\left(C_T, C_1^\varepsilon\right),$$

we have established the estimate

$$\Big\| \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J)\,\mathrm{d}\mu \Big\|_{\ell^1} \leqslant \Big\| \int_{t_n}^{\sigma} P^\perp \widetilde{S}_{1,\varepsilon}(\mu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} + \varepsilon|\sigma - t_n| C\left(C_T, C_1^\varepsilon\right). \tag{5.65}$$

Step 2. The remaining term of (5.65)

$$\Big\| \int_{t_n}^{\sigma} P^\perp \widetilde{S}_{1,\varepsilon}(\mu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \tag{5.66}$$

has to be treated in a similar way as (4.24) in the proof of Proposition 4.1.4. With $\Delta_1(\varepsilon k) = \Lambda_1(\varepsilon k) - \Lambda_1(0)$, we obtain by the definition (5.17) that

$$P^\perp \widetilde{S}_{1,\varepsilon}(\mu,k) = \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\Lambda_1(0)\right) P^\perp \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\Delta_1(\varepsilon k)\right) \Psi_1^*(\varepsilon k),$$

because $P^\perp$ commutes with every diagonal matrix. Hence, the term (5.66) can be expressed as

$$\Big\| \int_{t_n}^{\sigma} \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\Lambda_1(0)\right) P^\perp f_\varepsilon(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1},$$

where

$$f_\varepsilon(\mu,k) = \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\Delta_1(\varepsilon k)\right) \Psi_1^*(\varepsilon k)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu,k).$$

However, the diagonal matrix $\Lambda_1(0)$ is not invertible because $\lambda_{11}(0) = 0$. This problem is compensated by the projection $P^\perp$. As in Step 2 of the proof of Proposition 4.1.4, we replace the eigenvalue $\lambda_{11}(0)$ by 1 or any other nonzero number and consider a new diagonal matrix $\widetilde{\Lambda}_1(0) = \operatorname{diag}\big(1, \lambda_{12}(0), \dots, \lambda_{1s}(0)\big)$ instead of $\Lambda_1(0)$. This matrix is invertible because $\lambda_{1\ell}(0) \neq 0$ for all $\ell = 2,\dots,s$ by Assumption 4.1.1. Integrating by parts yields

$$\begin{aligned}
\Big\| \int_{t_n}^{\sigma} \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^\perp f_\varepsilon(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} &\leqslant \Big\| \tfrac{\varepsilon^2}{\mathrm{i}} \left(\widetilde{\Lambda}_1(0)\right)^{-1} \Big[\exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^\perp f_\varepsilon(\mu)\Big]_{\mu=t_n}^{\sigma} \Big\|_{\ell^1} \\
&\quad + \Big\| \tfrac{\varepsilon^2}{\mathrm{i}} \left(\widetilde{\Lambda}_1(0)\right)^{-1} \int_{t_n}^{\sigma} \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^\perp \partial_t f_\varepsilon(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \\
&\leqslant C\varepsilon^2\big(\|f_\varepsilon(\sigma)\|_{\ell^1} + \|f_\varepsilon(t_n)\|_{\ell^1}\big) + C\varepsilon^2 \int_{t_n}^{\sigma} \|\partial_t f_\varepsilon(\mu)\|_{\ell^1}\,\mathrm{d}\mu.
\end{aligned}$$

The goal is to show that these terms are uniformly bounded for all $\sigma \in \Gamma_n$. For the first two terms this follows immediately from (5.26), i.e.

$$\|f_\varepsilon(\sigma)\|_{\ell^1} = \|\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\sigma)\|_{\ell^1} \leqslant C_T\|\mathcal{P}_\varepsilon\widehat{v}_1(\sigma)\|_{\ell^1}^3 \leqslant C\left(C_T, C_1\right).$$

It remains to bound the integral term. Since Λ_1 is globally Lipschitz continuous by Assumption 3.7.2 it follows that

$$|\tfrac{\mathrm{i}}{\varepsilon^2}\Delta_1(\varepsilon k)|_2 = \tfrac{1}{\varepsilon^2}|\Lambda_1(\varepsilon k) - \Lambda_1(0)|_2 \leqslant C\tfrac{1}{\varepsilon}|k|_1 \tag{5.67}$$

with a constant which does not depend on ε and k. With

$$\partial_t f_\varepsilon(\mu,k) = \tfrac{\mathrm{i}}{\varepsilon^2}\Delta_1(\varepsilon k) f_\varepsilon(\mu,k) + \exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\Delta_1(\varepsilon k)\right)\Psi_1^*(\varepsilon k)\partial_t \mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_1, \mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu,k)$$

this yields

$$\begin{aligned}
\varepsilon^2\int_{t_n}^{\sigma}\|\partial_t f_\varepsilon(\mu)\|_{\ell^1}\,\mathrm{d}\mu &\leqslant \varepsilon^2|\sigma-t_n|\max_{\mu\in\Gamma_n}\|\partial_t f_\varepsilon(\mu)\|_{\ell^1}\\
&\leqslant \varepsilon|\sigma-t_n|\Bigg(\max_{\mu\in\Gamma_n}\sum_{k\in\mathbb{Z}^d}|k|_1\left|\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu,k)\right|_2\\
&\qquad + \varepsilon\max_{\mu\in\Gamma_n}\sum_{k\in\mathbb{Z}^d}\left|\partial_t\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu,k)\right|_2\Bigg)\\
&\leqslant \varepsilon|\sigma-t_n|C_T\max_{\mu\in\Gamma_n}\left(3\|\mathcal{P}_\varepsilon\widehat{v}_1(\mu)\|_{\ell^1_1}^3 + 3\varepsilon\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(\mu)\|_{\ell^1}\|\mathcal{P}_\varepsilon\widehat{v}_1(\mu)\|_{\ell^1_1}^2\right).
\end{aligned}$$

Since $\varepsilon\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_1(\mu)\|_{\ell^1}$ is uniformly bounded according to Lemma 5.4.2 a), it follows that (5.66) is bounded by

$$\begin{aligned}
\left\|\int_{t_n}^{\sigma} P^\perp\widetilde{S}_{1,\varepsilon}(\mu)\mathcal{T}\left(\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_1,\mathcal{P}_\varepsilon\widehat{v}_{-1}\right)(\mu)\,\mathrm{d}\mu\right\|_{\ell^1} &= \left\|\int_{t_n}^{\sigma}\exp\left(\tfrac{\mathrm{i}\mu}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right)P^\perp f_\varepsilon(\mu)\,\mathrm{d}\mu\right\|_{\ell^1}\\
&\leqslant \varepsilon|\sigma-t_n|C(C_T,C_1).
\end{aligned}$$

Combining this estimate with (5.65), we obtain for all multi-indices J with $\#J=1$ the desired estimate

$$\sum_{\#J=1}\left\|\int_{t_n}^{\sigma}P^\perp F_\varepsilon(\mu,z,J)\,\mathrm{d}\mu\right\|_{\ell^1} \leqslant \left(\varepsilon|\sigma-t_n|+\varepsilon^2\right)C\left(C_T,C_1^\varepsilon\right).$$

■

With this result, we prove that the term (5.63) is $\mathcal{O}\left(\varepsilon\tau^2+\varepsilon^2\tau\right)$, which is better than $\mathcal{O}\left(\tau^2\right)$ for large time-steps $\varepsilon<\tau$.

Lemma 5.4.9. *Under the assumptions of Lemma 5.4.8 we have for all $n\geqslant 1$*

$$\left\|\int_{t_{n-1}}^{t_{n+1}}\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P^\perp}(\sigma,z,z(t_n),J)\,\mathrm{d}\sigma\right\|_{\ell^1} \leqslant \left(\varepsilon\tau^2+\varepsilon^2\tau\right)C, \tag{5.68}$$

where C depends on C_T, C_1^ε, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

Proof. Since the nonlinearity $\mathcal{T}$ in (5.61) with $\square = P^\perp$ contains one term $P^\perp\partial_t z_{\pm1}$, we estimate the

left-hand side of (5.68) in a straightforward way in order to apply Lemma 5.4.8. We obtain

$$\begin{aligned}
\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P^\perp}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} &\leqslant \int_{t_{n-1}}^{t_{n+1}} \|\mathcal{G}_{2P^\perp}(\sigma,t_n,z,J)\|_{\ell^1}\,\mathrm{d}\sigma \\
&\leqslant C_T \int_{t_{n-1}}^{t_{n+1}} \|Pz_{j_1}(t_n)\|_{\ell^1} \Big\| \int_{t_n}^{\sigma} P^\perp \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1} \|\mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\|_{\ell^1}\,\mathrm{d}\sigma \\
&\leqslant C_T \max_{\sigma\in\Gamma_n} \|Pz_1(\sigma)\|_{\ell^1}^2 \int_{t_{n-1}}^{t_{n+1}} \Big\| \int_{t_n}^{\sigma} P^\perp \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu \Big\|_{\ell^1}\,\mathrm{d}\sigma \\
&\leqslant C_T \max_{\sigma\in\Gamma_n} \|Pz_1(\sigma)\|_{\ell^1}^2 \int_{t_{n-1}}^{t_{n+1}} \sum_{\#J_2=j_2} \Big\| \int_{t_n}^{\sigma} P^\perp F_\varepsilon(\mu,z,J_2)\,\mathrm{d}\mu \Big\|_{\ell^1}\,\mathrm{d}\sigma \\
&\leqslant C\left(C_T, C_1^\varepsilon\right) \int_{t_{n-1}}^{t_{n+1}} \left(\varepsilon|\sigma - t_n| + \varepsilon^2\right)\,\mathrm{d}\sigma \\
&\leqslant \left(\tau^2\varepsilon + \varepsilon^2\tau\right) C\left(C_T, C_1^\varepsilon\right),
\end{aligned}$$

where we apply (5.26), (5.43) and Lemma 5.4.8. This yields the assertion and completes the proof. ∎

Therefore, so far we have estimated

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_2(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} + \left(\tau^2\varepsilon + \varepsilon^2\tau\right) C \quad \text{for } \#J = 1,$$

which refines (5.60) for $\#J = 1$ to

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma,z(\sigma),J) - F_\varepsilon(\sigma,z(t_n),J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\quad\leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{G}_{1P}(\sigma,t_n,z,J) + \mathcal{G}_{2P}(\sigma,t_n,z,J) + \mathcal{G}_{3P}(\sigma,t_n,z,J)\big]\,\mathrm{d}\sigma \Big\|_{\ell^1} + \left(\tau^2\varepsilon + \varepsilon^2\tau\right) C.
\end{aligned} \tag{5.69}$$

The elaborate part is to show

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \left(\tau^3 + \tau^2\varepsilon + \tau\varepsilon^2\right) C. \tag{5.70}$$

If (5.70) holds and therefore also for $\mathcal{G}_{1P}$ and $\mathcal{G}_{3P}$, then we have established the estimate

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma,z(\sigma),J) - F_\varepsilon(\sigma,z(t_n),J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant \left(\tau^3 + \tau^2\varepsilon + \tau\varepsilon^2\right) C, \qquad \text{for } \#J = 1$$

and Proposition 5.4.5 follows directly. In order to prove (5.70), we first split

$$\begin{aligned}
\Big\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} &\leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\quad + \Big\| \int_{t_{n-1}}^{t_{n+1}} P^\perp\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1},
\end{aligned} \tag{5.71}$$

because the two terms require different proof techniques. The main challenge is to show that

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant C\left(\varepsilon\tau^2 + \tau^3\right), \quad \text{for } \#J = 1 \tag{5.72}$$

and

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} P^\perp\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant C\left(\varepsilon\tau^2 + \varepsilon^2\tau\right), \quad \text{for } \#J = 1. \tag{5.73}$$

The estimate (5.73) will be considered in Lemma 5.4.10, whereas (5.72) will be investigated in Lemma 5.4.11. Since we can proceed similarly for proving (5.73) as in the proof of Lemma 5.4.8, we state the following result.

Lemma 5.4.10. *Under the assumptions of Lemma 5.4.8, we have for all $n \geqslant 1$*

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} P^{\perp} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, d\sigma \Big\|_{\ell^1} \leqslant C\left(\varepsilon\tau^2 + \varepsilon^2\tau\right), \quad \text{for } \#J = 1,$$

where C depends on C_T, C_1^{ε}, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

Proof. The proof is similar to Step 2 in the proof of Lemma 5.4.8. Because $P^{\perp}$ commutes with every diagonal matrix, we again rewrite

$$\int_{t_{n-1}}^{t_{n+1}} P^{\perp} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, \mathrm{d}\sigma = \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\Lambda_1(0)\right) P^{\perp} f_{\varepsilon}(\sigma, k)\, \mathrm{d}\sigma,$$

where with (5.61) and $\square = P$

$$f_{\varepsilon}(\sigma, k) = \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\Delta_1(\varepsilon k)\right) \Psi_1^*(\varepsilon k) \mathcal{G}_{2P}(\sigma, t_n, z, J)(k). \tag{5.74}$$

The diagonal matrix $\Lambda_1(0)$ is not invertible, but this is compensated by the projection $P^{\perp}$. Hence, we consider the modified matrix $\widetilde{\Lambda}_1(0)$ instead of $\Lambda_1(0)$, which is invertible. Thus, we estimate

$$\begin{aligned}
\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^{\perp} f_{\varepsilon}(\sigma)\, \mathrm{d}\sigma \Big\|_{\ell^1} &\leqslant \Big\| \tfrac{\varepsilon^2}{\mathrm{i}} \left(\widetilde{\Lambda}_1(0)\right)^{-1} \Big[\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^{\perp} f_{\varepsilon}(\sigma) \Big]_{\sigma=t_{n-1}}^{t_{n+1}} \Big\|_{\ell^1} \\
&\quad + \Big\| \tfrac{\varepsilon^2}{\mathrm{i}} \left(\widetilde{\Lambda}_1(0)\right)^{-1} \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\widetilde{\Lambda}_1(0)\right) P^{\perp} \partial_t f_{\varepsilon}(\sigma)\, \mathrm{d}\sigma \Big\|_{\ell^1} \\
&\leqslant C\varepsilon^2 \left(\|f_{\varepsilon}(t_{n-1})\|_{\ell^1} + \|f_{\varepsilon}(t_{n+1})\|_{\ell^1} \right) + C\varepsilon^2 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t f_{\varepsilon}(\sigma)\|_{\ell^1}\, \mathrm{d}\sigma,
\end{aligned}$$

where C depends on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$. The goal is to show that these terms are uniformly bounded. For the first two terms we obtain in a straightforward way for $\sigma \in \Gamma_n$ with (5.28) and (5.43) that

$$\begin{aligned}
\|f_{\varepsilon}(\sigma)\|_{\ell^1} &\leqslant \|\mathcal{G}_{2P}(\sigma, t_n, z, J)\|_{\ell^1} \\
&\leqslant C_T \max_{t \in \Gamma_n} \|P z_1(t)\|_{\ell^1}^2 \int_{t_n}^{\sigma} \|P \partial_t z_1(\mu)\|_{\ell^1}\, \mathrm{d}\mu \\
&\leqslant |\sigma - t_n| C_T \max_{t \in \Gamma_n} \|P z_1(t)\|_{\ell^1}^2 \max_{\mu \in \Gamma_n} \|P \partial_t z_1(\mu)\|_{\ell^1} \\
&\leqslant |\sigma - t_n| C\left(C_T, C_1\right). \qquad (5.75)
\end{aligned}$$

It remains to bound the integral term. Taking the time derivative of (5.74) yields

$$\partial_t f_{\varepsilon}(\sigma, k) = \mathcal{R}_{1,\varepsilon}(\sigma, k) + \mathcal{R}_{2,\varepsilon}(\sigma, k),$$

where

$$\begin{aligned}
\mathcal{R}_{1,\varepsilon}(\sigma, k) &= \tfrac{\mathrm{i}}{\varepsilon^2}\Delta_1(\varepsilon k) f_{\varepsilon}(\sigma, k), \\
\mathcal{R}_{2,\varepsilon}(\sigma, k) &= \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\Delta_1(\varepsilon k)\right) \Psi_1^*(\varepsilon k) \partial_t \mathcal{G}_{2P}(\sigma, t_n, z, J).
\end{aligned}$$

Since Λ_1 is globally Lipschitz continuous, it follows with (5.67) and the product rule for $\mathcal{R}_{1,\varepsilon}$ that

$$\begin{aligned}\varepsilon^2 \int_{t_{n-1}}^{t_{n+1}} \|\mathcal{R}_{1,\varepsilon}(\sigma)\|_{\ell^1} \,\mathrm{d}\sigma &= \int_{t_{n-1}}^{t_{n+1}} \|\Delta_1(\varepsilon\cdot) f_\varepsilon(\sigma)\|_{\ell^1} \,\mathrm{d}\sigma \\ &\leqslant \varepsilon \int_{t_{n-1}}^{t_{n+1}} 3|\sigma - t_n| C_T \max_{t\in\Gamma_n} \|Pz_1(t)\|_{\ell^1_1}^2 \max_{\mu\in\Gamma_n} \|P\partial_t z_1(\mu)\|_{\ell^1_1} \,\mathrm{d}\sigma \\ &\leqslant 3C_T \varepsilon \tau^2 \max_{t\in\Gamma_n} \|Pz_1(t)\|_{\ell^1_1}^2 \max_{\mu\in\Gamma_n} \|P\partial_t z_1(\mu)\|_{\ell^1_1} \\ &\leqslant \varepsilon\tau^2 C\left(C_T, C_1\right), \end{aligned} \tag{5.76}$$

since all $|j_i| = 1$. For the term $\max_{\mu\in\Gamma_n} \|P\partial_t z_1(\mu)\|_{\ell^1_1}$ we use Lemma 5.1.6 with $r = 1$. Next, we bound the term $\mathcal{R}_{2,\varepsilon}$. For $\sigma \in \Gamma_n$ we have with the product rule and the estimate (5.26)

$$\begin{aligned}\|\mathcal{R}_{2,\varepsilon}(\sigma)\|_{\ell^1} &= \|\partial_t \mathcal{G}_{2P}(\sigma, t_n, z, J)\|_{\ell^1} = \left\|\partial_t \mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\right)\right\|_{\ell^1} \\ &\leqslant C_T \Bigg[\|\partial_t \widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n)\|_{\ell^1} \int_{t_n}^{\sigma} \|P\partial_t z_{j_2}(\mu)\|_{\ell^1} \,\mathrm{d}\mu \, \|Pz_{j_3}(\sigma)\|_{\ell^1} \\ &\quad + \|Pz_{j_1}(t_n)\|_{\ell^1} \left\|\partial_t \left(\widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu\right)\right\|_{\ell^1} \|Pz_{j_3}(\sigma)\|_{\ell^1} \\ &\quad + \|Pz_{j_1}(t_n)\|_{\ell^1} \int_{t_n}^{\sigma} \|P\partial_t z_{j_2}(\mu)\|_{\ell^1} \,\mathrm{d}\mu \, \|\partial_t \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\|_{\ell^1} \Bigg]. \end{aligned}$$

Next, we aim to bound every single term separately. With definition (5.17) and $\Lambda_{j_i}(\varepsilon k) = \Lambda_{j_i}(0) + \Delta_{j_i}(\varepsilon k)$, we obtain

$$\partial_t \widetilde{S}^*_{j_i,\varepsilon}(\sigma, k) = -\tfrac{\mathrm{i}}{\varepsilon^2} \Psi_{j_i}(\varepsilon k) \Big[\Lambda_{j_i}(0) + \Delta_{j_i}(\varepsilon k)\Big] \exp\left(-\tfrac{\mathrm{i}\sigma}{\varepsilon^2} \Lambda_{j_i}(\varepsilon k)\right).$$

Together with the product rule this yields

$$\begin{aligned}\partial_\sigma \left(\widetilde{S}^*_{j_i,\varepsilon}(\sigma, k) P \int_{t_n}^{\sigma} \partial_t z_{j_i}(\mu)\,\mathrm{d}\mu\right) &= -\tfrac{\mathrm{i}}{\varepsilon^2} \Psi_{j_i}(\varepsilon k) \Big[\Lambda_{j_i}(0) P + \Delta_{j_i}(\varepsilon k) P\Big] \exp\left(-\tfrac{\mathrm{i}\sigma}{\varepsilon^2} \Lambda_{j_i}(\varepsilon k)\right) \int_{t_n}^{\sigma} \partial_t z_{j_i}(\mu)\,\mathrm{d}\mu \\ &\quad + \widetilde{S}^*_{j_i,\varepsilon}(\sigma, k) P \partial_t z_{j_i}(\sigma), \end{aligned}$$

because P commutes with every diagonal matrix. Furthermore, since $\lambda_{\pm 11}(0) = 0$ we have

$$|\Lambda_{\pm 1}(0) P f|_2 = |\lambda_{\pm 11}(0) f_1| = 0,$$

and with (5.49)

$$\tfrac{1}{\varepsilon^2} \left|\Delta_{\pm 1}(\varepsilon k) P f\right|_2 = \tfrac{1}{\varepsilon^2} |\lambda_{\pm 11}(\varepsilon k) - \lambda_{\pm 11}(0)| \left|Pf\right|_1 \leqslant \tfrac{1}{\varepsilon} C|k|_1 \left|Pf\right|_2.$$

Hence, we bound

$$\begin{aligned}\left\|\partial_t \widetilde{S}^*_{j_i,\varepsilon}(\sigma) P z_{j_i}(t_n)\right\|_{\ell^1} &\leqslant \tfrac{1}{\varepsilon^2} \left\|\Lambda_{j_i}(0) P \exp\left(-\tfrac{\mathrm{i}\sigma}{\varepsilon^2} \Lambda_{j_i}(\varepsilon\cdot)\right) z_{j_i}(t_n)\right\|_{\ell^1} \\ &\quad + \tfrac{1}{\varepsilon^2} \left\|\Delta_{j_i}(\varepsilon\cdot) P \exp\left(-\tfrac{\mathrm{i}\sigma}{\varepsilon^2} \Lambda_{j_i}(\varepsilon\cdot)\right) z_{j_i}(t_n)\right\|_{\ell^1} \\ &\leqslant \tfrac{C}{\varepsilon} \|Pz_{j_i}(t_n)\|_{\ell^1_1} \leqslant \tfrac{C}{\varepsilon} C_1, \end{aligned} \tag{5.77}$$

and

$$\begin{aligned}\left\|\partial_t \left(\widetilde{S}^*_{j_i,\varepsilon}(\sigma, k) P \int_{t_n}^{\sigma} \partial_t z_{j_i}(\mu)\,\mathrm{d}\mu\right)\right\|_{\ell^1} &\leqslant \frac{C}{\varepsilon} |\sigma - t_n| \max_{\mu\in\Gamma_n} \|P\partial_t z_{j_i}(\mu)\|_{\ell^1_1} + \|P\partial_t z_{j_i}(\sigma)\|_{\ell^1} \\ &\leqslant \left(\frac{|\sigma - t_n|}{\varepsilon} + 1\right) C\left(C_T, C_1\right), \end{aligned} \tag{5.78}$$

where we use Lemma 5.1.6 with $r = 0$ and $r = 1$ for the last estimate.
For $\sigma \in \Gamma_n$ we estimate with the bounds (5.77), (5.78) and Lemma 5.1.6 with $r = 0$

$$\begin{aligned}\left\|\partial_t \mathcal{G}_{2P}(\sigma, t_n, z, J)\right\|_{\ell^1} &\leqslant C_T \Bigg[\left\|\partial_t \widetilde{S}^*_{j_1,\varepsilon}(\sigma) P z_{j_1}(t_n)\right\|_{\ell^1} |\sigma - t_n| C\left(C_T, C_0\right) \\ &\quad + \left(C_0\right)^2 \left\|\partial_t \left(\widetilde{S}^*_{j_2,\varepsilon}(\sigma) \int_{t_n}^{\sigma} P \partial_t z_{j_2}(\mu)\, \mathrm{d}\mu \right)\right\|_{\ell^1} \\ &\quad + |\sigma - t_n| C\left(C_T, C_1\right) \left\|\partial_t \mathcal{P}_\varepsilon \widehat{v}_{j_3}(\sigma)\right\|_{\ell^1} \\ &\leqslant \left(\frac{|\sigma - t_n|}{\varepsilon} + 1 \right) C\left(C_T, C_1\right), \end{aligned} \tag{5.79}$$

since $\max_{\sigma \in \Gamma_n} \|\partial_t \mathcal{P}_\varepsilon \widehat{v}_1(\sigma)\|_{\ell^1} \leqslant C\varepsilon^{-1}$.
This yields with (5.76) and (5.79)

$$\begin{aligned}\varepsilon^2 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t f_\varepsilon(\sigma)\|_{\ell^1}\, \mathrm{d}\sigma &\leqslant \int_{t_{n-1}}^{t_{n+1}} \|\Delta_1(\varepsilon\cdot) f_\varepsilon(\sigma)\|_{\ell^1}\, \mathrm{d}\sigma + \varepsilon^2 \int_{t_{n-1}}^{t_{n+1}} \|\partial_t \mathcal{G}_{2P}(\sigma, t_n, z, J)\|_{\ell^1}\, \mathrm{d}\sigma \\ &\leqslant \tau^2 \varepsilon C\left(C_T, C_1\right) + \left(\tau^2 \varepsilon + 2\tau\varepsilon^2\right) C\left(C_T, C_1\right). \end{aligned} \tag{5.80}$$

Thus, combining (5.75) and (5.80) it follows that

$$\begin{aligned}\left\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2} \Lambda_1(0)\right) P^\perp f_\varepsilon(\sigma)\, \mathrm{d}\sigma \right\|_{\ell^1} &\leqslant C\varepsilon^2 \left(\|f_\varepsilon(t_{n-1})\|_{\ell^1} + \|f_\varepsilon(t_{n+1})\|_{\ell^1} + \int_{t_{n-1}}^{t_{n+1}} \|\partial_t f_\varepsilon(\sigma)\|_{\ell^1}\, \mathrm{d}\sigma \right) \\ &\leqslant \left(\varepsilon\tau^2 + \varepsilon^2 \tau\right) C\left(C_T, C_1\right), \end{aligned}$$

which proves Lemma 5.4.10. ∎

Thus, so far we have refined (5.71) to

$$\left\| \int_{t_{n-1}}^{t_{n+1}} \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, \mathrm{d}\sigma \right\|_{\ell^1} \leqslant \left\| \int_{t_{n-1}}^{t_{n+1}} P \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, \mathrm{d}\sigma \right\|_{\ell^1} + \left(\tau^2 \varepsilon + \varepsilon^2 \tau\right) C.$$

The remaining term will be investigated in the last lemmata.

Lemma 5.4.11. *Let be z the exact solution of (5.18) with initial data $z(0) \in \ell^1_2$. Under the Assumptions 4.1.1, 3.2.2 and 3.7.2, we have for all $n \geqslant 1$*

$$\left\| \int_{t_{n-1}}^{t_{n+1}} P \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, \mathrm{d}\sigma \right\|_{\ell^1} \leqslant C\left(\varepsilon\tau^2 + \tau^3\right), \quad \text{for } \#J = 1,$$

where C depends on C_T and C_2^ε but not on ε.

Proof. We remark that the assumptions allow us to apply Lemma 5.4.2 b) and 5.4.4. The proof is different from the proofs of Lemma 5.4.9 and Lemma 5.4.10. In

$$\int_{t_{n-1}}^{t_{n+1}} P \widetilde{S}_{1,\varepsilon}(\sigma) \mathcal{G}_{2P}(\sigma, t_n, z, J)\, \mathrm{d}\sigma \tag{5.81}$$

we have the projection P in front of the transformation $\widetilde{S}_{1,\varepsilon}(\sigma)$ and in every component of the term $\mathcal{G}_{2P}$. Hence, the main idea of this proof is to use a Taylor expansion for the exponential terms and to apply the estimate (5.53) in order to handle the factor ε^{-2}. We divide the proof into several steps.

Step 1. In this step we split the term $P\widetilde{S}_{1,\varepsilon}(\sigma)$ into two terms. For this purpose we expand the transformation $\widetilde{S}_{1,\varepsilon}(\sigma)$ by $\exp\left(\frac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)$. Thus, we obtain

$$P\widetilde{S}_{1,\varepsilon}(\sigma,k) = P\exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)\Psi_1(\varepsilon k)^* = \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\widetilde{S}_{1,\varepsilon}(\sigma - t_n), \tag{5.82}$$

since P commutes with every diagonal matrix. Next, we use that

$$\exp\left(\tfrac{\mathrm{i}(\sigma-t_n)}{\varepsilon^2}\Lambda_j(\varepsilon k)\right) = I + \tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_j(\varepsilon k)\int_{t_n}^{\sigma}\exp\left(\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_j(\varepsilon k)\right)\mathrm{d}\mu. \tag{5.83}$$

Substituting this expansion into the exponential part of $\widetilde{S}_{1,\varepsilon}(\sigma - t_n)$ yields

$$\begin{aligned}\exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\widetilde{S}_{1,\varepsilon}(\sigma-t_n) &= \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\Psi_1^*(\varepsilon k) \\ &\quad + \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon k)P\int_{t_n}^{\sigma}\exp\left(\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)\mathrm{d}\mu\,\Psi_1^*(\varepsilon k).\end{aligned} \tag{5.84}$$

In the following step, we investigate the second term of (5.84) applied on $\mathcal{G}_{2P}$ in (5.81). We aim to bound the resulting integral term by $\mathcal{O}(\tau^3)$. The first term of (5.84) applied on $\mathcal{G}_{2P}$ will again be split into more terms. Again some of those new terms can be bounded by $\mathcal{O}(\tau^3)$.

Step 2. We note that with the estimate (5.53) the factor ε^{-2} cancels in the second term of (5.84). We consider the summands in (5.84) separately. For the second term it follows with (5.53) and (5.50) applied to the nonlinearity that

$$\begin{aligned}&\left\|\int_{t_{n-1}}^{t_{n+1}}\exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right)\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)P\int_{t_n}^{\sigma}\exp\left(\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right)\mathrm{d}\mu\Psi_1^*(\varepsilon\cdot)\mathcal{G}_{2P}(\sigma,t_n,z,J)\mathrm{d}\sigma\right\|_{\ell^1} \\ &\leqslant \int_{t_{n-1}}^{t_{n+1}}\left\|\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)P\int_{t_n}^{\sigma}\exp\left(\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right)\mathrm{d}\mu\Psi_1^*(\varepsilon\cdot)\mathcal{G}_{2P}(\sigma,t_n,z,J)\right\|_{\ell^1}\mathrm{d}\sigma \\ &\leqslant C\int_{t_{n-1}}^{t_{n+1}}|\sigma-t_n|\left\|\mathcal{G}_{2P}(\sigma,t_n,z,J)\right\|_{\ell^1_2}\mathrm{d}\sigma \\ &= C\int_{t_{n-1}}^{t_{n+1}}|\sigma-t_n|\left\|\mathcal{T}\left(\widetilde{S}_{j_1,\varepsilon}^*(\sigma)Pz_{j_1}(t_n),\widetilde{S}_{j_2,\varepsilon}^*(\sigma)\int_{t_n}^{\sigma}P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu,\mathcal{P}_\varepsilon\widehat{v}_{j_3}(\sigma)\right)\right\|_{\ell^1_2}\mathrm{d}\sigma \\ &\leqslant C\left(C_T\right)\int_{t_{n-1}}^{t_{n+1}}|\sigma-t_n|\left\|\int_{t_n}^{\sigma}\partial_t Pz_{j_2}(\mu)\,\mathrm{d}\mu\right\|_{\ell^1_2}\|Pz_{j_3}(\sigma)\|_{\ell^1_2}\mathrm{d}\sigma\,\|Pz_{j_1}(t_n)\|_{\ell^1_2} \\ &\leqslant C\left(C_T,C_2\right)\max_{\mu\in\Gamma_n}\|\partial_t z_1(\mu)\|_{\ell^1_2}\int_{t_{n-1}}^{t_{n+1}}(\sigma-t_n)^2\,\mathrm{d}\sigma \\ &\leqslant \tau^3 C\left(C_T,C_2\right),\end{aligned} \tag{5.85}$$

where we use Lemma 5.1.6 with $r=2$ for $\max\limits_{\mu\in\Gamma_n}\|\partial_t z_1(\mu)\|_{\ell^1_2}$.
The first term of (5.84) applied on $\mathcal{G}_{2P}$ is

$$\begin{aligned}&\exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\Psi_1^*(\varepsilon k)\mathcal{G}_{2P}(\sigma,t_n,z,J) \\ &= \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\Psi_1^*(\varepsilon k)\mathcal{T}\left(\widetilde{S}_{j_1,\varepsilon}^*(\sigma)Pz_{j_1}(t_n),\widetilde{S}_{j_2,\varepsilon}^*(\sigma)\int_{t_n}^{\sigma}P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu,\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)+\int_{t_n}^{\sigma}\mathcal{P}_\varepsilon\partial_t\widehat{v}_{j_3}(\mu)\,\mathrm{d}\mu\right) \\ &= \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right)P\Psi_1^*(\varepsilon k)\Bigg[\mathcal{T}\left(\widetilde{S}_{j_1,\varepsilon}^*(\sigma)Pz_{j_1}(t_n),\widetilde{S}_{j_2,\varepsilon}^*(\sigma)\int_{t_n}^{\sigma}P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu,\mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right) \\ &\qquad + \mathcal{T}\left(\widetilde{S}_{j_1,\varepsilon}^*(\sigma)Pz_{j_1}(t_n),\widetilde{S}_{j_2,\varepsilon}^*(\sigma)\int_{t_n}^{\sigma}P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu,\int_{t_n}^{\sigma}\mathcal{P}_\varepsilon\partial_t\widehat{v}_{j_3}(\mu)\,\mathrm{d}\mu\right)\Bigg],\end{aligned} \tag{5.86}$$

where we use in the last component of the nonlinearity the fundamental theorem of calculus and the linearity. We obtain for the second term in (5.86) with (5.26)

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right) P\Psi_1^*(\varepsilon\cdot)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)Pz_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \int_{t_n}^{\sigma}\mathcal{P}_\varepsilon\partial_t\widehat{v}_{j_3}(\mu)\,\mathrm{d}\mu\right)\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\quad\leqslant C_T\int_{t_{n-1}}^{t_{n+1}}(\sigma-t_n)^2\,\mathrm{d}\sigma\,\|Pz_{j_1}(t_n)\|_{\ell^1}\max_{\mu\in\Gamma_n}\|\partial_t z_{j_2}(\mu)\|_{\ell^1}\max_{\mu\in\Gamma_n}\|\partial_t\mathcal{P}_\varepsilon\widehat{v}_{j_3}(\mu)\|_{\ell^1} \\
&\quad\leqslant \tau^3 C\left(C_T, C_2\right), \qquad (5.87)
\end{aligned}$$

where we use Lemma 5.1.6 with $r=0$ and Lemma 5.4.2 b) for the last inequality.
Thus, so far we estimate

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma,t_n,z,J)\,\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\quad\leqslant \Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right) P\Psi_1^*(\varepsilon\cdot)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)Pz_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right)\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\qquad + C\tau^3.
\end{aligned}$$

In the next two steps we investigate the remaining term.

Step 3. For the first term of (5.86), we now proceed in the same way as in Step 1, but this time we make the expansion (5.82) of the transformation $\widetilde{S}^*_{j_i,\varepsilon}(\sigma)$ for $i=1,2$ by $\exp\left(-\frac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_i}(\varepsilon k^{(i)})\right)$ in the nonlinearity. Next, we substitute the expansion (5.83) for the exponential function into the exponential part of $\widetilde{S}^*_{j_i,\varepsilon}(\sigma-t_n)$ in every component of the nonlinearity. In total, we obtain

$$\begin{aligned}
&\exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right) P\Psi_1^*(\varepsilon k)\mathcal{T}\left(\widetilde{S}^*_{j_1,\varepsilon}(\sigma)Pz_{j_1}(t_n), \widetilde{S}^*_{j_2,\varepsilon}(\sigma)\int_{t_n}^{\sigma} P\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right) \\
&\quad= \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right) P\Psi_1^*(\varepsilon k)\Big[\widetilde{R}_1(\sigma,t_n,z,J)+\widetilde{R}_2(\sigma,t_n,z,J)+\widetilde{R}_3(\sigma,t_n,z,J)\Big],
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{R}_1(\sigma,t_n,z,J) &= \mathcal{T}\left(\mathcal{I}_{j_1}(t_n,z(t_n)), \mathcal{I}_{j_2}(\sigma,t_n,z), \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right), \\
\widetilde{R}_2(\sigma,t_n,z,J) &= \mathcal{T}\left(\widetilde{\mathcal{I}}_{j_1}(\sigma,t_n,z(t_n)), \mathcal{I}_{j_2}(\sigma,t_n,z), \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right) \\
&\quad + \mathcal{T}\left(\mathcal{I}_{j_1}(t_n,z(t_n)), \widetilde{\mathcal{I}}_{j_2}(\sigma,t_n,z), \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right), \\
\widetilde{R}_3(\sigma,t_n,z,J) &= \mathcal{T}\left(\widetilde{\mathcal{I}}_{j_1}(\sigma,t_n,z(t_n)), \widetilde{\mathcal{I}}_{j_2}(\sigma,t_n,z), \mathcal{P}_\varepsilon\widehat{v}_{j_3}(t_n)\right), \\
\mathcal{I}_{j_1}(t_n,z(t_n)) &= \Psi_{j_1}(\varepsilon\cdot)P\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)\right) z_{j_1}(t_n), \\
\mathcal{I}_{j_2}(\sigma,t_n,z) &= \Psi_{j_2}(\varepsilon\cdot)P\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\right)\int_{t_n}^{\sigma}\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu, \\
\widetilde{\mathcal{I}}_{j_1}(\sigma,t_n,z(t_n)) &= -\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)P\int_{t_n}^{\sigma}\exp\left(-\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)\right)\mathrm{d}\mu\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)\right) z_{j_1}(t_n), \\
\widetilde{\mathcal{I}}_{j_2}(\sigma,t_n,z) &= -\tfrac{\mathrm{i}}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)P\int_{t_n}^{\sigma}\exp\left(-\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\right)\mathrm{d}\mu\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\right)\int_{t_n}^{\sigma}\partial_t z_{j_2}(\mu)\,\mathrm{d}\mu.
\end{aligned}$$

The index m of $\widetilde{R}_m$ indicates how many integrals every nonlinear term contains. The goal of this step is to show for $m=2,3$

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right) P\Psi_1^*(\varepsilon\cdot)\widetilde{R}_m(\sigma,t_n,z,J)\mathrm{d}\sigma\Big\|_{\ell^1} \leqslant \tau^3 C.$$

Similarly to Step 2, the factor ε^{-2} in the terms $\widetilde{\mathcal{I}}_{j_1}$ and $\widetilde{\mathcal{I}}_{j_2}$ is canceled by means of the estimate (5.53). Hence, we have the bounds

$$\begin{aligned}
\big\|\mathcal{I}_{j_1}(t_n, z(t_n))\big\|_{\ell^1} &= \|z_{j_1}(t_n)\|_{\ell^1} \\
\big\|\mathcal{I}_{j_2}(\sigma, t_n, z)\big\|_{\ell^1} &\leqslant |\sigma - t_n| \max_{\mu\in\Gamma_n} \|\partial_t z_{j_2}(\mu)\|_{\ell^1}, \\
\big\|\widetilde{\mathcal{I}}_{j_1}(\sigma, t_n, z(t_n))\big\|_{\ell^1} &\leqslant \Big\| \int_{t_n}^{\sigma} \exp\Big(-\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)\Big)\,\mathrm{d}\mu \exp\big(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_1}(\varepsilon\cdot)\big)\, z_{j_1}(t_n)\Big\|_{\ell^1_2} \\
&\leqslant |\sigma - t_n| \, \big\|z_{j_1}(t_n)\big\|_{\ell^1_2}, \\
\big\|\widetilde{\mathcal{I}}_{j_2}(\sigma, t_n, z)\big\|_{\ell^1} &\leqslant \Big\| \int_{t_n}^{\sigma} \exp\Big(-\tfrac{\mathrm{i}(\mu-t_n)}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\Big)\,\mathrm{d}\mu \exp\big(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\big) \int_{t_n}^{\sigma} \partial_t z_{j_2}(\mu)\,\mathrm{d}\mu\Big\|_{\ell^1_2} \\
&\leqslant (\sigma - t_n)^2 \max_{\mu\in\Gamma_n} \|\partial_t z_{j_2}(\mu)\|_{\ell^1_2},
\end{aligned}$$

where for the estimates of the terms $\widetilde{\mathcal{I}}_{j_1}$ and $\widetilde{\mathcal{I}}_{j_2}$ we use (5.53) and (5.50).

The terms $\widetilde{R}_2$ and $\widetilde{R}_3$ which contain two and three integrals, respectively, can be bounded straightforwardly at least by $\mathcal{O}(\tau^3)$. As an example we only consider one term of $\widetilde{R}_2$. We obtain with (5.26) and the previously calculated bounds

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\big(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\big)\, P\Psi_1^*(\varepsilon\cdot)\mathcal{T}\Big(\mathcal{I}_{j_1}(t_n, z(t_n)), \widetilde{\mathcal{I}}_{j_2}(\sigma, t_n, z), \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\Big)\,\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\quad\leqslant C_T \int_{t_{n-1}}^{t_{n+1}} \big\|\mathcal{I}_{j_1}(t_n, z(t_n))\big\|_{\ell^1} \big\|\widetilde{\mathcal{I}}_{j_2}(\sigma, t_n, z)\big\|_{\ell^1} \big\|\mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\big\|_{\ell^1}\,\mathrm{d}\sigma \\
&\quad\leqslant C_T \|P z_{j_1}(t_n)\|_{\ell^1} \max_{\mu\in\Gamma_n} \|\partial_t z_{j_2}(\mu)\|_{\ell^1_2} \|\mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\|_{\ell^1} \int_{t_{n-1}}^{t_{n+1}} (\sigma - t_n)^2 \mathrm{d}\sigma \\
&\quad\leqslant \tau^3 C\,(C_T, C_2)\,,
\end{aligned} \tag{5.88}$$

where we use Lemma 5.1.6 with $r = 2$ and Lemma 5.4.2 b) for the last inequality. The other term of $\widetilde{R}_2$ and the term $\widetilde{R}_3$ can be treated similarly.

Step 4. The remaining term is

$$\widetilde{R}_1(\sigma, t_n, z, J) = \mathcal{T}\big(\mathcal{I}_{j_1}(t_n, z(t_n)), \mathcal{I}_{j_2}(\sigma, t_n, z), \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\big) \tag{5.89}$$

and the goal is to show

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\big(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\big)\, P\Psi_1^*(\varepsilon\cdot)\widetilde{R}_1(\sigma, t_n, z, J)\mathrm{d}\sigma\Big\|_{\ell^1} \leqslant C\left(\tau^3 + \varepsilon\tau^2\right). \tag{5.90}$$

The term $\mathcal{I}_{j_2}$ is the only term of (5.89) which depends on σ. All the other expressions are independent of σ. Next, we rewrite

$$\partial_t z_{j_2}(\mu) = \partial_t z_{j_2}(t_n) + \big(\partial_t z_{j_2}(\mu) - \partial_t z_{j_2}(t_n)\big).$$

If we substitute this expression into $\mathcal{I}_{j_2}$ we have

$$\begin{aligned}
\widetilde{R}_1(\sigma, t_n, z, J) = \mathcal{T}\Big(&\mathcal{I}_{j_1}(t_n, z(t_n)), \Psi_{j_2}(\varepsilon\cdot) P \exp\big(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\big) \int_{t_n}^{\sigma} \partial_t z_{j_2}(t_n)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\Big) \\
&+ \mathcal{T}\Big(\mathcal{I}_{j_1}(t_n, z(t_n)), \Psi_{j_2}(\varepsilon\cdot) \exp\big(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\big) \int_{t_n}^{\sigma} P\big(\partial_t z_{j_2}(\mu) - \partial_t z_{j_2}(t_n)\big)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\Big),
\end{aligned} \tag{5.91}$$

where P commutes with the diagonal matrix and the time derivative. We consider the first term of (5.91). We observe that the term under the inner integral is now independent of μ. Therefore, we obtain a factor $(\sigma - t_n)$. Furthermore, all expressions are now independent of σ, except $(\sigma - t_n)$. Thus, it follows for the integral over $[t_{n-1}, t_{n+1}]$ that

$$\int_{t_{n-1}}^{t_{n+1}} (\sigma - t_n)\,\mathrm{d}\sigma\, \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon k)\right) P\Psi_1^*(\varepsilon k)\mathcal{T}\Big(\mathcal{I}_{j_1}(t_n, z(t_n)), \tag{5.92}$$
$$\Psi_{j_2}(\varepsilon\cdot)P\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\right)\partial_t z_{j_2}(t_n), \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\Big)(k) = 0.$$

Here, we take advantage of the symmetry of the integral around the midpoint t_n, i.e.

$$\int_{t_{n-1}}^{t_{n+1}} (\sigma - t_n)\,\mathrm{d}\sigma = 0.$$

Next, we substitute (5.91) into the left-hand side of (5.90) and use the fact that (5.92) vanishes. With (5.26) and Lemma 5.4.4 we obtain

$$\begin{aligned}
&\Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right) P\Psi_1^*(\varepsilon\cdot)\widetilde{R}_1(\sigma, t_n, z, J)\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\quad = \Big\| \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_1(\varepsilon\cdot)\right) P\Psi_1^*(\varepsilon\cdot)\mathcal{T}\Big(\mathcal{I}_{j_1}(t_n, z(t_n)), \\
&\qquad \Psi_{j_2}(\varepsilon\cdot)\exp\left(-\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Lambda_{j_2}(\varepsilon\cdot)\right)\int_{t_n}^{\sigma} P\big(\partial_t z_{j_2}(\mu) - \partial_t z_{j_2}(t_n)\big)\,\mathrm{d}\mu, \mathcal{P}_\varepsilon \widehat{v}_{j_3}(t_n)\Big)\,\mathrm{d}\sigma\Big\|_{\ell^1} \\
&\quad \leqslant \int_{t_{n-1}}^{t_{n+1}} \int_{t_n}^{\sigma} \big\|P\big(\partial_t z_{j_2}(\mu) - \partial_t z_{j_2}(t_n)\big)\big\|_{\ell^1}\,\mathrm{d}\mu\,\mathrm{d}\sigma\, \|Pz_{j_3}(t_n)\|_{\ell^1}^2 \\
&\quad \leqslant C\left(C_T, C_2^\varepsilon\right) \int_{t_{n-1}}^{t_{n+1}} \int_{t_n}^{\sigma} \left(|\mu - t_n| + \varepsilon\right)\mathrm{d}\mu\,\mathrm{d}\sigma \\
&\quad \leqslant C\left(C_T, C_2^\varepsilon\right)\left(\tau^3 + \varepsilon\tau^2\right).
\end{aligned} \tag{5.93}$$

Combining all the different bounds (5.85), (5.87), (5.88) and (5.93) yields

$$\Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{2P}(\sigma, t_n, z, J)\,\mathrm{d}\sigma\Big\|_{\ell^1} \leqslant C\left(C_T, C_2^\varepsilon\right)\left(\tau^3 + \varepsilon\tau^2\right).$$

■

With the estimates from Lemma 5.4.7, Lemma 5.4.9, Lemma 5.4.10 and Lemma 5.4.11, we prove Proposition 5.4.5.

Proof of Proposition 5.4.5.
For the local error we have for $n \geqslant 1$

$$\|\mathbf{d}^{n+1}\|_{\ell^1} \leqslant \sum_{\#J=1} \big\|\mathbf{F}(\mathbf{z}, t_n, J) - \mathbf{F}(\mathbf{z}(t_n), t_n, J)\big\|_{\ell^1} = \sum_{\#J=1} \Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma\Big\|_{\ell^1}.$$

From (5.69) together with the splitting (5.71), but now for every $\mathcal{G}_{iP}$, we know that we can estimate

$$\begin{aligned}
&\sum_{\#J=1} \Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\quad \leqslant \sum_{\#J=1} \Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{G}_{1P}(\sigma, t_n, z, J) + \mathcal{G}_{2P}(\sigma, t_n, z, J) + \mathcal{G}_{3P}(\sigma, t_n, z, J)\big]\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\qquad + \sum_{\#J=1} \Big\| \int_{t_{n-1}}^{t_{n+1}} P^\perp\widetilde{S}_{1,\varepsilon}(\sigma)\big[\mathcal{G}_{1P}(\sigma, t_n, z, J) + \mathcal{G}_{2P}(\sigma, t_n, z, J) + \mathcal{G}_{3P}(\sigma, t_n, z, J)\big]\,\mathrm{d}\sigma \Big\|_{\ell^1} + \left(\tau^2\varepsilon + \varepsilon^2\tau\right) C \\
&\quad \leqslant \sum_{\#J=1} \sum_{i=1}^{3} \Big\| \int_{t_{n-1}}^{t_{n+1}} P\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{iP}(\sigma, t_n, z, J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\qquad + \sum_{\#J=1} \sum_{i=1}^{3} \Big\| \int_{t_{n-1}}^{t_{n+1}} P^\perp\widetilde{S}_{1,\varepsilon}(\sigma)\mathcal{G}_{iP}(\sigma, t_n, z, J)\,\mathrm{d}\sigma \Big\|_{\ell^1} + \left(\tau^2\varepsilon + \varepsilon^2\tau\right) C.
\end{aligned}$$

Now, by combining Lemma 5.4.10 and Lemma 5.4.11, but now for every $\mathcal{G}_{iP}$, we obtain

$$\sum_{\#J=1} \Big\| \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_n), J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \leqslant C\left(C_T, C_2^\varepsilon\right)\left(\tau^3 + \varepsilon\tau^2 + \tau\varepsilon^2\right).$$

For the first step we obtain with Proposition 5.3.3

$$\begin{aligned}
\|\mathbf{d}^1\|_{\ell^1} &\leqslant \sum_{\#J=1} \big\|\mathbf{F}(\mathbf{z}, t_0, J) - \mathbf{F}(\mathbf{z}(t_0), t_0, J)\big\|_{\ell^1} = \sum_{\#J=1} \Big\| \int_{t_0}^{t_1} F_\varepsilon(\sigma, z(\sigma), J) - F_\varepsilon(\sigma, z(t_0), J)\,\mathrm{d}\sigma \Big\|_{\ell^1} \\
&\leqslant \tau^2 C.
\end{aligned}$$

■

Now by means of the Proposition 5.3.4 and 5.3.5, we are in a position to show the global error result by combining the stability result and the local error bound with the classical telescoping sum argument of Lady Windermere's fan.

Proof of Theorem 5.3.2 b). Similarly to the proof of Theorem 5.3.2 a), we estimate

$$\begin{aligned}
\|z^n - z(t_n)\|_{\ell^1} \leqslant \|\mathbf{z}^n - \mathbf{z}(t_n)\|_{\ell^1} &\leqslant e^{t_{\mathrm{end}} C_T \mathcal{C}_\star^2} \left(\sum_{m=1}^{n-1} \|\Phi_{\tau,t_m}\left(\mathbf{z}(t_m)\right) - \mathbf{z}(t_{m+1})\|_{\ell^1} + \|\mathbf{d}^1\|_{\ell^1} \right) \\
&\leqslant e^{t_{\mathrm{end}} C_T \mathcal{C}_\star^2} \left((n-1)\left(\varepsilon^2\tau + \tau^2\varepsilon + \tau^3\right)\tilde{C} + \tau^2\tilde{C} \right) \\
&\leqslant e^{t_{\mathrm{end}} C_T \mathcal{C}_\star^2} \left(t_{\mathrm{end}}\left(\varepsilon^2 + \tau\varepsilon + \tau^2\right)\tilde{C} + \tau^2\tilde{C} \right) \\
&\leqslant e^{t_{\mathrm{end}} C_T \mathcal{C}_\star^2} \left(\tau^2 + \varepsilon^2\right) t_{\mathrm{end}}\tilde{C},
\end{aligned}$$

because of the Proposition 5.4.5 and $\tau\varepsilon \leqslant \max\{\tau^2, \varepsilon^2\}$. ■

Conclusion

For different ranges of the step-size τ, one of the two parameters τ or ε is the dominant one.

- For step-sizes $\tau \geqslant \varepsilon$ the global error bound (see Theorem 5.3.2 b)) becomes smaller as the step-size decreases because we have

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant \tau^2 C.$$

- For step-sizes $\varepsilon^2 \leqslant \tau < \varepsilon$ the global error is dominated by the parameter ε because

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant \varepsilon^2 C.$$

- Now we consider the case $\tau < \varepsilon^2$. For those step-sizes the ratio between ε and τ is so small that we can apply standard theory and observe second-order accuracy again. In other words, we proceed for the proof of the local error bound as in the classical way and apply Taylor expansion, similarly to the proof of Lemma 5.4.11. We note that in general we can estimate

$$\tfrac{1}{\varepsilon^2}|\Lambda_{\pm 1}(\varepsilon k)|_2 \leqslant \tfrac{1}{\varepsilon^2}\big(|\Lambda_{\pm 1}(\varepsilon k) - \Lambda_{\pm 1}(0)|_2 + |\Lambda_{\pm 1}(0)|_2\big)$$

and $$\|\partial_t z_{\pm 1}(\mu) - \partial_t z_{\pm 1}(t_n)\|_{\ell^1} \leqslant (\mu - t_n) \max_{\nu \in \Gamma_n} \|\partial_t^2 z_{\pm 1}(\nu)\|_{\ell^1} = \mathcal{O}\big((\mu - t_n)\varepsilon^{-2}\big)\,.$$

  Adapting the proof of Lemma 5.4.11 results in a local error estimate of $\mathcal{O}\Big(\frac{\tau^3}{\varepsilon^2}\Big)$ and, thus, for the global error bound we obtain

$$\|z^n - z(t_n)\|_{\ell^1} \leqslant \frac{\tau^2}{\varepsilon^2} C \qquad \text{for } \tau < \varepsilon^2.$$

  We omit the details because for $\tau < \varepsilon^2$ the total error of the analytical and numerical approximations is dominated by the analytical error, which is $\mathcal{O}\big(\varepsilon^2\big)$.

Hence, we conclude that it does not make sense to use a step-size τ which is smaller than $\mathcal{O}(\varepsilon)$.

5.5 Cherry Picking

The drawback of reformulating the system (3.16) in terms of the transformed system (5.18) is the multiple sums hidden in the term $F_\varepsilon(t, z, J)$, cf. (5.19) combined with the definition (5.12) of $\mathcal{T}$. From a numerical point of view the nested sum structure makes the evaluations of the nonlinearity more costly. For this reason, we investigate the nonlinear term in more detail with the aim to hopefully reduce the numerical work for the one- and two-step method. In the following, we only consider the two-step method as an example.

5.5.1 Observations

We recall that the two-step method is given by

$$z^{n+1}(k) = z^{n-1}(k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z^n, J)(k)\,\mathrm{d}\sigma,$$

cf. (5.37). For the numerical calculation we have to evaluate each integral exactly. In order to do this, we take a closer look at the individual entries of the integral. With the notation (3.52) the m-th entry of the integral term is

$$\begin{aligned}\int_{t_{n-1}}^{t_{n+1}} F_{\varepsilon,m}(\sigma, z^n, J)(k)\,\mathrm{d}\sigma = \sum_M \sum_{\#K=k} \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\left[\lambda_{1m}(\varepsilon k) - \lambda_{JM}(\varepsilon K)\right]\right)\,\mathrm{d}\sigma \\ \times Z^n_{JM}(K)\psi^*_{1m}(\varepsilon k) T\left(\psi_{JM}(\varepsilon K)\right).\end{aligned}$$

Let j, J, m, M, k, K be fixed and define

$$\Delta\lambda = \lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K) = \lambda_{jm}(\varepsilon k) - \sum_{i=1}^{3} \lambda_{j_i m_i}(\varepsilon k^{(i)}).$$

Then, we have

$$\begin{aligned}\int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\left[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)\right]\right)\,\mathrm{d}\sigma &= \int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}\Delta\lambda\right)\,\mathrm{d}\sigma \\ &= \frac{\varepsilon^2}{|\Delta\lambda|}\left(\exp\left(\tfrac{\mathrm{i}t_{n+1}}{\varepsilon^2}\Delta\lambda\right) - \exp\left(\tfrac{\mathrm{i}t_{n-1}}{\varepsilon^2}\Delta\lambda\right)\right) \\ &= \tau \exp\left(\tfrac{\mathrm{i}t_n}{\varepsilon^2}\Delta\lambda\right)\left(\varphi_1\left(\tfrac{\mathrm{i}\tau}{\varepsilon^2}\Delta\lambda\right) + \varphi_1\left(-\tfrac{\mathrm{i}\tau}{\varepsilon^2}\Delta\lambda\right)\right),\end{aligned} \tag{5.94}$$

where we use the definition of the φ_1-function from [20].
However, this is computationally expensive. While the φ_1-functions have to be calculated only once at the beginning and then saved, it has to be done for every possible combination of j, J, m, M, k, K. The total number of operations is proportional to the number of different indices j, J, m, M, k, K. We know that there are three different multi-indices J with $\#J = 1$ in the case $j_{\max} = 1$. Since $m \in \{1, \dots, s\}$ and $M \in \{1, \dots, s\}^3$, there are s different choices for m and s^3 different choices for M. We define the set

$$\mathbb{G} := \{-\frac{k_{max}}{2}, \dots, \frac{k_{max}}{2} - 1\}, \tag{5.95}$$

where $k_{max} \in 2\mathbb{N}$ is a fixed number. After a space discretization by Fourier collocation (see Subsection 5.6.1) the total number of $K = (k^{(1)}, k^{(2)}, k^{(3)}) \in \mathbb{G}^d \times \mathbb{G}^d \times \mathbb{G}^d$ is $d^3 k_{max}^3$. The total number of

combinations $K = (k^{(1)}, k^{(2)}, k^{(3)}) \in \mathbb{G}^d \times \mathbb{G}^d \times \mathbb{G}^d$ under the constraint $\#K = k \in \mathbb{G}^d$ is only a subset and, hence, the total number of operations is smaller than

$$3 \cdot s \cdot s^3 \cdot d^3 k_{max}^3 = 3s \cdot (s \cdot d \cdot k_{max})^3 . \tag{5.96}$$

Consequently, our methods are only useful if (5.96) is sufficiently small, and furthermore, are of interest only for small ε. For (5.96) to be small, we need few space discretization points. The parameters d and s are typically not so large. For a Klein–Gordon equation in $d = 1$, we have $s = d + 1 = 2$. However, for the Maxwell-Lorentz system in $d = 3$, it follows that $s = 12$. Thus, we expect that our methods are only applicable for $d = 1$.

It is not surprising that the workload depends on $s \cdot d \cdot k_{max}$, but the cubic scaling is a severe problem. Moreover, the diagonalisation has to be computed and stored for $j = 1$ and every $k \in \mathbb{G}^d$.

The multiple sum structure in the numeric schemes (one-step method and two-step method) is the key limiting factor for the competitiveness of these methods in comparison with the Schrödinger approximation. Compared to standard methods like splitting methods, the advantage is not having a restriction on the step-size τ.

In order to reduce the numerical work, we introduce a strategy which we call *cherry picking*. This ansatz does not change the cubic scaling but is used to reduce the computational work considerably.

Idea. The idea of cherry picking is to reduce the numerical workload by omitting terms which are a priori "small enough" compared to the accuracy. For this purpose we break the right-hand side of (5.37) down to the level of single entries. The m-th entry of the term under the integral is given by

$$F_{\varepsilon,m}(t,z,J)(k) = \sum_{M} \sum_{\#K=k} \exp\Big(\tfrac{\mathrm{i}t}{\varepsilon^2}\big[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)\big]\Big) Z_{JM}(t,K) \psi_{jm}^*(\varepsilon k) T\big(\psi_{JM}(\varepsilon K)\big),$$

where $J \in \mathcal{J}^3$ with $\#J = 1$. If the m-th term of the following integral satisfies

$$\int_{t_{n-1}}^{t_{n+1}} F_{\varepsilon,m}(\sigma, z, J)(k)\, \mathrm{d}\sigma = \mathcal{O}(\tau\varepsilon^2),$$

then the corresponding term can be omitted, because the total accuracy of our approach cannot be better than $\mathcal{O}(t_{\mathrm{end}}\varepsilon^2)$, anyway. For the exact solution we know by Assumption 5.4.1 (ii) that

$$z_{11}(t,k) = \mathcal{O}(1), \quad z_{1m}(t,k) = \mathcal{O}(\varepsilon) \quad \text{if } m \neq 1.$$

Therefore, we only include terms Z_{JM} for which at least two entries of M are 1. This means for example the multi-index $M = (1,1,1)$ or $M = (m_1, 1, 1)$ for $m_1 = 2, \ldots, s$ and its permutations.

5.5.2 Construction of the cherry picking method

The first goal is to state the cherry picking method in a rigorous way. We start with the exact two-step equation (5.36)

$$z(t_{n+1}, k) = z(t_{n-1}, k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k)\, \mathrm{d}\sigma.$$

With the decomposition (5.41) we have

$$F_\varepsilon(t, z(t), J) = F_\varepsilon(t, Pz(t), J) + \mathcal{N}_1^\perp(t, z(t), J) + \mathcal{N}_2^\perp(t, z(t), J),$$

cf. (5.45). We observe that $\mathcal{N}_2^\perp$ includes the pairs of multi-indices J and M where

$$\int_{t_{n-1}}^{t_{n+1}} F_{\varepsilon,m}(\sigma, z(\sigma), J)(k)\,\mathrm{d}\sigma = \mathcal{O}(\tau\varepsilon^2)$$

is fulfilled for the exact solution. An approximation for the exact solution at time t_{n+1} can be obtained by omitting the term $\mathcal{N}_2^\perp$ and freezing $z(\sigma)$ in the remaining terms of the nonlinearity at $\sigma = t_n$. We define

$$F_\varepsilon^{cher}(t, z(t), J) = F_\varepsilon(t, Pz(t), J) + \mathcal{N}_1^\perp(t, z(t), J) \tag{5.97}$$

and, thus, an approximation is given by

$$z(t_{n+1}, k) \approx z(t_{n-1}, k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon^{cher}(\sigma, z(t_n), J)(k)\,\mathrm{d}\sigma.$$

This yields the cherry picking two-step method for $n \geqslant 1$

$$z^{n+1,cher}(k) = z^{n-1,cher}(k) + \sum_{\#J=1} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon^{cher}(\sigma, z^{n,cher}, J)(k)\,\mathrm{d}\sigma, \tag{5.98}$$

where for the starting step, i.e $n = 0$, we use the one-step method

$$z^{1,cher}(k) = z^{0,cher}(k) + \sum_{\#J=1} \int_{t_0}^{t_1} F_\varepsilon^{cher}(\sigma, z^{0,cher}, J)(k)\,\mathrm{d}\sigma.$$

Equivalent one-step method of the cherry picking method. In the following we introduce the abbreviation

$$\sum_{\#J=1} \mathbf{F}^{cher}(\mathbf{z}^{n,cher}, t_n, J)(k) = \begin{pmatrix} \int_{t_{n-1}}^{t_{n+1}} \sum_{\#J=1} F_\varepsilon^{cher}(\sigma, z^{n,cher}, J)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix}.$$

Thus, we obtain for the cherry picking method the one-step formulation

$$\mathbf{z}^{n+1,cher}(k) = \mathcal{M}\mathbf{z}^{n,cher}(k) + \sum_{\#J=1} \mathbf{F}^{cher}(\mathbf{z}^{n,cher}, t_n, J)(k).$$

5.5.3 Error analysis for the cherry picking two-step method

Inserting the exact solution $\mathbf{z}(t_n, k)$ into the numerical scheme and subtracting from the exact solution at time t_{n+1} yields

$$\begin{aligned} \mathbf{z}(t_{n+1}, k) - \mathbf{z}^{n+1,cher}(k) &= \sum_{\#J=1} \left(\mathbf{F}(\mathbf{z}, t_n, J)(k) - \mathbf{F}^{cher}(\mathbf{z}(t_n), t_n, J)(k)\right) \\ &= \sum_{\#J=1} \begin{pmatrix} \int_{t_{n-1}}^{t_{n+1}} F_\varepsilon(\sigma, z(\sigma), J)(k) - F_\varepsilon^{cher}(\sigma, z(t_n), J)(k)\,\mathrm{d}\sigma \\ 0 \end{pmatrix}. \end{aligned}$$

Therefore, we define the local error for $n \geqslant 1$

$$\mathbf{d}^{n+1,cher}(k) = \sum_{\#J=1} \left(\mathbf{F}(\mathbf{z}, t_n, J)(k) - \mathbf{F}^{cher}(\mathbf{z}(t_n), t_n, J)(k)\right).$$

The global error of the cherry picking two-step method applied to the system (5.18) satisfies the following bound.

Theorem 5.5.1. *Let $z \in C^2([0,t_{end}];\ell^1) \cap C^1([0,t_{end}];\ell^1_1) \cap C([0,t_{end}];\ell^1_2)$ be the solution of (5.18). If Assumptions 3.2.1, 4.1.1, 3.2.2, 3.2.6, 3.7.2, and 4.3.1 hold, then for sufficiently small step-sizes τ the global error of the scheme (5.98) is bounded by*

$$\|z^{n,cher} - z(t_n)\|_{\ell^1} \leqslant \left(\tau^2 + \varepsilon^2\right) C,$$

where C depends on t_{end}, C_T, $\|z(0)\|_{\ell^1_2}$, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.

We state the following local error result for the cherry picking method.

Proposition 5.5.2 (Local error cherry)**.** *If $\mathbf{z}(0) \in \ell^1_2$ and the Assumptions 3.2.1, 4.1.1, 3.2.2 and 3.7.2 hold, then the local error of the equivalent one-step method applied to (5.18) satisfies*

$$\|\mathbf{d}^{n+1,cher}\|_{\ell^1} \leqslant \left(\varepsilon^2\tau + \tau^2\varepsilon + \tau^3\right)\tilde{C}, \qquad (n+1)\tau \leqslant t_{end}, \qquad n \in \mathbb{N},$$

where $\tilde{C}$ depends on C_T, $\|z(0)\|_{\ell^1_2}$, on the inverse of the nonzero eigenvalues of $\Lambda_1(0)$, and on the Lipschitz constant in Assumption 3.7.2, but not on ε.
For $n = 0$ we have

$$\|\mathbf{d}^{1,cher}\|_{\ell^1} \leqslant \left(\tau^2 + \tau\varepsilon^2\right)\tilde{C},$$

where $\tilde{C}$ depends on C_T, $\|z(0)\|_{\ell^1_2}$, but not on ε.

Proof. First, we observe that (5.97) is equivalently to

$$F^{cher}_\varepsilon(t,z(t),J) = F_\varepsilon(t,z(t),J) - \mathcal{N}^\perp_2(t,z(t),J).$$

Hence, the difference in the local error can be written as

$$\sum_{\#J=1}\left(F_\varepsilon(\sigma,z(\sigma),J) - F^{cher}_\varepsilon(\sigma,z(t_n),J)\right) = \sum_{\#J=1}\left(F_\varepsilon(\sigma,z(\sigma),J) - F_\varepsilon(\sigma,z(t_n),J) + \mathcal{N}^\perp_2(\sigma,z(t_n),J)\right).$$

Together with the definition (5.38) we obtain

$$\mathbf{d}^{n+1,cher}(k) = \mathbf{d}^{n+1}(k) + \mathbf{d}^{n+1,new}(k),$$

where

$$\mathbf{d}^{n+1,new}(k) = \sum_{\#J=1}\begin{pmatrix}\int_{t_{n-1}}^{t_{n+1}} \mathcal{N}^\perp_2(\sigma,z(t_n),J)(k)\,\mathrm{d}\sigma \\ 0\end{pmatrix}.$$

With the estimate (5.47), we directly obtain

$$\sum_{\#J=1}\int_{t_{n-1}}^{t_{n+1}} \|\mathcal{N}^\perp_2(\sigma,z(t_n),J)\|_{\ell^1}\,\mathrm{d}\sigma \leqslant \tau\varepsilon^2 C\left(C_T, C^\varepsilon_0\right).$$

Together with Proposition 5.4.5 the assertion follows.
For $n = 0$ the claimed estimate follows with the same reasoning, whereby now we apply Proposition 5.2.4.
■

The stability and the boundedness of the solution of the cherry picking two-step method can be proved in the same way as for the full two-step method in Section 5.3. This allows us to prove Theorem 5.5.1 via the telescoping sum argument of Lady Windermere's fan. We omit the details at this point because this argument has already be presented in detail in the proofs of Theorem 5.3.2 a) and b).

5.5.4 Further reducing of the workload

If we are only interested in step-sizes $\tau > \varepsilon$, we make another observation. If $|\Delta\lambda| \geqslant c > 0$, then we have for (5.94)

$$\int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)]\right) \mathrm{d}\sigma = \mathcal{O}(\varepsilon^2).$$

Furthermore, in this case

$$\int_{t_{n-1}}^{t_{n+1}} \exp\left(\tfrac{\mathrm{i}\sigma}{\varepsilon^2}[\lambda_{jm}(\varepsilon k) - \lambda_{JM}(\varepsilon K)]\right) \mathrm{d}\sigma Z_{JM}^n(K) = \mathcal{O}(\varepsilon^3), \text{ whenever } m_i > 1 \text{ for at least one } i \in \{1,2,3\}.$$

Thus, we can omit terms in the cases with $M = (m_1, 1, 1)$ for $m_1 = 2, \ldots, s$ and its permutations, if $|\Delta\lambda| \geqslant c > 0$. This could be done in a query during implementation.

5.6 Numerical experiments

In this section we illustrate the behavior of the one-step and the two-step method by numerical examples. In the classical method of lines the PDE is first discretized in space. This leads to the approximation of the PDE by a system of ODEs. This system of ODEs is then solved by a time-integrator as constructed in the previous sections.

5.6.1 Space discretization by Fourier collocation

Discretizing the PDE in space means that we approximate the derivatives in space. In this thesis, the implementation of the spectral method is accomplished with a collocation approach. The explanations are based on [31, Chapter III]. The idea of spectral methods is to write the solution of the differential equation as a sum of certain basis functions with coefficients, where these coefficients are time-dependent if we apply spectral methods to time-dependent PDEs. In our case the exact solution is represented (formally) by the semidiscrete Fourier transform (5.6), i.e.

$$v_j(t,x) = \sum_{k\in\mathbb{Z}^d} \widehat{v}_j(t,k) e^{\mathrm{i}k\cdot x}.$$

Substituting this ansatz into the PDE yields a system of ODEs for the coefficients, cf. (5.13).
As mentioned before, for the discretization in space, we assume periodic boundary conditions and choose a fixed number $k_{max} \in 2\mathbb{N}$ which denotes the number of grid points. We define the set $\mathbb{G}$ as in (5.95). We aim for an approximation of the exact solution in terms of

$$v_j(t,x) \approx \sum_{k\in\mathbb{G}^d} \widetilde{\widehat{v}}_j(t,k) e^{\mathrm{i}k\cdot x}.$$

In other words, in this approach we approximate the infinite Fourier series in (5.6) by a finite sum which corresponds to the truncation

$$\widetilde{\widehat{v}}_j(t,k) = 0 \quad \text{for } k \notin \mathbb{G}^d$$

or after transforming into the variables $\widetilde{z}_j$

$$\widetilde{z}_j(t,k) = 0 \quad \text{for } k \notin \mathbb{G}^d.$$

The condition of collocation methods is that the differential equation is satisfied at all collocation points. After all, we obtain for the coefficients $\widetilde{\widehat{v}}_j(t,k)$ and $k \in \mathbb{G}^d$ the (finite) system of ODEs

$$\partial_t \widetilde{\widehat{v}}_j(t,k) + \frac{\mathrm{i}}{\varepsilon^2}\widetilde{\mathcal{L}}_j(\varepsilon k)\widetilde{\widehat{v}}_j(t,k) = \sum_{\#J=j} \mathcal{T}\left(\widetilde{\widehat{v}}_1, \widetilde{\widehat{v}}_2, \widetilde{\widehat{v}}_3\right)(t,k).$$

Analogously, we obtain after transformation a (finite) system of ODEs for the coefficients $\widetilde{z}_j(t,k)$

$$\partial_t \widetilde{z}_j(t,k) = \sum_{\#J=j} F_\varepsilon(t, \widetilde{z}, J)(k).$$

Remark 5.6.1. *For any numerical method presented in this thesis, we focus solely on the error analysis of the semi-discretization in time. Nevertheless, in all subsequent numerical examples the (pseudo-)spectral collocation method introduced in this section is employed.*

5.6.2 Model problem

We consider the one-dimensional Klein–Gordon system (1.3) with $\kappa = 1.2$, $\upsilon = 0.7$, $\mathcal{M} = E$, $\omega = \max\{\omega_1(\kappa), \omega_2(\kappa)\}$, where ω_m is the m-th eigenvalue of $\mathcal{L}(0,\kappa)$, and with initial data $p(x) = \psi_{11}(0)\exp\left(-(x-0.5)^2\right)$. We set $t_{\mathrm{end}} = 1$ and consider 2^6 equidistant grid points in the interval $[-2,2]$. The reference solution is computed by the Strang splitting method with a small step-size $\tau \approx 10^{-5}$, where we note that the step-size is chosen small enough in comparison to the choice of ε which we consider. For the Strang splitting method we split the PDE (5.5) with $j_{\max} = 1$ into the linear and the nonlinear part. This results in the linear subproblem

$$\partial_t v_1^\bullet(t) = -\mathcal{B}_1 v_1^\bullet(t) \qquad \text{with given } v_1^\bullet(0), \tag{5.99}$$

where $\mathcal{B}_1 := \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}(\omega,\kappa) + \frac{1}{\varepsilon}B(\partial)$, and the nonlinear subproblem

$$\partial_t v_1^{\bullet\bullet}(t) = \sum_{\#J=1} T(v_{j_1}^{\bullet\bullet}, v_{j_2}^{\bullet\bullet}, v_{j_3}^{\bullet\bullet})(t) \qquad \text{with given } v_1^{\bullet\bullet}(0). \tag{5.100}$$

The operator $\mathcal{B}_1$ generates a strongly continuous group on $W(\mathbb{T}^d)$. Thus, for $t \geqslant 0$ we obtain a solution of the linear subproblem via

$$v_1^\bullet(t) = e^{-t\mathcal{B}_1} v_1^\bullet(0) \tag{5.101}$$

and, hence, (5.99) can be solved exactly in Fourier space. Since we cannot solve the nonlinear subproblem (5.100) exactly, we approximate the solution with Heun's method which is a Runge-Kutta method of order two.

In total we obtain an approximation $v_1^n \approx v_1(t_n)$ recursively, meaning by solving the subproblems (5.99) and (5.100) in alternating fashion. In order to calculate v_1^{n+1} we first approximate the solution of (5.100) via Heun's method with one half time-step $\frac{\tau}{2}$ and initial data v_1^n which yields $v_1^{n+1,-}$. Next, we compute $v_1^{n+1,+}$ by taking a full time-step τ of the exact solution (5.101), where now $v_1^{n+1,-}$ is the initial data. Finally, we approximate the solution of (5.100) via Heun's method again with one half time-step and initial data $v_1^{n+1,+}$ which yields the approximation v_1^{n+1}.

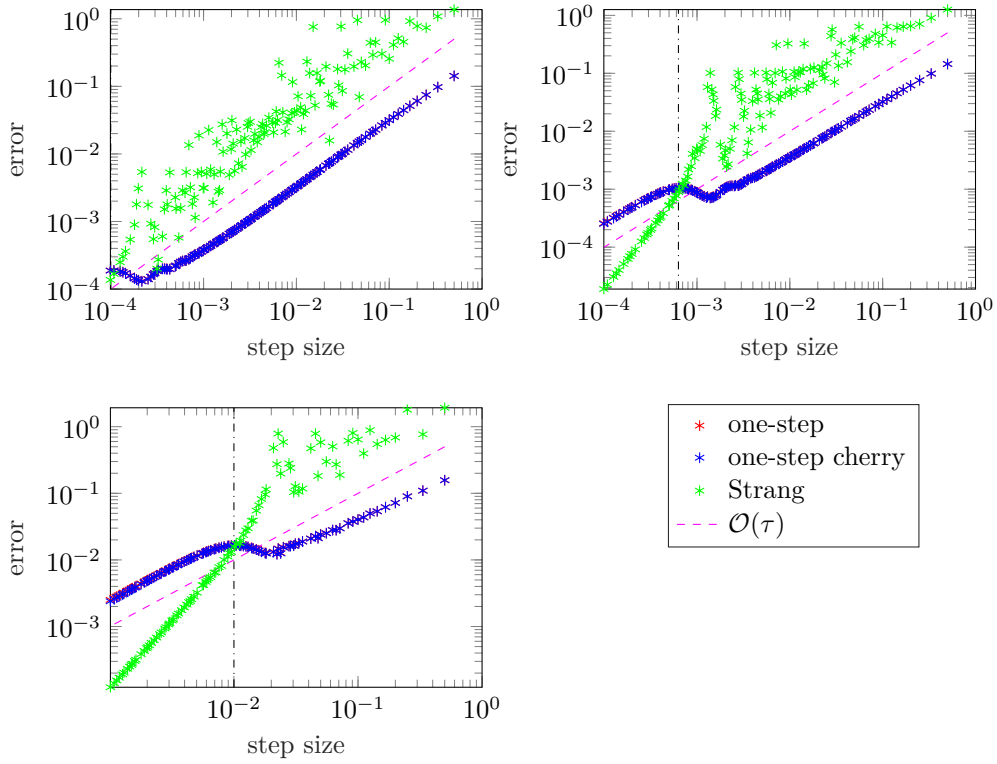


Figure 5.1: Accuracy of the one-step method for $\varepsilon = 0.01$ (top left), $\varepsilon = 10^{-1.6}$ (top right) and $\varepsilon = 0.1$ (bottom left). Additionally, the accuracy of the Strang splitting is shown. The dashed magenta line is a reference line for order one. The black vertical line is at $\tau = \varepsilon^2$.

Figure 5.1 and Figure 5.2 show the accuracy of the one-step method (5.30) and the two-step method (5.37), respectively, together with the corresponding cherry picking method. For comparison, the accuracy of the Strang splitting method is shown. The dashed line in Figure 5.1 is a reference line for order one and the vertical line highlights the value $\tau = \varepsilon^2$, whereas in Figure 5.2 the dashed line is of order two and the value $\tau = \varepsilon$ is highlighted. We observe the familiar erratic behavior of the Strang splitting method for $\tau > \varepsilon^2$. Figure 5.1 illustrates the first-order convergence of the one-step method for $\tau > \varepsilon^2$ and Figure 5.2 illustrates the second-order convergence of the two-step method for $\tau > \varepsilon$, in accordance with Theorem 5.2.1 and to Theorem 5.3.2, respectively. We note that in the regime $\varepsilon^2 < \tau < \varepsilon$ the plot in Figure 5.1 turns into a plateau, which roughly speaking marks the difference between the oscillatory and the non-oscillatory situation. This is a typical phenomenon. In the regime $\varepsilon^2 < \tau < \varepsilon$ the global error is dominated by the parameter ε in each panel of Figure 5.2. The code to reproduce the plots is available on https://www.doi.org/10.5445/IR/1000149721.

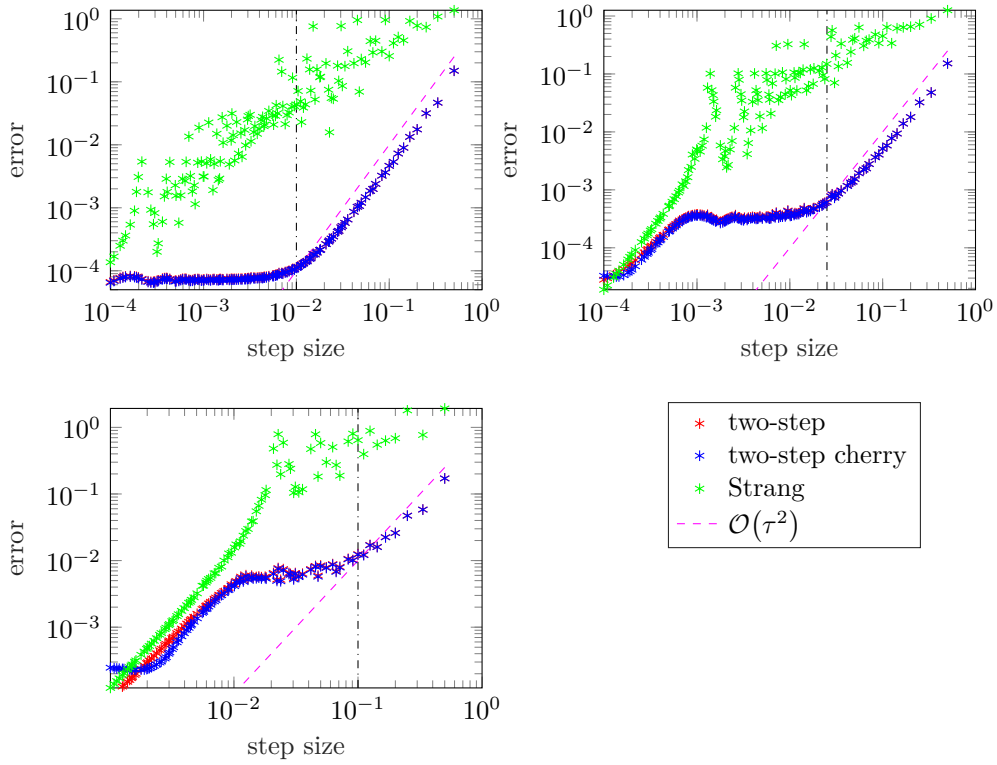


Figure 5.2: Accuracy of the two-step method for $\varepsilon = 0.01$ (top left), $\varepsilon = 10^{-1.6}$ (top right) and $\varepsilon = 0.1$ (bottom left). Additionally, the accuracy of the Strang splitting is shown. The dashed magenta line is a reference line for order two. The black vertical line is at $\tau = \varepsilon$.

CHAPTER 6

Modulated Fourier expansion

In this chapter, we investigate a complementary approach of constructing an approximation to the solution $\mathbf{u}$ of (1.4a). In Chapter 3 the ansatz (3.15) was plugged into (1.4a) which led to the system of PDEs (3.16). We recall that the main advantage of (3.16) over (1.4a) was that the solutions of (3.16) did not oscillate in space if the initial data had the form (1.4b). However, the solutions still oscillated in time which caused difficulties for standard methods. Therefore, in Chapter 5 we constructed tailor-made time integrators which handled these oscillations appropriately. Moreover, we presented a special and elaborate error estimate. We emphasize that compared to Chapter 5 we again consider $\mathbb{R}^d$ and not $\mathbb{T}^d$, and an arbitrary positive odd integer $j_{\max}$ and not $j_{\max} = 1$.

Compared to the previous chapters, the question arises whether there is a possibility to construct analytical approximations which additionally avoid the oscillations in time. The hope is that the resulting *smooth* functions of this approximation will be easier to treat numerically because we have ε-induced oscillations neither in space nor in time. It will turn out that such an analytical approximation is possible, but not for arbitrary initial data. This is the motivation to search approximations to $\mathbf{u}$ of the form (3.15), where now the coefficients u_j are smooth and the initial data is specially constructed for this purpose. More precisely, the coefficients satisfy the constraint $\sup_{t\in[0,t_{\text{end}}/\varepsilon]} \|u_j(t)\|_W = \mathcal{O}\big(\varepsilon^{j-1}\big)$, and the same conditions for space and time derivatives.

Analogous to Subsection 5.1.1, we transform the system (1.4a) in space and time, and we refer to Remark 5.1.1 that after transforming the system the variables t and x are different variables than in the previous Chapters 1-4. The chapter is structured as follows. In Section 6.1, we introduce the modulated Fourier expansion (MFE). The idea of the modulated Fourier expansion is to combine the ansatz (5.4) with a formal asymptotic expansion in powers of ε. More precisely, the coefficient functions v_j in (5.4) admit an expansion of the form

$$v_j(t,x) \approx \sum_{\ell} \varepsilon^{\ell} v_j^{\ell}(t,x), \qquad j \in \mathcal{J}_+, \tag{6.1}$$

where ℓ is a power on ε and a superscript on v_j. This expansion leads to a set of equations for the corre-

sponding coefficient functions v_j^ℓ which are now independent of ε. The majority of Section 6.1 is devoted to the construction of the coefficients of the MFE. In Section 6.2, we show the boundedness of these constructed coefficient functions, which is crucial for the main result in this chapter. In Theorem 6.2.3 we make a statement about the accuracy of the MFE. For more details on the MFE we refer to [10, 18] and [19, Chapter XIII.5].

Note. The results of this chapter have been published with Prof. Dr. Tobias Jahnke and Prof. Dr. Christian Lubich in the preprint [5], whereby the presentation is slightly different.

6.1 Construction of the modulated Fourier expansion

We start this section by introducing the general problem setting.

6.1.1 Problem setting and approach for the MFE

Co-moving coordinate system and rescaling of time. Analogously to Chapter 5, for the analysis it is convenient to change to the new variables

$$t' = \varepsilon t, \qquad x' = x - tc_g, \qquad \mathbf{v}(t', x') = \mathbf{v}(\varepsilon t, x - tc_g) = \mathbf{u}(t, x),$$

where c_g is the group velocity, cf. (3.5). We recall that the change of variables leads to the system (5.1) and the approximation (5.4), where we quote Remark 5.1.1 which states that we omit in the following the prime and write again t and x instead of t' and x', respectively. Furthermore, the variables t and x are different variables for the rest of this chapter than in the Chapters 1-4.

Assumptions and definitions. For the construction of approximations, we make similar definitions and assumptions as in Section 3.2. More precisely, the pair (ω, κ) fulfills the dispersion relation (3.4) and we choose $\omega = \omega_1(\kappa) \neq 0$, where $\omega_1(\kappa)$ is one of the eigenvalues of $A(\kappa) - \mathrm{i}E$. We assume a smooth eigendecomposition of the matrix $\mathcal{L}(0, \beta)$ which corresponds to Assumption 3.2.2. In contrast to Chapter 3, we define the fixed eigenprojector $\mathcal{P} \in C^\infty(\mathbb{R}^d \backslash \{0\}, \mathbb{C}^{s\times s})$ associated to ω, which in comparison to (3.65) does not depend on the spatial variable. By definition it is the orthogonal projection onto the nontrivial kernel of $\mathcal{L}_1 = \mathcal{L}(\omega, \kappa)$, i.e.

$$\mathcal{P}\mathcal{L}_1 = \mathcal{L}_1\mathcal{P} = 0. \tag{6.2}$$

In the following we set $\mathcal{P}^\perp = I - \mathcal{P}$.

Remark 6.1.1. *Because of the dispersion relation (3.4), $\mathcal{L}_1$ is not invertible, but the restricted mapping $\mathcal{L}_1|_{\mathcal{P}^\perp\mathbb{C}^s} : \mathcal{P}^\perp\mathbb{C}^s \to \mathcal{P}^\perp\mathbb{C}^s$ has an inverse. For the sake of simplicity, we denote this inverse by $\mathcal{L}_1^{-1} = (\mathcal{L}_1|_{\mathcal{P}^\perp\mathbb{C}^s})^{-1}$, but we emphasize that $\mathcal{L}_1^{-1}$ can only be applied to vectors in $\mathcal{P}^\perp\mathbb{C}^s$.*

Using these introduced projections, we cite [29, Lemma 2.9 and Lemma 2.12] without proofs, and state them together in one lemma.

Lemma 6.1.2. *If the pair (ω, κ) is smooth and satisfies the dispersion relation (3.4), then we have for every $v(x) \in \mathbb{C}^s$*

$$\mathcal{P}A(\partial)\mathcal{P}v(x) = \mathcal{P}(c_g \cdot \nabla)v(x), \tag{6.3}$$

$$\mathcal{P}A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^\perp A(\partial)\mathcal{P}v(x) = -\frac{1}{2}(\nabla \cdot H\nabla)\mathcal{P}v(x) = -\frac{1}{2}\sum_{k=1}^{d}\sum_{\mu=1}^{d} H_{k\ell}\partial_k\partial_\mu \mathcal{P}v(x), \tag{6.4}$$

where $c_g = \nabla w_1(\kappa)$, and $H = (H_{k\mu})_{k,\mu} = \nabla^2\omega_1(\kappa)$ is the Hessian of ω_1 in κ.

We define for an odd number $j_\star > 0$ the sets

$$\begin{aligned} \mathcal{J}(j_\star) &= \big\{j \in 2\mathbb{Z} - 1, \quad |j| \leqslant j_\star\big\}, \\ \mathcal{J}_+(j_\star) &= \mathcal{J}(j_\star) \cap \mathbb{N}. \end{aligned} \tag{6.5}$$

In addition, we adopt Assumption 3.2.6.

Aim. We aim to construct a solution of (5.1) of the form

$$\mathbf{v}(t,x) = \sum_{j \in \mathcal{J}(j_\star)} e^{\mathrm{i}j\kappa\cdot x/\varepsilon} e^{\mathrm{i}j(\kappa\cdot c_g - \omega)t/\varepsilon^2} v_j(t,x) + \mathcal{O}\big(\varepsilon^{j_\star+1}\big), \qquad v_{-j} = \overline{v_j} \tag{6.6}$$

with $\mathcal{P}v_1(0,x) = p(x)$, where $p(x)$ is a given smooth function which satisfies

$$\mathcal{P}p(x) = p(x) \text{ for all } x \in \mathbb{R}^d.$$

We call the solution of the form (6.6) a *polarized solution*. Polarized in the sense that the solution only depends on the frequency ω, where the pair (ω, κ) satisfies the dispersion relation. The other eigenvalues $\omega_m(\kappa)$ with $m \neq 1$ of $\mathcal{L}(0,\kappa)$ do not appear in the highly oscillatory exponentials $e^{\mathrm{i}j\kappa\cdot x/\varepsilon}e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t/\varepsilon^2}$ of (6.6). We remark that it is not *a priori* clear that this can be achieved for the semilinear system (5.1). Furthermore, the constructed coefficient functions v_j of (6.6) are smooth with the properties that for $t \in [0, t_{\mathrm{end}}]$

$$\mathcal{P}v_1(t) = \mathcal{O}(1)\,, \tag{6.7a}$$

$$\mathcal{P}^\perp v_1(t) = \mathcal{O}(\varepsilon)\,, \tag{6.7b}$$

$$v_j(t) = \mathcal{O}\big(\varepsilon^{j-1}\big) \quad \text{for} \quad 3 \leqslant j \in \mathcal{J}_+(j_\star). \tag{6.7c}$$

The same properties are true for the space and time derivatives of the coefficient functions v_j up to some fixed order.

We will show in Section 6.2 that the approximations of the coefficient functions constructed below do indeed satisfy (6.7) for all $t \in [0, t_{\mathrm{end}}]$. Next, in order to construct the coefficients v_j for $j \in \mathcal{J}_+(j_\star)$ of (6.6) with the above mentioned properties, we derive the corresponding PDEs for v_j.

PDEs for v_j. We proceed similarly to Section 3.3. Substituting (5.4) with $j \in \mathcal{J}_+(j_\star)$ into (5.1) and discarding higher harmonics, i.e. terms with prefactor

$$e^{\mathrm{i}j\kappa\cdot x/\varepsilon}e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t)/\varepsilon^2}, \qquad |j| > j_\star,$$

yields the system (5.3) with $j \in \mathcal{J}_+(j_\star)$. According to Assumption 3.2.6, we know that $\mathcal{L}_j$ is invertible for $3 \leqslant j \in \mathcal{J}_+(j_\star)$. Hence, we reformulate (5.3) to

$$v_j = \mathrm{i}\varepsilon\mathcal{L}_j^{-1}\Big(\varepsilon\partial_t v_j + B(\partial)v_j - \varepsilon \sum_{\#J=j} T(v_J)\Big), \qquad 3 \leqslant j \in \mathcal{J}_+(j_\star). \tag{6.8}$$

As in the previous chapters, the case $j = 1$ is special. We distinguish

$$\begin{aligned} y_1 &= \mathcal{P}v_1, & y_{-1} &= \overline{y_1} = \overline{\mathcal{P}}v_{-1}, \\ z_1 &= \mathcal{P}^\perp v_1, & z_{-1} &= \overline{z_1} = \overline{\mathcal{P}^\perp}v_{-1}, \end{aligned} \tag{6.9}$$

due to the partition in (6.7). In order to derive equations for y_1 and z_1, we use the relations (6.2) and (6.3). Together with definitions (5.2), (6.9), $\mathcal{P}^2 = \mathcal{P}$ and $\mathcal{P}\mathcal{P}^\perp = 0$, it follows that

$$\begin{aligned} \mathcal{P}B(\partial)y_1 &= \mathcal{P}A(\partial)\mathcal{P}v_1 - \mathcal{P}(c_g \cdot \nabla)\mathcal{P}v_1 = 0, \\ \mathcal{P}B(\partial)z_1 &= \mathcal{P}A(\partial)\mathcal{P}^\perp v_1 - \mathcal{P}(c_g \cdot \nabla)\mathcal{P}^\perp v_1 = \mathcal{P}A(\partial)z_1. \end{aligned}$$

Hence, multiplying (5.3) for $j = 1$ by $\mathcal{P}$ and $\mathcal{P}^\perp$ yields

$$\partial_t y_1 + \frac{1}{\varepsilon}\mathcal{P}A(\partial)z_1 = \sum_{\#J=1} \mathcal{P}T(v_J), \tag{6.10}$$

$$\partial_t z_1 + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}_1 z_1 + \frac{1}{\varepsilon}\mathcal{P}^\perp B(\partial)(y_1 + z_1) = \sum_{\#J=1} \mathcal{P}^\perp T(v_J), \tag{6.11}$$

respectively. With Remark 6.1.1 the equation (6.11) is equivalent to

$$z_1 = \mathrm{i}\varepsilon\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(\varepsilon\partial_t z_1 + B(\partial)(y_1 + z_1) - \varepsilon \sum_{\#J=1} T(v_J)\Big). \tag{6.12}$$

We note that the equations (6.8), (6.10) and (6.12) still depend on ε. The idea is to introduce a formal expansion as in (6.1) with respect to ε for the coefficients v_j with $j \in \mathcal{J}_+(j_\star)$, truncate the expansion and then collect the terms of the same powers of ε.

6.1.2 Asymptotic expansion

The next goal is to approximate the solution of the PDE system (5.3) with coefficient functions $\widetilde{v}_j$ which fulfill the required properties (6.7) and which satisfy the initial condition $\mathcal{P}\widetilde{v}_1(0) = p$ for a given smooth function p. Because of the conditions (6.7), we specify (6.1) and suggest an approximation of the form

$$v_j(t,x) \approx \widetilde{v}_j(t,x) := \sum_{\ell=j-1}^{j_\star} \varepsilon^\ell v_j^\ell(t,x), \qquad j \in \mathcal{J}_+(j_\star). \tag{6.13}$$

Here, $j_\star$ is the same number as in the definition (6.5) of $\mathcal{J}(j_\star)$. As a notational remark we point out that ℓ is a power on ε and a superscript on v_j.

If we substitute (6.13) into (5.4), we obtain a *modulated Fourier expansion*

$$\widetilde{\mathbf{v}}^{(j_\star)}(t,x) := \sum_{j \in \mathcal{J}(j_\star)} e^{\mathrm{i}j\kappa\cdot x/\varepsilon} e^{\mathrm{i}j(\kappa\cdot c_g - \omega)t/\varepsilon^2} \widetilde{v}_j(t,x) \tag{6.14}$$

$$= \sum_{j \in \mathcal{J}(j_\star)} e^{\mathrm{i}j\kappa\cdot x/\varepsilon} e^{\mathrm{i}j(\kappa\cdot c_g - \omega)t/\varepsilon^2} \sum_{\ell=j-1}^{j_\star} \varepsilon^\ell v_j^\ell(t,x) \tag{6.15}$$

with $v_{-j}^{\ell} = \overline{v_j^{\ell}}$. In the following we construct the coefficients v_j^{ℓ} such that the MFE is an approximation to the solution of (5.1).

As an example for the reader we consider the asymptotic expansion (6.13) for $j_\star = 5$. In this case we have $\mathcal{J}_+(j_\star) = \{1,3,5\}$ and (6.13) reads

$$\begin{aligned} v_1 &\approx v_1^0 + \varepsilon v_1^1 + \varepsilon^2 v_1^2 + \varepsilon^3 v_1^3 + \varepsilon^4 v_1^4 + \varepsilon^5 v_1^5, \\ v_3 &\approx \varepsilon^2 v_3^2 + \varepsilon^3 v_3^3 + \varepsilon^4 v_3^4 + \varepsilon^5 v_3^5, \\ v_5 &\approx \varepsilon^4 v_5^4 + \varepsilon^5 v_5^5. \end{aligned}$$

As before, negative subscripts mean complex conjugation for $j \in \mathcal{J}_+(j_\star)$. Hence, in accordance with (6.9) we set

$$y_1^{\ell} = \mathcal{P} v_1^{\ell}, \quad y_{-1}^{\ell} = \overline{y_1^{\ell}}, \quad z_1^{\ell} = \mathcal{P}^{\perp} v_1^{\ell}, \quad z_{-1}^{\ell} = \overline{z_1^{\ell}} \tag{6.16}$$

and obtain by multiplying v_1 with $\mathcal{P}$ and $\mathcal{P}^{\perp}$

$$\mathcal{P}\widetilde{v}_1 = y_1^0 + \varepsilon y_1^1 + \varepsilon^2 y_1^2 + \ldots + \varepsilon^{j_\star} y_1^{j_\star}, \tag{6.17}$$

$$\mathcal{P}^{\perp}\widetilde{v}_1 = \ 0 \ + \varepsilon z_1^1 + \varepsilon^2 z_1^2 + \ldots + \varepsilon^{j_\star} z_1^{j_\star}, \tag{6.18}$$

respectively. Moreover, we set

$$\widetilde{v}_j = \ 0 \ + \ldots + 0 + \varepsilon^{j-1} v_j^{j-1} + \varepsilon^j v_j^j + \ldots + \varepsilon^{j_\star} v_j^{j_\star} \qquad \text{for } j > 1, \tag{6.19}$$

$$z_{\pm 1}^0 = 0, \qquad v_{\pm j}^{\ell} = 0 \quad \text{for } \ell < |j| - 1, \tag{6.20}$$

which implies, in particular, that $v_{\pm 1}^0 = y_{\pm 1}^0$. Furthermore, in accordance with the initial condition $\mathcal{P}\widetilde{v}_1(0) = p$ and (6.17), there is also an expansion

$$p(x) = \sum_{\ell=0}^{j_\star} \varepsilon^{\ell} p^{\ell}(x).$$

For multi-indices $J = (j_1, j_2, j_3) \in \mathcal{J}^3(j_\star)$ and $L = (\ell_1, \ell_2, \ell_3) \in \mathbb{N}_0^3$ we introduce the abbreviated notation

$$v_J^L = \big(v_{j_1}^{\ell_1}, v_{j_2}^{\ell_2}, v_{j_3}^{\ell_3}\big).$$

For the rest of the section we derive equations for the coefficient functions v_j^{ℓ} for $j \in \mathcal{J}_+(j_\star)$ and $\ell \in \{j-1, \ldots, j_\star\}$. The strategy is divided into two parts. First, we plug the expansions (6.13), (6.17), (6.18), and (6.19) into (6.10), (6.12) and (6.8). Then, we collect the terms of the same order in ε.

Equations for y_1^0 and z_1^1. For the first two coefficients y_1^0 and z_1^1 we obtain

$$\partial_t y_1^0 + \mathcal{P} A(\partial) z_1^1 = \sum_{\substack{\#J=1 \\ |L|_1=0}} \mathcal{P} T(v_J^L), \tag{6.21}$$

$$z_1^1 = \mathrm{i}\mathcal{L}_1^{-1} \mathcal{P}^{\perp} B(\partial) y_1^0. \tag{6.22}$$

Here, we compare for (6.21) powers of $\varepsilon^0 = 1$ in (6.10) and for (6.22) powers of ε^1 in (6.12). Substituting the algebraic equation (6.22) into (6.21) yields

$$\partial_t y_1^0 + \mathrm{i}\mathcal{P} A(\partial) \mathcal{L}_1^{-1} \mathcal{P}^{\perp} B(\partial) y_1^0 = \sum_{\substack{\#J=1 \\ |L|_1=0}} \mathcal{P} T(v_J^L). \tag{6.23}$$

By definition (5.2) of $B(\partial)$ and because $\mathcal{P}^\perp y_1^0 = 0$ according to (6.16), it follows that

$$\mathcal{P}^\perp B(\partial) y_1^0 = \mathcal{P}^\perp A(\partial) y_1^0 - (c_g \cdot \nabla)\mathcal{P}^\perp y_1^0 = \mathcal{P}^\perp A(\partial) y_1^0. \tag{6.24}$$

Next, substituting (6.24) and (6.4) from Lemma 6.1.2 into (6.23) leads to the nonlinear Schrödinger equation

$$\partial_t y_1^0 - \frac{\mathrm{i}}{2}\nabla \cdot H \nabla y_1^0 = \sum_{\substack{\#J=1 \\ |L|_1=0}} \mathcal{P} T(v_J^L). \tag{6.25}$$

We note that in comparison to the classical nonlinear Schrödinger equation there is no imaginary unit in front of the nonlinearity. Since $|L|_1 = 0$ is only true for the multi-index $L = (0,0,0)$, we observe that the nonlinearity on the right-hand side of (6.25)

$$\sum_{\substack{\#J=1 \\ |L|_1=0}} T(v_J^L) = T(y_1^0, y_1^0, y_{-1}^0) + T(y_1^0, y_{-1}^0, y_1^0) + T(y_{-1}^0, y_1^0, y_1^0)$$

depends only on y_1^0 and $y_{-1}^0 = \overline{y_1^0}$. Consequently, (6.25) is independent of z_1^1 and, thus, we first solve (6.25) with initial value $y_1^0(0,x) = p^0(x)$ to determine y_1^0. Then, we compute z_1^1 from (6.22) because the algebraic equation only depends on y_1^0.

Equations for y_1^1 and z_1^2. Comparing in (6.10) powers of ε^1 and in (6.12) powers of ε^2 yield for the next coefficients

$$\partial_t y_1^1 + \mathcal{P} A(\partial) z_1^2 = \sum_{\substack{\#J=1 \\ |L|_1=1}} \mathcal{P} T(v_J^L), \tag{6.26}$$

$$z_1^2 = \mathrm{i}\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(B(\partial)(y_1^1 + z_1^1) - \sum_{\substack{\#J=1 \\ |L|_1=0}} T(v_J^L)\Big). \tag{6.27}$$

Plugging the algebraic equation (6.27) into (6.26) and proceeding as before leads to

$$\partial_t y_1^1 - \frac{\mathrm{i}}{2}\nabla \cdot H \nabla y_1^1 + \mathrm{i}\mathcal{P} A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(B(\partial) z_1^1 - \sum_{\substack{\#J=1 \\ |L|_1=0}} T(v_J^L)\Big) = \sum_{\substack{\#J=1 \\ |L|_1=1}} \mathcal{P} T(v_J^L). \tag{6.28}$$

Next, we consider the sum on the right-hand side of (6.28) without the projection $\mathcal{P}$ in more detail. We observe that

$$\begin{aligned}\sum_{\substack{\#J=1 \\ |L|_1=1}} T(v_J^L) = {} & T(v_1^1, v_1^0, v_{-1}^0) + T(v_1^1, v_{-1}^0, v_1^0) + T(v_{-1}^0, v_1^1, v_1^0) + T(v_1^0, v_1^1, v_{-1}^0) + T(v_1^0, v_{-1}^0, v_1^1) \\ & + T(v_{-1}^0, v_1^0, v_1^1) + T(v_1^0, v_1^0, v_{-1}^1) + T(v_1^0, v_{-1}^1, v_1^0) + T(v_{-1}^1, v_1^0, v_1^0)\end{aligned}$$

depends only on y_1^0 and $v_1^1 = y_1^1 + z_1^1$. Since v_1^1 appears exactly once in each evaluation of the trilinearity and since z_1^1 does not depend on y_1^1, cf. (6.22), the PDE (6.28) is a *linear* inhomogeneous Schrödinger equation for y_1^1. Again we note that there is no imaginary unit on the right-hand side. Thus, first we solve (6.28) with initial value $y_1^1(0,x) = p^1(x)$ to obtain y_1^1 and, then we compute z_1^2 from the algebraic equation for (6.27). The procedure to date is summarized in Figure 6.1.

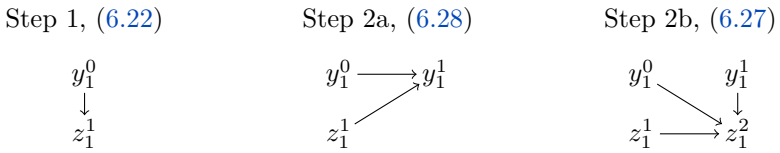


Figure 6.1: Illustration for the first two steps, showing which coefficients are needed for the construction and in which order.

Equations for y_1^2, z_1^3, and v_3^2. Similarly to the first two steps, we obtain

$$\partial_t y_1^2 + \mathcal{P}A(\partial) z_1^3 = \sum_{\substack{\#J=1 \\ |L|_1=2}} \mathcal{P}T(v_J^L), \tag{6.29}$$

$$z_1^3 = \mathrm{i}\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(\partial_t z_1^1 + B(\partial)(y_1^2 + z_1^2) - \sum_{\substack{\#J=1 \\ |L|_1=1}} T(v_J^L)\Big), \tag{6.30}$$

where we compare in (6.10) powers of ε^2 and in (6.12) powers of ε^3. We emphasize that there appears a structural change for the first time because there is now a contribution from the time-derivative $\partial_t z_1$ on the right-hand side of (6.12). With the same reasoning as before, substituting the algebraic equation (6.30) into (6.29) leads to

$$\partial_t y_1^2 - \frac{\mathrm{i}}{2}\nabla \cdot H\nabla y_1^2 + \mathrm{i}\mathcal{P}A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(\partial_t z_1^1 + B(\partial) z_1^2 - \sum_{\substack{\#J=1 \\ |L|_1=1}} T(v_J^L)\Big) = \sum_{\substack{\#J=1 \\ |L|_1=2}} \mathcal{P}T(v_J^L). \tag{6.31}$$

Since v_1^2 appears exactly once in each evaluation of the trilinearity and since z_1^2 does not depend on y_1^2, cf. (6.27), the PDE (6.31) is again a linear inhomogeneous Schrödinger equation for y_1^2. However, in contrast to the first two steps the sum

$$\sum_{\substack{\#J=1 \\ |L|_1=2}} \mathcal{P}T(v_J^L)$$

on the right-hand side of (6.31) involves not only terms $v_{\pm 1}^\ell$ with subscript ± 1. For $|L|_1 = 2$ the multi-index $L = (2,0,0)$ occurs and combined with $J = (3,-1,-1)$ there are $\mathcal{P}T(v_J^L) = \mathcal{P}T(v_3^2, v_{-1}^0, v_{-1}^0)$ and two other terms with permuted arguments. Hence, now v_3^2 has to be computed from (6.8) before we are able to solve (6.31). Since by construction some terms are assumed to be zero, cf. (6.20), it follows that $\partial_t v_3^0 = 0$ and $v_3^1 = 0$. Hence, we obtain from (6.8)

$$v_3^2 = -\mathrm{i}\mathcal{L}_3^{-1} \sum_{\substack{\#J=3 \\ |L|_1=0}} T(v_J^L). \tag{6.32}$$

The sum

$$\sum_{\substack{\#J=3 \\ |L|_1=0}} T(v_J^L) = T(v_1^0, v_1^0, v_1^0) = T(y_1^0, y_1^0, y_1^0)$$

on the right-hand side of (6.32) depends only on y_1^0, which is already available, by solving the nonlinear Schrödinger equation (6.25). Thus, after computing v_3^2, we solve (6.31) with initial value $y_1^2(0,x) = p^2(x)$

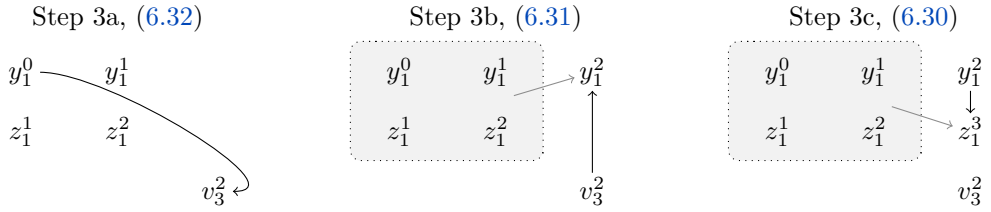


Figure 6.2: Illustration for the third step, showing which coefficients are needed for the construction and in which order. The gray shading means that all containing coefficients are required.

to determine y_1^2, and then, we compute z_1^3 from (6.30), cf. Figure 6.2.

From this point on, it becomes clear how we iteratively construct the remaining coefficients for $\ell = 3, \dots, j_\star$.

Equations for y_1^ℓ, $z_1^{\ell+1}$, and v_j^ℓ. For $\ell = 3, \dots, j_\star$ and $j \in \mathcal{J}_+(j_\star)$ with $3 \leqslant j \leqslant \ell + 1$ we set

$$v_j^\ell = \mathrm{i}\mathcal{L}_j^{-1}\Big(\partial_t v_j^{\ell-2} + B(\partial) v_j^{\ell-1} - \sum_{\substack{\#J=j \\ |L|_1=\ell-2}} T(v_J^L)\Big). \tag{6.33}$$

All the required coefficients in the sum of (6.33) are already constructed because $|L|_1 = \ell - 2$ and, hence, we compute v_j^ℓ with (6.33) for $j \neq \pm 1$. Further, we obtain y_1^ℓ for $\ell = 3, \dots, j_\star$ by solving the PDE

$$\partial_t y_1^\ell - \frac{\mathrm{i}}{2}\nabla \cdot H \nabla y_1^\ell + \mathcal{P}A(\partial)\Big(\mathrm{i}\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(\partial_t z_1^{\ell-1} + B(\partial) z_1^\ell - \sum_{\substack{\#J=1 \\ |L|_1=\ell-1}} T(v_J^L)\Big)\Big) = \sum_{\substack{\#J=1 \\ |L|_1=\ell}} \mathcal{P}T(v_J^L) \tag{6.34}$$

with initial value $y_1^\ell(0,x) = p^\ell(x)$. This is possible because y_1^ℓ appears not more than once in each evaluation of the trilinearity on the right-hand side of (6.34), and since z_1^ℓ is independent of y_1^ℓ and already computed in the previous step. Thus, y_1^ℓ is available and we set

$$z_1^{\ell+1} = \mathrm{i}\mathcal{L}_1^{-1}\mathcal{P}^\perp\Big(\partial_t z_1^{\ell-1} + B(\partial)(y_1^\ell + z_1^\ell) - \sum_{\substack{\#J=1 \\ |L|_1=\ell-1}} T(v_J^L)\Big). \tag{6.35}$$

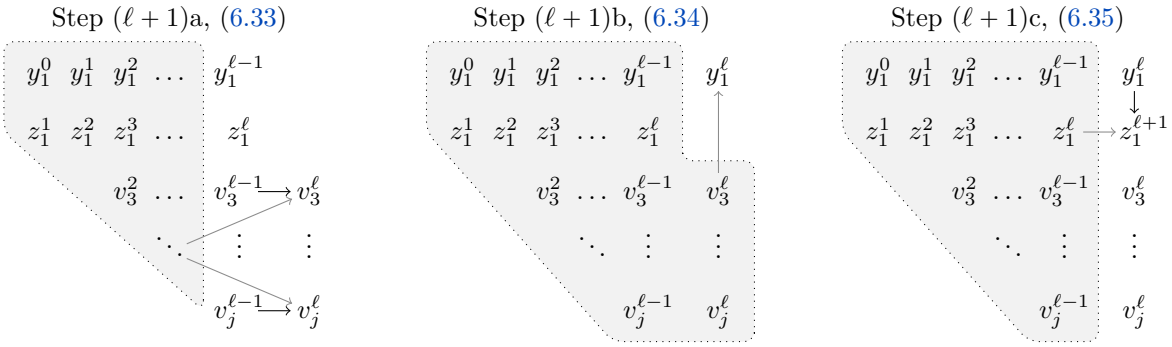


Figure 6.3: Illustration for the three main steps and arbitrary $\ell \geqslant 3$, showing which coefficients are needed for the construction and in which order. The gray shading means that all containing coefficients are required.

If $\ell = j_\star$, we point out that (6.35) does not have to be computed, because $z_1^{j_\star+1}$ appears neither in the ansatz (6.13) nor in the construction of the other quantities, see also Figure 6.3.

Remark 6.1.3. *We observe the crucial advantage of this approach. In comparison to the equations* (6.8), (6.10) *and* (6.12), *the equations* (6.22), (6.25), (6.27), (6.28), (6.30)–(6.35) *do not depend on* ε. *Thus, there are no* ε*-induced oscillations in time for* y_1^ℓ, z_1^ℓ *and* v_j^ℓ *for* $3 \leqslant j \in \mathcal{J}_+(j_\star)$ *and* $\ell = 0, \dots, j_\star$.

On the initial condition. We emphasize that the part of the constructed solution which satisfies the initial condition $\mathcal{P}\widetilde{v}_1(0) = p$, i.e. $\mathcal{P}v_1^\ell(0) = y_1^\ell(0)$, $\ell = 0, \dots, j_\star$, can be choosen arbitrarily because we choose the smooth function p. For this reason, we could also simply set $y_1^0(0) = p$. This would lead to the fact that all the linear inhomogeneous Schrödinger equations for y_1^ℓ, $\ell \geqslant 1$ in the above construction need only be solved with zero initial value, i.e. $y_1^\ell(0) = 0$ for $\ell \geqslant 1$. However, by construction this choice has a crucial impact on the other coefficients. It determines the initial data for $z_1^\ell(0)$ as well as for all $v_j^\ell(0)$ with $3 \leqslant j \in \mathcal{J}_+(j_\star)$ due the algebraic equations (6.22), (6.27), (6.30), (6.32), (6.35) and (6.33). This relation between the different parts of the initial data can be seen as a *nonlinear polarization condition*. We remark that the parts of the initial data which are determined by p are only of size $\mathcal{O}(\varepsilon^\ell)$ with $\ell \geqslant 1$ and that the initial data

$$\mathbf{v}(0,x) = \sum_{j \in \mathcal{J}(j_\star)} e^{\mathrm{i}j\kappa\cdot x/\varepsilon}\widetilde{v}_j(0,x) = \mathcal{P}\widetilde{v}_1(0,x)e^{\mathrm{i}\kappa\cdot x/\varepsilon} + c.c. + \mathcal{O}(\varepsilon) = p(x)e^{\mathrm{i}\kappa\cdot x/\varepsilon} + c.c. + \mathcal{O}(\varepsilon)$$

is determined by p up to $\mathcal{O}(\varepsilon^{j_\star+1})$. We call this special initial data $\mathbf{v}(0)$ a *polarized initial data* since for those initial data the other eigenvalues $\omega_m(\kappa)$ with $m \neq 1$ of $\mathcal{L}(0,\kappa)$ do not enter the oscillatory exponentials of the MFE. Hence, we indeed achieve polarized solutions of the form (6.6) for the semilinear system (5.1).

6.2 Accuracy of the modulated Fourier expansion

In this section, we analyze the accuracy of the modulated Fourier expansion (6.15) in the Wiener algebra, cf. Section 3.4.

6.2.1 Boundedness of the coefficient functions

The MFE (6.15) can only provide a reasonable approximation if all coefficient functions y_1^ℓ, z_1^ℓ, v_j^ℓ with $3 \leqslant j \in \mathcal{J}_+(j_\star)$ and for $\ell = 1, \dots, j_\star$, remain bounded on $[0, t_{\mathrm{end}}]$. We recall that the first step in the construction from Section 6.1.2 is to compute y_1^0 by solving the nonlinear Schrödinger equation (6.25). Via classical arguments such as the variation of constants formula and Banach's fixed point theorem the existence of a mild solution of (6.25) can be shown. With the standard argument that we can glue and shift solutions, the mild solution can then be extended to a maximal time interval. We note that this maximal time interval could, in principle, be smaller than the interval $[0, t_{\mathrm{end}}]$ where the exact solution of (5.1) exists. However, in the following we assume that the mild solution exists on $[0, t_{\mathrm{end}}]$. Under stronger regularity assumptions on the initial data the mild solution is in fact a classical solution with a

certain degree of regularity. Therefore, we further assume for the solution of (6.25) that

$$y_1^0 \in X^r := \bigcap_{i=0}^{\lfloor r/2 \rfloor} C^i([0,t_{\mathrm{end}}], W^{r-2i})$$

for a sufficiently large number $r \in \mathbb{N}$, which will be specified at the end of this subsection. With $\lfloor r/2 \rfloor$ we denote the largest integer which is not larger than $r/2$.

We recall that the construction of the coefficient functions is done iteratively with increasing ℓ and hence, all coefficient functions can be traced back to y_1^0. This implies that their regularity also depends on r in the sense that the larger ℓ becomes, the more regularity is required which is seen next. We start with z_1^1 which is given by the algebraic equation (6.22). Because of the operator $B(\partial)$ on the right-hand side of (6.22) it follows that $z_1^1 \in X^{r-1}$ if $y_1^0 \in X^r$. Following the construction, the next step was to compute y_1^1 by solving the linear inhomogeneous Schrödinger equation (6.28). In order to investigate the regularity of y_1^1 the following classical result from semigroup theory is helpful; cf. Section 4.2 in [36].

Lemma 6.2.1. *Let A be the generator of a strongly continuous semigroup on a Banach space $\mathcal{X}$. Let $D(A) \subset \mathcal{X}$ denote the domain of A and suppose that $u_0 \in D(A)$. If*

$$f \in C^1([0,t_{end}], \mathcal{X}) \qquad or \qquad f \in C([0,t_{end}], D(A)),$$

then the inhomogeneous abstract Cauchy problem

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

has a unique classical solution

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

Taking a closer look at (6.28), we note that the inhomogeneity contains the crucial term

$$-\mathrm{i}\mathcal{P}A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^{\perp}B(\partial)z_1^1 \in X^{r-3}$$

and other terms such as

$$-\mathrm{i}\mathcal{P}A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^{\perp}T(y_1^0, y_1^0, y_{-1}^0) \in X^{r-1}$$

with higher regularity. If we assume $y_1^1(0) \in W^{r-3} = D(A)$ and set $\mathcal{X} = X^{r-5}$ with $r \geqslant 5$, then applying Lemma 6.2.1 yields $y_1^1 \in X^{r-3}$. This in turn has an impact on z_1^2 defined by the algebraic equation (6.27). The right-hand side of (6.27) involves $B(\partial)y_1^1$ plus other terms of higher regularity, such that we have $z_1^2 \in X^{r-4}$.

The same reasoning can be used to treat the coefficient function y_1^2. In contrast to the first two steps of the construction, now the coefficient function v_3^2 comes into play. However, since the function v_3^2 defined by (6.32) depends only on $y_1^0 \in X^r$, we directly have $v_3^2 \in X^r$. Similarly as before, the crucial term in the inhomogeneity of the linear Schrödinger equation (6.31) has the form

$$-\mathrm{i}\mathcal{P}A(\partial)\mathcal{L}_1^{-1}\mathcal{P}^{\perp}B(\partial)z_1^2 \in X^{r-6}.$$

If we assume $y_1^2(0) \in W^{r-6} = D(A)$ and set $\mathcal{X} = X^{r-8}$ with $r \geqslant 8$, we obtain $y_1^2 \in X^{r-6}$ by means of Lemma 6.2.1. Because of the operator $B(\partial)$ an immediate consequence of the algebraic equation (6.30) is that $z_1^3 \in X^{r-7}$.

By continuing this procedure iteratively, we know how to treat the remaining coefficient functions for $\ell = 3, \dots, j_\star$. Consequently, all coefficient functions involved in the approximation (6.15) are bounded if r is sufficiently large and all initial data $y_1^\ell(0)$ for $\ell = 0, \dots, j_\star$ are sufficiently smooth. Comparing $y_1^0 \in X^r$ with $y_1^1 \in X^{r-3} = X^{r-3\cdot 1}$ and $y_1^2 \in X^{r-6} = X^{r-3\cdot 2}$ implies, roughly speaking, that three orders of regularity are required if the upper index of y_1^ℓ is increased by one. In consequence we obtain $y_1^\ell \in X^{r-3\ell}$, such that we have $y_1^{j_\star} \in X^0$, if $y_1^0 \in X^{3j_\star}$. However, since we need a classical solution and boundedness of $\partial_t v_j^{j_\star}$ the number r has to be $r = 3j_\star + 2$.

6.2.2 Error analysis for the MFE

In this section, we state the main result of this chapter in Theorem 6.2.3. In order to prove that the modulated Fourier expansion (6.15) with $j_\star = j_{\max}$ approximates a solution of (5.1) up to $\mathcal{O}\big(\varepsilon^{j_{\max}+1}\big)$, we require the following assumption. We note that the assumption includes coefficients of the MFE (6.15) with $j_\star = j_{\max} + 2$. In the proof for the error bound it becomes clear why we need this assumption and not just for the coefficients with $j \in \mathcal{J}_+(j_{\max})$ and $\ell = 0, \dots, j_{\max}$.

Assumption 6.2.2. *The initial conditions satisfy $p^\ell = y_1^\ell(0) \in W^{3(j_{max}+2-\ell)+2}$ for $\ell = 0, \dots, j_{max}+2$ and there exist constants C independent of ε such that*

$$\sup_{t\in[0,t_{end}]} \|v_j^\ell(t)\|_W \leqslant C, \qquad \ell = 0, \dots, j_{max} + 2,$$

for all $j \in \mathcal{J}(j_{max}+2)$ with v_j^ℓ constructed as in Subsection 6.1.2. The same bounds are true for the mixed spatio-temporal partial derivatives of these coefficient functions up to a fixed order depending on j_{max}.

Because of the previous subsection and Lemma 6.2.1, we know that this assumption on the boundedness of the coefficients is reasonable and can be shown with high technical effort since the initial data $y_1^\ell(0)$ for $\ell = 0, \dots, j_{\max} + 2$ are sufficiently smooth in space.

Theorem 6.2.3. *Let $\widetilde{\mathbf{v}}^{(j_{max}+2)}$ be the constructed MFE (6.15) with $j_\star = j_{max}+2$ and with polarized initial data $\widetilde{\mathbf{v}}^{(j_{max}+2)}(0)$. Assume that a unique solution $\mathbf{v}$ of (5.1) with initial data $\mathbf{v}(0) = \widetilde{\mathbf{v}}^{(j_{max})}(0)$ exists on the time interval $[0, t_{end}]$, and that*

$$\sup_{t\in[0,t_{end}]} \|\mathbf{v}(t)\|_W \leqslant C_{\mathbf{v}} \tag{6.36}$$

uniformly in $\varepsilon \in (0, 1]$. Under Assumptions 3.2.2, 3.2.6 and 6.2.2, there exists a constant such that

$$\sup_{t\in[0,t_{end}]} \|\mathbf{v}(t) - \widetilde{\mathbf{v}}^{(j_{max})}(t)\|_W \leqslant C\varepsilon^{j_{max}+1}, \tag{6.37}$$

$$\sup_{t\in[0,t_{end}]} \|\mathbf{v}(t) - \widetilde{\mathbf{v}}^{(j_{max})}(t)\|_{L^\infty} \leqslant C\varepsilon^{j_{max}+1}, \tag{6.38}$$

where the constant C is independent of ε.

Proof. The second bound (6.38) is an immediate consequence of the embedding $W \hookrightarrow L^\infty$ and, thus, we only have to show (6.37). First, under Assumption 6.2.2 we estimate

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)\|_W \leqslant C_{\mathbf{v}}, \tag{6.39}$$

in addition to (6.36), possibly with a larger constant $C_{\mathbf{v}}$.
By the triangle inequality we obtain

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|\mathbf{v}(t)-\widetilde{\mathbf{v}}^{(j_{\max})}(t)\|_W \leqslant \sup_{t\in[0,t_{\mathrm{end}}]} \|\mathbf{v}(t)-\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)\|_W + \sup_{t\in[0,t_{\mathrm{end}}]} \|\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)-\widetilde{\mathbf{v}}^{(j_{\max})}(t)\|_W.$$

A detailed justification of why this decomposition is necessary is stated after the proof in Remark 6.2.4.
The goal is now to show

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|\mathbf{v}(t)-\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)\|_W \leqslant C\varepsilon^{j_{\max}+1} \tag{6.40}$$

and

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)-\widetilde{\mathbf{v}}^{(j_{\max})}(t)\|_W \leqslant C\varepsilon^{j_{\max}+1}. \tag{6.41}$$

Since the proof of Theorem 6.2.3 is rather lengthy, we subdivide it into several steps.

Step 1. We denote the error between the exact solution $\mathbf{v}$ and the approximation (6.15) with $j_\star = j_{\max}+2$ by

$$\delta = \mathbf{v} - \widetilde{\mathbf{v}}^{(j_{\max}+2)}. \tag{6.42}$$

The first goal is to derive an evolution equation for δ. The approximation (6.15) with $j_\star = j_{\max}+2$ solves (5.1) up to the residual

$$\begin{aligned} R(t,x) = \partial_t\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) + \frac{1}{\varepsilon}B(\partial)\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) + \frac{1}{\varepsilon^2}E\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) \\ - T(\widetilde{\mathbf{v}}^{(j_{\max}+2)},\widetilde{\mathbf{v}}^{(j_{\max}+2)},\widetilde{\mathbf{v}}^{(j_{\max}+2)})(t,x). \end{aligned}$$

Substituting (6.14) into the left-hand side of (5.1) yields

$$\begin{aligned} &\partial_t\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) + \frac{1}{\varepsilon}B(\partial)\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) + \frac{1}{\varepsilon^2}E\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t,x) \\ &\qquad = \sum_{j\in\mathcal{J}(j_{\max}+2)} e^{\mathrm{i}j\kappa\cdot x/\varepsilon}e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t/\varepsilon^2}\left(\partial_t\widetilde{v}_j(t,x) + \frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}_j\widetilde{v}_j(t,x) + \frac{1}{\varepsilon}B(\partial)\widetilde{v}_j(t,x)\right). \end{aligned} \tag{6.43}$$

Comparing (6.43) with

$$\begin{aligned} T(\widetilde{\mathbf{v}}^{(j_{\max}+2)},\widetilde{\mathbf{v}}^{(j_{\max}+2)},\widetilde{\mathbf{v}}^{(j_{\max}+2)})(t,x) &= \sum_{J\in\mathcal{J}^3(j_{\max}+2)} e^{\mathrm{i}\#J(\kappa\cdot x)/\varepsilon}e^{\mathrm{i}\#J(\kappa\cdot c_g-\omega)t/\varepsilon^2}T(\widetilde{v}_{j_1},\widetilde{v}_{j_2},\widetilde{v}_{j_3})(t,x) \\ &= \sum_{\substack{j \text{ odd}\\ |j|\leqslant 3j_{\max}+6}} e^{\mathrm{i}j(\kappa\cdot x)/\varepsilon}e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t/\varepsilon^2}\sum_{\#J=j} T(\widetilde{v}_{j_1},\widetilde{v}_{j_2},\widetilde{v}_{j_3})(t,x) \end{aligned}$$

shows that the residual R has the representation

$$R(t,x) = \sum_{\substack{j \text{ odd}\\ |j|\leqslant 3j_{\max}+6}} e^{\mathrm{i}j(\kappa\cdot x)/\varepsilon}e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t/\varepsilon^2}R_j(t,x) \tag{6.44}$$

with $R_{-j}=\overline{R_j}$. The defects R_j have two different origins depending on j. For $j\leqslant j_{\max}+2$ the defects R_j are caused by solving the PDEs (5.3), (6.10), (6.11) only approximately. The defects R_j with $j > j_{\max}+2$

originate from the fact that only the terms with index $j \in \mathcal{J} = \{\pm 1, \pm 3, \dots, \pm j_{\max} \pm 2\}$ are used in the ansatz (5.4). Hence, we obtain

$$R_j(t,x) = \begin{cases} \partial_t \widetilde{v}_j(t,x) + \left[\frac{\mathrm{i}}{\varepsilon^2}\mathcal{L}_j + \frac{1}{\varepsilon}B(\partial)\right]\widetilde{v}_j(t,x) - \sum_{\#J=j} T(\widetilde{v}_{j_1}, \widetilde{v}_{j_2}, \widetilde{v}_{j_3})(t,x) & \text{if } 1 \leqslant j \leqslant j_{\max}+2, \\ -\sum_{\#J=j} T(\widetilde{v}_{j_1}, \widetilde{v}_{j_2}, \widetilde{v}_{j_3})(t,x) & \text{if } j_{\max}+2 < j \leqslant 3j_{\max}+6. \end{cases}$$

By definition (6.44) of the residual R, the error $\delta = \mathbf{v} - \widetilde{\mathbf{v}}^{(j_{\max}+2)}$ solves the evolution equation

$$\partial_t \delta = -\frac{1}{\varepsilon}B(\partial)\delta - \frac{1}{\varepsilon^2}E\delta + \left[T(\mathbf{v},\mathbf{v},\mathbf{v}) - T(\widetilde{\mathbf{v}}^{(j_{\max}+2)}, \widetilde{\mathbf{v}}^{(j_{\max}+2)}, \widetilde{\mathbf{v}}^{(j_{\max}+2)})\right] - R. \tag{6.45}$$

Step 2. In this step, we aim to bound δ via Gronwall's lemma. We observe that for every $\varepsilon > 0$ the operator

$$\mathcal{A}_\varepsilon = -\frac{1}{\varepsilon}B(\partial) - \frac{1}{\varepsilon^2}E \quad \text{with domain} \quad D(\mathcal{A}_\varepsilon) = W^1$$

generates a strongly continuous group $\big(\exp(t\mathcal{A}_\varepsilon)\big)_{t\in\mathbb{R}}$ on W. The group operators are explicitly given by

$$\mathcal{F}\big(\exp(t\mathcal{A}_\varepsilon)f\big)(k) = \exp\left(-\frac{t}{\varepsilon^2}\big(\mathrm{i}\varepsilon B(k) + E\big)\right)\widehat{f}(k)$$

for every $f \in W$ and all $t \in \mathbb{R}$. We note that the matrix $\mathrm{i}\varepsilon B(k) + E$ is skew-Hermitian for every k, because $E \in \mathbb{R}^{s\times s}$ is skew-symmetric and $\mathrm{i}B(k)$ is skew-Hermitian. Hence, it follows that for every $t \in \mathbb{R}$ the group operator $\exp(t\mathcal{A}_\varepsilon) : W \to W$ is an isometry, because

$$\|\exp(t\mathcal{A}_\varepsilon)f\|_W = \big\|\mathcal{F}\big(\exp(t\mathcal{A}_\varepsilon)f\big)\big\|_{L^1} = \int_{\mathbb{R}^d} \Big|\exp\Big(-\tfrac{t}{\varepsilon^2}\big(\mathrm{i}\varepsilon B(k)+E\big)\Big)\widehat{f}(k)\Big|_2\, dk = \int_{\mathbb{R}^d} |\widehat{f}(k)|_2\, dk = \|f\|_W.$$

With the variation-of-constants formula applied to (6.45) and with the trilinear estimate (3.31), we obtain

$$\|\delta(t)\|_W \leqslant \|\delta(0)\|_W + 3C_T C_{\mathbf{v}}^2 \varepsilon \int_0^t \|\delta(\sigma)\|_W \,\mathrm{d}\sigma + \int_0^t \|R(\sigma)\|_W \,\mathrm{d}\sigma,$$

where $C_{\mathbf{v}}$ is the constant from the assumptions (6.36) and (6.39). By assumption we have for the initial data $\mathbf{v}(0) = \widetilde{\mathbf{v}}^{(j_{\max})}(0)$ such that $\delta(0) = \widetilde{\mathbf{v}}^{(j_{\max})}(0) - \widetilde{\mathbf{v}}^{(j_{\max}+2)}(0)$. Thus, with (6.14) and (6.15) at $t = 0$ it follows

$$\delta(0) = \sum_{j\in\mathcal{J}(j_{\max})} e^{\mathrm{i}jk\cdot x/\varepsilon} \sum_{\ell=j_{\max}+1}^{j_{\max}+2} \varepsilon^\ell v_j^\ell(0,x) + \Big(e^{\mathrm{i}(j_{\max}+2)\kappa\cdot x/\varepsilon}\widetilde{v}_{j_{\max}+2}(0,x) + c.c.\Big).$$

By Assumption 6.2.2 all coefficients are uniformly bounded and we obtain with the condition (6.7) for $j = j_{\max}+2$

$$\|\delta(0)\|_W \leqslant C\varepsilon^{j_{\max}+1}.$$

It remains to show that

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|R(t)\|_W \leqslant C\varepsilon^{j_{\max}+1} \tag{6.46}$$

with a constant C which does not depend on ε. If (6.46) holds, then we obtain

$$\|\delta(t)\|_W \leqslant C\varepsilon^{j_{\max}+1} + 3C_{\mathbf{v}}^2\varepsilon \int_0^t \|\delta(\sigma)\|_W \,\mathrm{d}\sigma + Ct_{\mathrm{end}}\varepsilon^{j_{\max}+1}$$

such that Gronwall's lemma and (6.42) yields

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|\mathbf{v}(t) - \widetilde{\mathbf{v}}^{(j_{\max}+2)}(t)\|_W \leqslant C\varepsilon^{j_{\max}+1}.$$

Step 3. In this step, the goal is to show (6.46) by estimating the single contributions in (6.44) with

$$\sup_{t\in[0,t_{\mathrm{end}}]} \|R_j(t)\|_W \leqslant C\varepsilon^{j_{\max}+1} \tag{6.47}$$

for all odd $j = 1, \dots, 3j_{\max}+6$ for some constant C. First, we consider the case $j > j_{\max}+2$. Assumption 6.2.2 and the expansion (6.13) imply that $\|\widetilde{v}_j(t)\|_W = \mathcal{O}\big(\varepsilon^{j-1}\big)$ for all $t \in [0, t_{\mathrm{end}}]$. Together with the trilinear estimate (3.30), we obtain

$$\|T(\widetilde{v}_{j_1}, \widetilde{v}_{j_2}, \widetilde{v}_{j_3})(t)\|_W \leqslant c\,\varepsilon^{j_1-1}\varepsilon^{j_2-1}\varepsilon^{j_3-1} = c\,\varepsilon^{\#J-3} \tag{6.48}$$

for some constant c which depends on $C_{\mathcal{T}}$ from (3.30). Since j and $j_{\max}+2$ are both odd numbers and we consider $j > j_{\max}+2$, we have $j \geqslant j_{\max}+4$ and we obtain with (6.48)

$$\|R_j(t)\|_W \leqslant \sum_{\#J=j} \|T(\widetilde{v}_{j_1}, \widetilde{v}_{j_2}, \widetilde{v}_{j_3})(t)\|_W \leqslant C\varepsilon^{j-3} \leqslant C\varepsilon^{j_{\max}+1} \qquad \text{for } j > j_{\max}+2.$$

The constant C depends on c from (6.48) and on the number of multi-indices $J \in \mathcal{J}^3(j_{\max}+2)$ with $\#J = j$.

Next, we consider the case $j \leqslant j_{\max}+2$. Here, compared to (6.13) with $j_\star = j_{\max}+2$, the defect R_j is also given by the expansion

$$R_j(t,x) = \sum_{\ell=j-1}^{j_{\max}+2} \varepsilon^\ell R_j^\ell(t,x) \tag{6.49}$$

with

$$R_j^\ell(t,x) = \partial_t v_j^\ell(t,x) + \mathrm{i}\mathcal{L}_j v_j^{\ell+2}(t,x) + B(\partial)v_j^{\ell+1}(t,x) - \sum_{\substack{\#J=j \\ |L|_1=\ell}} T(v_J^L)(t,x).$$

By construction of the coefficients y_1^ℓ, z_1^ℓ, v_j^ℓ with $3 \leqslant j \leqslant j_{\max}+2$, cf. Subsection 6.1.2, it follows that $R_j^\ell = 0$ for $0 \leqslant \ell \leqslant j_{\max}$ and $1 \leqslant j \leqslant j_{\max}+2$. Hence, all R_j^ℓ vanish except for

$$R_j^{j_{\max}+1}(t,x) = \partial_t v_j^{j_{\max}+1}(t,x) + B(\partial)v_j^{j_{\max}+2}(t,x) - \sum_{\substack{\#J=j \\ |L|_1=j_{\max}+1}} T(v_J^L)(t,x),$$

$$R_j^{j_{\max}+2}(t,x) = \partial_t v_j^{j_{\max}+2}(t,x) - \sum_{\substack{\#J=j \\ |L|_1=j_{\max}+2}} T(v_J^L)(t,x).$$

$\|R_j^{j_{\max}+1}(t)\|_W$ and $\|R_j^{j_{\max}+2}(t)\|_W$ remain uniformly bounded under Assumption 6.2.2 for all $t \in [0, t_{\mathrm{end}}]$. Hence, together with (6.49) the estimate (6.47) follows for all $j \leqslant j_{\max}+2$.
Combining the estimates from the two cases $j > j_{\max}+2$ and $j \leqslant j_{\max}+2$ proves (6.46), and thus (6.40).

Step 4. It remains to show (6.41). By construction of the coefficient functions v_j^ℓ for $j \in \mathcal{J}(j_{\max}+2)$ and $\ell = 0, \dots, j_{\max}+2$, the definitions (6.15) and (6.13), and by Assumption 6.2.2, we obtain

$$\begin{aligned}
&\sup_{t\in[0,t_{\mathrm{end}}]} \|\widetilde{\mathbf{v}}^{(j_{\max}+2)}(t) - \widetilde{\mathbf{v}}^{(j_{\max})}(t)\|_W \\
&\quad\leqslant \sup_{t\in[0,t_{\mathrm{end}}]} \Big\| \sum_{j\in\mathcal{J}(j_{\max})} e^{\mathrm{i}j\kappa\cdot x/\varepsilon} e^{\mathrm{i}j(\kappa\cdot c_g-\omega)t/\varepsilon^2} \sum_{\ell=j_{\max}+1}^{j_{\max}+2} \varepsilon^\ell v_j^\ell(t,x) \\
&\qquad\qquad + \Big(e^{\mathrm{i}(j_{\max}+2)\kappa\cdot x/\varepsilon} e^{\mathrm{i}(j_{\max}+2)(\kappa\cdot c_g-\omega)t/\varepsilon^2} \widetilde{v}_{j_{\max}+2}(t,x) + c.c. \Big) \Big\|_W \\
&\quad\leqslant C\varepsilon^{j_{\max}+1}.
\end{aligned}$$

This estimate shows (6.41) and yields together with (6.40) the assertion. ∎

We end this section with a remark.

Remark 6.2.4. *Consequently, all terms $y_1^\ell, z_1^\ell, v_j^\ell$ with $\ell \geqslant j_{max}+1$ and $j > j_{max}$ are actually not required in the approximation $\widetilde{\mathbf{v}}^{(j_{max})}$. We emphasize that omitting these terms does* not *change the terms $y_1^\ell, z_1^\ell, v_j^\ell$ with $\ell < j_{max}+1$ because for the construction of the coefficients we use in general (6.33)–(6.35). All these equations depend only on the previously constructed coefficients or, as in (6.34), additionally on coefficients with the same superscript ℓ. However, these terms cannot be omitted in (6.13) from the very beginning, meaning that $j_{max}+2$ cannot be replaced by j_{max} in the construction and the Assumption 6.2.2, since they are crucial for the defects in Step 3 of the proof of Theorem 6.2.3. We recall that by construction of the coefficient functions y_1^ℓ, z_1^ℓ, v_j^ℓ with $3 \leqslant j \leqslant j_{max}+2$ the defects R_j^ℓ are zero for $0 \leqslant \ell \leqslant j_{max}$ and $1 \leqslant j \leqslant j_{max}+2$. If we replace in (6.40) $\widetilde{\mathbf{v}}^{(j_{max}+2)}$ by $\widetilde{\mathbf{v}}^{(j_{max})}$ and proceed as in Step 3 of the proof, the right-hand side of the estimates (6.47) and (6.46) change from $\mathcal{O}\big(\varepsilon^{j_{max}+1}\big)$ to $\mathcal{O}\big(\varepsilon^{j_{max}-1}\big)$. Without the coefficient functions $v_j^{j_{max}+1}$ and $v_j^{j_{max}+2}$, only the defects R_j^ℓ for $0 \leqslant \ell \leqslant j_{max}-2$ and $1 \leqslant j \leqslant j_{max}$ vanish. The remaining defects $R_j^{j_{max}-1}$ and $R_j^{j_{max}}$ are given by*

$$\begin{aligned}
R_j^{j_{max}-1}(t,x) &= \partial_t v_j^{j_{max}-1}(t,x) + B(\partial) v_j^{j_{max}}(t,x) - \sum_{\substack{\#J=j \\ |L|_1=j_{max}-1}} T(v_J^L)(t,x), \\
R_j^{j_{max}}(t,x) &= \partial_t v_j^{j_{max}}(t,x) - \sum_{\substack{\#J=j \\ |L|_1=j_{max}}} T(v_J^L)(t,x)
\end{aligned}$$

and are nonzero.

CHAPTER 7

Summary and Outlook

In this thesis we provided analytical and numerical approximations to the specific semilinear hyperbolic system (1.4). We introduced in Chapter 3 a natural extension of the SVEA. The main advantage of this ansatz was that the solutions of the corresponding PDEs did not oscillate in space anymore. In previous works in the literature the SVEA offered a possibility to approximate the solution of (1.4) up to $\mathcal{O}(\varepsilon)$ which means that the accuracy of the SVEA is fixed a priori by the parameter ε. In some instances, however, a more accurate approximation is required. We were able to significantly improve the error bound of the SVEA under slightly stronger assumptions in our first main result in Chapter 4 up to $\mathcal{O}\big(\varepsilon^2\big)$ (Theorem 4.3.4), which made the SVEA attractive for numerical computations. These assumptions included non-resonance conditions that were important in order to apply integration by parts. However, the non-resonance conditions represented a limitation of the technique of proof for extensions to approximations with higher accuracy in $d > 1$, depending on the structure of the underlying problem. An interesting question would be whether it is possible to get rid of the limiting non-resonance conditions in order to achieve higher accuracy after all.

In Chapter 5 we constructed numerical methods for the SVEA. We first introduced a one-step method and then refined the time-integration method to obtain a two-step method. For both methods we provided a rigorous error analysis for the semi-discretization in time. One of our main results was Theorem 5.3.2, where we showed that approximating solutions of the SVEA by the two-step method with step-sizes $\tau > \varepsilon$ yields approximations of $\mathcal{O}(\tau^2)$. An increase in computational cost due to nested multiple sums was the price to pay, which we reduced by our cherry picking strategy. Here, the question of constructing better schemes without the nested multiple sums can be pursued, which are also applicable to problems with dimension $d > 1$, as e.g. the Maxwell–Lorentz system. Furthermore, we considered the torus $\mathbb{T}^d$ only for simplification. The incorporation of non-reflecting or absorbing boundary conditions and their investigation would be an extension for future work.

In Chapter 6 we complemented this thesis by constructing polarized solutions to the system (1.4). Their construction and analysis was done by means of modulated Fourier expansions and we also provided

an error bound in Theorem 6.2.3. The goal was to construct smooth coefficients that did not oscillate in space or time. The numerical approximation of these constructed coefficient functions by suitable methods remain an interesting problem for future research.

APPENDIX A

Gronwall lemma and function spaces

A.1 Gronwall lemma

For the convenience of the reader, we state a standard integral version of Gronwall's lemma without proof. The result and its proof is given in more general form in [32, Chapter 1, Theorem 8.1].

Lemma A.1.1 (Gronwall)**.** *Let $T_\star > 0$, $c_1, c_2 \geqslant 0$ and let u be a continuous, nonnegative function on $[0, T_\star]$. If*

$$u(t) \leqslant c_1 + c_2 \int_0^t u(\sigma)\,\mathrm{d}\sigma \quad \textit{for all } t \in [0, T_\star],$$

then

$$u(t) \leqslant c_1 e^{c_2 t} \quad \textit{for all } t \in [0, T_\star].$$

A.2 Function spaces

We give a short introduction to the functional analytical background used in this thesis. We define important spaces and state only those properties which are important for the thesis. All these results and proofs can be found in [6, 21, 34, 40].

The spaces L^1 and L^∞. Let $\left(\mathbb{R}^d, \mathcal{F}, \mu\right)$ denote a σ-finite measure space with $\mathcal{F}$ being the σ-algebra of measurable sets and μ equals the Lebesgue measure.

First, we define the space which consists of all (real-valued) absolutely Lebesgue integrable functions by

$$L^1(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} : f \text{ measurable}, \int_{\mathbb{R}^d} |f(x)|\,\mathrm{d}x < \infty \right\}$$

with norm

$$\|f\|_{L^1(\mathbb{R}^d)} := \int_{\mathbb{R}^d} |f(x)| \, \mathrm{d}x.$$

Moreover, we define

$$L^\infty(\mathbb{R}^d) := \big\{ f : \mathbb{R}^d \to \mathbb{R} : f \text{ measurable}, \, \|f\|_{L^\infty(\mathbb{R}^d)} < \infty \big\}$$

with the essential supremum norm

$$\|f\|_{L^\infty(\mathbb{R}^d)} := \inf \big\{ c \geqslant 0 \, : \, |f(x)| \leqslant c \text{ almost everywhere} \big\}.$$

For $p \in \{1, \infty\}$, $\| \cdot \|_{L^p}$ describes a norm on $L^p(\mathbb{R}^d)$ and the spaces $\big(L^p(\mathbb{R}^d), \| \cdot \|_{L^p(\mathbb{R}^d)}\big)$ are Banach spaces for $p \in \{1, \infty\}$.

Sequence space ℓ^1. If we take $\mathbb{Z}^d$ instead of $\mathbb{R}^d$ and μ equal to the counting measure, then we obtain a "discrete" version of the L^1 space which is denoted by ℓ^1. The space ℓ^1 will be useful when we introduce spaces on the torus instead of the full space. We define

$$\ell^1(\mathbb{Z}) = \big\{ x = (x_n)_{n \in \mathbb{Z}} \, : \, x_n \in \mathbb{C} \text{ for all } n \in \mathbb{Z} \text{ and } \|x\|_{\ell^1} < \infty \big\}$$

with norm

$$\|x\|_{\ell^1} = \sum_{n=-\infty}^{\infty} |x_n|.$$

We remark that $(\ell^1, \| \cdot \|_{\ell^1})$ is a Banach space.

Fourier transform on $L^1(\mathbb{R}^d)$ and the space $\mathcal{S}(\mathbb{R}^d)$. Recall that $L^1(\mathbb{R}^d)$ denotes the Banach space of functions that are absolutely integrable. For a function $f \in L^1(\mathbb{R}^d)$ its Fourier transform is defined by

$$(\mathcal{F}f)(k) = \widehat{f}(k) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-\mathrm{i}k \cdot x} \, \mathrm{d}x, \qquad k \in \mathbb{R}^d. \tag{A.1}$$

This and all following (in)equalities are to be understood for almost all $k \in \mathbb{R}^d$.
The inverse Fourier transform is given by

$$(\mathcal{F}^{-1}\widehat{f})(x) = f(x) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{f}(k) e^{\mathrm{i}k \cdot x} \, \mathrm{d}k, \qquad x \in \mathbb{R}^d. \tag{A.2}$$

For further investigations the Schwartz space

$$\mathcal{S}(\mathbb{R}^d) = \big\{ f \in C^\infty(\mathbb{R}^d) \, : \, \sup_{x \in \mathbb{R}^d} |x|_2^m |\partial^\alpha f(x)| < \infty \ \text{ for all } m \in \mathbb{N}_0, \alpha \in \mathbb{N}_0^d \big\}$$

turns out to be very useful. The Schwartz space contains all functions f which decay rapidly since all derivatives of f decay faster than $|x|_2^{-m}$ for any $m \in \mathbb{N}$, as $|x|_2 \to \infty$.

The dual of $\mathcal{S}(\mathbb{R}^d)$. The dual of $\mathcal{S}(\mathbb{R}^d)$ is denoted by $\mathcal{S}'(\mathbb{R}^d)$ and consists all continuous linear maps $f : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$. The elements of $\mathcal{S}'(\mathbb{R}^d)$ are called tempered distributions. The main advantage of tempered distributions is that they have a Fourier transform which is itself a tempered distribution. The Fourier transform of a distribution $f \in \mathcal{S}'(\mathbb{R}^d)$ is denoted either by $\mathcal{F}f$ or $\widehat{f}$ as before.

Bibliography

[1] D. Alterman and J. Rauch. Diffractive short pulse asymptotics for nonlinear wave equations. *Phys. Lett. A*, 264(5): 390–395, 2000. URL https://doi.org/10.1016/S0375-9601(99)00822-1.

[2] D. Alterman and J. Rauch. Diffractive nonlinear geometric optics for short pulses. *SIAM J. Math. Anal.*, 34(6): 1477–1502, 2003. URL https://doi.org/10.1137/S0036141002403584.

[3] K. Barrailh and D. Lannes. A general framework for diffractive optics and its applications to lasers with large spectrums and short pulses. *SIAM J. Math. Anal.*, 34(3):636–674, 2002. URL https://doi.org/10.1137/S0036141001398976.

[4] J. Baumstark and T. Jahnke. Approximation of high-frequency wave propagation in dispersive media. CRC 1173-Preprint 2022/9, Karlsruhe Institute of Technology, 2022. URL https://www.waves.kit.edu/downloads/CRC1173_Preprint_2022-9.pdf.

[5] J. Baumstark, T. Jahnke, and C. Lubich. Polarized high-frequency wave propagation beyond the nonlinear Schrödinger approximation. CRC 1173-Preprint 2022/28, Karlsruhe Institute of Technology, 2022. URL https://www.waves.kit.edu/downloads/CRC1173_Preprint_2022-28.pdf.

[6] H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, NY, 2011. URL https://zbmath.org/?q=an:1220.46002.

[7] R. Carles, E. Dumas, and C. Sparber. Multiphase weakly nonlinear geometric optics for Schrödinger equations. *SIAM Journal on Mathematical Analysis*, 42(1):489–518, 2010. URL https://doi.org/10.1137/090750871.

[8] P. Chartier, N. Crouseilles, M. Lemou, and F. Méhats. Uniformly accurate numerical schemes for highly oscillatory Klein-Gordon and nonlinear Schrödinger equations. *Numer. Math.*, 129(2):211–250, 2015. URL https://doi.org/10.1007/s00211-014-0638-9.

[9] Y. Chung, C. K. R. T. Jones, T. Schäfer, and C. E. Wayne. Ultra-short pulses in linear and nonlinear media. *Nonlinearity*, 18(3):1351–1374, 2005. URL https://doi.org/10.1088/0951-7715/18/3/021.

[10] D. Cohen, E. Hairer, and C. Lubich. Long-time analysis of nonlinearly perturbed wave equations via modulated Fourier expansions. *Archive for Rational Mechanics and Analysis*, 187:341–368, 2008. URL https://doi.org/10.1007/s00205-007-0095-z.

[11] M. Colin and D. Lannes. Short pulses approximations in dispersive media. *SIAM J. Math. Anal.*, 41(2):708–732, 2009. URL https://doi.org/10.1137/070711724.

[12] T. Colin. Rigorous derivation of the nonlinear Schrödinger equation and Davey-Stewartson systems from quadratic hyperbolic systems. *Asymptot. Anal.*, 31(1):69–91, 2002. URL https://content.iospress.com/articles/asymptotic-analysis/asy511.

[13] T. Colin, G. Gallice, and K. Laurioux. Intermediate models in nonlinear optics. *SIAM J. Math. Anal.*, 36(5):1664–1688, 2005. URL https://doi.org/10.1137/S0036141003423065.

[14] P. Donnat, J.-L. Joly, G. Metivier, and J. Rauch. Diffractive nonlinear geometric optics. In *Séminaire sur les Équations aux Dérivées Partielles, 1995–1996*, Sémin. Équ. Dériv. Partielles, pages Exp. No. XVII, 25. École Polytech., Palaiseau, 1996. URL http://www.numdam.org/article/SEDP_1995-1996____A17_0.pdf.

[15] P. Donnat and J. Rauch. Modeling the dispersion of light. In *Singularities and oscillations (Minneapolis, MN, 1994/1995)*, volume 91 of *IMA Vol. Math. Appl.*, pages 17–35. Springer, New York, 1997. URL https://doi.org/10.1007/978-1-4612-1972-9_2.

[16] P. Donnat and J. Rauch. Dispersive nonlinear geometric optics. *J. Math. Phys.*, 38(3):1484–1523, 1997. URL https://doi.org/10.1063/1.531905.

[17] W. Dörfler, M. Hochbruck, J. Köhler, A. Rieder, R. Schnaubelt, and C. Wieners. *Wave Phenomena : Mathematical Analysis and Numerical Approximation*. Oberwolfach seminars ; 49. Birkhäuser, Basel, in print.

[18] L. Gauckler, E. Hairer, and C. Lubich. Long-term analysis of semilinear wave equations with slowly varying wave speed. *Communications in Partial Differential Equations*, 41(12):1934–1959, 2016. URL https://doi.org/10.1080/03605302.2016.1235581.

[19] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration : structure-preserving algorithms for ordinary differential equations*. Springer series in computational mathematics ; 31. Springer, Berlin, 2nd edition, 2006. URL https://zbmath.org/?q=an:1094.65125.

[20] M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numerica*, 19:209–286, 2010. URL https://doi.org/10.1017/S0962492910000048.

[21] L. Hörmander. *Distribution theory and Fourier analysis : The analysis of linear partial differential operators*, volume 1 of *Classics in mathematics*. Springer, Berlin, 2nd edition, 2003. URL https://zbmath.org/?q=an:1028.35001.

[22] J.-L. Joly, G. Métivier, and J. Rauch. Rigorous resonant 1-d nonlinear geometric optics. *Journées équations aux dérivées partielles*, art. 7, 1990. URL http://www.numdam.org/item/JEDP_1990____A7_0/.

[23] J.-L. Joly, G. Métivier, and J. Rauch. Generic rigorous asymptotic expansions for weakly nonlinear multidimensional oscillatory waves. *Duke Math. J.*, 70(2):373–404, 1993. URL https://doi.org/10.1215/S0012-7094-93-07007-X.

[24] J. L. Joly, G. Metivier, and J. Rauch. Global solvability of the anharmonic oscillator model from nonlinear optics. *SIAM J. Math. Anal.*, 27(4):905–913, 1996. URL https://doi.org/10.1137/S0036141094273672.

[25] J.-L. Joly, G. Metivier, and J. Rauch. Diffractive nonlinear geometric optics with rectification. *Indiana Univ. Math. J.*, 47(4):1167–1241, 1998. URL https://doi.org/10.1512/iumj.1998.47.1526.

[26] J.-L. Joly, G. Metivier, and J. Rauch. Transparent nonlinear geometric optics and Maxwell-Bloch equations. *J. Differential Equations*, 166(1):175–250, 2000. URL https://doi.org/10.1006/jdeq.2000.3794.

[27] P. Kirrmann, G. Schneider, and A. Mielke. The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A*, 122(1-2):85–91, 1992. URL https://doi.org/10.1017/S0308210500020989.

[28] D. Lannes. Dispersive effects for nonlinear geometrical optics with rectification. *Asymptot. Anal.*, 18(1-2):111–146, 1998. URL https://content.iospress.com/articles/asymptotic-analysis/asy318.

[29] D. Lannes. High-frequency nonlinear optics: from the nonlinear Schrödinger approximation to ultrashort-pulses equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 141(2):253–286, 2011. URL https://doi.org/10.1017/S030821050900002X.

[30] J. Liesen and V. Mehrmann. *Linear Algebra*. Springer Undergraduate Mathematics Series. Springer, Cham, 1st edition, 2015. URL https://doi.org/10.1007/978-3-319-24346-7.

[31] C. Lubich. *From quantum to classical molecular dynamics : reduced models and numerical analysis*. Zurich lectures in advanced mathematics. European Mathematical Society, Zürich, 2008. URL https://doi.org/10.4171/067.

[32] X. Mao. *Stochastic differential equations and applications*. Woodhead Publishing, Oxford [u.a.], 2nd edition, 2011.

[33] R. I. McLachlan and G. R. W. Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002. URL https://doi.org/10.1017/S0962492902000053.

[34] D. Mitrea. *Distributions, Partial Differential Equations, and Harmonic Analysis*. Universitext. Springer International Publishing, Cham, 2nd edition, 2018. URL https://doi.org/10.1007/978-3-030-03296-8.

[35] G. Métivier. Chapter 3 - the mathematics of nonlinear optics. In C. Dafermos and M. Pokorný, editors, *Handbook of Differential Equations*, volume 5 of *Handbook of Differential Equations: Evolutionary Equations*, pages 169–313. North-Holland, 2009. URL https://www.sciencedirect.com/science/article/pii/S1874571708002107.

[36] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983. URL https://doi.org/10.1007/978-1-4612-5561-1.

[37] J. Rauch. *Hyperbolic partial differential equations and geometric optics*, volume 133 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. URL https://doi.org/10.1090/gsm/133.

[38] S. Roth and A. Stahl. *Optik : Experimentalphysik – anschaulich erklärt*. Springer eBook Collection. Springer, Berlin, Heidelberg, 1st edition, 2019. URL https://doi.org/10.1007/978-3-662-59337-0.

[39] G. Schneider and H. Uecker. *Nonlinear PDEs*, volume 182 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2017. URL https://doi.org/10.1090/gsm/182.

[40] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer Spektrum, Berlin, Heidelberg, 8th edition, 2018. URL https://doi.org/10.1007/978-3-662-55407-4.

[41] S. Zain. *Techniques of classical mechanics : from Lagrangian to Newtonian mechanics*. IOP expanding physics. IOP Publishing, Bristol, UK, 2019. URL https://dx.doi.org/10.1088/2053-2563/aae1b9.