UiT The Arctic University of Norway

Faculty of Science and Technology — Department of Mathematics and Statistics

Nonlinear optics

Dávid Juhász

A dissertation for the degree of *Philosophiae Doctor*…July 2021

Dedicated to my great-grandmother Béres Ilona (1919-2012)

Abstract

Nonlinear light-matter interactions have been drawing attention of physicists since the 1960's. Quantum mechanics played a significant role in their description and helped to derive important formulas showing the dependence on the intensity of the electromagnetic field. High intensity light is able to generate second and third harmonics which translates to generation of electromagnetic field with multiples of the original frequency. In comparison with the linear behaviour of light, the nonlinear interactions are smaller in scale. This makes perturbation methods well suited for obtaining solutions to equations in nonlinear optics. In particular, the method of multiple scales is deployed in paper 3, where it is used to solve nonlinear dispersive wave equations. The key difference in our multiple scale solution is the linearity of the amplitude equation and a complex valued frequency of the mode. Despite the potential ill-posedness of the amplitude equation, the multiple scale solution remained a valid approximation of the solution to the original model. The results showed great potential of this method and its promising wider applications.

Other methods use pseudo-spectral methods which require an orthogonal set of eigenfunctions (modes) used to create a substitute for the usual Fourier transform. This mode transform is only useful if it succeeds to represent target functions well. Papers 1 and 2 deal with investigating such modes called resonant and leaky modes and their ability to construct a mode transform. The modes in the first paper are the eigenvalues for a quantum mechanical system where an external radiation field is used to excite an electron trapped in an electrical potential. The findings show that the resonant mode expansion converges inside the potential independently of its depth. Equivalently, leaky modes are obtained in paper 2 which are in close relation to resonant modes. Here, the modes emerge from a system where a channel is introduced with transparent boundaries for simulation of one-directional optical beam propagation. Artificial index material is introduced outside the channel which gives rise to leaky modes associated with such artificial structure. The study is showing that leaky modes are well suited for function representation and thus solving the nonlinear version of this problem. In addition, the transparent boundary method turns out to be useful for spectral propagators such as the unidirectional pulse propagation equation in contrast to a perfectly matched layer.

Acknowledgements

I would like to thank my supervisor Prof. Per Kristen Jakobsen for his dedicated support, availability, time and guidance. Per continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout my doctoral studies. I would also like to thank Prof. Miroslav Kolesik for providing advice for my first research project. Finally, many thanks to my family who have been a great source of support and enabled my studies to be possible.

Contents

1 Introduction

1.1 Nonlinear optics

One of the most unifying understanding of large and diverse phenomena in science was Maxwell's theory of electric and magnetic fields. It was believed in the second half of the nineteenth century that only a few new fundamental discoveries needed to be made. This belief was shattered by the failure to explain the physical phenomena such as the photoelectric effect, x-rays, radioactivity or the radiation spectrum. To understand these phenomena, a new idea of the nature of light must have been revived. This led to the birth of quantum theory and optical science ended up under the wings of a resolved science. As new progress was made, the linear superposition of simpler solutions represented an important role in the development. But the intensity of the light seemed to play no significant part. However, there were still some observations that could not be explained on a linear basis such as double refraction in isotropic media. This suggested that maybe the intensity of the field was important after all.

The theoretical experiment of two-photon absorption from 1931 was the first spark for the rise of the field of nonlinear optics. The German-born theoretical physicist Maria Goeppert Mayer proposed the theory of possible simultaneous two-photon absorption by atoms [1]. The probability of this absorption is proportional to the square of the light intensity making it a nonlinear optical process. This nonlinear process remained in a theoretical level until 1961, when exactly 30 years later, almost simultaneous two-photon absorption was observed [2] and second-harmonic generation was discovered by Franken and co-workers [3]. Nonlinear effects in the process when light propagates through and interacts with matter occur with high-intensity light. The two-photon absorption was therefore observed thanks to the construction of the first laser prior to the experiment by Theodore Maiman [4][5]. This new source of coherent radiation could be focused to achieve extremely high local intensities. To be a somewhat young field of nonlinear optics of that time, it was scientifically fruitful and promising to be one of the most essential areas of science for the next decades. Many fundamental new phenomena and phenomena familiar to other fields are following and it continues to grow, becoming richer by the day.

The use of lasers in modern technology is enormous, ranging from high-density data storage in memory disks to neurosurgery and dermatology in medicine. These high-intensity devices are, however, not by any means simple models. They are highly complex systems of dynamical nature that can reveal a behaviour of both fixed and chaotic attractors. Note, that a dynamical system tends to evolve to a set of numerical values called attractor regardless of the starting conditions of the system. Also, many smaller lasers can be coupled together providing a coherent high-power output. In this case, this array will not behave as a single consistent unit but will produce a mosaic of patterns [6]. But such a complicated behaviour is yet to be categorized and understood. In optical fibers, these laser arrays could serve as a source for the pulses.

Another great potential of nonlinear optics lies in telecommunication technologies. Due to the small cross-section in the waveguide of the fiber and long interaction lengths, the low-energy optical pules can reach high intensities. Nowadays, with linear propagation methods, the speed at which the information passes across continents in thin optical fibers reaches rates of gigabits per second [7]. Some say that within a decade, this linear technology will be taken over by a nonlinear one where light pulses are transmitted as solitary waves.

There are several reasons why nonlinear optics is an ideal subject for a theoretician

who is interested in nonlinear behaviour and model building. Firstly, it is very diverse and provides a variety of behaviours associated with nonlinear equations, development of singularities, pattern formation, bifurcation processes or turbulence. All of these are familiar to the theoretician in many contexts. Another reasons are the parallelism with other fields and some new concepts of nonlinear science that are located on the different side of dynamical behaviour such as soliton or strange attractor. These concepts often show up and require more mathematical knowledge. Especially the strange attractor that continues to have a large impact in optical feedback devices such as lasers [8]. The last reason is the variety of mathematical tools a theoretician uses in computer simulation. The ideas and theories are tested using a wide range adjustable parameters and useful qualitative understanding is obtained through simple models. These reasons together with an eager experimentalist and physical intuition will continue to steer the development of nonlinear optics.

In order to describe what do we exactly mean by optical nonlinearity, we need to start with polarization. In the classical approach, any material is made up of atoms. Each atom consists of a "cloud" of negative charge and a positive charge. Usually, the nucleus of an atom is positively charged because of the protons and the electrons around the nucleus contribute to the negative charge. When there are no other forces at play, these two charges cancel each other and the atom has no net charge. However, the mutual position of these opposite charges is not constant but changes all the time due to a presence of electromagnetic forces around. This causes these clouds to be distorted and asymmetric which leads to a so called electric dipole. A dipole is characterized by its dipole moment that is defined as the charge times the vector of displacement of the two charged clouds. By definition, the direction is from negative to positive charges. The *polarization* $\mathbf{P}$ is then the density of atomic electric dipole moments. The distortion or separation of the charges in an atom can be caused by a presence of a changing electromagnetic field (EM field). The creation of this separation itself creates an EM field around itself which, in return, causes other separations of charges in the material. It is a kind of chain reaction. Therefore, polarization is a source of an EM field and its units are coulomb per meter squared ($\mathrm{C/m^2}$). There are altogether four sources of EM fields: polarization, magnetization, free charge and free current. There can be more than one various source of EM fields at a point in space and they can overlap continuously in time. They produce EM fields but, at the same time, they are also influenced by these fields. It means that sources can convert from one form to another.

Two more vector fields appear besides the electric and magnetic field in classical electrodynamics; electric displacement $\mathbf{D}$ and magnetic induction $\mathbf{B}$. The polarization plays a role in the electric displacement through the equation

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}, \tag{1.1.1}$$

where ε_0 is the permittivity of free-space. It is a measure of the ability of a material to be polarized and thus store electric potential energy. $\mathbf{E}$ is the electric field. The term $\varepsilon_0\mathbf{E}$ is called the induced polarization of free-space. It means that the electric field acts locally on the vacuum creating electric dipole moments. The volume density of these moments is proportional to the electric field with the proportionality constant being ε_0.

As explained earlier, the electric field displaces the electrons in the atom from the nucleus inducing electric dipoles that collectively give rise to polarization. This electric dipole is modelled by a powerful mass-and-spring model, also called the Lorentz oscillator model, first proposed by the Dutch physicist H. A. Lorentz [9]. It gives a quite precise

picture of how the electric field depends on polarization for wide range of materials such as dense gases, absorptive liquids and solids including dielectrics, semiconductors and metals. To generalize the result obtained from this model, the polarization at a given location $\mathbf{x}$ in an isotropic and linearly-polarizable material excited by an electric field is expressed as

$$\mathbf{P}(\mathbf{x}, t) = \varepsilon_0 C(\omega)\mathbf{E}(\mathbf{x}, t), \tag{1.1.2}$$

where $C(\omega)$ is the polarizability coefficient that depends on the frequency of the electric field and three material parameters: plasma frequency, resonance frequency and the damping coefficient of spring in the model. A typical material is made from atoms that have many electrons connected to their nuclei. Therefore each electron can be represented by a mass-and-spring system so that $C(\omega)$ becomes a sum of the polarizability coefficients of all the electrons.

The Lorentz oscillator model may be also used to describe the conduction electrons in a material which move freely and are not bound to any atom. Some parameters in the model will be changed, but a similar result can be obtained since the conduction electron may still respond to an oscillating electric field. The proportionality constant in this case is between the electric field and polarization is now represented by $\chi_e(\omega)$ instead of $C(\omega)$ and it is called *electric susceptibility* of the conduction electrons. The expression for the electric susceptibility gained in this way is called the Drude model of the conduction electrons. However, in practical situations, the electric susceptibility refers to susceptibility of the material, not only its conduction electrons and it is denoted by $\chi(\omega)$. One can derive the light propagation equation through a medium using general χ from Maxwell's equations. In this wave equation, a quantity called *the refracting index* n will appear which is defined as $n = \sqrt{1 + \chi}$. It is a dimensionless number greater than unity. It defines how the light ray bends when entering a medium.

The dependence of $\mathbf{P}$ and ε_0 on time history, spatial inhomogeneities, changing medium density, and field intensity leads to nontrivial behaviour in the propagation of light. For small electric field amplitude and no resonance between the electric field and the medium, $\mathbf{P}$ depends linearly on $\mathbf{E}$ which is expressed as

$$\mathbf{P}(\mathbf{x}, t) = \varepsilon_0 \chi(\omega)\mathbf{E}(\mathbf{x}, t). \tag{1.1.3}$$

This is true for an isotropic medium. The susceptibility $\chi(\omega)$ is in general not a constant and depends on the oscillation frequency of the electric field. When light enters a medium, it does not respond instantaneously, but rather captures the electric field at previous times. This memory effect, which is the embodiment of causality, is in optics called *temporal dispersion*, or just dispersion. In an isotropic and homogeneous medium the memory effect is captured by writing

$$\mathbf{P}(\mathbf{x}, t) = \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t' \chi(t - t')\mathbf{E}(\mathbf{x}, t'), \tag{1.1.4}$$

or it can be written in a shorter form as $\mathbf{P} = \varepsilon_0 \chi \mathbf{E}$. In nonlinear optics, the optical response can often be described by generalizing (1.1.4) by expressing the polarization $\mathbf{P}$ as a power series in the electric field as

$$\mathbf{P} = \varepsilon_0 \chi^{(1)} : \mathbf{E} + \varepsilon_0 \chi^{(2)} : \mathbf{EE} + \varepsilon_0 \chi^{(3)} : \mathbf{EEE} + \ldots, \tag{1.1.5}$$

where $\chi^{(n)}$ is known as the n-th order nonlinear optical susceptibility. In a vector notation we are using, the terms $\chi^{(n)}$ become tensors of rank $n+1$ and the tensor operation : is called a contraction of tensors leaving a vector as an outcome. One can also write the equation (1.1.5) as convolution integrals, where the first term on the right hand side would have the same form as (1.1.4), the second term would involve two convolution integrals, the third term would have three and so on. Writing it in such form represents the most general relationship between the electric field and the polarization. The first term defines the usual linear susceptibility, the second term is the lowest order nonlinear susceptibility and so on. Because the optical nonlinearities are small, this procedure is useful, for example when applying perturbation theory. The experimental discovery of the equation (1.1.5) had to wait until the development of powerful lasers as we discussed earlier. Physicists like Maxwell, Hertz, Lorentz or Drude lacked the experimental stimulation, but the stimulated emission of light changed this. There are many generalizations of the classical laws of optics to the regime of intensities where nonlinear effects are not negligible. The study of nonlinear susceptibilities of the structure of matter is of intrinsic interest as well.

It should be noted that the power series expansion expressed by (1.1.5) does not necessarily need to converge. In such circumstances different procedures must be used to express the relationship between the material response and the applied electric field amplitude. A resonant excitation of an atomic system is one such circumstance, where a significant fraction of the atoms can be excited from the ground state. There are, however, some non-resonant conditions, under which a strong photoionization can occur and equation (1.1.5) loses its validity. This happens because the strength of the applied laser field becomes comparable to the characteristic atomic field strength.

The linear, or the 0-th order nonlinear susceptibility, determines the paths of light rays. The knowledge of this behaviour gives insight about the nature of the material. One important thing to emphasize is that rays will turn away from regions of smaller and toward regions of greater refractive index. This property is the basis for Snell's law in optics. We have defined earlier the refractive index from the susceptibility to be $n(\omega) = \sqrt{1 + \chi(\omega)}$. This formula suggests that as the susceptibility depends, in general, on the oscillation frequency of the electric field, so does the refractive index. For light generated by a laser, one finds that the refractive index depends on the light intensity or in other words the amplitude of the electric field. This can be seen from formula (1.1.5) as well. Recalling the property of light bending towards higher refractive index, we can conclude that the intensity dependent index will tend to focus light into areas of high light intensity. This increases the intensity, creating an even larger index, which focuses light even more strongly. One can see that this runaway effect tends to quickly create a local high intensity and is capable of destroying the material.

As we mentioned, for an isotropic, homogeneous medium, only the first term of (1.1.5) is significant and the polarization varies linearly with the electric field. However, for materials with non-negligible change in the refractive index in response to an applied electric field, also called *Kerr effect*, the third term in (1.1.5) $\chi_e^{(3)}$ is significant. The even-order terms typically dropping out due to inversion symmetry of the medium. More details on this subject will be provided in the summary of paper 3. It turns out that the second-order nonlinear optical interactions can occur only in noncentrosymmetric crystals. In other words, in crystals that do not display inversion symmetry. Materials like liquids, gases or glass and many crystals that have the inversion symmetry, can not produce second-order nonlinear optical interactions. For such materials, $\chi_e^{(2)}$ vanishes. On the other hand, third-order nonlinear optical interactions can occur for both centrosymmetric

and noncentrosymmetric media. In this case it is usually assumed that the nonlinear polarization is restricted to the Kerr effect. This leads to the simplified formula

$$\mathbf{P} = \varepsilon_0 \chi_e \mathbf{E} + \varepsilon_0 \eta \mathbf{E} \cdot \mathbf{E}\mathbf{E}, \tag{1.1.6}$$

where η is the Kerr coefficient, which is a material property. The expression (1.1.6) goes directly in this form into the light propagation equation derived from Maxwell's equation, where the unknown function is the electric field. It thus becomes a nonlinear partial differential equation.

Now, how is it, that nonlinear optics is so accessible to theoretical analysis when compared with other areas of nonlinear physics? The main reasons are that at the light intensities available nowadays, the coupling coefficients from (1.1.5) are small. Next reason is that the frequency spectrum of the EM field is concentrated around a discrete frequency (narrow band spectrum). Thanks to these properties one can remove fast space and time scales from the equation using perturbation techniques which leads to significant simplifications. The light and matter can be thought of as a system of uncoupled oscillators up to the first order of approximation. The variables of light and matter obey linear equations and light consists of wavetrains of the form

$$A(\mathbf{x}, t) e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \tag{1.1.7}$$

where A is the amplitude, $\mathbf{k}$ is the propagation direction of the light, $\mathbf{x}$ is the position vector, ω is the oscillation frequency and t denotes time. The sources of oscillation in the matter are atomic and molecular vibrations, rotations, acoustic waves or conduction electrons as it was discussed earlier. The nonlinear terms are one or more order of magnitude smaller, but it does not mean they are negligible and can have long-time and distance effects. However, only a certain subset of all possible linear and nonlinear interactions between different oscillators are important, especially those that satisfy the resonant conditions. This subset represents a finite number combinations of oscillator models. Since the solutions can be represented as a sum of discrete wavepackets (1.1.7) with localized frequency spectrum, the oscillation term represented by the exponential varies much faster in space and time than the amplitude. This allows us to conclude the inequalities $\partial^2 A/\partial t^2 \ll \omega \partial A/\partial t$ and $\partial^2 A/\partial z^2 \ll k_z \partial A/\partial z$ (k_z is the z-components of the vector $\mathbf{k}$). Consequently it means that the amplitude A of the wavepacket satisfies an equation containing only low powers of derivatives, typically the first order derivatives. The usual distance and time units for light waves are $2\pi/\omega \approx 10^{-15}$ s (order of femtoseconds) and $2\pi/k_z \approx 10^{-6}$ m (order of micrometers). The amplitude can vary in times between 10^{-9} s (nanoseconds) and 10^{-12} s (picoseconds).

One of the primary goals of this theory is to write down the equations that govern the amplitude A. These equation are in general nonlinear, but they often have the form of nonlinear equation about which is much known. One such equation is the nonlinear Schrödinger equation (NSLE), that is useful for fiber optics and nonlinear waveguides. There are many perturbation techniques that help us to obtain these amplitude equations. We are going to introduce one standard perturbation procedure, called *method of multiple scales*, or MMS in short. It is used to derive the amplitude equation from the governing Maxwell's equations.

1.2 The method of multiple scales

Physicists, engineers and applied mathematicians face today many problems involving nonlinear equations, variable coefficients or nonlinear boundary conditions at complex

boundaries which hinders the enquiry of exact solutions. New phenomena occur in nonlinear problems that we do not see in the corresponding linear problems. Therefore, the purpose in the study of nonlinear problems is to aim attention on the features of nonlinearities that are the grounds for new phenomena, rather than to introduce methods to improve the accuracy of linear methods. Developing a comprehensive theory of nonlinear phenomena is often out of the question because of the complexity of mathematical problems associated with nonlinearities.

While learning about differential equations, one quickly exhausts the few types that can be solved analytically, or in closed form using elementary functions. Linear higher-order partial differential equations are reducible to first or higher-order linear, homogeneous equations with constant coefficients by separation of variables. There are several directions after this: approximation of solution using formulas or advanced theory. In order to solve these problems, we are forced to reach for a sort of approximations, numerical solutions or a combination of both. For something less than complete generality, one practical approach was to settle. Instead of studying the global behaviour of solutions of nonlinear problems, one seeks nonlinear solutions in the vicinity of (or in other words, perturbations around) a known linear solution. This is the basic idea behind perturbative solution of a nonlinear problem, or *perturbation methods*. Perturbation theory was at the beginning used to solve otherwise unmanageable problems in the calculation of the motion of planets in the solar system [10]. This motion was described by Newton's gravitational equations which explained the phenomenon with two astronomical bodies. But when a third body was introduced, a computational problem arose. Rising demands in the accuracy of solutions to Newton's gravitational equations were a consequence of, among other things, astronomical observations. This led to several notable mathematicians in the 18th and 19th century, such as Lagrange or Laplace, who also generalized the perturbation methods. These two mathematicians promoted the view that the constants involved in the motion of the planets around a star are perturbed by other planets and that these constants are a function of time [10]. Accordingly, the name for the theory became perturbation theory. It was first studied by Laplace, Poisson and Gauss and the calculations could be done with a very high accuracy. A big triumph of perturbation theory was the discovery of the planet Neptune by Urbain Le Verrier. The calculations were based on the deviations in motion caused by the planet Uranus [10].

Gradually, to solve new problems arising, perturbation methods were even more developed, adapted and used. Especially during the development of quantum mechanics in the early 20th century. A quantum perturbation theory was established by Paul Dirac in 1927 in order to find out when a particle is emitted from radioactive elements. This theory became later known as Fermi's golden rule [11]. The quantum notation in this perturbation theory allowed to write the expressions and formulas in a more compact form, which made the theory more accessible. This led to its much wider applications. It was known in the late 20th century, when chaos theory was developed, that unperturbed systems were integrable systems and perturbed systems were not. This immediately led to investigation of nearly integrable systems. Prior to that, nonlinear systems that were solvable only with perturbation theory, were, in fact, integrable. It represented a climatic discovery because it allowed to obtain exact solutions. One could now compare the results of perturbation series with the exact solutions which helped to resolve the meaning of perturbation series.

Many who are using perturbation theory successfully, view it as a bag of given formulas that work even though they are not always justifiable. Its infamous reputation is largely

known also by those who refuse to use it because it allegedly lacks mathematical rigour. But the separation between perturbation methods and other approaches (e.g. geometrical analysis) is a deception of the true essence of the subject. It was due to the definition of asymptotic series that this was realized. Before applying a perturbation method, one should always at the same time consider the existence and the uniqueness of the solution, the geometry of the solution, bifurcations or other factors that might have an impact on the solution.

It is common to teach perturbation theory to people that have not yet mastered the methods of proof that are needed to show the existence of solutions to which approximations are found. It is because the demands for perturbation theory are from the extremely applied end of mathematical spectrum and because of the nature of the theory that is almost entirely informal in content. It is where mathematical theory interacts with practical computational methods. One might say that perturbation theory makes it possible to appreciate the conditions where these practical computational methods fail. From this point of view, a user of mathematics is able to acquire important mathematical skills as an ability to read a theorem and extract the significance of its applications without a need of a proof. And from the other end, it can help a mathematician who has some appreciation for a need of a proof, while studying problem solving with proofs coming later, to feel at ease when the proofs are come across.

As the name perturbation methods suggest, there are more than one method used in this theory. According to these methods, the solution to the nonlinear problem is represented by the first few terms in the perturbation expansion. The perturbation expansions can be useful for both qualitative and quantitative representations of a solution, although they can be divergent. Sometimes even more useful than uniformly or absolutely convergent expansions [12]. A straightforward expansion in powers of a parameter can break down in some regions of nonuniformity which is more of a rule rather than an exception. To remedy this problem, a number of techniques have been developed by physicists, engineers and applied mathematician working in different branches of expertise. Some of these techniques can be viewed as a different interpretation of the same idea while others are entirely different. The idea behind perturbation methods is that a perturbative solution becomes applicable if it is in a close proximity of another problem that we know how to solve. A solution to a simpler problem is then studied and utilized to express the solution to a more difficult problem in terms of the simpler one with a small correction. This procedure creates a sequence of problems easier to solve in the sense that when we find a corrected approximation, the process is repeated to obtain a better approximation.

A common feature of them is that the solution is represented by the first few terms in the asymptotic expansion, as mentioned earlier. The expansions are carried out in terms of a parameter ε which is usually small and appears in the equation. It may be introduced artificially in the equations where it represents a dimensionless amplitude of a perturbation. The perturbation expansion is then with respect to the small parameter ε. It is then easy to see, that the accuracy of this expansion gets better for smaller values of ε. Such expansions are called *parameter perturbations*. The zeroth order terms are typically a solution to the linearized version of the nonlinear problem. The coefficients in the asymptotic expansions are obtained as solutions of sequences of linear problems. A uniformly convergent Taylor series in ε in its domain of analyticity is used to develop the relevant quantities. The expectation of the solution to be analytically dependent on the parameter ε is justified by the parameter involved in the differential equation together with the boundary condition in an analytic way. Solutions to linear equation that contain

inhomogeneities involving previously calculated lower order quantities, produce the higher order quantities. Alternatively, instead of a parameter, the expansion may be based on a coordinate. These expansions are called *coordinate perturbations*.

Perturbation problems are often divided into regular and singular perturbation problems. The difference between the two is in the role of the perturbation parameter in the equation.

In a regular problem, the asymptotic expansion is obtained from a straightforward procedure that leads to a hierarchy of differential equations with boundary conditions for each term in the expansion. It is called the *perturbation hierarchy*. The hierarchy, or system, is solved term by term recursively with a gradual improvement in the accuracy as ε gets smaller in the whole domain of interest. The perturbation expansion must also be valid uniformly. Otherwise if the expansion is nonuniform and it persists in the approximations of higher orders, then it leads to a singular perturbation problem.

In a singular perturbation problem, the perturbation parameter ε multiplies the highest order derivative in the differential equation, or the highest power in an algebraic equation. For this reason, the solution of the leading order of the equation obeys a lower order equation that is not satisfying the boundary conditions. This, of course, causes a failure in the perturbation hierarchy where at some layers at the boundary or inside of the domain, the procedure fails. Therefore is this kind of a problem also called layer-type problem. Problems, mainly in dynamical astronomy, were the starting points of singular perturbation problems for over a century. It helped to develop and resolve the issues involved within the singular case.

If the system is solved over an infinite domain and contains small terms that are cumulatively building up, then regular perturbation problem also fails. There are, however, perturbation methods that work even in these kind of situations. They are the method of averaging and the method of multiple scales both of which are used to derive asymptotic expansions while remaining valid in the far field.

There are many different perturbation methods, out of which the most known are:

- the method of strained coordinates
- the method of matched composite and asymptotic expansions
- the method of averaging
- the method of multiple scales

Usually, in various textbooks, these techniques are presented through simple but useful examples from physics and applied mathematics. The examples mostly involve ordinary and partial differential equations from solid mechanics, fluid dynamics, quantum mechanics, plasma physics and nonlinear optics. These differential equation include equations or are included in problems like the van der Pol oscillator, the Duffing's equation, Klein-Gordon equation, Earth-Moon spaceship problem, supersonic flow past a thin airfoil, the time-dependent and the nonlinear Shrödinger equation (NLSE) [13][14][15][16][17]. The last two examples are from quantum theory which is arguably the best description of reality we have so far. Perturbation methods are tools most used in quantum theory, for quantum electrodynamics, these methods are essentially the only tool available. Quantum theories are more or less known for their perturbation expansions.

But let us return to introduction of one particular perturbation method that we use in one of our papers. We are going to focus on this method, which is widely used in many

areas of applied science. Method of multiple scales, or MMS, is one of the two prominent methods that take into account small cumulative effects of perturbations over a larger period of time. As opposed to other perturbation methods, MMS features the nonexistence of a limit process expansion for long times. It leads to writing the solution in a form of a general asymptotic expansion. Historically, an astronomer Lindstedt proposed a method, also known as Poincare-Lindstedt method, for calculation of periodic solutions whose generalization is MMS [18]. One such periodic solution comes from a problem of a pendulum as a function of amplitude when the amplitude is small. MMS allows us to calculate the period of such pendulum. It can be also used to determine how the fundamental frequency of an oscillator varies with the nonlinearity or calculate the exchange of energy between weakly coupled oscillators through nonlinearities in the equations [19].

However, one can find different views on the applications of the method and its limitations. It can therefore appear that the various descriptions of the method are quite different, depending on the author's views. One view, for example, presents the method in a way which is very effective and allows to take the perturbation expansions to a higher order in terms of the small perturbation parameter that would otherwise be not possible.

Whether one finds MMS is successful or not, in any given situation, does it not only depend on the nature of the problem, but also on what one asks the method to provide. Most perturbation methods, including MMS, were originally designed to find some form of analytic solution to problems of interest. This is in particular true for application to problems involving ordinary differential equations. Amongst all the perturbation methods, MMS was not the one that received most of the focus from the authors for many years. Although it was always regarded as with a great potential. Like all the other perturbation methods, MMS also underwent many changes and adjustments during the years which made it one of the lightest accessible and comprehensible tools. One of these changes was to extending the number of scales from two to as many as one likes. This adjustment made it possible to reach the desired accuracy of the asymptotic solution and to widen the applicability of the method.

The method of multiple scales is, among other things, a tool for investigating dispersive wave equations. As such a tool, it has a long pedigree and deep roots. It appears in most textbooks on general perturbation methods and in all respectable textbooks focused on singular perturbation methods. There are other perturbation methods that have been applied to the problem of dispersive wave propagation, but most such methods rely on a deeper and more extensive mathematical machinery as compared to MMS. In order to apply these various perturbation methods to any given situation, very similar restrictions has to be imposed on the wave equations of interest. Thus, to decide which of these methods to apply, is to some extent a matter of personal taste and mathematical sophistication.

One of the methods that was developed in order to unify, systematize approaches with more mathematical rigour to solve problems that were previously solved by perturbation methods is center manifold theory (CMT). This theory arose to remedy the nature of perturbation methods in general, that is its vague domain of applicability, where the logical relations between the formulas are not altogether clear. To introduce more rigour, CMT developed an approach backed by geometry and useful mathematical theory. Consequently, it is interesting to compare the CMT with MMS. Perhaps the most obvious difference is that the CMT puts very little importance on the size of various physical effects. In MMS are these effects expressed with the help of the perturbation parameter ε. However, center manifold analysis requires to pinpoint the dominant terms in a given

linear operator and the perturbation terms. These terms are various nonlinear terms or effects that vary slowly and regarded as perturbation terms. Next, the center manifold approach uses iterative refinements to generate higher order approximations. Nevertheless, some derivations such as evolution equations for spatial patterns or for wave modulation are traditionally done using MMS, but can also be handled by CMT.

When it comes to complexity and the level of education one must have, to be able to use these methods, the method of multiple scales appears much more available and approachable with a comparably wide range of use, as we will demonstrate. CMT studies equilibrium points of the dynamical system on which a center manifold is based. One starts with representing the equation in a form of a differential equation operator. The solution space of the linearized equation is decomposed into a stable subspace. This is done through identifying the eigenvalues and eigenvectors of the linearized dynamical system. Based on the eigenvalues, the stability of the eigenspace is determined and the systems dynamics near the equilibrium is completely characterized. The center manifold is calculated by iteration and at the end, the original system is restricted on the obtained manifold and solved as a lower dimensional problem.

Nayfeh in his paper [20] analyzed the nature of Hopf bifurcations in retarded systems modelled by nonlinear homogeneous ordinary differential equations with discrete time delay. The analysis was done using both the CMT and MMS in order to compare the two methods. He concluded that the method of multiple scales seemed to be simpler in a sense that it could be directly applied to the retarded differential equations. To compare with CMT, the retarded equations need to be converted into operator equation in a Banach space. Since the Banach space does not have a natural inner product with a norm, one also needs to find a tool that acts like an inner product. Then one has to define the adjoint of the linearized operator, perform the projection on the center manifold and obtain the normal form of the dynamical system on the center manifold. It is not hard to see the clear difference in the complexity of the two methods, their practical use and approachability.

In contrast to CMT, the method of multiple scales starts with a generalized version of an expansion. Coordinates (variables) for each region (in time and space) that are independent of each other, are separated. These separated coordinates are called scales. Scales are introduced to be either fast-scale or slow-scale variables. It is important to emphasize, that these variables are independent of one another. The given equation is then transformed into a sequence of partial differentials equation even if the original equation was an ordinary differential equation. The sequence of differential equation is the previously mentioned perturbation hierarchy. Each level in the hierarchy is then solved recursively where each solution represents a correction to the solution found earlier. If one would carry on with this procedure indefinitely, the exact solution would be obtained. In the process of solution, the independent variables introduce some degree of freedom which is then used to remove so called *secular terms* from the equations in the hierarchy. Secular terms impose constraints on the approximate solution. When these secular terms are untreated, we find that these corrections can exhibit an unbounded growth in time. In this case, the small perturbation has no longer a small effect but a larger one if time gets large enough. At every step in the perturbation hierarchy, we get one equation in a form of a condition for the amplitude that we introduced as a part of the solution to the linearized problem. These equations come from the effort to eliminate secular terms. At the end of the procedure, the equation are joined into one using the perturbation expansion and we get a simpler equation for the amplitude which is easier to solve than

the original problem.

To gain a better understanding for the method of multiple scales, we will now present a demonstration of the method on a simple initial value problem for a 2nd order ordinary differential equation. The problem is picked from the lecture notes by Per Jakobsen [21]. The procedure presented is the way the method is used in one of our papers.

Consider a cubic oscillator also known as the Duffing's equation

$$\begin{aligned} y''(t) + y(t) &= \varepsilon y^3(t), \quad t > 0, \\ y(0) &= 1, \\ y'(0) &= 0, \end{aligned} \tag{1.2.1}$$

where ε is the small perturbation parameter. The first step in MMS is to introduce a function with scales

$$y(t) = h(t_0, t_1, t_2, \ldots)|_{t_j = \varepsilon^j t}\,, \tag{1.2.2}$$

together with the expansions

$$\frac{\mathrm{d}}{\mathrm{d}t} = \partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots, \tag{1.2.3}$$

$$h = h_0 + h_1 + h_2 + \ldots\ . \tag{1.2.4}$$

These expansions say that a function h_j varies on the time scale $T_j = \varepsilon^{-j}$. In other words $h_j = y\left(\varepsilon^j t\right)$. The more functions h_j we obtain, the better can $y(t)$ be represented for larger times and thus, better approximation for the solution in a form of asymptotic expansion.

Inserting these expansions into the differential equation (1.2.1) and expanding everything in sight, we get

$$\begin{aligned} &\left(\partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots\right)\left(\partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots\right)\left(h_0 + h_1 + h_2 + \ldots\right) \\ &\qquad = \varepsilon\left(h_0 + h_1 + h_2 + \ldots\right)^3, \\ &\qquad \Downarrow \\ &\partial_{t_0t_0}h_0 + h_0 + \varepsilon\left(\partial_{t_0t_0}h_1 + h_1 + \partial_{t_0t_1}h_0 + \partial_{t_1t_0}h_0\right) \\ &+ \varepsilon^2\left(\partial_{t_0t_0}h_2 + h_2 + \partial_{t_0t_1}h_1 + \partial_{t_1t_0}h_1 + \partial_{t_0t_2}h_0 + \partial_{t_1t_1}h_0 + \partial_{t_2t_0}h_0\right) + \ldots \\ &\qquad = \varepsilon h_0^3 + 3\varepsilon^2 h_0^2 h_1. \end{aligned} \tag{1.2.5}$$

Matching the expressions for the different orders of ε from both sides of the equation gives us the following perturbation hierarchy to second order in ε

$$\begin{aligned} &\text{order } \varepsilon^0: \ \partial_{t_0t_0}h_0 + h_0 = 0, \\ &\text{order } \varepsilon^1: \ \partial_{t_0t_0}h_1 + h_1 = h_0^3 - \partial_{t_0t_1}h_0 - \partial_{t_1t_0}h_0, \\ &\text{order } \varepsilon^2: \ \partial_{t_0t_0}h_2 + h_2 = 3h_0^2h_1 - \partial_{t_0t_1}h_1 - \partial_{t_1t_0}h_1 - \partial_{t_0t_2}h_0 - \partial_{t_1t_1}h_0 - \partial_{t_2t_0}h_0. \end{aligned} \tag{1.2.6}$$

The sequence of equations above is the perturbation hierarchy. Observe a common differential operator appearing in the equations $\mathcal{L} = \partial_{t_0t_0} + 1$.

There are many different approaches how to deal with these calculations depending on the authors that practice the method of multiple scales. It comes down to how do we treat the function $h(t_0, t_1, \ldots)$. A particular way of doing these calculations is followed if one takes the function h seriously. It is done in most of the textbooks. The approach we

will use is not the same since the function we actually want is y, not h, where the relation between the two is defined in (1.2.2). This will make the calculations distinct from what we can find elsewhere about this subject. Our approach is efficient and allows to go to orders beyond ε^2 avoiding the huge amount of algebra involved in the calculations.

We will not consider h as a serious multivariable function and one consequence of that is keeping the higher order partial derivatives $\partial_i\partial_j h_k$ and $\partial_j\partial_i h_k$ separate from each other, as we can see in the perturbation hierarchy. The equality of these terms is not used to simplify the expressions. The other consequence is that we disregard the initial conditions at this stage when we solve the hierarchy. The initial conditions are going to be used at the end of the calculations.

Let us proceed to solve the equations in the perturbation hierarchy. At order ε^0 we have the solution

$$h_0(t_0, t_1, \ldots) = A_0(t_1, t_2, \ldots)e^{it_0} + (*), \tag{1.2.7}$$

where $(*)$ is the complex conjugate of all the preceding terms. One feature of multiple scales is the use of general solution of differential equation at order ε^0. For the other orders, where partial differential equations appear, it will be different.

Note that the formula (1.2.7) tells us how h_0 depends on t_0, but not on the other variables $t_1, t_2, \ldots$. Also, the integration constant A_0 depends only on the other variables. The form of the formula tells us as well that the oscillatory term e^{it_0} varies slower than the function A_0 because the variables t_0 and t_1 are defined as $t_0 = t$ and $t_1 = \varepsilon t$. Thus $t_1, t_2, \ldots$ are slower time scales than t_0.

Proceeding to the next order in (1.2.6), we insert (1.2.7) into the right hand side of the ε^1 order equation and obtain

$$\partial_{t_0t_0}h_1 + h_1 = \left(3|A_0|^3A_0 - 2i\partial_{t_1}A_0\right)e^{it_0} + A_0^3e^{3it_0} + (*). \tag{1.2.8}$$

Observe that this equation is a harmonic oscillator driven by a force with frequency 3 but also 1 which is the resonant frequency of the oscillator. Solving the equation including the homogeneous solution means linear growth and eventually breakdown. Breakdown is what we try to avoid with multiple scale approach. The factor in front of the resonant oscillatory term is called *secular term*. These terms cause growth and breakdown. Fortunately, we have the freedom to remove the secular term by postulating that

$$\partial_{t_1}A_0 = -\frac{3i}{2}|A_0|^2A_0. \tag{1.2.9}$$

With this condition, the ε^1 order equation simplifies to

$$\partial_{t_0t_0}h_1 + h_1 = A_0^3e^{3it_0} + (*). \tag{1.2.10}$$

One of the next consequences of treating h differently is that the general solution of the homogeneous part of the equation (1.2.10) is disregarded. In fact, it is disregarded for all the equations in the hierarchy except the ε^0 order. Solving (1.2.10) we take only the particular solution

$$h_1 = -\frac{1}{8}A_0^3e^{3it_0} + (*). \tag{1.2.11}$$

We now insert h_1 into the next order equation in the perturbation hierarchy, the ε^2 order and find

$$\partial_{t_0t_0}h_2 + h_2 = \left(-\frac{3}{8}|A_0|^4A_0 - 2i\partial_{t_2}A_0 - \partial_{t_1t_1}A_0\right)e^{it_0} + (*) + NST, \tag{1.2.12}$$

where NST stands for nonsecular terms. The reason why we do not bother calculating precisely these terms, is that we are not planning to go beyond the ε^2 order. The only terms needed at this order are the secular terms we want to remove. To remove them, we postulate

$$\partial_{t_2} A_0 = \frac{3i}{16}|A_0|^4 A_0 + \frac{i}{2}\partial_{t_1 t_1} A_0. \tag{1.2.13}$$

To summarize, what we have so far is

$$h(t_0, t_1, t_2, \ldots) = A_0(t_1, t_2, \ldots)e^{it_0} - \frac{1}{8}A_0^3 e^{3it_0} + (*), \tag{1.2.14}$$

and

$$\partial_{t_1} A_0 = -\frac{3i}{2}|A_0|^2 A_0, \tag{1.2.15}$$

$$\partial_{t_2} A_0 = \frac{3i}{16}|A_0|^4 A_0 + \frac{i}{2}\partial_{t_1 t_1} A_0. \tag{1.2.16}$$

We started with one differential equation and ended up with two coupled differential equations. There would be even more if we decided to go to higher orders. The equation for $\partial_{t_2} A_0$ can be simplified by removing the derivatives in the right hand side using the equation for $\partial_{t_1} A_0$ by taking its derivative with respect to t_1. This leads to the system

$$\begin{aligned} \partial_{t_1} A_0 &= -\frac{3i}{2}|A_0|^2 A_0, \\ \partial_{t_2} A_0 &= -\frac{15i}{16}|A_0|^4 A_0. \end{aligned} \tag{1.2.17}$$

One thing to observe about this system is that it is overdetermined. We have one function A_0 and two equations. Going to higher orders, the system would get even more overdetermined. Typically, overdetermined systems have no solutions, which means that under normal circumstances, the function $h(t_0, t_1, \ldots)$ does not exist. But we treat the function h differently in our setting. For systems of 1st order partial differential equations like (1.2.17) we can do a cross derivative test to check whether a solution exists. Taking the ∂_{t_2}-derivative of the first equation and ∂_{t_1}-derivative of the second equation we get

$$\begin{aligned} \partial_{t_2 t_1} A_0 &= -\frac{15i}{16}\left(2A_0 \partial_{t_2} A_0 A_0^* + A_0^2 \partial_{t_2} A_0^*\right) = -\frac{45}{32}|A_0|^6 A_0, \\ \partial_{t_1 t_2} A_0 &= -\frac{15i}{16}\left(3A_0^2 \partial_{t_1} A_0 \left(A_0^*\right)^2 + 2A_0^3 A_0^* \partial_{t_1} A_0^*\right) = -\frac{45}{32}|A_0|^6 A_0. \end{aligned} \tag{1.2.18}$$

The system is solvable according to the test, so the function h does exist. Or at least in two variables t_1, t_2. To check if it exists also for t_3, we would want to remove the secular terms for the ε^3 order equation and perform the cross derivative test with all 3 equations for A_0. In fact, if this was done, one would find that no matter how many orders one would take, the system would be solvable and the function h would exist. This is, of course, thanks to the current example we are solving. In general, we would not be so lucky with the existence of the function h. This fact is the reason why we are not taking h seriously as a multivariable function.

On the other hand, the nonexistence of the solution to the perturbation hierarchy is not a serious issue because it is actually the function $y(t)$ we care about, not h. Our aim is

to find the solution to the original equation and the existence of h is only of a theoretical interest. Inspired by this realization we define the *amplitude* $A(t)$ by

$$A(t) = A_0\left(t_1, t_2, \ldots\right)\big|_{t_j=\varepsilon^j t}. \tag{1.2.19}$$

Using (1.2.2) and (1.2.14), the perturbation expansion for $y(t)$ yields

$$y(t) = A(t)e^{it} - \varepsilon\frac{1}{8}A^3(t)e^{3it} + (*) + \mathcal{O}(\varepsilon^2). \tag{1.2.20}$$

And to get the equation for the amplitude $A(t)$, we multiply the first equation in (1.2.17) by ε and the second equation by ε^2 and add them up. We get

$$\begin{gathered}
\varepsilon\left(\partial_{t_1}A_0 + \frac{3i}{2}|A_0|^2A_0\right) + \varepsilon^2\left(\partial_{t_2}A_0 + \frac{15i}{16}|A_0|^4A_0\right) = 0,\\
\left(\partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots\right)A_0 + \varepsilon\frac{3i}{2}|A_0|^2A_0 + \varepsilon^2\frac{15i}{16}|A_0|^4A_0 = 0,\\
\Downarrow\\
\frac{\mathrm{d}}{\mathrm{d}t}A = -\varepsilon\frac{3i}{2}|A|^2A - \varepsilon^2\frac{15i}{16}|A|^4A.
\end{gathered} \tag{1.2.21}$$

This is an amplitude equation. The amplitude equation determines the amplitude which then determines the perturbation expansion for our solution to the original equation through (1.2.20) that is uniform for $t \lesssim \varepsilon^{-3}$. Observe that the amplitude equation (together with a given set of initial conditions that are yet to be determined) has a unique solution regardless of the solvability of the overdetermined system (1.2.17). Thus, the cross derivative test is unnecessary to perform.

At this point we can see that instead of solving a 2nd order nonlinear ODE for a real unknown function $y(t)$, we need to solve a 1st order ODE for a complex function $A(t)$. There are two reasons why the second alternative is better. Firstly, it is possible to solve (1.2.21) analytically whereas it is not the case for (1.2.1). On the other hand, this feature of the amplitude equation may disappear for higher orders in ε because of the more terms in the amplitude equation. Amplitude equations for many different equations share the same mathematical structure. Therefore solving more problems through this method helps to get insight into an amplitude equation and can be useful in other different situations.

Secondly, there is a difference between solving (1.2.1) and (1.2.21) from a numerical point of view. Numerical solutions require a carefully chosen time step that is bound by the physical context of the problem. In our case, the time step for the linearized form of (1.2.1) is constrained by the oscillation period which is of order 1. For the amplitude equation, the time step is constrained by the period ε^{-1}. This makes it much quicker to numerically solve (1.2.21) since we can take ε^{-1}-times less time steps. This difference can be significant since ε is small. It also makes MMS a tool for reformulation of a problem such that it is possible to solve a weakly nonlinear ODE or PDE with a help of a fast numerical method.

The last thing left to do is fitting the initial conditions. We use (1.2.20) truncating at

order ε together with the conditions from (1.2.1) to get the following equations

$$\begin{aligned} y(0) &= A(0) - \varepsilon\frac{1}{8}A^3(0) + (*) = 1, \\ y'(0) &= iA(0) - \varepsilon\left(\frac{3i}{2}|A(0)|^2A(0) + \frac{3i}{8}A^3(0)\right) \\ -\varepsilon^2&\left(\frac{15i}{16}|A(0)|^4A(0) - \frac{9i}{16}|A(0)|^2A^3(0)\right) + (*) = 0, \end{aligned} \tag{1.2.22}$$

where we used the amplitude equation to simplify the condition for the derivative and disregarded the ε^3 term after the substitution. Since the unknown $A(0)$ is a complex number, we can treat it as two unknown variables; the real, and the imaginary part. With the two equations, the system (1.2.22) represents a 2 by 2 system of nonlinear algebraic equations. It can be solved numerically, for example, using Newton's iteration starting with the solution for $\varepsilon = 0$, where the solution is easily obtained being $A(0) = 1/2$. This will give us the correct initial condition for the amplitude equation up to order ε^2.

We have demonstrated MMS on an example of a nonlinear cubic oscillator. This is only one face of the method from many. Similarly, other examples are solved introducing more than one amplitude. In the case of a nonlinear system of ODEs or PDEs, one must use a help from linear algebra, in particular Fredholm's alternative theorem, or obtaining the eigensystem of a matrix in order to solve the perturbation hierarchy. For singular perturbation problems, one needs to introduce appropriate transformation of the variable and then apply multiple scales.

It is clear that the MMS can be applied to many different situations where direct approach of perturbation expansion creates nonuniform expansions. In various problems with nonlinear PDEs and ODEs, different amount of algebra is required to construct and solve the perturbation hierarchy. In textbook examples, like the one we did here, the algebra is manageable. However, often in real world situations, it can be more challenging. In our third paper presented in this work, it is such a case. It illustrates the possible difficulty when applying MMS to nonlinear optics through the derivation of the amplitude equation for nonlinearly polarized light pulse in a dispersive medium. In our third paper, the nonlinear equation we are solving comes from the Maxwell's equations.

MMS became so popular in the 1970's, that it has been discovered over and over again nearly every half year. It has been done so in many parts of science such as physics, engineering or applied mathematics. Let us list some of the areas where it has a prominent role.

The problems that were analyzed in the 1960's and 70's are weakly linear and nonlinear vibrations governed by 2nd or 3rd order ODEs [22][23][24][25], nonlinear oscillations in differential equation with slowly varying coefficients [26], turning point problems for linear ODEs [27], linear equation with variable coefficients [28], the effect of the scales on an aging spring [29]. Another problems include the effect of cosmological expansion on particles described by inhomogeneous equations with slowly varying coefficients [30], problem of passing through resonance for an oscillator with slowly varying frequency [31], boundary value problems for nonlinear differential equations [32] or solving the Orr-Sommerfeld equation [33].

The earth-moon spaceship problem is one example that is used frequently in the textbooks to demonstrate MMS [34]. Within orbital mechanics, the problem of a satellite with a circular or elliptical orbits and a small thrust or drag were analyzed [35][36]. This led to a problem of a motion of satellite around three bodies [37] or the stability of the

triangular points in the elliptic problem of three bodies [38]. To analyze the motion of satellites in systems of bodies became popular and using MMS, even higher order terms were obtained in this motion taking into account the effects of eccentricity and inclination [39]. Furthermore, the motion of a satellite with a period comparable to the period of rotation of the gravitational primary was also studied [40].

As well as motion of satellites, the motion of missiles with asymmetries was also investigated. In particular, the nonlinear resonances in their motion [41]. A rolling missile with variable roll rate was studied by Nayfeh [42], with linear and nonlinear aerodynamics.

From a different part of physics, nonlinear dynamic buckling of imperfect and elastic dynamic columns of solid was analyzed [43]. MMS was also used to resolve a problem of travelling wave on a cylindrical and a spherical shell [44] as well as the nonlinear panel and membrane flutter [45]. The propagation of waves was also investigated in an inhomogeneous rod [46]. The Klein-Gordon equation was a popular equation to study using MMS as well [47][48].

As one can see from all the examples above, MMS is widely used to treat numerous problems from orbital, flight and solid mechanics, various other kinds of differential equations within wave interactions, atmospheric science, plasma physics, hydrodynamics, fluid dynamics, statistical mechanics or general physics. Nonlinearities tend to appear in almost every aspect of physics if one goes deep enough into investigating the given phenomenon. Methods like MMS help to gain more understanding precisely in situations like this.

2 Summary of the papers

In this section we introduce the three papers presented in this thesis. In particular, the leading idea behind them, the methods that are used and the main results.

2.1 Paper 1 - Convergence and completeness for square-well Stark resonant state expansions

The purpose of this paper is to investigate the completeness of the Stark resonant states for a quantum particle. In order to explain the problem and how it is related to nonlinear optics, we start with the physical setup that has a quantum nature. Consider an atom and an associated electron, which is a quantum particle, located in a square-well electric potential $V(x)$, say, due to the nucleus. This system is then exposed to a homogeneous external electric field with a constant fixed strength ε. In classical electrodynamics, the electric scalar potential has as a source, electric charge and is related to the electric field through the gradient of the potential. So there is an electric field present in the system at all times.

Since it is a quantum-mechanical system, we are using Schrödinger's equation (SE)

$$i\hbar\frac{\partial}{\partial_t}\psi(\mathbf{x},t) = H\psi(\mathbf{x},t), \tag{2.1.1}$$

that governs the wave function $\psi(\mathbf{x},t)$ of a such quantum-mechanical system. In order to apply SE, one needs to obtain the Hamiltonian H for the system. In other words, the energies of the particles in the system, accounting for the kinetic and potential energies. The unknown complex valued function $\psi(\mathbf{x},t)$ that solves SE is interpreted as a wave function for the system, a function of space and time. It contains an information about the system. The practical interpretation of the wave function is that it defines a probability density function when taken its square of the absolute value at each point. The wave function itself is a product of an exponential depending only on time with a complex frequency and a function of spatial variables called the eigenstates.

In the context of our system, the external electric field which could be produced by a laser, is providing energy into the system and drives the particle away from the atom. Eventually, the electron escapes from the confining but unstable potential. Given this phenomenon, the eigenstates are growing functions in space and thus not normalizable. This tells us that it is more likely to find the particle far from the nucleus than closer because it is more likely that it escaped earlier in time than later. On the other hand, the corresponding wave functions are exponentially decaying in time. As the name suggests, a wave function can also be interpreted as a "wave" in a sense that the electron generates it and creates the probability density field. Since the electron is the only source of this field, the waves should only be outgoing from the system. Therefore, no incoming waves are expected. We have just deduced two important properties of the quantum states: they are unstable and only outgoing. The wave functions satisfying the outgoing wave condition became known as *resonant states*, or also Stark resonant states. We have thus established that the resonant states decay exponentially in time and at each point in time, they grow in space in the direction of the laser source.

As stated in the beginning of this section, the aim of investigation of this paper is the convergence and completeness of the eigenstates corresponding to resonant states. An orthogonal set of functions $\{\phi_k\}$, where k is an index, can be used to represent square

integrable functions through the expansion

$$f(x) = \sum_k \frac{(f, \phi_k)}{(\phi_k, \phi_k)} \phi_k(x), \tag{2.1.2}$$

where $(\cdot, \cdot)$ is the inner product associated to the domain and the context in which the problem is set. According to mathematical analysis, an orthonormal set is said to be complete if the Parseval's equality $\|f\|^2 = \sum_k (f, \phi_k)^2$ holds for all square integrable functions, where $\| \cdot \|$ represents the norm associated to the inner product. When an orthonormal set is complete, then the expansion (2.1.2) for any square integrable function f converges in the mean square sense to f. In our case, the set of functions are the eigenstates. It was mentioned earlier that the eigenstates are not normalizable so in the standard von Neumann sense, they can not be used in such expansions to represent functions. However, an extended mathematical foundation was developed to remedy this problem involving rigged Hilbert spaces. This space is flexible enough to also contain resonant states. Nevertheless, the resonant states can be made normalizable by evaluating them on a line in the complex plane, as we will later see. In our paper, the completeness of the resonant states is determined through certain properties and the completeness of the scattering states. These states are obtained in the same way as the resonant states with considering both outgoing and incoming waves to the right of the well. Thus they are in a close relation to the resonant states. In particular, the scattering states are self-adjoint and the Hamiltonian has an absolutely continuous spectrum that is equal to the real line. Using the expansion (2.1.2) we also directly investigate its convergence rate through the asymptotic form of the terms in the expansion.

Let us explain, how is the problem we are studying related to nonlinear optics. We have learned at the start of this work how nonlinear response to an EM field arises. A present EM field in a material made up from atoms is induced by the atoms which generate electric dipoles. The density of these electric dipole moments defines polarization. Then we found the nonlinear relation between the electric field and the polarization in the material through optical susceptibilities (1.1.5). If we go deeper into understanding the issue with the dipoles, we eventually need to reach for quantum mechanics and show how the susceptibility depends on the dipole transition moments and atomic levels.

An electron in an atom occupies so called atomic levels depending on its energy. In a ground state, the electron has its minimal energy and can most likely be found in the first level nearest to the nucleus. Through EM radiation, the electron can gain energy absorbing photons and jump "higher" (higher means further from the nucleus in this context) into the next atomic level. These energy gains come in quantified amounts meaning that there is a precise amount of energy the electron must gain in order to jump. In the case of jumping to a "lower" level, the electron releases a photon with the same amount of energy required to jump. Besides the energy eigenlevels of the free atom, there are virtual levels which represent the combined energy of one energy eigenstate of the atom and one or more photons from the radiation field.

There is a technique called resonant enhancement that allows us to obtain large values of the nonlinear susceptibility because the energy levels of free atoms tend to be very sharp. The whole situation can be visualized by considering the interactions in terms of exchange of photons between the electrons and the radiation field. For example, an electron absorbs two photons of frequency ω, jumps two virtual levels and on the way down, it releases a photon with frequency 2ω. In resonant enhancement, one of the real atomic levels is almost coincident with one of the virtual levels and creates a strong

coupling between the radiation and the atom which than leads to a large nonlinear optical susceptibility.

Thanks to SE, all of the properties of the atomic system can be described in terms of the atomic wave function $\psi(\mathbf{x}, t)$. We express the appropriate Hamiltonian of the system as $H = H_0 + V(t)$, where H_0 is the Hamiltonian for a free atom and $V(t)$ describes the interaction of the atom with the EM field. It is usually taken as

$$V(t) = -e\ \mathbf{x} \cdot \hat{\mathbf{E}}(t), \tag{2.1.3}$$

where $-e$ is the charge of the electron, $-e\mathbf{x}$ is the electric dipole moment and $\hat{\mathbf{E}}(t)$ is the external electric field. We obtain the wave function $\psi(\mathbf{x}, t)$ by solving the SE. It contains the information from both the atom itself and the external field. The square of its absolute value of the normalized wave function provides the probability density distribution for the system. It is also used to obtain the expectation value of observables. By expectation value we mean the probabilistic expectation value of some measurement. One can think of it as an average of all possible outcomes of the measurement, where the outcomes are weighted by their likelihoods which are provided by the wave function. Let us consider A to be an observable. Mathematically, it can be a self-adjoint operator with respect to the Hilbert space. Then the expectation value of A in the state described by the normalized vector ψ (wave function) is defined as

$$\langle A \rangle_\psi = \langle \psi \, | A | \, \psi \rangle, \tag{2.1.4}$$

where $\langle A \rangle_\psi$ denotes the expectation value of A in the state ψ. The notation $|\psi\rangle$ is the Dirac notation of a normalized state vector. Let us say, the operator A is the position operator in 1D, $(A\psi)(x) = x\psi(x)$. Then the expectation value of A is calculated as

$$\langle A \rangle_\psi = \langle \psi \, | A | \, \psi \rangle = \int_{-\infty}^{\infty} \psi(x) x \psi^*(x) \mathrm{d}x = \int_{-\infty}^{\infty} x \, |\psi(x)|^2 \, \mathrm{d}x. \tag{2.1.5}$$

In general, the atom is exposed to an EM field such that the SE can not be solved exactly. In such cases, perturbation theory is often adequate to use to deal with the problem. In order to do so, the Hamiltonian H is replaced with

$$H = H_0 + \alpha V(t), \tag{2.1.6}$$

where α is a parameter that varies continuously ranging from zero to one. It characterizes the strength of the interaction between the atom and the EM field. For $\alpha = 1$ we have the full physical situation. Following the perturbation theory, we are now looking for the solution to the SE in the form of a power series in α

$$\psi(\mathbf{x}, t) = \psi^{(0)}(\mathbf{x}, t) + \alpha \psi^{(1)}(\mathbf{x}, t) + \alpha^2 \psi^{(2)}(\mathbf{x}, t) + \dots. \tag{2.1.7}$$

The solution of this form makes sure that $\psi^{(N)}$ will be the part of the solution which is of order N in the interaction energy V [49].

Carrying on the perturbation scheme, one finds the solutions to all orders. We then use these results to describe the linear optical properties of the material. In particular, the expectation value of the electric dipole moment can be described which is given by

$$\langle \mathbf{p} \rangle = \langle \psi \, | \mathbf{m} | \, \psi \rangle, \tag{2.1.8}$$

where $\mathbf{m} = -e\mathbf{x}$ is our observable, the electric dipole moment and ψ is given by the perturbation expansion (2.1.7) with $\alpha = 1$. We therefore find that the lowest order contribution to $\langle \mathbf{p} \rangle$ is given by

$$\langle \mathbf{p}^{(1)} \rangle = \langle \psi^{(0)} \, |\mathbf{m}| \, \psi^{(1)} \rangle + \langle \psi^{(1)} \, |\mathbf{m}| \, \psi^{(0)} \rangle, \tag{2.1.9}$$

where we assumed that the exact closed forms of all the quantities involved in the above expression were previously obtained. The expectation value (2.1.9) is then used to calculate the form of the linear polarization as

$$\mathbf{P}^{(1)} = N \langle \mathbf{p}^{(1)} \rangle, \tag{2.1.10}$$

where N is the number density of atoms. From the relation obtained above it is also possible to calculate the form of the linear susceptibility where one can identify the resonant and antiresonant contributions to the susceptibility.

Analogously, the expression for the second order polarization is derived in a similar way. The contribution of the second order to the induced dipole moment for an atom is given by

$$\langle \mathbf{p}^{(2)} \rangle = \langle \psi^{(0)} \, |\mathbf{m}| \, \psi^{(2)} \rangle + \langle \psi^{(1)} \, |\mathbf{m}| \, \psi^{(1)} \rangle + \langle \psi^{(2)} \, |\mathbf{m}| \, \psi^{(0)} \rangle, \tag{2.1.11}$$

and the corresponding polarization becomes $\mathbf{P}^{(2)} = N \langle \mathbf{p}^{(2)} \rangle$. This procedure is possible to carry on further to an arbitrary order.

Taking the sum of all the obtained orders of polarization, one arrives at the expression (1.1.5) with concrete formulas and finds that the first order contains the first power of the electric field, the second order term contains the second power, and so on [49]. The expressions derived using this method can also be used for more accurate predictions of a nonresonant response for atomic systems. But it fails to describe relaxation processes such as nearresonant response.

We have demonstrated the idea behind the connection between quantum mechanics and nonlinear optics and how nonlinear optics arises directly from the quantum level.

Let us now proceed to the details of the actual paper. We have already discussed the physical system which we are considering. The Hamiltonian is expressed by

$$H = -\frac{1}{2}\partial_{xx} + V(x) - \varepsilon x, \tag{2.1.12}$$

where $\varepsilon > 0$ is the strength of the external field and $V(x)$ is the atomic potential that is modelled by a square well of width $2d$ and depth V_0

$$V(x) = \begin{cases} -V_0, & |x| < d, \\ 0, & |x| > d \end{cases} . \tag{2.1.13}$$

According to the brief characterization of the resonant states we made earlier, i.e. they must only be outgoing waves, the solution we seek to the SE (2.1.1) has the form

$$\psi(x,t) = \psi_\omega(x) e^{-i\omega t}. \tag{2.1.14}$$

The functions $\psi_\omega(x)$ are the eigenstates corresponding to the eigenproblem $H\psi_\omega = \omega\psi_\omega$ and which satisfy the boundary conditions

$$\begin{aligned} &\psi_\omega(x) \to 0 \quad \text{when} \quad x \to -\infty \\ &\psi_\omega(x) \ \text{ and } \ \psi'_\omega(x) \ \text{ are continuous at } \ x = -d, d \\ &\psi_\omega(x) \ \text{ is a purely outgoing wave at } \quad x = \infty. \end{aligned} \tag{2.1.15}$$

The solutions we found to the eigenproblem are expressed using Airy functions and their combinations. An Airy function is a solution to the nonlinear ODE $y(x)'' - xy(x) = 0$. One can see that this form emerges in our calculations because of the external field term. Hence the eigenstates are

$$\psi_p(x) \equiv \psi_{\omega_p}(x) = \begin{cases} a_1 \mathrm{Ai}(y_1(x,\omega_p)) & x < -d, \\ a_2 \mathrm{Ai}(y_2(x,\omega_p)) + a_3 \mathrm{Bi}(y_2(x,\omega_p)) & -d < x < d, \\ a_4 \mathrm{Ci}^+(y_1(x,\omega_p)) & d < x < \infty, \end{cases} . \tag{2.1.16}$$

The function in the region $d < x$ is defined as $\mathrm{Ci}^+ = \mathrm{Bi} + i\mathrm{Ai}$. It is a combination of the Airy functions that asymptotically behaves as a purely outgoing wave. The functions $y_1(x,\omega)$ and $y_2(x,\omega)$ are parametrizations of the arguments such that the Airy functions are in fact solutions to the eigenproblem.

The eigenvalues ω_p are obtained from the continuity conditions for the eigenstate and its derivative at both ends of the square-well. Applying these conditions, one gets a 4 by 4 homogeneous system of linear equations for the coefficients a_i. The system has nontrivial solution if the determinant is zero $\det M(\omega_p) = 0$, where M is the 4 by 4 matrix. The determinant can be expressed in the form

$$\det M(\omega) = (A_0A_1' - A_0'A_1)(B_2C_3' - B_2'C_3) - (A_0B_1' - A_0'B_1)(A_2C_3' - A_2'C_3), \tag{2.1.17}$$

where the quantities involved are the Airy functions evaluated at different arguments depending on the parameters d, ε, V_0 and the variable ω, for example
$A_1 = \mathrm{Ai}\left(-(2\varepsilon)^{1/3} - d + (\omega + V_0)/\varepsilon\right)$. To find the zero points of (2.1.17) is anything but easy. What appears as one equation, is, in fact, two because of the complex nature of the determinant. Also, the count of the unknowns is two, namely the real and imaginary part of ω_p. The first thing that comes across the mind is to plot the real and imaginary part of the determinant as a contourplot $\det M(\omega_p) = 0$. The points where they cross in the complex plane, are the resonant eigenvalues ω_p. A system with a particle confined only in a square-well without an external field would have a finite number of eigenvalues on the negative real axis corresponding to the bound states. The external field with arbitrary nonzero strength has a large effect on the energetic spectrum, or the number and locations of eigenvalues. It causes all the finite eigenvalues to shift a small amount downwards in the complex plane close to the negative real axis and creates two infinite families of new eigenvalues. The graph of the contourplot reveals us two infinite families of eigenvalues, both of them in the lower half-plane. One is located along the positive real axis, we call it the A-series and they are long-living states since the imaginary part of ω_p is small. The other family is located along the ray $\arg(z) = -2\pi/3$, which we call the C-series and which corresponds to fast-decaying states since here, the imaginary part of ω_p is large. The finite number of zero points close to the negative real axis are the previously mentioned perturbed bound states. The same behaviour and locations of the eigenvalues after "switching on" the external field is also observed with the Dirac delta potential [50], so we reckon, this phenomenon is generic.

It was possible to obtain the asymptotic formula for the zero points from the asymptotic forms of Airy functions. The formula was, of course, not exact, but it provided excellent initial points that were used in Newton's iteration method to obtain the actual ones with arbitrary accuracy.

Now that we have both the resonant states and the corresponding eigenvalues at our disposal, we focus on the goal of the paper. That is to investigate the possibility of the

resonant states to be used for the expansion

$$f(x) = \sum_p \frac{(f, \psi_p)}{(\psi_p, \psi_p)} \psi_p(x). \tag{2.1.18}$$

As it turns out, the eigenproblem $H\psi_\omega = \omega\psi_\omega$ with (2.1.15) is not self-adjoint and because of the complex nature of the eigenvalues, the eigenstates are not normalizable on the real line. In order to normalize them, we introduce a complex contour $\mathcal{L}$, on which the resonant states decay and thus, it is possible to normalize them. The contour has the following form

$$\mathcal{L} = z(x) = \begin{cases} x & x < x_c \\ x_c + e^{i\theta}(x - x_c) & x > x_c \end{cases}, \tag{2.1.19}$$

where $x_c > 0$ and $0 < \theta \leq \pi/2$ are chosen parameters. Note that the contour serves its purpose for any chosen parameter within its range. The resonant states decay on the negative real axis but grow exponentially on the positive real axis. However, evaluated on the contour, they also decay for $x > x_c$ and for any angle θ within its range. In our paper, we chose $\theta = \pi/2$ and $x_c > d$. In this way, the contour $\mathcal{L}$ is used to normalize the states. The usual Hermitian inner product for complex functions on the real line is replaced by a bilinear inner product for complex valued function on the contour $\mathcal{L}$. In situations like this, it is common to define a bilinear inner product for complex valued functions (Φ, Ψ) for a complex contour $\mathcal{C}$ as

$$(\Phi, \Psi) = \int_C \Phi(z)\overline{\Psi}(z)\mathrm{d}z. \tag{2.1.20}$$

From complex analysis we know that taking the complex conjugate of a complex function removes its analyticity. Therefore, a natural generalization is a complex conjugate defined as $\overline{\Psi}(z) = \overline{\Psi(\bar{z})}$. Inspired by this, we identified $\varphi(x) = \Phi(z(x))$ and $\psi(x) = \overline{\Psi}(z(x))$, where $z(x)$ is defined in (2.1.19) and with $\theta = \pi/2$ we get the following explicit formula for the inner product

$$(\varphi, \psi) = \int_{-\infty}^{x_c} \varphi(x)\psi(x)\mathrm{d}x + i\int_{x_c}^{\infty} \varphi(x)\psi(x)\mathrm{d}x. \tag{2.1.21}$$

Here, the functions φ and ψ are continuous at $x = x_c$ but not smooth in a sense that the first derivatives satisfy $\lim_{x\to x_c^+} \partial_x\varphi = i \lim_{x\to x_c^-} \partial_x\varphi$ and the same goes for ψ. In the case of an operator with real eigenvalues and eigenfunctions, the operator is self-adjoint. The eigenfunctions belong to a real vector space where the usual inner product is defined. But because of the discontinuity of the derivative of the functions, their complex nature and the nonpositivness of the inner product (2.1.21), they belong to a somewhat different vector space than usual. Our goal is to show completeness of the states and without a self-adjoint operator, we would be in trouble. Having a self-adjoint operator has its advantage in the form of orthogonality of its eigenfunctions. And without the orthogonality, the expansion (2.1.18) would not be a reality. Indeed, this would be the case if not for the bilinear complex inner product (2.1.21). As it turns out, on the current vector space equipped with this inner product, the Hamiltonian (2.1.12) is, in fact, self-adjoint $(H\varphi, \psi) = (\varphi, H\psi)$. Under the process of showing this, one must use the discontinuity conditions at $x = x_c$. Thus, the states corresponding to different eigenvalues are orthogonal. We would like the reader to pause at this point and appreciate this fact

since it is not obvious at all that given such unusual vector space and inner product, one would be led to orthogonality and self-adjointness.

Usually, this would prove the completeness but we do not have a Hilbert space. One would prove it for the vector space we are currently working in, using functional analysis and constructing topologies. In this paper we are not following this path but rather a one that leads to weaker convergence.

Having defined the most important mathematical tools, we are ready to perform numerical tests in order to show, how well the expansion (2.1.18) represents functions. The main focus was on Gaussians and wavepackets. From previous work done on this topic using the Dirac delta potential [50] we expected the expansion to converge to the left of the well. Based on our observations from the numerical tests, the expansion appears to do so left to the square well but does not converge at the same rate inside the well. In fact, the convergence seemed slower than exponential because many more terms had to be included to achieve a satisfactory match. This constituted our hypothesis that the resonant state expansion converges inside the well.

We next proceeded to the actual proof of this assumption. The way we approached to the execution of the proof is somewhat more physical than mathematical. To prove the completeness for the resonant states, we use the completeness for the scattering states of our system. Scattering states arise when we allow both incoming and outgoing waves as oppose to only outgoing waves for the resonant states. So in a physical sense, scattering and resonant states are close to each other. For the Stark Hamiltonian (2.1.12), the scattering states are real valued functions and have real eigenvalues and they constitute a continuous real spectrum. Thus we have completeness for them expressed as

$$\int_{-\infty}^{\infty} \psi_\omega(x)\psi_\omega(x')\mathrm{d}\omega = \delta(x-x'), \tag{2.1.22}$$

where now $\psi_\omega(x)$ are the scattering states with ω being real. One can illustrate how does the formula (2.1.22) help to obtain function representation (2.1.18). For eigenstates $\phi_k(x)$ of a system with a discrete spectrum, the completeness takes the discrete form

$$\sum_k \phi_k(x)\phi_k(x') = \delta(x-x'), \tag{2.1.23}$$

where it is assumed that the eigenstates $\phi_k(x)$ are normalized. Multiplying (2.1.23) by a function with a compact support $f(x)$ and integrating over the real x'-axis, we get

$$\begin{aligned} \sum_k \phi_k(x) \int_{-\infty}^{\infty} f(x')\phi_k(x')\mathrm{d}x' &= \int_{-\infty}^{\infty} f(x')\delta(x-x')\mathrm{d}x', \\ \sum_k \phi_k(x)(f,\phi_k) &= f(x), \end{aligned} \tag{2.1.24}$$

where the inner product (f,ψ_k) is defined as $\int_{-\infty}^{\infty} f(x)\phi_k(x)\mathrm{d}x$. It is thus demonstrated this way that completeness leads to an expansion equivalent to (2.1.18).

The structure of the scattering states is similar to the resonant states (2.1.16) except for the region $d < x$, where we have a linear combination of the outgoing Ci^+ and incoming

Ci^- part:

$$\psi_\omega(x) = \chi \begin{cases} \frac{2}{\pi^2 |\det \mathbf{M}(\omega)|} \mathrm{Ai}\left(\mu\left(x + \frac{\omega}{\varepsilon}\right)\right) & x < -d, \\ \frac{2}{\pi |\det \mathbf{M}(\omega)|} \left[(B_1' A_0 - B_1 A_0') \mathrm{Ai}\left(\mu\left(x + \frac{\omega+V_0}{\varepsilon}\right)\right)\right. & \\ \qquad \left. + (A_1 A_0' - A_1' A_0) \mathrm{Bi}\left(\mu\left(x + \frac{\omega+V_0}{\varepsilon}\right)\right)\right] & -d < x < d, \\ i \left(\frac{\overline{\det \mathbf{M}}(\omega)}{\det \mathbf{M}(\omega)}\right)^{\frac{1}{2}} \mathrm{Ci}^+\left(\mu\left(x + \frac{\omega}{\varepsilon}\right)\right) & \\ -i \left(\frac{\det \mathbf{M}(\omega)}{\overline{\det \mathbf{M}}(\omega)}\right)^{\frac{1}{2}} \mathrm{Ci}^-\left(\mu\left(x + \frac{\omega}{\varepsilon}\right)\right) & d < x, \end{cases} \tag{2.1.25}$$

where $\mu = -(2\varepsilon)^{\frac{1}{3}}$. By playing with the formulas, it is possible to split na scattering states into an incoming and outgoing part, so by this construction we have $\psi_\omega = \psi_\omega^+ + \psi_\omega^-$, where ψ_ω^+ is the outgoing part that includes only the combination Ci^+ and ψ_ω^- is the incoming part with Ci^-. Next, we can use this splitting to split any function $f(x)$ into outgoing and incoming part $f(x) = f^+(x) + f^-(x)$, where

$$f^\pm(x) = \frac{1}{2} \int_{-\infty}^{\infty} \mathrm{d}\omega' \, a(\omega') \left[\mp i \frac{|\det \mathbf{M}(\omega')|}{p(\omega')}\right] \int_{-\infty}^{\infty} \mathrm{d}\omega \, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' \mp i\xi}, \tag{2.1.26}$$

with $a(\omega') = \int_{-\infty}^{\infty} \mathrm{d}x \, \psi_{\omega'}(x) f(x)$ being the energy representation of $f(x)$. $\mathbf{M}(\omega')$ is the matrix of the system of equations coming from the continuity conditions at the boundaries of the well and $p(\omega')$ is an expression involving the Airy functions. The parameter ξ was introduced artificially as a correction variable to ensure exponential convergence of the functions $\mathrm{Ci}^\pm(x)$ for large values of x which are involved in the inner integral in (2.1.26). These functions decay algebraically, but not fast enough for the integral to converge. For this reason we introduced a correction in the variable ω'. Calculations will be performed treating ξ to be finite and at the end we remove this regularization by letting it approach zero. The formula (2.1.26) was obtained using the completeness (2.1.22) as well as special rules for the antiderivatives for Airy functions and their combinations.

The goal with the formula (2.1.26) is to obtain an expression for $f^\pm(x)$ as a linear combination of resonant states which would represent the desired expansion (2.1.18). First, we decide to focus only on the outgoing part $f^+(x)$ and the incoming part is treated in a similar way. From the form of the expression in (2.1.26) we see that in order to get an expansion, the inner integral must represent functions that are proportional to the resonant states (2.1.16). The corresponding integrand is complex and has a pole at the point $\omega = \omega' + i\xi$ in the upper complex half-plane. Through the functions ψ_ω (2.1.25), it also has poles both in the lower and upper half-plane because of $\det \mathbf{M}(\omega)$ and its complex conjugate, respectively. These determinants are part of the coefficients in the scattering states that ensure the continuity, where they appear in the denominator.

Having an integral over the real axis of a complex valued function with complex poles immediately invites for the use of Cauchy's residue theorem. Briefly put, it states, that if Γ is a simple closed positively oriented contour in the complex plane and g is a complex valued function which is analytic inside and on Γ except at the points $z_1, z_2, \ldots, z_n$, then

$$\int_\Gamma g(z) \mathrm{d}z = 2\pi i \sum_{j=1}^{n} \mathrm{Res}(g, z_j), \tag{2.1.27}$$

where $\mathrm{Res}(z_j)$ is the residue of g at the point z_j and is defined as

$$\mathrm{Res}(g, z_j) = \lim_{z \to z_j} \frac{1}{(m-1)!} \frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}} \left[(z - z_j)^m g(z)\right], \tag{2.1.28}$$

where m is the order of the pole z_j. Applying Cauchy's residue theorem to our problem, we chose a closed contour Γ being a half-circle negatively oriented in the lower half-plane, where the center of the circle is the origin and its radius is R. Choosing to integrate the integrand found in the inner integral in (2.1.26) along Γ, we can split the integral into two parts: along the arc and the diameter in the following way

$$\int_\Gamma \mathrm{d}\omega\ \psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'\mp i\xi} = \lim_{R\to\infty}\left[\int_{-R}^{R}\mathrm{d}\omega\ \psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'\mp i\xi} + \int_{C_R}\mathrm{d}\omega\ \psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'\mp i\xi}\right], \tag{2.1.29}$$

where C_R denotes the arc part of the half-circle. Since the scattering states include the term $\det\mathbf{M}(\omega)$ in the denominator of the coefficients, the integrand has poles being exactly the energy eigenvalues ω_j calculated for the resonant states. Using Cauchy's residue theorem, we get from (2.1.29)

$$\begin{aligned}\int_{-\infty}^{\infty}\mathrm{d}\omega\ \psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'\mp i\xi} &= -2\pi i\sum_j \mathrm{Res}\left(\psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'-i\xi},\omega_j\right)\\ &\quad - \lim_{R\to\infty}\int_{C_R}\mathrm{d}\omega\ \psi_\omega(x)\frac{\psi_\omega(d)}{\omega-\omega'\mp i\xi},\end{aligned} \tag{2.1.30}$$

where there is a minus sign by the sum of the residues term because the orientation of Γ is negative. Upon evaluating the residues, the regulator ξ can be safely removed. After evaluating the limits in the residues, the final outcome will include the scattering states at the resonant eigenvalues ω_j, due to the definition of the residue terms (2.1.28). This removes the incoming part of the scattering states leaving only the outgoing wave and turning the scattering states into functions that are proportional to resonant states. Furthermore, if the integral over C_R in (2.1.30) vanishes in the limit when R approaches infinity, then the inner integral in (2.1.26) can be written as a linear combination of the resonant states and thus, the function $f^+(x)$ itself can be written as such. At this point, we have reached the main point of the proof. The function f^+ then can be written as

$$f^+(x) = \sum_j c_j\psi_j(x), \tag{2.1.31}$$

where the coefficients c_j include the factors from the residues and the outer integral in (2.1.26) and $\psi_j(x)$ are functions proportional to resonant states corresponding to the eigenvalue ω_j. A similar expression can be found for the incoming part f^-. The overall expansion would then be the sum of the expansions for both f^+ and f^-.

The rest of the proof in the article involves evaluating the integral over C_R in (2.1.30) and showing that under certain conditions, it vanishes in the limit. It is shown by parametrizing the circular arc with $\omega = Re^{i\theta}$. Since we let R to be arbitrary large, the Airy functions involved in the integrand can be replaced by their asymptotic formulas for large arguments. These asymptotic formulas also change depending on the phase θ of the complex argument dividing the circular arc into two angular sectors. We found that in the sector $-\pi < \theta < -2\pi/3$, the integrand decays exponentially for all x. In the second sector $-2\pi/3 < \theta < 0$, however, decays exponentially for $x < -d$ and for $x > d$, the limit does not exist. For inside the well, $-d < x < d$, we found that the integrand decays independently of the depth of the well V_0. Despite its appearance in the final expression at multiple places, the depth V_0 disappears at the end. This may be the main result of the paper.

Using the completeness of the scattering states, we have shown the completeness of the resonant states by obtaining an expansion of a function (2.1.31). In the last chapter of the paper, we investigate the convergence of the resonant state expansion (2.1.18) with a more direct method. First, we determine what contributes the most to the expansion. The energy eigenvalues are divided into three different series, as we discussed. The perturbed bound state eigenvalues is a finite set and therefore, it does not contribute. Eigenvalues belonging to the C-series are located along the $\arg(z) = -2\pi/3$ ray in the complex plane and tend to have very large imaginary part which means, they decay exponentially very quickly. Thus we can also rule them out. A detailed investigation of the contribution from this series confirmed this. So this leaves us with the A-series.

The question we started with was how does the expansion (2.1.18) behave for large indices p. To answer this, one needs the asymptotic forms of the resonant states. They are made up only of Airy functions whose different asymptotic forms are well known. First, the asymptotic form of the A-series eigenvalues was found ω_p for $p \gg 1$. To confirm how accurate the asymptotic formulas are, we compared them with the high precision numerical eigenvalues found by some iteration method. The region of interest for the convergence was, of course, inside of the well. Since the well is placed symmetrically around the origin, we investigated the expansion at the point $x = 0$. The function represented by the expansion $f(x)$ is assumed to have its support inside the well too. With this said, the resonant expansion for f then becomes

$$f = \sum_p \frac{b_p}{N_p}, \text{ where } b_p = \psi_{\omega_p}(0) \int_{-d}^{d} f(x)\psi_{\omega_p}(x)\mathrm{d}x, N_p = \int_{-d}^{d} \psi_{\omega_p}(x)^2 \mathrm{d}x. \tag{2.1.32}$$

The asymptotic formulas were then obtained for b_p and N_p using a Gaussian function for f and, again, compared with the numerically calculated values with high precision. To finish with, we analysed the rate of convergence with some estimates choosing a particular function f. For the simplest choice for f being a constant 1 inside and 0 outside the well, we could obtain the analytic expression for b_p. The resulting expression depended inversely on V_0. Next, we assumed a general function f with its support inside the well and being zero at the boundaries of the well that is n times continuously differentiable. The asymptotic formula for b_p contains the integral as seen from (2.1.32), where the resonant states ψ_{ω_p} are a combination of cosine and sine functions. Solving the mentioned integral using integration by parts, we get the factor by the x term in the arguments of the trigonometric functions in the denominator each time we perform integration by parts. That factor behaves asymptotically as $\sim p^{\frac{1}{3}}$. Thus, multiplying the asymptotic expression by $p^{-\frac{1}{3}}$ enough times causes the series to converge algebraically. In fact the rate of convergence can be estimated to be

$$\left|\frac{b_p}{N_p}\right| \lesssim p^{-\frac{n+1}{3}} \frac{\beta M_n}{V_0 \left|\cos\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|}, \tag{2.1.33}$$

where β is a constant not depending on p or V_0 and $\left|f^{(n)}(x)\right| \leq M_n$ for $-d < x < d$. With the formula (2.1.33) we have obtained the dependence of the terms in the series on the smoothness of the target function f and the depth V_0. The rate of convergence can then be acquired by dividing two consecutive terms (2.1.33). One would find that the dependence on the depth V_0 disappears.

In this paper we have investigated the completeness of the Stark resonant states for a system with square well potential and external field. We found that they converge

pointwise to the left and inside of the well independently of the well depth V_0. Inside of the well, the resonant state expansion series converge algebraically and the size of the terms grows as V_0^{-1}. This means that for a small value for V_0, one would need to include more terms in the expansion. The algebraic convergence also explains the observation that the expansion needed to include more terms in order to achieve better match with the target function.

Exploiting the fact that the resonant state expansion converges inside the well for any nonzero depth V_0, we propose a similar result for a general potential with a compact support, e.g. a Gaussian. Such potential could be approximated using several square wells with different nonzero depths and as such, the resonant state expansion would be expected to converge here as well, no matter how shallow the potential would be. At the end, a limit would be taken of the joined square wells approaching a smooth potential.

2.2 Paper 2 - Constructing a partially transparent computational boundary for UPPE using leaky modes

The propagation of optical pulses in a weakly nonlinear dispersive medium is usually investigated through the nonlinear Schrödinger equation (NLSE) [51]. It provides a robust description and is based on the assumption of a weak instantaneous nonlinearity and that the pulse is composed of a carrier wave with a slowly varying envelope. It can be derived directly from Maxwell's equations as an asymptotic expansion in a small parameter including higher order correction terms [19]. This can be done using the method of multiple scales, which we widely described earlier in this work. The small parameter, in this case, may represent the Kerr coefficient which stands in front of the third power of the electric field term (1.1.6). Over time, the development of better laser technologies allowed to use shorter pulses with higher intensities which makes the foundation of NLSE and envelope equations no longer valid even with correction terms included. Derivation of improved and corrected equation has been done by several studies [52][53]. But for high-power optical pulses in dispersive nonlinear media, solving the Maxwell's equations directly is not feasible. Therefore, NLSE is in this case replaced by a pulse propagation equation which is a smooth transition from Maxwell's equations to envelope based models and takes advantage of well-defined direction of the pulse propagation. There are number of such pulse evolution equations that differ in assumptions and approximations [54]. However, there is one that stands out and applies to all equations that are designed to treat the optical pulse propagating in one direction. The main assumption in this equation is that the nonlinear polarization response (which we discussed to large extent) can be approximated by the response from portion of the optical field that propagates forward. This replacement is called the unidirectional approximation. The equation that is derived using this approximation explicitly is the *unidirectional pulse propagation equation*, or UPPE [55][56]. It captures the linear and nonlinear response of real materials over a spectrum that is physically relevant. UPPE provides a correct description of extreme linear and nonlinear focusing events in the scale of order of magnitude approaching the wavelength of light, unlike the earlier approaches. Some of the advantages of UPPE is that envelope equation can be derived from it and can be effectively and numerically implemented.

It was given relatively little attention to, whether it is actually possible to clarify if a given situation can be treated with the unidirectional approximation. Limits to unidirectional propagation were investigated by Kinsler [57] where he explored the one-dimensional model and pointed out the existence of backward propagating waves that affect the non-

linear response for the forward propagating component. This problem was picked up by Jakobsen [58] where he introduced a method called bidirectional pulse propagation which is an extension to the unidirectional method and takes over when the basic assumptions behind UPPE become invalid.

In this paper, we create transparent computational domain to simulate unidirectional propagation of optical pulses. The transparent computational domain is used to mimic infinitely large domain. Solution of any scattering problem in such unbounded region is problematic to tackle numerically since one needs to truncate the domain without having too much of an effect on the original problem. In order to do so, the truncation should be easy to implement (numerically speaking) and efficient. The first typical condition the numerical solution must satisfy for a correct representation of a propagating optical pulse, is the radiation condition at infinity. Consider a source of radiation (EM waves) located at the origin of a plane. We assume that this source is the only originator of all waves and has a compact support. From this basic assumption it is obvious, that no left moving waves should be to the right of the source's support and likewise, no right moving wave to the left of the source. All wave phenomena must satisfy the condition that waves move away from their source. This condition is called the radiation condition at infinity, or outgoing waves at infinity and it is the correct boundary condition when describing wavelike phenomena.

Implementing an infinite computational domain involves some key elements, each method is bound to consider. When solving a PDE numerically by discretization of the domain, the computational grid must be truncated in some way. So the main question is how to truncate without introducing unwanted artefacts into the computation. Introducing periodic boundary conditions is sometimes natural to do when dealing with periodic structures. In problems whose solution rapidly decay with exponential rate, truncating might be unnecessary as long as the computational domain is large enough. In contrast, solutions that vary slower at larger distances, one can simply use coordinate transformation to remap the infinite interval $(-\infty, \infty)$ to a finite interval $(-1, 1)$ using, for example, hyperbolic tangent. But the domain in problems involving wave equations are very challenging because the solutions typically oscillate and decay slowly at great distances. Here, the trick of remapping from infinite to finite domain does not hold because the solutions will oscillate infinitely fast at the boundary and this a serious problem for the finite numerical resolution. Instead of a transparent boundary, one gets a hard reflecting wall. This problem requires a different approach that makes the boundary in such a way that the waves are being absorbed when striking it, without any reflections and with achievable resolution.

Numerical methods such as boundary element method, infinite element method or methods based on truncating Fourier expansions were deployed to deal with this problem [59][60][61]. The absorbing boundary conditions method (ABC) counts as computationally effective [62][63]. Lowers order ABCs have a simple implementation, but because the truncated boundary of the domain is not fully non-reflecting, good accuracy is produced only for higher order ABCs [64]. Consequently, high accuracy means considerable computational cost and increase in the difficulty of the implementation.

An alternative approach was found when, in 1994, Berenger transformed the problem of truncation of unbounded domains and introduced absorbing boundaries for the wave equations [65]. The main idea was that instead of finding a boundary condition describing absorption, an absorbing boundary layer was used. This absorbing layer would be consisting of an artificially absorbing material placed at the edges of the computational

grid independent of the boundary conditions. It would work the following way: when a wave hits the absorbing layer, it is constricted by the absorption and exponentially decays. When the wave is reflected, it will be exponentially small when it returns again. There is however a problem with this approach. Because of the construct treating the boundary as the meet of two materials, one expects three reflecting waves. This problem was solved by constructing the layer such that waves do not reflect at the interface, the so called perfectly matched layer (PML) [66]. This method is applicable both for electromagnetism and other wave equations although it was initially created for Maxwell's equations and then used for Helmholtz equation and other problems in acoustics [67], elasticity or hyperbolic problems.

There are, however, problems where good boundary treatment is still missing. For example high intensity pulses with few cycles in nonlinear optics which have very broad spectra and spectral propagators are used to solve them. PML uses a finite difference solver which do not go well together with spectral methods. In this case of spectral propagators, any finite difference approach is not applicable and boundary treatment is still a problem. The interaction between highly nonlinear optical pulses and matter tends to send significant energy toward the boundaries where it must be absorbed to mimic the propagation into infinite space. In this paper we address this problem in connection with spectral-based numerical simulation.

The key element of the method in this paper is the introduction of a material outside the interface of the boundary with an artificial refracting index. Unlike PML where the layer absorbs, our layer is made as transparent as possible making its refracting index as close to the index inside the computational box as possible. Thus we create a waveguide that differs from its outside only by the refracting index. This artificial structure creates modes that propagate comparatively long along the propagation direction, or close to the waveguide axis (paraxial) until the waveguide is unable to retain the radiation from the source prior to its eventual escape, or leakage. The portion of the electromagnetic energy starts to leak out and is absorbed by the surrounding environment. Modes that exhibit such behaviour are called the *leaky modes*. They are characterized by a unique set of complex valued frequencies where the imaginary part represents the decay rate of the leaky mode.

The appearance of leaky modes in electromagnetics has a long history. In the previous paper we described modes called resonant states which are, in fact, a form of leaky modes. Their common feature is that they both include purely outgoing waves at infinity. Leaky modes exhibit the same behaviour as resonant states starting with the fact that they are also eigenstates decaying in time. They have had an important role in quantum mechanics since they were first described by Thomson in 1884 [68]. Just as in the case of resonant states, leaky modes being decaying eigenstates means that the corresponding operator in the eigenvalue problem is not self-adjoint. That means any projecting of functions into a sum of leaky modes and the completeness of the resulting expansions are not part of any general theory. Another significant feature of leaky modes is their unstable nature since they exponentially grow in space. Thus, as resonant states, they cannot be placed in any vector space with an inner product and be normalized. For such cases, a general theory is challenging to create and so the matter of completeness and projection must be handled case by case. This matter is addressed partially in our current paper. We introduce a method where a projection for leaky modes is possible based on a well-known technique of shifting them over to a contour in the complex plane just as we did with the resonant states.

Let us proceed with introducing the model. The propagation direction of the pulse we wish to propagate is chosen to be along the z-axis. We choose therefore a straight channel oriented along the z-axis with a uniform width across the x-axis (the transverse direction). This will be our waveguide. We assume that the EM field and the channel are independent of the y-variable, so the geometry of the whole model can be drawn on the xz plane with two straight lines at $x = -a$ and $x = a$ and the beginning of the channel at $z = 0$.

Before we continue, we need to decide what type of boundary conditions we want to have for the waveguide. Let us therefore briefly discuss Maxwell's boundary conditions in general. Consider a surface that divides 3D space into two parts, say, region 1 and 2. One can think of the surface as an interface of a material surrounded by free space, where region 1 would correspond to the material and region 2 to the free space or vice versa. Another example could be a hypothetical cut through a volume filled by one material medium dividing it into two region on the opposite sides of the cut. For simplicity, let us consider this surface to be smooth, such that for any point $\mathbf{x}_0$ at the surface, there would be a flat and smooth region around $\mathbf{x}_0$ containing $\mathbf{x}_0$ inside of it. Next, consider a short line segment through $\mathbf{x}_0$ and perpendicular to the surface. Both ends of this line segment define two points $\mathbf{x}_1$ and $\mathbf{x}_2$, where the indices indicate at which region the point is. Maxwell's equation are then used to find the boundary condition that relate the magnitudes of all four field components $\mathbf{E}_\parallel, \mathbf{H}_\parallel, \mathbf{D}_\perp$ and $\mathbf{B}_\perp$, where $\mathbf{E}$ is the electric field, $\mathbf{D}$ is the electric displacement field, $\mathbf{H}$ is the magnetic field and $\mathbf{B}$ is the magnetic induction. The subscripts $\parallel$ and $\perp$ indicate the parallel and perpendicular part of the field to the surface, respectively. In order to apply Maxwell's equations at the point $\mathbf{x}_0$ to obtain the boundary conditions, for example for $\mathbf{E}_\parallel$, the correct equation must be considered, in this case Maxwell's 3rd equation $\nabla \times \mathbf{E}(\mathbf{x}, t) = -\partial \mathbf{B}(\mathbf{x}, t)/\partial_t$. It includes the curl operator, therefore the equation is applied to a small rectangular loop centred at $\mathbf{x}_0$ and perpendicular to the surface. The longer sides of the rectangle containing $\mathbf{x}_1$ and $\mathbf{x}_2$ are taken to be parallel to the surface and the shorter ones to be perpendicular. Taking the integral over the surface of the small rectangle of Maxwell's 3rd equation and taking the limit $\mathbf{x}_1 \to \mathbf{x}_2$, when the loop shrinks and disappears at the surface, we find that the flux of the $\mathbf{B}$ field through the loop vanishes. The surface integral of a curl turns to a line integral through Stoke's theorem and we find $\mathbf{E}_\parallel(\mathbf{x}_1, t) = \mathbf{E}_\parallel(\mathbf{x}_2, t)$. In a similar manner, without the presence of any free charges and currents, the boundary conditions for the other fields are found to be $\mathbf{D}_\perp(\mathbf{x}_1, t) = \mathbf{D}_\perp(\mathbf{x}_2, t)$, $\mathbf{H}_\parallel(\mathbf{x}_1, t) = \mathbf{H}_\parallel(\mathbf{x}_2, t)$, $\mathbf{B}_\perp(\mathbf{x}_1, t) - \mathbf{B}_\perp(\mathbf{x}_2, t) = \sigma_{free}(\mathbf{x}_0, t)$ with σ_{free} being the free surface charge.

When solving Maxwell's equations in isotropic, homogeneous, linear media, a simply but powerful method is used that involves going back and forth between the space-time domain $(\mathbf{x}, t)$ and the Fourier domain $(\mathbf{k}, \omega)$. The sources can be as well transformed into their spectral domains and therefore, both the EM field and the sources can be expressed as a superposition of plane-waves. A plane electromagnetic wave has the electric and magnetic fields defined by

$$\begin{aligned} \mathbf{E}(\mathbf{x}, t) &= \mathrm{Re}\left[\mathbf{E}_0 \exp(i(\mathbf{k} \cdot \mathbf{x} - \omega t))\right], \\ \mathbf{H}(\mathbf{x}, t) &= \mathrm{Re}\left[\mathbf{H}_0 \exp(i(\mathbf{k} \cdot \mathbf{x} - \omega t))\right], \end{aligned} \tag{2.2.1}$$

where the vector quantities $\mathbf{k}, \mathbf{E}_0$ and $\mathbf{H}_0$ and the frequency ω are all complex valued. The vector $\mathbf{k}$ is associated with the propagation direction of the plane-wave and the vectors $\mathbf{E}_0, \mathbf{H}_0$ associated with the polarization state of the corresponding vector field. The sources such as induced polarization or magnetization are also expressed as plane-waves. If we

now substitute all the fields into Maxwell's equation and assume a material medium with no free charges or currents, we find simple equations containing dot products and vector products without any differential operators

$$\mathbf{k} \cdot \mathbf{E}_0 = 0, \tag{2.2.2a}$$
$$\mathbf{k} \times \mathbf{H}_0 = -\omega\varepsilon_0\varepsilon(\omega)\mathbf{E}_0, \tag{2.2.2b}$$
$$\mathbf{k} \times \mathbf{E}_0 = \omega\mu_0\mu(\omega)\mathbf{H}_0, \tag{2.2.2c}$$
$$\mathbf{k} \cdot \mathbf{H}_0 = 0, \tag{2.2.2d}$$

where μ_0, ε_0 and μ, ε are the permeability and permittivity of the free space and the material the field resides in, respectively. From these equations (2.2.2a)-(2.2.2d) one can derive the relation between the vector magnitude $|\mathbf{k}| = k$ and the frequency ω, called the *dispersion relation*. The propagation vector $\mathbf{k}$ is, in general, a 3D vector in a Cartesian coordinate system with components $\mathbf{k} = (k_x, k_y, k_z)$. The dispersion relation tells us that if, for example, k_x and k_y are known, then k_z can be obtained using this constraint. Similarly, through Maxwell's equations in the Fourier domain, one can calculate E_{0z} from the knowledge of E_{0x} and E_{0x}, where $\mathbf{E}_0 = (E_{0x}, E_{0y}, E_{0z})$ using (2.2.2a). Other field components are then easily determined using the rest of the equations.

In our setting, where the channel resides in the xz-plane, we assumed that the pulse will not propagate in the y-direction, so we can set $k_y = 0$. For a moment, let us forget about that the pulse starts propagating at $z = 0$. The goal is to find the boundary conditions at $z = 0$. Consider a plane-wave that is incident from above to the flat interface, in this situation being the xy-plane. We expect the part of it being reflected and a part being transmitted. From the discussion earlier about the Maxwell's boundary conditions, we found that the parallel part of the electric field must be continuous at the interface. In our case, the parallel part is $\mathbf{E}_{\parallel} = (E_x, E_y)$ such that these vector components must be continuous. Similarly, we also find that the continuity of D_z, H_x and H_y must be required at all interfacial points $(x, y, 0)$. This is only possible if the following three conditions are satisfied:

- $\omega^i = \omega^r = \omega^t$, where the upper subscripts correspond to the incident, reflected, and transmitted wave, respectively.
- the x-components of the vector $\mathbf{k}$ to all three waves must be equal, namely $k_x^i = k_x^r = k_x^t$. We recognize this as a generalization of Snell's law of optics.
- the y-components of the vector $\mathbf{k}$ to all three waves must be equal, namely $k_y^i = k_y^r = k_y^t$.

Once the tangential field components E_x, E_y and H_x, H_y are made continuous, Maxwell's equations in the Fourier domain make automatically the components D_z and B_z continuous.

In the preceding discussion we set $k_y = 0$ which leaves the unknowns in the dispersion relation to be k_x and k_z. So knowing k_x determines k_z. We have also established that knowing E_{0x} and E_{0y} determines E_{0z}. Totally, the quantities $\omega^i, k_x^i, E_{0x}^i$ and E_{0y}^i uniquely determine the EM field of the incident plane-wave. Because of the three conditions listed above together with the dispersion relation, the frequencies and the $\mathbf{k}$ vectors of all 3 waves (incident, reflected and transmitted) are determined. Since we need to ensure the continuity of four components E_x, E_y and H_x, H_y, totally we have four equations. We

know the quantities for the incident wave so the four unknowns are going to be E^r_{0x}, E^r_{0y} and E^t_{0x}, E^t_{0y}.

We can observe by writing down all four of the Maxwell's equations in Fourier domain

$$E_{0z} = -\frac{k_x}{k_z} E_{0x}, \tag{2.2.3a}$$

$$H_{0x} = -\frac{k_z}{\omega\mu_0\mu} E_{0y}, \tag{2.2.3b}$$

$$H_{0y} = \frac{k_z E_{0x} - k_x E_{0z}}{\omega\mu_0\mu}, \tag{2.2.3c}$$

$$H_{0z} = \frac{k_x}{\omega\mu_0\mu} E_{0y}, \tag{2.2.3d}$$

where μ_0 and μ are the permeability of the free space and the material the field resides in, that they can be separated into two independent groups:

- knowing E_{0x} in equations (2.2.3a) and (2.2.3c), one needs only these two equations to determine E_{0z} and H_{0y}. The incident wave is therefore specified by $(k_x, k_z, E_{0x}, E_{0z}, H_{0y})$ and is commonly referred to as *Transverse Magnetic* (TM), or p-polarized wave. Here we assume that for the incident wave we have $k^i_y = E^i_{0y} = 0$. The only independent variables are k^i_x and E^i_{0x}. All other field components are determined through (2.2.3a)-(2.2.3d).

- knowing E_{0y} in equations (2.2.3b) and (2.2.3d), one needs only these two equations to determine H_{0z} and H_{0x}. The incident wave is therefore specified by $(k_x, k_z, E_{0y}, H_{0x}, H_{0z})$ and is commonly referred to as *Transverse Electric* (TE), or s-polarized wave. Here we assume that for the incident wave we have $k^i_y = E^i_{0x} = E^i_{0z} = 0$. The only independent variables are k^i_x and E^i_{0y}. All other field components are determined through (2.2.3a)-(2.2.3d).

For these two groups of waves one can calculate the field components for the reflected and transmitted waves, but we are not going to pursue that here.

With the current geometry of the model in our paper, we want the electric field to be parallel to the boundaries of the slab. According to this requirement, we are choosing the electric field to be transverse electric, or TE, because the only nonzero component is the y-component and thus, parallel to the planes $x = \pm a$ at the boundary. In the case of TM waves, the x-component would be nonzero which is in the transverse direction to the boundary. We have also assumed that the field does not change in the y-direction. So we have

$$\begin{aligned} \mathbf{E}(\mathbf{r}, \omega) &= (0, e(x, z, \omega), 0), \\ \mathbf{P}(\mathbf{r}, \omega) &= (0, p(x, z, \omega), 0), \end{aligned} \tag{2.2.4}$$

where $\mathbf{E}$ is the electric field and $\mathbf{P}$ is the polarization. The electric field in the time domain is then simply obtained using inverse Fourier transform in $\omega \to t$.

The propagation equation is readily derived using Maxwell's equations with no free charges or currents. In the time frequency domain it has the form

$$\partial_{zz} e(x, z, \omega) + \partial_{xx} e(x, z, \omega) + \left(\frac{\omega}{c}\right)^2 n^2(x, \omega) e(x, z, \omega) = p(x, z, \omega), \tag{2.2.5}$$

where $n(x,\omega)$ is the refracting index and c is the speed of light. The refractive index $n(x,\omega)$ is designed to have the value 1 (vacuum) inside the channel and value > 1 outside:

$$n(x,\omega) = \begin{cases} 1 & |x| < a \\ n & |x| > a \end{cases}, \tag{2.2.6}$$

where n is some constant larger than 1.

As we discussed the Maxwell's boundary conditions, we found that the tangential part of the electric field to the boundary must be continuous. Our field is TE which leaves the tangential part to be $e(x,z,\omega)$ and the continuity equations become

$$\begin{aligned} e(\pm a_-, z, \omega) &= e(\pm a_+, z, \omega), \\ \partial_x e(\pm a_-, z, \omega) &= \partial_x e(\pm a_+, z, \omega). \end{aligned} \tag{2.2.7}$$

To write down a suitable spectral propagator, such as UPPE for this model, the goal is to find the leaky modes for the linearized version of (2.2.5) which are then used to develop a leaky mode transform. This transform will be a substitute for the usual Fourier transform. We get the linearized version of (2.2.5) by letting $p = 0$.

When solving the linearized model we assume the wave starts propagating at $z = 0$. As the wave propagates, it eventually hits the boundaries where a part of it is reflected and a part transmitted. For this reason, the solution includes only right travelling wave for $x > a$ and only left travelling waves for $x < -a$. Such functions have the form

$$\begin{aligned} e(x,z,\omega) &= De^{i\beta z}e^{i\xi x}, && x > a, \\ e(x,z,\omega) &= e^{i\beta_0 z}\left(Be^{i\xi_0 x} + Ce^{-i\xi_0 x}\right), && -a < x < a, \\ e(x,z,\omega) &= Ae^{i\beta z}e^{-i\xi x}, && x < -a, \end{aligned} \tag{2.2.8}$$

where β_0, ξ_0 and β, ξ are the propagation constants inside and outside the channel, respectively. These would correspond to the propagation vector $\mathbf{k} = (\xi, 0, \beta)$ we discussed earlier. The constant ξ represents therefore the propagation in the transverse direction and β in the paraxial direction (along the axis of the propagation direction). Due to the reflection and transmission of the propagating wave from the boundaries, the wave is disturbed and receives gradually more and more of these disturbances. Through the dispersion relation that connect the propagation constant and the frequency ω we find

$$\beta = \left(\left(\frac{\omega}{c}\right)^2 n^2 - \xi^2\right)^{\frac{1}{2}}, \quad \beta_0 = \left(\left(\frac{\omega}{c}\right)^2 - \xi_0^2\right)^{\frac{1}{2}}. \tag{2.2.9}$$

The dispersion relation can be readily found by inserting $\exp(i(\mathbf{k}\cdot\mathbf{x} - \omega t))$, where $\mathbf{x} = (x,y,z)$, into the linearized form of (2.2.5).

To determine the constants A to D in (2.2.8), we need to use the boundary conditions (2.2.7). As a consequence, we get a homogeneous linear system of four equations which has nontrivial solutions only if the determinant is zero. This is expressed by the equation

$$\tan(2a\xi_0) + i\frac{2\xi\xi_0}{\xi^2 + \xi_0^2} = 0, \tag{2.2.10}$$

where we have the following identity

$$\xi^2 = \left(\frac{\omega}{c}\right)^2 (n^2 - 1) + \xi_0^2, \tag{2.2.11}$$

which is also a consequence of the boundary conditions. The relation (2.2.11) represents Snell's law. Observe that upon substituting (2.2.11) into (2.2.9), we get $\beta = \beta_0$ which is in agreement with the second condition from page 31. Solving (2.2.10) for ξ_0 using (2.2.11) gives us the leaky modes.

The form of the equation (2.2.10) with the linear system it originated from, reveals certain symmetries which allow us to consider the solutions ξ_0 only in one quadrant of the complex plane. These symmetries can be expressed as

$$\{\{\xi, \xi_0\}, (A, B, C, D)\} \to \{\{\xi, -\xi_0\}, (A, C, B, D)\}, \tag{2.2.12a}$$

$$\{\{\xi, \xi_0\}, (A, B, C, D)\} \to \{\{-\xi^*, -\xi_0^*\}, (A^*, B^*, C^*, D^*)\}. \tag{2.2.12b}$$

Any solution in the other quadrants can be thus generated by using the symmetries. Note, that the variable ξ_0 is complex valued. We chose to find to solutions in the second quadrant. Here, expressing ξ in (2.2.11) results in two different sets depending on the positive or negative square root. The two separate systems we get, are

$$\tan(2a\xi_0) = -i\frac{2\xi_0\sqrt{\alpha + \xi_0^2}}{\xi^2 + \xi_0^2}, \qquad \xi = \sqrt{\alpha + \xi_0^2}, \tag{2.2.13a}$$

$$\tan(2a\xi_0) = i\frac{2\xi_0\sqrt{\alpha + \xi_0^2}}{\xi^2 + \xi_0^2}, \qquad \xi = -\sqrt{\alpha + \xi_0^2}, \tag{2.2.13b}$$

where we defined $\alpha = (\omega/c)^2(n^2 - 1)$. In order to grasp, where the solutions to (2.2.10) are located in the second quadrant, we plot the real and imaginary part of (2.2.10) as a contour plot and look where the two contours meet. The contour plot showed that there is infinitely many solutions lying on one branch in the second quadrant with slow logarithmic growth. From the mentioned symmetries of the problem we get the solutions in the fourth quadrant once the solutions in the second quadrant are obtained. A similar approach is then made for the solutions in first and third quadrant.

Since the branch the solutions lie on, grows at a logarithmic rate, it is safe to assume that for solutions located further to the left, will have their real part dominating the imaginary part. We use this observation to obtain the asymptotic formula for the solutions. So assuming $|\xi_0| \gg \sqrt{\alpha}$ and consequently $\xi_0 = x + iy$, where $|x| \gg |y|$ we arrive at the following formula

$$\xi_{0p} = -\frac{p\pi}{2a} + \frac{i}{4a}\mathrm{Log}\left[\frac{16\left(-\frac{p\pi}{2a} + \frac{i}{4a}\ln\left[\frac{p^4\pi^4}{\alpha^2 a^4}\right]\right)^4}{\alpha^2}\right], \tag{2.2.14}$$

where p is the index of the solution. This is the asymptotic approximation for the solutions of (2.2.10), (2.2.11) in the second quadrant. Using the symmetries (2.2.12a), (2.2.12b), similar formulas can be found for the other three quadrants. From the locations of ξ_{0p} based on (2.2.14), it is evident that the values for ξ_p through the relation (2.2.11) have positive real and negative imaginary part. If we substitute such value for ξ in the modes (2.2.8), we conclude that the modes in the second quadrant are outgoing and exponentially growing in the transverse (x) direction. Likewise, we can use the formula (2.2.9) to determine that these modes are decaying in the propagation direction. This classifies them to be leaky modes. The same can be concluded for the modes in the fourth quadrant. However, if we do the same analysis for the modes in the first and third quadrant, we find that these are incoming and exponentially growing in the transverse direction and growing in the propagation direction. These are thus not leaky modes but gaining modes.

Formula (2.2.14) turned out to be a really good approximation even for p of order 1. It can, however, break down or not stay in the same quadrant as intended. This happens if the natural logarithm inside the complex valued logarithm becomes negative. Clearly, it happens when $p \lesssim \alpha\sqrt{\alpha}/\pi$. So p must be smaller than some constant for the formula (2.2.14) to not be correct. This is understandable since it was designed for large values of p. Going back to the initial assumption where the magnitude of ξ_0 was larger than $\sqrt{\alpha}$, we can now assume the opposite, namely $|\xi_0| \ll \sqrt{\alpha}$. In this case, we obtain a recursive formula for ξ_0 which can be iterated once to get

$$\xi_{0p} \approx -\frac{\pi p}{2a} + i\frac{\pi p}{2a^2\sqrt{\alpha}}. \tag{2.2.15}$$

Since we are still in the second quadrant, we assume that the imaginary part must be a small correction to the real part, or in other words $\pi p/(2a) \gg \pi p/(2a^2\sqrt{\alpha})$, or equivalently, $p \ll a\sqrt{\alpha}$, which is precisely when formula (2.2.14) is no longer valid. The formula (2.2.15) is therefore an approximation for small values of p.

Having obtained the asymptotic formulas for the propagation constants, we can analyse which modes propagate at which angles with respect to the propagation direction along the z-axis. The angle is measured between the z-axis and the propagation vector (ξ_0, β_0). Modes for which this angle is small, are called paraxial. Using simple trigonometry, one finds the formula for the angle to be $\theta_p = \arctan\left(\mathrm{Re}[\xi_{0p}]/\mathrm{Re}[\beta(\xi_{0p})]\right)$. Only the real part is considered because the imaginary part corresponds to decay while the real part to the actual propagation. The angle must be small for paraxial modes and this holds only if $\mathrm{Re}\left[\xi_{0p}\right] \ll \mathrm{Re}\left[\beta\left(\xi_{0p}\right)\right]$. Using the asymptotic formulas for ξ_0 and the formula for β, we get the condition

$$p \ll \frac{a\sqrt{2}\omega}{\pi c}. \tag{2.2.16}$$

This formula tells us for which mode index p is the corresponding mode paraxial.

The equation (2.2.10) becomes exponentially small well away from the real axis and it can be very challenging to see solutions from the plots even if some might be there. For this reason we investigated the possibility of the existence of solutions at different parts of the second quadrant that were not covered. We found asymptotic formulas for the solutions using the assumption $\mathrm{Re}[\xi_0] \gg \mathrm{Im}[\xi_0]$. Here, we considered two different conditions to see if we can find more solutions, namely $\mathrm{Re}[\xi_0] \ll \mathrm{Im}[\xi_0]$ and $\mathrm{Re}[\xi_0] \sim \mathrm{Im}[\xi_0] \gg 1$. After careful investigations we concluded that no additional solutions are present, so (2.2.14) are the only ones.

One of the goal of this paper is to propagate the initial pulse as long as possible without any disturbances which may come as a result of reflection from the boundaries of the channel. Thus, we are interested in minimizing these reflections and so we want to choose the difference of the artificial refractive index between the inside and the outside as small as possible. This difference is depicted in the parameter $\alpha = (\omega/c)^2(n^2 - 1)$. Thus α decreases as we minimize the desired difference. We observed what happens with the zeros ξ_{0p} as α decreases. All the zeros in the second quadrant were moving up as expected from the logarithmic dependence of α in the formula (2.2.14). The behaviour of the first zero is particularly interesting, because it approaches the imaginary axis and eventually disappears at a finite value $\alpha = \alpha_c$. For $\alpha < \alpha_c$, our asymptotic formula (2.2.14) produces a double zero. As we let α decrease even more, the next zero disappears and our formula produced two double zeros. This goes on as α approaches zero. Careful numerical

examination revealed that no double zeros are actually present. The production of double zeros can be explained by the attempt of the formula to preserve the original zeros even though they vanished in reality. This sudden change in the formula with varying α made us think about the possible reason for this doubling being the crossing of a branch cut for the complex valued logarithm in (2.2.14). To explore this option, we used the standard branch of logarithm being the negative real axis. The idea is that when the imaginary part of the argument, let us denote it z, of the complex logarithm in (2.2.14) is zero, we will have a crossing of the branch cut since the real part is negative. Using this condition we found that there is a crossing of the branch cut whenever α crosses the value

$$\alpha_p = \frac{p^2\pi^2}{a^2}\exp(-p\pi). \tag{2.2.17}$$

So the disappearing of the first zero happens when α crosses the value $\pi^2\exp(-\pi)/a^2$. The exponential decay of the critical values for α when another zero disappears indicate that α must decrease even faster for the next zero to vanish. Numerical investigations confirmed the obtained result (2.2.17) and we observed the double zeros precisely at the calculated critical values.

We now proceed to the form of the modes. We discussed essentially two types of modes: leaky modes and gaining modes both corresponding to two families of zeros of the determinant (2.2.13a) and (2.2.13b), respectively. It is important to explain how are the symmetries (2.2.12a) and (2.2.12b) related to the type of the mode. Starting with the zeros in the second quadrant, we know that they are leaky modes. Using the symmetry (2.2.12a), we get the zeros in fourth quadrant and these behave in the same way. Both modes from second and fourth quadrant are thus outgoing, leaky modes. A simple analysis shows that they are, in fact, the same modes and we can therefore disregard those from fourth quadrant. Similarly, we consider zeros in the first quadrant that generate incoming, gaining modes. Using the same symmetry (2.2.12a) we find the modes from third quadrant to be the same which we disregard. The symmetry (2.2.12b) turns the zeros from second quadrant into zeros from the first quadrant. This concludes that the symmetry (2.2.12a) switches between modes of the same type while the symmetry (2.2.12b) switches between different types. Let us denote the leaky modes with ξ_0 from the second quadrant $u_p^-(x)$ and the gaining modes with zeros from the first quadrant $u_p^+(x)$. They are given by the formulas

$$u_p^-(x) = \begin{cases} De^{i\xi_p x}, & x > a \\ Be^{i\xi_{0p}x} + Ce^{-i\xi_{0p}x}, & -a < x < a \\ Ae^{-i\xi_p x}, & x < -a \end{cases}, \quad \xi_p = \left(\alpha + (\xi_{0p})^2\right)^{1/2}, \tag{2.2.18a}$$

$$u_p^+(x) = \begin{cases} D^*e^{i\xi_p x}, & x > a \\ B^*e^{-i\xi_{0p}^* x} + C^*e^{i\xi_{0p}^* x}, & -a < x < a \\ A^*e^{-i\xi_p x}, & x < -a \end{cases}, \quad \xi_p = -(\alpha + \xi_{0p}^{*2})^{1/2}. \tag{2.2.18b}$$

One notable observation is the relation $(u_p^+)^* = u_p^-$. Both leaky and gaining modes are exponentially growing in x which makes them not-normalizable. We recognize this problem from the previous paper with the resonant states. Here, we perform the same procedure to fix this issue, namely evaluating the modes at a complex contour. It is an analytical continuation into a complexified spatial domain. Depending on the type of the

mode, the contour will be different. They are of the form

$$z^+(x) = \begin{cases} a - i(x-a), & x > a \\ x, & |x| < a \\ -a - i(x+a), & x < -a \end{cases}, \quad z^-(x) = \begin{cases} a + i(x-a), & x < a \\ x, & |x| < a \\ -a + i(x+a), & x < -a \end{cases}, \tag{2.2.19}$$

where z^+ is used for the gaining modes and z^- for the leaky modes. The different types of modes behave differently outside the channel. One is of outgoing and the other of incoming type which is expressed by different sings in the exponents. This is the reason why we need different contours to make both of them decay.

The modes are then evaluated at the contours (2.2.19) resulting into the decaying functions $\psi_p^+(x) = u_p^+(z^+(x))$ and $\psi_p^-(x) = u_p^-(z^-(x))$ in both directions of the x-axis. For these complexified modes, we also have the relation $\left(\psi_p^+(x)\right)^* = \psi_p^-(x)$. From now on, we will refer to the complexified mode whenever we mention a mode.

The contours (2.2.19) are evidently singular at $x = \pm a$ which, in turn, makes the modes $\psi_p^\pm$ not continuously differentiable at the boundaries. In addition, the modes are satisfying different boundary conditions at $x = \pm a$ since the contours have different shapes. The two types of modes are therefore categorized by different spaces of functions to which they belong. These spaces of functions are spaces of smooth functions that satisfy the corresponding boundary condition depending on the type of the mode. Let us denote V^+ the space where the functions satisfy the boundary conditions for ψ^+ and V^- the space where the functions satisfy the boundary conditions for ψ^-.

In order to use the modes as a generalized Fourier series to expand functions, we need to have a notion of inner product and orthogonality. A suitable inner product defined for complex values functions is constructed in the following way

$$(\Phi, \Psi) = \int_{\mathcal{C}} \Phi(z)\overline{\Psi}(z)\mathrm{d}z \in \mathbb{C}, \tag{2.2.20}$$

where $\mathcal{C}$ is any contour in the complex plane and the function $\overline{\Psi}(z) = \Psi^*(z^*)$ is analytic. It follows from complex analysis that taking the complex conjugate of a function by itself makes the function non-analytic, but the generalized version of the complex conjugate used in this inner product (2.2.20) fixes this problem. If we now apply (2.2.20) to the contour z^-, we get a complex valued scalar product on the space V^-

$$(\psi, \phi)^- = i\int_{-\infty}^{-a} \psi(x)\phi(x)\mathrm{d}x + \int_{-a}^{a} \psi(x)\phi(x)\mathrm{d}x + i\int_{a}^{\infty} \psi(x)\phi(x)\mathrm{d}x, \tag{2.2.21}$$

for ψ, ϕ both belonging to V^-. A similar expression can be obtained for the inner product corresponding to the space V^+. The main focus in this paper is, however, on the leaky modes so we are using mostly them to demonstrate calculations.

In addition to the inner product we also need to show that the modes are orthogonal. It turns out, the modes $\psi^\pm$ are eigenfunctions to a differential operator $\mathcal{L}_x$

$$\mathcal{L}_x = \begin{cases} \partial_{xx} + \left(\frac{\omega}{c}\right)^2, & |x| < a \\ \\ -\partial_{xx} + \left(\frac{\omega}{c}\right)^2 (n^2 - 1), & |x| > a \end{cases}, \tag{2.2.22}$$

with different eigenvalues, depending on the type of the mode. Hence, the orthogonality is assured if the operator $\mathcal{L}_x$ is self-adjoint with respect to the inner product defined in

(2.2.21). Due to the fact that the modes and their eigenvalues are complex, this may seem as impossible to achieve at first glance, but now it is straightforward to show that the differential operator $\mathcal{L}_x$ is indeed self-adjoint on the space V^-. The orthogonality of the modes ψ^- is then easy to check and the same goes for $\psi^+ \in V^+$.

There are few properties we can observe here. The relation we mentioned between the two types of modes $\left(\psi_p^+(x)\right)^* = \psi_p^-(x)$ and their corresponding boundary conditions imply that if $\psi^- \in V^-$, then $(\psi^-)^* \in V^+$ so the conjugation operator maps between the two spaces. One can also check that the conjugation maps between the inner products on the two spaces as well. A simple analysis of V^+ and V^- shows that they are linear spaces and complex algebras as well, because any product $\phi\psi$ for $\phi, \psi \in V^-$ also preserves the boundary conditions at $x = \pm a$.

The establishment of orthogonality and normalizability now allows us to consider an expansion of any function f in V^- in terms of the generalized Fourier series using the leaky modes ψ_p^- of the form

$$f(x) = \sum_{p=1}^{\infty} \frac{\left(f(x), \psi_p^-\right)^-}{\left(\psi_p^-, \psi_p^-\right)^-} \psi_p^-(x). \tag{2.2.23}$$

The next logical step would be to test whether the series (2.2.23) is able to accurately represent functions. In the case of the resonant states we have also pursued the issue of completeness, but we will not do that with leaky modes. On the other hand, we will investigate the question of convergence. First of all, in order to use the expansion (2.2.23) in practical modelling of transparent boundary conditions, for example for UPPE, the expansion must represent well two types of function. These are physically reasonable initial conditions and products of functions. By physically reasonable functions which represent initial data, we mean functions with compact support such as Gaussian wave packets. Products of functions are important to be in the span of leaky modes because of the implementation of spectral propagators (UPPE) on the full nonlinear problem (2.2.5). The nonlinear term may very well include powers of the electric field which results in finding the expansions for powers of leaky modes themselves. We conducted, therefore, numerical tests where we compared Gaussian wave packets and squares of leaky modes (in particular of those corresponding to $p = 10$ and $p = 50$) with their generalized Fourier series. The index outside the channel was set to $n = 1+10^{-12}$. The results were satisfying. The original functions and their series overlapped perfectly using only 30 terms. Due to the high oscillatory behaviour of the functions ψ_{10}^2 and ψ_{50}^2, we needed to use 60 and 200 terms in the series, respectively.

After having established the functionality of the series (2.2.23) numerically, we next proceeded to verify how well did the leaky modes expansion linearly propagate a Gaussian beam in contrast to finite and infinity Fourier series solutions. The boundary conditions for the leaky modes are already known and for the regular Fourier modes we used perfectly reflecting boundary conditions at $x = \pm a$. Both are then compared to the exact, infinite domain solution using Fourier modes applied in a much larger domain in x. In this comparison we used $n = 1 + 10^{-12}$ outside the channel for leaky modes and a Gaussian in the optical regime of infrared light. We expect the leaky modes solution to go hand in hand with the exact infinite Fourier solution until the reflection from the boundaries in the case of leaky modes makes large enough disturbances. When the accumulative effect of the reflection gets large enough, the leaky modes solution starts to deviate from the infinite Fourier solution at some value for z. We expect the finite Fourier solution to

deviate from the other two solutions at the moment when the wave hits the boundary of the computational domain. And this was indeed what we observed. Approximately at $z \approx 4 \times 10^4$, the finite Fourier solution gained much of a disturbance from the reflections and was no longer a feasible solution. At this distance, the leaky modes and infinite Fourier modes overlapped nicely. The leaky modes managed to keep up until around $z \approx 4 \times 10^5$. Thus we can conclude that leaky modes were able to propagate around 10 times further than finite Fourier modes. The longer we propagate, the more reflection leaky mode solution gets since the channel is not perfectly transparent. The difference in the refractive index between the inside and the outside is nonzero. One would suggest making the difference even smaller resulting in even further possible propagation, but this turned out not to be entirely the case. Numerical test with smaller difference, in particular $n = 1 + 10^{-15}$ showed that the leaky modes series did not represent the target function well at all. However, it seemed to converge pointwise but to a different function than the target Gaussian. Including more terms in the expansion did not remedy the issue. Two possible explanations were proposed: either the series diverges but so slowly, it cannot be detected numerically, or it converges pointwise but not to the target function. Explanation number two was observed to be correct based on extensive numerical investigations. It would implicate that the series actually never converges to the target function. This, however, does not make the leaky modes expansion any less useful for the task it was designed for.

One observation we made concerning this issue was tied to a dimensionless number $\eta = a^2\alpha = a^2(\omega/c)^2\,(n^2 - 1)$. We found that if $\eta \gtrsim 0.8$ then the series gives a practically satisfying representation of the function. Note, that two parameters in η depend on characteristics of the channel. These are the width a and the refracting index on the outside n. Making the difference of the indices smaller results in the term $n^2 - 1$ being approximately that difference. Mathematically, the quality of the series representation depends on a and the index difference inversely. In other words, larger width implies smaller difference and vice versa. This observation has important consequences. When applying leaky modes expansion as a part of a spectral propagator for the nonlinear equation (2.2.5), all the support of the nonlinear interactions should be confined well within the boundary walls. This affects the choice of the channel width and, consequently, the index difference.

The question why does the series lose its ability to accurately represent functions under certain circumstances, in particular small index difference, was further investigated. We noticed that making the index difference smaller is also associated with a different phenomena we observed earlier, specifically the loss of the first zero when α crosses the value α_1 (2.2.17). These two phenomena (loss of a zero and bad convergence) indeed are connected. Losing the first eigenvalue leads to the loss of the first leaky mode ψ_1^-. Since the first term in an expansion is the most important one, the loss of the entire term is threatening for the accuracy of the whole series. The reason why this loss affects the accuracy can be explained as follows. The shape of the leaky modes suggest that they alternate between being odd and even functions depending on the index p. The leaky mode belonging to the first zero ψ_1^- is an even function. We know that for $\alpha \gtrsim \alpha_1$, the expansion is a good representation of an even function like Gaussian. The expansion starts to fail for $\alpha < \alpha_1$ which is a clear indication that the loss of the first zero is in play. Instead of the first term in the series being an even function, it is now odd and understandably fails to represent an even Gaussian. To further test this explanation, we tried to represent an odd function (first derivative of the Gaussian) for $\alpha < \alpha_1$. Before,

the series represented the Gaussian badly for $n = 1 + 10^{-15}$. The odd function, in this case, was accurately represented for an index step as small as $n = 1 + 10^{-24}$. Thus this makes our explanation even more plausible.

At the end of the paper we looked closely at the convergence rate of the series (2.2.23). We conjectured that the mode expansion does not converge to the target function for small index step. The investigation of the convergence would not prove the conjecture entirely, but it would certainly support it.

We start first by assuming that $\sqrt{\alpha} \ll |\xi_{0p}|$ which defines our asymptotic regime as we wish to obtain the terms in the series for large p and small α. In this regime, we have the asymptotic formulas for the zeros (2.2.14). Using this formula, we are interested in finding the asymptotic formula for the terms in the series (2.2.23). Before we proceed to the actual terms, the continuity coefficients in (2.2.8) A, B, C, D need to be checked how do they depend on p asymptotically. In order to obtain them, we will use the matrix $\mathbf{M}$ whose nullspace contains the vector $(A, B, C, D)^T$. The matrix $\mathbf{M}$ comes from applying the boundary conditions. To compute the nullspace of $\mathbf{M}$, we row reduce it and find the last row to be of the form $(0, 0, 0, \zeta(\xi_0))$, where the function $\zeta(\xi_0)$ includes the determinant of $\mathbf{M}$. So for $\xi_0 = \xi_{0p}$, the last row becomes zero and the basis of the nullspace is easily found. Using the asymptotic formulas for ξ_{0p} and taking the limit as p grows, the continuity coefficients are approximated as $A = B = (-1)^{p+1}$ and $C = D = 0$, thus they do not contribute to the convergence rate of the series.

We continue to write the mode expansion in the following form

$$f(x) = \sum_p \frac{b_p(x)}{N_p} = \sum_p c_p(x), \text{ where} \tag{2.2.24a}$$

$$b_p(x) = \int_{-a}^{a} f(x)\psi_p^-(x)\mathrm{d}x\psi_p^-(x), \; N_p = \left(i\int_{-\infty}^{-a} + \int_{-a}^{a} + i\int_{a}^{\infty} \right) \left(\psi_p^-(x)\right)^2 \mathrm{d}x. \tag{2.2.24b}$$

The normalization term behaves in the asymptotic limit as $N_p \approx 4a(-1)^{p+1}$. The terms $b_p(x)$ include an integral of the product $f\psi_p^-$ over the interval $[-a, a]$. In this region, the leaky modes are exponential functions with $\pm i\xi_{0p}$ in the exponent. Using integration by parts to solve this integral, we end up with the factor $i\xi_{0p}$ in the denominator, each time the integration by parts is performed. Here again, we can see a similarity to the convergence investigation of the resonant state expansion from the previous paper. At this point we can deduce that the convergence rate will depend on the smoothness of the target function f. The factor $1/\xi_{0p}^n$ we get after n consecutive integrations by parts can be, again, approximated using the asymptotic formula and we get the following expression for $b_p(x)$

$$b_p(x) \approx \psi_p^-(x) \left(\frac{-2ai}{p\pi}\right)^n \int_{-d}^{d} f^{(n)}(x) \left(\psi_p^-(x)\right)^{\bigstar(n)} \mathrm{d}x, \tag{2.2.25}$$

where $\left(\psi_p^-(x)\right)^{\bigstar(n)}$ is the leaky mode in the region $[-a, a]$ with a plus or minus sign between the terms depending whether n is even or odd, respectively, where $\bigstar(n)$ is a superscript.

The newly obtained asymptotic formulas in (2.2.24a) were tested using two sample functions: a triangle function and a Gaussian wave packet. In the case of the triangle function we were able to identify the corresponding series as a polylogarithm function $Li(n, z) = \sum_p z^p/p^n$ for some complex number z defined from the asymptotic expressions we found and n being one of the following expressions: $2 \pm x/a, 2 \pm (d + x)/a$ and $2 \pm$

$(d-x)a$, where $[-d,d]$ is the width of the triangle function with $0<d<a$. So clearly, the convergence rate depends on the location x. After careful analysis of the various inequalities we found that we have absolute convergence for the triangle function in the region $-a+d<x<a-d$ which corresponds to the inside of the triangle. In the rest of the region outside the triangle but still inside the channel, or in other words $[-a,-d]$ and $[d,a]$, we found convergence as well but only a conditional one.

For the second sample function we chose a Gaussian wave packet with its support well inside the channel. The coefficients c_p in (2.2.24a) decayed exponentially and thus ensured that the leaky mode expansion converged for all x inside the channel. One can also dedicate this to the infinite smoothness of the sample functions as opposed to the triangle function which lacked this property.

Even though we were able to say something about the convergence rate of the expansion, it is still not enough to fully resolve the problem of bad representation. However, we attempted to solve it using the same approach we used to prove the completeness of the resonant states. We start with the completeness for the scattering states for our system which have real eigenvalues

$$\int_{-\infty}^{\infty}\varphi_{\xi_0}(x)\varphi^*_{\xi_0}(x')\mathrm{d}\xi_0=\delta(x-x'). \tag{2.2.26}$$

The scattering states include the determinant of $\mathbf{M}$ in the denominator, thus they have poles at the leaky modes eigenvalues ξ_{0p} in the second and the fourth quadrant. A complex integration contour $\mathcal{C}$ is then introduced such that it includes the poles ξ_{0p} inside of it. The determinant in the denominator of the scattering modes also contain the quantity $\xi=\sqrt{\alpha+\xi_0^2}$ and as a complex square root, it has branch cuts on the negative real axis. This has to be taken into consideration as well.

After the contour $\mathcal{C}$ is chosen, the completeness relation (2.2.26) is multiplied by $f(x')$, integrated over the real x'-axis and using Cauchy's theorem, integrated over the contour $\mathcal{C}$. Due to the residue theorem, the resulting equation has a sum over the residues on the left hand side which upon evaluating, turn the scattering modes into resonant leaky modes. The right hand side becomes the function $f(x)$. The sum of residues, as a function of x, should be then equal to the leaky modes expansion of f. The residues can be calculated and checked. The resulting equation is of the form

$$2\pi i\sum_{p=0}^{\infty}\mathrm{Res}\left(\varphi_{\xi_0}(x)\varphi^*_{\xi_0}(x'),\xi_{0p}\right)\left(f(x),\psi^-_{\xi_{0p}}(x)\right)^-=\sum_{p=1}^{\infty}\frac{\left(f(x),\psi^-_{\xi_{0p}}(x)\right)^-}{\left(\psi^-_{\xi_{0p}}(x),\psi^-_{\xi_{0p}}(x)\right)^-}\psi^-_{\xi_{0p}}(x). \tag{2.2.27}$$

Up to this point, everything would be perfect. However, the scattering modes include also one free parameter $a^+(\xi_{0p})$ as a result of the boundary conditions. It must be chosen such that the sum of residues is equal to the leaky mode expansion in (2.2.27). The unfortunate thing is that the parameter a^+ is not analytic because it is zero inside $\mathcal{C}$ and contains two families of countably many zeros which produce branch cuts. The integration over $\mathcal{C}$ must therefore include the contributions of the integrals over all these possible branch cuts. This results into the conclusion that a function $f(x)$ equals to its leaky mode expansion plus some additional terms originating from the integrals along the branch cuts. From our numerical observations, we know that as long as α is not smaller than a certain critical value, the series representation is good. It is thus explained by

the negligible contribution of those integral terms. When α gets smaller, the contribution grows and some of them will dominate while not affecting the convergence of the series. This extra contribution, however, makes the series converge to a different function than the one generating it. It is possible to identify the extra integral terms along the branch cuts, find their asymptotic forms in the limit when α goes to zero and pick out the ones with the largest contribution. This way, one could add this term to the leaky mode expansion resulting into the correction of the expansion and an accurate representation of the target function. In spite of the effort, we were not able to find the integral terms, but we believe they could be found. Nonetheless, even if they are found, their form and complexity could make the leaky mode expansion too hard to use for practical purposes.

This paper introduces a new approach to mimic infinite domain by constructing a transparent computational boundary for wave equations. The modes of the system, called leaky modes are supported by the artificial index of the outside domain. A generalized Fourier series using leaky modes is possible to use in solving the nonlinear wave equation using spectral propagators such as UPPE. We have seen that the expansion (2.2.23) is able to represent a function as long as the parameter values are chosen such that $\eta = a^2\alpha \gtrsim 0.8$. This condition is plentiful for all practical purposes. We demonstrated by more detailed investigation concerning the completeness of the leaky modes by identifying finitely many integral terms from branch cuts, that the leaky mode expansion does not converge to the function used to construct it. In order to resolve the issue of convergence totally, one has to find the contributing integral terms, but this definitely belongs to future work.

2.3 Paper 3 - Modelling pulse propagation in complex index materials using the method of multiple scales

This paper is perhaps the one where nonlinear optics truly manifests itself. In the first chapter of the introduction we briefly discussed the origins of nonlinearity in optics. Terms such as electric susceptibility or refracting index were introduced from the optical perspective. Shortly said, nonlinearity originates in the polarization term which includes higher orders of power of the electric field. These higher powers become more apparent the higher the intensity of the optical pulse. The nonlinear polarization term then finds its way into the wave equation derived from Maxwell's equations. In this paper, we are solving such nonlinear dispersive wave equations using the method of multiple scales (MMS) which was broadly introduced in chapter 1.2.

The model equation we are using is the simplest nontrivial wave equation from nonlinear optics. Before we derive it, let us explain little bit more about the nonlinear polarization we are using here. In the introduction of our first paper, we went through the process of obtaining the nonlinear polarization terms from scratch, i.e. the Schrödinger equation. The main idea was that electrons in an atom occupy certain energy levels. The 0-th energy level is called the ground state and it is closest to the nucleus. The more energy the electron receives, the higher level it occupies (further away from the nucleus). In the presence of an external electric field photons are being absorbed and released by the electrons in the atom and jump back and forth between different energy levels. The energy of a photon is defined by its oscillation frequency which can get enhanced by the electrons if they require the same amount in order to jump and cause resonance. To illustrate this process, assume an electron in the ground state needs to absorb a photon of frequency ω to jump to the first energy level. In the process of returning back to the ground state, a photon is released later with the same frequency ω and contributes to the electric field in

return. This happens at low intensity optical pulses and is called the induced polarization which we discussed in chapter 1.1. It is also called the *first harmonic generation*. When the intensity gets larger, the electron absorbs two photons of frequency ω and jumps two levels, later creating a 2ω frequency photon in a single quantum-mechanical process. This phenomenon is called the *second harmonic generation*.

There can be two different types of second harmonic generation depending on the frequencies of the photons that are getting absorbed and released. The process described above required two photons with the same frequency ω. Think of a situation where the two input photons are at different frequencies ω_1 and ω_2. In one case, the electron absorbs both, jumps two levels and on the way down creates a single photon with frequency $\omega_3 = \omega_1 + \omega_2$. This is the sum-frequency generation. In the other case, the atom first absorbs a photon of frequency ω_1 and jumps to the second level. The presence of the other frequency ω_2 then stimulates a process, where the second level decays by creating two photons with frequencies ω_2 and $\omega_3 = \omega_1 - \omega_2$. This is called the difference-frequency generation and is also known as optical parametric amplification.

A similar process occurs when applying three photons with the same or different frequencies, where the electron jumps between three different energy levels and produces photons with new frequencies. This is the *third harmonic generation*. Note, that to obtain higher order harmonic generation, the intensity of the external electric field must be larger and larger. The amount of resulting types of different generations in this case gets also bigger.

The processes we have just described involve virtual energy levels which are distinct from the real energy levels of an atom. Virtual levels represent the sum of the energies coming from the real energy levels and one or more photons from the surrounding radiation field. This is the reason why are these processes of, so called, parametric type. Conversely, processes that involve transfer from one real energy level to another are called nonparametric processes. We have seen earlier that the different harmonic generations give rise to different orders of susceptibilities (1.1.5) through the formulas (2.1.9), (2.1.11) and so on. As a consequence of virtual levels, parametric processes can always be described by real susceptibilities while nonparametric processes are described by complex valued susceptibilities that can be derived using the method in the summary of paper 1. In parametric cases, the energy of the photons transferred in the process is always conserved and remains present only in the exchanges between the atoms and the radiation field. This is not true for the nonparametric processes where the photon energy can be transferred into or from the material medium. A simple distinction between parametric and nonparametric processes would be the real and complex valued refractive index, respectively. The real part of the index exists as a consequence of the parametric processes while the imaginary part is a consequence of the nonparametric processes. The imaginary part corresponds to absorption of the radiation which occurs following the transfer of the photons energy into the material.

From the formal definition of the n-th order of susceptibility $\chi^{(n)}(\omega)$ it follows that this quantity must be tensor of rank $n+1$, see (1.1.5). In the case of the second harmonic generation, the corresponding nonlinear polarization term is expressed [49] as

$$P_i^{(2)}(\omega_n + \omega_m) = \varepsilon_0 \sum_{j,k} \sum_{(n,m)} \chi_{ijk}^{(2)}(\omega_n + \omega_m, \omega_n, \omega_m) E_j(\omega_n) E_k(\omega_m), \tag{2.3.1}$$

where the optical field residing in a second-order nonlinear optical medium consists of two distinct frequency components $E(t) = E_1(\omega_1)\exp(-i\omega_1 t) + E_2(\omega_2)\exp(-i\omega_2 t) + (*)$. The

symbol $(*)$ represents the complex conjugate of the preceding terms. The formula (2.3.1) is in accordance with the process illustrated in the introduction of paper 1. The indices i, j, k are the Cartesian components of the fields. The indices n, m take values 1 and 2, as the field around consists of the frequencies ω_1 and ω_2. The notation (n, m) means that under the summation through n, m in (2.3.1), the sum $\omega_n + \omega_m$ in the argument should be kept unchanged as the nonlinear polarization oscillates at frequency $\omega_n + \omega_m$, even if ω_n and ω_m vary. The function $\chi^{(2)}$ is written with three arguments. Technically speaking, this is unnecessary, because thee first is always the sum of the other two. The polarization term (2.3.1) therefore represents the sum-frequency generation.

In order to completely describe this interaction resulting into second-harmonic generation, we need to know all the nonlinear polarization term components $P_i^{(2)}$. Denoting $\omega_3 = \omega_1 + \omega_2$, we therefore need to determine 12 tensors $\chi_{ijk}^{(2)}(\omega_3, \omega_1, \omega_2), \chi_{ijk}^{(2)}(\omega_1, \omega_3, -\omega_2), \ldots$ and so on, all the possible combinations. In addition, each tensor contains 27 Cartesian components since the rank of the tensor is 3. Totally, 324 different complex numbers need be determined. However, the form of (2.3.1) makes it possible to observe certain symmetries due to the restrictions that relate these symmetries to the components of $\chi^{(2)}$. Since i, j, k, n and m are dummy indices, they can be interchanged and useful properties of the susceptibility can be exploited. For example if we allow the nonlinear susceptibility to be unchanged if the last two arguments and the Cartesian indices are interchanged, then we get $\chi_{ijk}^{(2)}(\omega_n + \omega_m, \omega_n, \omega_m) = \chi_{ikj}^{(2)}(\omega_n + \omega_m, \omega_m, \omega_n)$. This is referred to as intrinsic permutation symmetry. We get full permutation symmetry if we can freely interchange all the arguments as long as the corresponding Cartesian indices are also interchanged simultaneously. Using these and various other types of symmetries, the number of components to determine greatly reduces. The intrinsic permutation symmetry only reduces the number to only 81 independent parameters.

The symmetry properties of the material medium puts constraints on the linear and nonlinear susceptibilities. In a crystal which is isotropic in the x and y direction, but different in the z direction, the optical response for an applied field polarized in either the x or the y direction would be the same. This would impose a symmetry on the second-order susceptibility such that $\chi_{zxx}^{(2)} = \chi_{zyy}^{(2)}$. It is thus important to determine all types of symmetry properties in a crystalline medium and its consequences on the form of the linear and nonlinear optical susceptibilities. This can be done using mathematical methods such as group theory.

In some crystals one observes a type of symmetry in reference to the nonlinear susceptibility, called centrosymmetry, or inversion symmetry. Let us consider a second-harmonic generation in such medium that responds instantaneously to the applied optical field. So no temporal dispersion. The nonlinear polarization is then given by $P^{(2)}(t) = \varepsilon_0 \chi^{(2)} E^2(t)$. Inversion symmetry means that if the sign of the applied electrical field changes, so does the sign of the polarization. Thus the formula for the polarization must be identical to $-P^{(2)}(t) = \varepsilon_0 \chi^{(2)}(-E)^2(t)$ which is equivalent to $-P^{(2)}(t) = \varepsilon_0 \chi^{(2)} E^2(t)$. By comparison we see that $\varepsilon_0 \chi^{(2)} E^2(t)$ must be equal to both $P^{(2)}(t)$ and $-P^{(2)}(t)$. Hence we get that $P^{(2)}(t) = -P^{(2)}(t)$ which implies that $P^{(2)}(t) = 0$ and likewise $\chi^{(2)} = 0$. The conclusion is that for a centrosymmetric material system, the $\chi^{(2)}$ nonlinear susceptibility must vanish identically. This fact is true in general in nature for all centrosymmetric materials. Intuitively, it can be understood at the atomic level. Consider an electron and its motion in a nonparabolic potential well $U(x)$ centred at the nucleus, or at the origin in the case of one dimension and that is an even function (symmetric under the operation $x \to -x$). The nonparabolic potential in 1D being a function of x, is approximated using Taylor

series around the origin as $U(x) \approx c_1 x^2 - c_2 x^4$, where c_1 and c_2 depend on parameters such as the mass of the electron or the resonant frequency of the atom. The fourth order term is considered as a correction to the second order term, consistent with the Taylor series. Using classical mechanics, the restoring force F_r generated from the potential is of the form $F_r = -\mathrm{d}U/\mathrm{d}x = -2c_1 x + 4c_2 x^3$. The restoring force is then a nonlinear function of the displacement of the electron from its equilibrium position. Because of this nonlinearity, the atomic response will show significant harmonic distortion. In more than one dimension, the restoring force takes the form $\mathbf{F}_r = -2c_1\mathbf{x} + 4c_2(\mathbf{x}\cdot\mathbf{x})\mathbf{x}$. The variable $\mathbf{x}$ represents the displacement of the electron from its equilibrium position. Next, the equation of motion for the electron is obtained where the electron is moving under the restoring force and the electromagnetic force coming from the external field. The equation essentially describes a motion of a nonlinear oscillator. Since the equation is nonlinear because of the restoring force, the solution is sought in the form of a power series in a parameter. A sequence of simpler equations are obtained which can be analytically solved [49]. The first order equation gives us the linear polarization term $\mathbf{P}^{(1)}$ and the susceptibility $\chi^{(1)}$. The second order equation turns out to be damped but not driven by any force, its steady state solution vanishes, thus $\chi^{(2)} = 0$. The third equation gives us the third order nonlinear polarization term that can be shown [49] to be of the form

$$\mathbf{P}^{(3)}(\omega_q) = Ne \sum_{mnp} \frac{be^3 \left[\mathbf{E}(\omega_m)\cdot\mathbf{E}(\omega_n)\right]\mathbf{E}(\omega_p)}{m^3 D(\omega_q)D(\omega_m)D(\omega_n)D(\omega_p)}. \tag{2.3.2}$$

Here, N is the number density of atoms, b is a parameter that characterizes the strength of the nonlinearity in the restoring force, D is defined as $D(\omega) = \omega_0^2 - \omega^2 - 2i\gamma\omega$ with ω_0 being the resonant frequency of the atom and γ is the strength of the damping force. The indices m, n, p go through the values 1,2 and 3 and the frequencies $\omega_{m,n,p}$ are the three different frequencies in the applied field given by $\mathbf{E}(t) = \sum_{n=1}^{3} \mathbf{E}_n(\omega_n)\exp(-i\omega_n t)$. The frequency ω_q is assumed to contain the combination of the other three frequencies $\omega_q = \omega_m + \omega_n + \omega_p$ (sum-frequency generation). Observe that the form of (2.3.2) is the same as the third order term in (1.1.5).

Conversely, material media possessing no inversion symmetry, are called noncentrosymmetric media. This property can also be explained in terms of the potential well $U(x)$ in the atom. In this case, the potential is also nonparabolic, but not symmetric around the y-axis. Therefore, the first two terms in its Taylor expansion around the origin are $U(x) \approx c_1 x^2 + c_2 x^3$. The first term corresponds to a harmonic potential and the second term corresponds to an anharmonic correction term. This model describes only noncentrosymmetric media, because the potential energy function $U(x)$ contains both even and odd powers of x, while the potential in centrosymmetric media consists of only even terms. The corresponding restoring force then becomes $F_r = -c_1 x - 3c_2 x^2$. Similar nonlinear equation of motion can be obtained for the electron and using the power series method, a sequence of equations is obtained. The difference here is that the second order equation is driven and so, has a nonzero solution resulting into a nonzero second order nonlinear susceptibility $\chi^{(2)}$. We can conclude that the lowest order nonlinear contribution to the polarization of a noncentrosymmetric material is second order in the applied field strength. Note that this analysis for both centro- and noncentrosymmetric media can be extended to include higher order effects.

In this paper, we assume a centrosymmetric, isotropic, homogeneous and lossy medium in which the applied field consists of only one frequency ω. Also, in an isotropic medium, the third order nonlinear susceptibility $\chi^{(3)}$ is assumed to respond instantaneously to

the applied optical field and is essentially independent of the frequencies of the applied waves whenever these frequencies are much smaller than the resonance frequency ω_0. We also assume that in our isotropic, homogeneous medium, the nonlinear susceptibility is constant in all directions. Such a third order nonlinear susceptibility is called the Kerr coefficient, denoted by η in this paper. Thus, in accordance with the formula (2.3.2) we have

$$\mathbf{P}_{NL} = \varepsilon_0 \eta (\mathbf{E} \cdot \mathbf{E})\mathbf{E}, \tag{2.3.3}$$

where $\mathbf{P}_{NL}$ stands for nonlinear polarization. The linear polarization term is assumed to not react instantaneously but rather depend on the electric field in all the previous times, called dispersion. It is expressed by

$$\mathbf{P}_L(\mathbf{x}, t) = \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t' \chi(t - t')\mathbf{E}(\mathbf{x}, t') = \varepsilon_0 \sqrt{2\pi}\hat{\chi}(i\partial_t)\mathbf{E}(\mathbf{x}, t), \tag{2.3.4}$$

where the last equality is an alternative form for the linear polarization. It can be readily derived using Fourier transform and Taylor series, where $\hat{\chi}(\omega)$ is the Fourier transform of $\chi(\omega)$. The overall polarization will then be the sum of the linear and nonlinear polarization

$$\mathbf{P} = \mathbf{P}_L + \mathbf{P}_{NL}. \tag{2.3.5}$$

The model equation is derived from the Maxwell's equations

$$\begin{aligned} \partial_t \mathbf{B} + \nabla \times \mathbf{E} &= 0, \\ \partial_t \mathbf{D} - \nabla \times \mathbf{H} &= 0, \\ \nabla \cdot \mathbf{D} &= 0, \\ \nabla \cdot \mathbf{B} &= 0. \end{aligned} \tag{2.3.6}$$

The quantities $\mathbf{E}, \mathbf{D}, \mathbf{H}$ and $\mathbf{B}$ are the electric field, electric displacement, magnetic field and magnetic induction, respectively. We assume no magnetic response in the material so we have the relations $\mathbf{H} = \mathbf{B}/\mu_0$ and $\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$. The constants μ_0 and ε_0 are the permeability and permittivity of the free space, respectively. We are also considering the propagation direction of the optical pulse to be the positive z direction and that it does not depend on x and y coordinate. In paper 2 we discussed the Maxwell's boundary conditions where two types of boundary condition were made possible: the transverse magnetic and transverse electric waves. Out of these two, transverse electric waves are more suitable here, because we want the electric field to be parallel to the boundaries which are, in this case, the x and y directions. According to these assumptions, we restrict ourselves to solutions of the form

$$\mathbf{E}(z, t) = E(z, t)\mathbf{e}_y,$$

$$\mathbf{B}(z, t) = B_1(z, t)\mathbf{e}_x + B_2(z, t)\mathbf{e}_z, \tag{2.3.7}$$

$$\mathbf{P}_{NL}(z, t) = P_{NL}(z, t)\mathbf{e}_y. \tag{2.3.8}$$

The nonlinear polarization term (2.3.3) then simply becomes $P_{NL} = \varepsilon_0 \eta E^3$. Using (2.3.4)-(2.3.8) one can eliminate the magnetic field components with cross derivatives and arrive at the equation

$$\partial_{tt} E - c^2 \partial_{zz} E + \sqrt{2\pi}\partial_{tt}\hat{\chi}(i\partial_t)E = -\eta \partial_{tt} E^3, \tag{2.3.9}$$

which is the basic model equation for this paper. Since we are solving it using a perturbation method and consequently comparing the solution to a high precision numerical solution, it is useful and practical to scale the equation (2.3.9). There are three variables which can be scaled, namely $z = Z_0 z', t = T_0 t'$ and $E = E_0 E'$. The scaled variables are then dimensionless by definition. The scales Z_0, T_0 and E_0 are picked in a convenient way to eliminate most of the factors that appear after the scaling is applied. Doing so, we get the relation $T_0 = Z_0/c$. Since Fourier transforms in both time and space play a prominent role, we need also to pick the scales for the wave number $k = K_0 k'$ and the frequency $\omega = \Omega_0 \omega'$ such that it does not affect the usual form of the associated Fourier transforms. Using the relation between T_0 and Z_0 one finds that the relation $\Omega_0 = cK_0$ assures this. The scaled equation has then the form

$$\partial_{tt} E - \partial_{zz} E + \sqrt{2\pi}\partial_{tt}\hat{\chi}(i\partial_t)E = -\varepsilon^2 \partial_{tt} E^3, \tag{2.3.10}$$

where we dropped all the primes and introduced a dimensionless quantity $\varepsilon = E_0\sqrt{\eta}$. The field strength E_0 is chosen to be the peak amplitude of the initial field which depends on the strength of the laser used to generate it. Thus ε can be varied over orders of magnitude, but typically it is always less than one in realistic situations. The parameter ε will also serve us as the small perturbation parameter when we apply MMS.

The material medium in which we wish to propagate the initial pulse is associated with strong dispersion. As a consequence, the pulse spreads quickly out but has a very narrow wave-number spectrum. Thus, we will assume that the initial field has a narrow spectrum centred on a wave length which is determined by the laser generating the pulse. So the equation (2.3.10) is solved as an initial value problem restricting the solution to a spectrally narrow one. Using MMS, one can derive equations for such narrow band solutions, called amplitude equations that give a good approximation to Maxwell's equations for these type of solutions. In this paper, we derive such an amplitude equation. Earlier, we introduced MMS and demonstrated the procedure on an example. The same procedure will be done on the equation (2.3.10). We start by introducing the expansions

$$\begin{aligned} \partial_t &= \partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \dots, \\ \partial_z &= \partial_{z_0} + \varepsilon\partial_{z_1} + \varepsilon^2\partial_{z_2} + \dots, \\ e &= e_0 + \varepsilon e_1 + \varepsilon^2 e_2 + \dots. \end{aligned} \tag{2.3.11}$$

The electric field is then given by

$$E(z,t) = e(z_0, t_0, z_1, t_1, \dots)\big|_{t_j = \varepsilon^j t, z_j = \varepsilon^j z}. \tag{2.3.12}$$

The function $\hat{\chi}(i\partial_t)$ is unknown at this point and we use the expansion for time (2.3.11) and Taylor series to get an expansion up to the order ε^2. The process continues in a usual way by writing down the perturbation hierarchy. The 0-th order equation has the wave packet solution

$$e_0(z_0, t_0, z_1, t_1, \dots) = A_0(z_1, t_1, \dots)e^{i\theta_0} + (*), \tag{2.3.13}$$

where $\theta_0 = kz_0 - \omega t_0$ and $\omega = \omega(k)$ is a complex function which is the solution to the dispersion relation $\omega^2 n^2(\omega) = k^2$, where the refracting index n is defined as $n^2(\omega) = 1 + \sqrt{2\pi}\hat{\chi}(\omega)$. The reason why $\omega(k)$ is a complex valued function is because the material media we are considering is of lossy type, or in other words the linear susceptibility has a nonzero imaginary type that corresponds to absorption of the applied field. Solving

the equation (2.3.10) as an initial value problem leads to the choice of the initial field at time $t = 0$. The wave number spectrum is then known for the initial field and is real which makes the time frequency ω complex. This has tremendous effect on the amplitude equation. Usually, the amplitude equations derived with MMS are nonlinear, such as the nonlinear Schrödinger equation or the complex Ginsburg-Landau equation. In these examples, the time frequency is, however, real valued. Recall that amplitude equation is obtained from removing the secular terms at every order of the perturbation hierarchy. In our case, the ε^2 order equation contains the third power of e_0. Taking the third power of (2.3.13) we get 4 nonlinear terms in A_0

$$e_0^3 = A_0^3 e^{i3\theta_0} + |A_0|^2 A_0 e^{i\theta_0} e^{2t_0\omega_i} + (*), \tag{2.3.14}$$

where $\omega_i = \text{Im}\ \omega$. The secular terms at all orders are the factors in front of the exponential term $e^{i\theta_0}$. Besides the terms in (2.3.14), the ε^2 order equation has only secular terms which are linear because no other source of nonlinearity exists except for the third power. This concludes that if ω was real valued, the second term from (2.3.14) would contribute to the secular terms resulting into a nonlinear amplitude equation. But since this is not the case, the amplitude equation remains linear. This makes the MMS based on decaying modes fundamentally different from stationary modes.

Carrying on with the calculations, we found the amplitude equation to be

$$\partial_t A + \omega'(k)\partial_z A - i\left(\beta - \alpha\left(\omega'(k)\right)^2\right)\partial_{zz} A = 0, \tag{2.3.15}$$

which is linear and where α and β are defined as

$$\alpha = \omega'(k)\frac{n^2(\omega) + 2\omega\sqrt{2\pi}\hat{\chi}'(\omega) + \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)}{2k}, \tag{2.3.16a}$$

$$\beta = \frac{\omega'(k)}{2k}. \tag{2.3.16b}$$

The nonlinearity comes in play through the relation between the electric field and the amplitude of the following form

$$E(z,t) = A(z,t)e^{i(kz-\omega t)} + \varepsilon^2(c_1 A^3(z,t)e^{i3(kz-\omega t)} + c_2|A(z,t)|^2 A(z,t)e^{i(kz-\omega t)}e^{2t\omega_i}) + (*), \tag{2.3.17}$$

where

$$c_1 = \frac{1}{n^2(\omega) - n^2(3\omega)}, \tag{2.3.18a}$$

$$c_2 = \frac{3(\omega + 2i\omega_i)^2}{k^2 - (1 + \hat{\chi}(\omega + i2\omega_i))\left(\omega + 2i\omega_i\right)^2}. \tag{2.3.18b}$$

In the case of stationary modes, or real valued ω the nonlinearity would be present also trough the amplitude equation.

At the first and second order equations, we took only the particular solutions and disregarded the homogeneous solutions. By considering them as well, new amplitudes would be introduced and for each of them, the amplitude equations would also be linear and decoupled. The defining electric field in terms of these amplitudes would be much more complicated. The deciding factor whether to include new amplitudes depends on the kind of solutions we wish to approximate. Here, the initial condition must be consistent

with perturbation assumptions made during the MMS. The amplitude equation (2.3.15) is linear so any narrow band initial condition for this equation will lead to a narrow band wave packet solution with dispersive properties. Our goal is, nevertheless, not to validate the amplitude equation with its relation to E. The goal is to find a solution to the wave equation (2.3.10) so the initial condition needs to be chosen there. It is also justified from the physical point of view since the initial field for E is determined by a laser which is spectrally narrow, and not for the amplitude A. The initial condition is then a narrow Gaussian centred at some particular wave length and nowhere else. From the relation (2.3.17) we see that the initial condition for E consists of two Gaussian centred at k and $3k$ which is not consistent with the wanted initial condition. It is also clear that in order for E to be a Gaussian centred on k, the amplitude A must be a Gaussian centred on $k = 0$. If we would impose the initial condition for E being only at k and calculate the initial condition for the amplitude A, it would result in two Gaussians again at $k = 0$ and $2k$. But in MMS we assumed only one wave packet centred at k (2.3.13). The Gaussian at $2k$ would be, however, smaller. To solve this issue, we have to introduce a homogeneous solution to the ε^2 order equation with a new amplitude

$$e_2(z_0, t_0, z_1, t_1, \ldots) = B_0(z_1, t_1, \ldots)e^{i(3kz-\omega(3k)t)} + (*), \tag{2.3.19}$$

centred at $3k$. This way we separate the two Gaussian obtained in the initial condition for A. Now we get two initial condition for A and B, each consisting of one Gaussian. The relation determining the electric field E from the amplitudes A and B is then of the form

$$\begin{aligned} E(z,t) = A(z,t)e^{i(kz-\omega t)} + \varepsilon^2((B + c_1A^3(z,t))e^{i3(kz-\omega t)} \\ +c_2|A(z,t)|^2A(z,t)e^{i(kz-\omega t)}e^{2t\omega_i}) + (*). \end{aligned} \tag{2.3.20}$$

We obtain the initial condition for B by letting $B(z,0) = -c_1A^3(z,0)$ and we get an equation from (2.3.20) with Gaussians on both sides centred at k. The initial condition for A is then easily found by solving

$$E(z,0) = A(z,0)e^{ikz} + \varepsilon^2 c_2|A(z,0)|^2A(z,0)e^{ikz} + (*), \tag{2.3.21}$$

by an iterative numerical method.

To get the amplitude equation for the new amplitude B, secular terms for the ε^3 order equation must be removed. This means a lot of work with tedious algebra but we can deduce them cleverly by observing the form of the right hand sides for the equations in the perturbation hierarchy. The right hand side of the ε^1 equation contains only derivatives of e_0. The ε^2 equation's right hand side contains both e_0 and e_1 but the terms with e_1 are exactly the same as for e_0 in the previous order equation. Therefore we can deduce that the ε^3 equation will contain the same terms only with e_2, where our new amplitude B_0 is. Thus the secular terms for B_0 in the ε^3 equations are the same as for A_0 in the ε^1 equation, except with $3k$ instead of k. We then conclude that the amplitude equation for B must take the form

$$\partial_t B + \omega'(3k)\partial_z B = 0. \tag{2.3.22}$$

The rest of the paper is about validating the MMS solution (2.3.20) using high precision numerical simulations. The narrow band solutions we are considering are challenging to solve numerically because they consist of wave packets with wide and slowly varying

envelope and fast oscillations. This implies very high resolution to capture the oscillations and large computational domain to include all the envelope which leads to long running time. Solving an amplitude equation instead requires much less computational time. The main reason is that for narrow band solutions, the carrier wave centred at the fast frequency does not need to be resolved, only the small deviations around the center frequency.

In order to validate our solutions, we need to know the formula for the electric susceptibility $\hat{\chi}(\omega)$ which plays a role in the dispersion relation. There are some polynomial formulas that approximate the real dispersion relation quite well if one is far from material resonance. For such approximations it is fairly easy to numerically solve the model equation and validate the amplitude equation. A much general class of approximations for the susceptibility are rational functions. To see how they are obtained, we need to go back to the atomic level of the material. The atoms in material media are composed of massive nuclei surrounded by electrons with small mass relative to the nucleus. The applied electric field $\mathbf{E}$ can excite the electrons and displaces them from the nucleus creating electric dipoles which give rise to the polarization $\mathbf{P}$. This effect can be described by a simple yet powerful mass-spring model. It was first proposed by H. A. Lorentz [9], a Dutch physicist. This model provides a surprisingly accurate description for the dependence of the electric field on polarization.

Consider an atom with a nucleus of mass M and charge $+q$ with an electron of mass m and charge $-q$ attached to it. The nucleus is assumed to have much larger mass than the mass of the electron. We are going to model the attachment through a spring between the nucleus and the electron with a spring constant α. This system has also a dynamic friction with a real valued, positive friction constant β. This will cause a damping effect in the system. Let us assume that the position of the nucleus is at the origin of the xy plane and the electron moves up and down in the y direction. When there is no excitation, the electron is in its equilibrium position which coincides with the position of the nucleus with the net charge of 0, so there is no electric dipole moment defined as $\mathbf{p} = p\ \mathbf{e}_y$. Under the influence of a monochromatic electric field $\mathbf{E}(t) = E_y \cos(\omega t)\mathbf{e}_y$ applied along the y-axis, the electron is driven from its equilibrium position creating en electric dipole moment $\mathbf{p} = -qy(t)\mathbf{e}_y$, where $y(t)$ is the displacement of the negative charge along the y-axis. From classical electrodynamics, we know that the force exerted on the electron is $-e\mathbf{E}(t)$. From classical mechanics, the restoring spring force becomes $-\alpha y(t)\mathbf{e}_y$ and the damping force reads $-\beta\dot{y}(t)\mathbf{e}_y$, where $\dot{y} = \mathrm{d}y/\mathrm{d}t$ is the velocity of the electron particle. Note that treating this quantum system with classical mechanics is by no means accurate from quantum mechanical point of view, but the goal is to obtain an approximation from a mechanical perspective. According to Newton's law of motion, the total force acting on the negative charge is equal to its mass times its acceleration, or in other words

$$\begin{aligned} m\ddot{y}(t) &= -qE(t) - \alpha y(t) - \beta\dot{y}(t), \\ &\Downarrow \\ \ddot{y}(t) + \gamma\dot{y}(t) + \omega_0^2 y(t) &= -\frac{q}{m}E(t), \end{aligned} \tag{2.3.23}$$

where we denoted $\gamma = \beta/m$ being the damping coefficient and $\omega_0 = \sqrt{\alpha/m}$ is the resonance frequency of the atom. We will look for a solution of the form
$y(t) = \mathrm{Re}[y_0 \exp(-i\omega t)]$, where y_0 is complex valued and write the electric field as $E(t) = \mathrm{Re}[E_{y0} \exp(-i\omega t)]$, where E_{y0} is complex. Then the equation (2.3.23) turns

to

$$-\omega^2 y_0 - i\gamma\omega y_0 + \omega_0^2 y_0 = -\frac{q}{m}E_y. \tag{2.3.24}$$

The solution to the above equation then yields

$$y_0 = \frac{-q/mE_y}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{2.3.25}$$

Our goal is to obtain the formula for the linear electric susceptibility of the material $\hat{\chi}(\omega)$. Recall that the linear polarization term is of the form $\mathbf{P}(t) = \varepsilon_0\hat{\chi}(\omega)\mathbf{E}(t)$, see (1.1.3). Thus we need to obtain the polarization $\mathbf{P}$. Inspired by this, we write the electric dipole moment p of the mass-spring system as $\mathbf{p} = -qy(t)\mathbf{e}_y = \mathrm{Re}[-qy_0\mathbf{e}_y\exp(-i\omega t)] = \mathrm{Re}[\mathbf{p}_0\exp(-i\omega t)]$, where we denoted

$$\mathbf{p}_0 = \frac{-q^2/mE_y}{\omega_0^2 - \omega^2 - i\gamma\omega}\mathbf{e}_y, \tag{2.3.26}$$

and we used the result (2.3.25). In order to obtain the electric dipole moment density which the polarization $\mathbf{P}$ really is, we need a constant N which stands for the number of such atoms with dipoles $\mathbf{p}$. The polarization is then defined as $\mathbf{P} = \mathrm{Re}[N\mathbf{p}_0\exp(-i\omega t)]$ that can be written in the form $\mathbf{P} = \mathrm{Re}[\varepsilon_0\hat{\chi}(\omega)E_y\exp(-i\omega t)\mathbf{e}_y]$, where we introduced the term

$$\hat{\chi}(\omega) = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega}, \tag{2.3.27}$$

with $\omega_p = \sqrt{Nq^2/(\varepsilon_0 m)}$ called the plasma frequency. The formula for the linear susceptibility (2.3.27) is referred to as the Lorentz model.

The type of approximations such as (2.3.27) in form of rational functions are commonly known as Sellmeier formulas in optics. Using these formulas in the dispersion relation turns the model equation, which is a differential equation in time, into a more complicated pseudo-differential equation in time. This is because the frequency variable ω is in the denominator, what in the time domain becomes an integral operator. The integral operator in our model equation (2.3.10) would be $\hat{\chi}(i\partial_t)$. For this situation, we introduce a transformation that allows us to turn the model equation into a differential equation in time, called Sellmeier transformation. The reason why we need such transformation is to be able to solve the model equation numerically and validate our MMS solution by comparison. The numerical solution of the model equation can be only obtained if the equation does not include integral operators.

Let us assume that the electric susceptibility of the material is approximated by a Sellmeier formula of the form $\hat{\chi}(\omega) = P(\omega)/Q(\omega)$, where P and Q are some polynomials in ω like the one we obtained (2.3.27). The idea behind the Sellmeier transformation is to transform the original pseudo-differential model equation (2.3.10) which can be written as $\mathcal{L}(E, E^3) = 0$ into a differential equation $\tilde{\mathcal{L}}(E, E^3) = 0$. The way we proceed is to take the Fourier transform of the model equation (in both space and time, although only time is sufficient) and rewrite it in the following way

$$\begin{aligned} \hat{\mathcal{L}}\left(E, E^3\right) &= 0, \\ \frac{1}{Q(\omega)}\left[Q(\omega)\hat{\mathcal{L}}\left(E, E^3\right)\right] &= 0, \\ \frac{1}{Q(\omega)}\hat{\tilde{\mathcal{L}}}\left(E, E^3\right) &= 0, \end{aligned} \tag{2.3.28}$$

where we denoted $\hat{\tilde{\mathcal{L}}} = Q(\omega)\hat{\mathcal{L}}$. At the end, the equation $\hat{\tilde{\mathcal{L}}}(E, E^3) = 0$ is transformed back using inverse Fourier transform. If Q was defined as $Q(\omega) = a\omega^2 + b\omega + c$, as is the form in (2.3.27), the operator $\tilde{\mathcal{L}} =$ becomes $\tilde{\mathcal{L}} = (c + ib\partial_t - a\partial_t^2) \circ \mathcal{L}$. The Sellmeier transformation of the model equation is then $\tilde{\mathcal{L}}(E, E^3) = 0$ which is a normal differential equation in time.

It is easy to observe from the transformed equation $\tilde{\mathcal{L}}(E, E^3) = 0$ that any of its solutions is also a solution to the original equation $\mathcal{L}(E, E^3) = 0$, but not vice versa. We can express this by stating that $S_D \subset S_M$, where S_M denotes the solution space of the original model equation and S_D is solution space of the transformed equation. One can argue that the Sellmeier transformation can break down when $Q(\omega)$ becomes zero, but this situation is rarely realized. We are, however, focusing on an even smaller subspace of S_D that are narrow band solutions described by the amplitude equation (2.3.15). Then the only question is whether the MMS applied to the Sellmeier transformed equation produces exactly the same amplitude equation as the one we got from the pseudo-differential equation (2.3.10). If this is true, then we can use the transformed equation to validate our MMS solution without the pseudo-differential operator in our way. We will choose two different examples of susceptibility, check whether the MMS solution matches for both transformed and original model equation and in a positive case, validate it by comparing with the numerical solution of the transformed equation.

The first model of material response function we are using is of the simple form

$$\hat{\chi}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{\gamma - ia\omega}, \tag{2.3.29}$$

where a and γ are real positive constants. This kind of susceptibility may not be physical at all, even though it satisfies the Kramers-Kronig relations. To briefly summarize what these relations mean, assume that we apply an electric field on a material at $t = 0$. An impulse-response $\mathbf{P}(t)$ (material polarization) in a material must therefore start at $t = 0$ as well and not earlier. Hence, $\mathbf{P}(t)$ is a causal function, that is $\mathbf{P}(t) = 0$ for $t < 0$. This is natural to assume and should be so in any physical system where a response to an impulse cannot occur before the impulse itself. The causality of the impulse-response is what intimately relates the real and imaginary part of the susceptibility $\hat{\chi}(\omega)$. These relationships caused by the causality are called Kramers-Kronig relations.

The model (2.3.29) not being physical does not need to stop us from using it for the validation of the amplitude equation. In fact, choosing a linear function in the denominator makes it the simplest possible choice. Using the procedure of the Sellmeier transformation we get the following transformed equation

$$\begin{aligned} (\gamma + a\partial_t)\left(\partial_{tt}E - \partial_{zz}E + \varepsilon^2\partial_{tt}E^3\right) + \partial_{tt}E = 0, \\ a\partial_{ttt}E + (\gamma+1)\partial_{tt}E - \gamma\partial_{zz}E - a\partial_{zzt}E + \varepsilon^2\gamma\partial_{tt}E^3 + \varepsilon^2 a\partial_{ttt}E^3 = 0, \end{aligned} \tag{2.3.30}$$

where we used the interchangeability between $\omega \leftrightarrow i\partial_t$ as a consequence of the Fourier transform. Observe that the transformed equation (2.3.30) is indeed differential equation in time and in this paper, we use pseudo-spectral method to solve it numerically. Recall that in order to use this transformation, the MMS solution to (2.3.30) must yield the same equations as (2.3.15), (2.3.17). The derivations are provided in Appendix B of the paper and it turns out, the MMS solutions are the same.

The equation (2.3.30), in its current form, cannot be solved with respect to its highest order time derivative since it contains a nonlinear term. Here, the small parameter ε can

be used to obtain an approximation to (2.3.30) by iteration such that it can be solved with respect to the highest derivative. This is achieved by expressing the highest time derivative from the linear term. The obtained expression is then substituted into the full equation (2.3.30) with $\partial_{ttt}E^3 = 6\left(\partial_t E\right)^3 + 18E\partial_t E\partial_{tt}E + 3E^2\partial_{ttt}E$. The terms of higher order than ε^2 are dropped, for it must be consistent with the MMS solution which is valid to order ε^2.

Before we proceed to the actual solving of the equation (2.3.30), we observe, that the amplitude equation is derived only for one amplitude, or one mode, to be precise, while (2.3.30) has three such modes since it is of third order derivative in time. The solution to the linear part of (2.3.30) can thus be expressed as inverse Fourier transform

$$E(z,t) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{3} \int_{-\infty}^{\infty} \mathrm{d}k A_i(k) e^{i(kz - \omega_i(k)t)}, \tag{2.3.31}$$

where $\omega_i(k)$ are the solutions to the dispersion equation which is a third order polynomial in ω. Due to a symmetry of the dispersion relation, namely that if $\omega(k)$ is a solution, then $-\omega^*(k)$ is a solution as well and that the electric field (2.3.31) must be real, one arrives at the following relations between the amplitudes $A_{1,2,3}(k)$:

$$\begin{aligned} A_1(k) &= A_2^*(-k), \\ A_3(k) &= A_3^*(-k). \end{aligned} \tag{2.3.32}$$

We see that the amplitudes for $k < 0$ are determined by their values for $k > 0$, so it is sufficient to specify them only for positive argument k. As mentioned earlier, the amplitude equation is correct for a narrow band solution of only one of the modes $A_i(k)$ centred around a positive wave number k_0. Let us assume that the mode, for which we have the amplitude equation, is $A_1(k)$. According to (2.3.32), specifying $A_1(k > 0)$ and $A_1(k < 0) = 0$ gives us $A_2(k > 0) = 0$. It is therefore consistent to choose $A_3(k) = 0$ for all k and $A_2(k > 0) = 0$. With these choices, the electric field (2.3.31) simplifies to

$$E(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k \underbrace{\left[A_1(k)e^{-i\omega(k)t} + A_1^*(-k)e^{i\omega^*(-k)t}\right]}_{\hat{E}(k,t)} e^{ikz} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k \hat{E}(k,t) e^{ikz}, \tag{2.3.33}$$

where we let $\omega_1(k) = \omega(k)$ to match the parameters in our MMS solution. As a part of our pseudo-spectral method of solving (2.3.30), we Fourier transform the equation into its k domain and obtain a third order ODE with a nonlinear right hand side. The third order implies we need three initial conditions $\hat{E}(k,0), \partial_t\hat{E}(k,0)$ and $\partial_{tt}\hat{E}(k,0)$ if we want to solve it as an initial value problem. These conditions can be acquired from (2.3.33) through the amplitude $A_1(k)$. If we denote the three initial conditions $\hat{f}(k), \hat{g}(k)$ and $\hat{h}(k)$, respectively, then we get

$$\hat{f}(k) = A_1(k) + A_1^*(-k), \tag{2.3.34a}$$

$$\hat{g}(k) = -i\omega(k)A_1(k) + i\omega^*(-k)A_1^*(-k), \tag{2.3.34b}$$

$$\hat{h}(k) = -\omega^2(k)A_1(k) - \left(\omega^*(-k)\right)^2 A_1^*(-k). \tag{2.3.34c}$$

Thus the choice of $A_1(k)$ is sufficient to determine all the initial conditions to solve (2.3.30). The ODE can be turned into a system of first order ODEs and solved numerically. To be

able to solve the amplitude equation as well, we need to compute the initial condition for (2.3.15) in terms of the chosen function $A_1(k)$. Taking the inverse Fourier transform in k of the right hand side of (2.3.34a) gives us $E(z,0)$. Consequently, we equate this function with the right hand side of (2.3.21) evaluated at $t=0$ and match separately the first part and the second part which is the complex conjugate of the first part. We arrive at the system

$$\begin{aligned}\mathcal{F}^{-1}\left\{A_1(k)\right\} &= A(z,0)e^{ik_0z} + c_2\eta E_0^2|A(z,0)|^2A^*(z,0)e^{ik_0z},\\ \mathcal{F}^{-1}\left\{A_1^*(-k)\right\} &= A^*(z,0)e^{-ik_0z} + c_2^*\eta E_0^2|A(z,0)|^2A(z,0)e^{-ik_0z}.\end{aligned} \tag{2.3.35}$$

This is a nonlinear system of algebraic equations with the unknowns $A(z,0)$ and $A^*(z,0)$. It can be solved numerically with Newton's method. The initial condition for the other amplitude $B(z,0)$ is readily obtained using the condition $B(z,0) = -c_1A^3(z,0)$.

We must be aware of the consistency with assumption for the narrow band solutions, while choosing the initial condition $A_1(k)$. Perhaps the most natural representation in the spectral domain k is a narrow Gaussian centred at $k_0 > 0$. The parameters of the Gaussian should be chosen such that the function is narrow enough. The assumption is that the amplitude $A(z,t)$ should be slowly varying in z, or in other words $\partial_z A \sim \mathcal{O}(\varepsilon)$. The width of the Gaussian needs to be therefore chosen so that its width becomes $\sim 1/\varepsilon$.

In the numerical test to validate the amplitude equation, the parameters for the material response model (2.3.29) were chosen to be $a = 20$ and $\gamma = 5$. With the choice of the perturbation parameter, we face a kind of a tradeoff. The underlying assumption is that $\varepsilon < 1$. Making the parameter ε as small as possible so that the separation of scales in the MMS solution is as clear as possible, is reasonable. On the other hand, making ε very small means, we have to propagate the wave equation a very long time for the nonlinearities to be detectable or to affect the spectrum of the waves. Also, it would be challenging numerically for the model equation due to long running times. During the MMS process, we removed the secular terms up to the ε^2 order which implies that the MMS solution should be valid up to $t \lesssim \varepsilon^{-2}$. So for a very small ε, the model equation should be solved up to ε^{-2} which could become very large.

In our simulations, we chose $\varepsilon = 10^{-1}$. This may not be a very small number, but it turns out, the amplitude equation does approximate the exact model quite well for this value of ε. The numerical test begins with computing the initial conditions for the amplitude equations $A(k,0)$ and $B(k,0)$. They are both Gaussian centred at $k = 0$ with widths $\sim 1/\varepsilon$, as expected. We have also calculated the initial condition without considering the other amplitude $B(k,0)$. The outcome was that $A(k,0)$ consisted of a Gaussian at $k = 0$ and also another Gaussian of order ε^2 centred at $2k_0$. This led to a contradiction since we assumed a narrow band initial condition.

Returning back to the numerical test, with $k_0 = 2\pi$, the dispersion relation yields the frequency $\omega_0 = 6.28 - i2.5 \times 10^{-2}$. The real part of the frequency ω_0 represents the oscillation frequency of the initial pulse while the imaginary part stands for its decaying rate. From the graph of $n(\omega)$, we would observe that the numerical value of the real part of ω_0 puts our initial pulse to the right of the material resonance. In our paper, we provided plots of the two solutions of (2.3.30) using a system of ODEs and using the amplitude equation, on top of each other at four different times $t = 0, 15, 40$ and $t = 50$. In addition, the functions were compared in their spectral domain k at $t = 50$. At all times, the two functions matched surprisingly well. The small deviations are not exceeding the ε^2 order. In their spectrum graphs, the functions were consisting of two Gaussians, one larger sitting at $k_0 = 2\pi$, as expected and one smaller at $3k_0 = 6\pi$ with height of order ε^2.

The one at $3k_0$ can be explained by the nonlinearity having the cubic form E^3. Inserting a plane wave for E in (2.3.30) would result into multiplying the wave numbers by a factor of 3, hence $3k_0$. It can be therefore regarded as the nonlinear effect of the third order polarization term.

However, the results were not always satisfactory during our extensive numerical tests with various parameters for the susceptibility and the initial wave number k_0. It was observed that in order to stay on the correct asymptotic regime, the parameters were subjected to certain constraints. The susceptibility $\hat{\chi}(\omega)$ in (2.3.10) was approximated by its Taylor expansion whose terms can be observed in the parameter α in (2.3.16a). We expect the Taylor series to converge, so the derivatives of $\hat{\chi}(\omega)$ involved here should not break the order of the preceding terms in the expansion. This depends mainly on the parameters a, γ and the frequency ω around which the Taylor series is expanded.

Judging from the results of the numerical tests, we may conclude that the amplitude equation accurately approximates the exact solutions for Maxwell. On the other hand, a stability analysis of the amplitude equation reveals that the equation is in fact, ill-posed. This might come as a shock since the numerical test suggests its validity. The stability of (2.3.15) can be easily investigated by substituting $A(z,t) = \exp(\lambda(k)t)\exp(ikz)$ into the equation and cancelling the common factors. At the end we find the expression for the function $\lambda(k)$ which serves as the growth curve. If $\mathrm{Re}\ \lambda > 0$, the solution to the amplitude equation grows exponentially and decays when $\mathrm{Re}\ \lambda < 0$. We find that the growth curve is of the form

$$\mathrm{Re}\ \lambda(k) = a_1 k^2 + a_2 k, \tag{2.3.36}$$

where the constants a_1, a_2 include some algebraic combinations of the real and imaginary parts of the quantities $\omega'(k_0), \alpha$ and β. The function (2.3.36) is a parabola passing through the origin. For our parameter choices in the previous numerical test, the constant a_1 is positive and a_2 negative which leads to positive growing of the parabola, thus exponential growth of the solution to the amplitude equation. We face two contradictory facts: on one hand, the amplitude equation approximates the exact solution very well and on the other hand, it is ill-posed which usually means bad news for any proposed mathematical model. There are many examples in mathematics where diverging models describe the reality accurately within certain asymptotic bounds [12][69]. Here, the constants a_1 and a_2 are of order 10^{-6} and 10^{-5}, respectively. The graph of the parabola is thus very shallow. If we plot the initial condition $A(k,0), B(k,0)$ together with the growth curve in the range $k \in [-10, 10]$, we find that the initial conditions have their support confined within $k \in [-2, 2]$ and the maximum of $\lambda(k)$ in the range $k \in [-10, 10]$ is around 5×10^{-4} at $k = -10$. In addition, the stability curve always passes zero, the initial conditions are centred at zero and they are narrow by definition. This suggests that the region where the solution to the amplitude equation resides, will grow at the slowest rate in case of a growth. The amplitude equation is also valid during the time $t \lesssim \varepsilon^{-2}$ which in our case is $t \lesssim 100$. All these facts contribute to the conclusion that for any solution satisfying the assumptions to derive the amplitude solution, will not grow large enough to affect to solution up to order ε^2 and greater.

In the rest of the paper, we present one more example of the susceptibility model, namely the Lorentz model we derived earlier in (2.3.27)

$$\hat{\chi}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{a\omega^2 + ib\omega + c}, \tag{2.3.37}$$

where $a = -1/\omega_p^2, b = -\gamma/\omega_p^2$ and $c = \omega_r^2/\omega_p^2$. Note, that the formula (2.3.37) is scaled according to the scaling introduced in the beginning of this paper as opposed to the formula calculated in (2.3.27). This model represents an actual physical model as opposed to (2.3.29) and also satisfies the Kramer-Kronig relations. The function we are using has an additional factor $1/\sqrt{2\pi}$ due to a Fourier transform convention. As with the previous mode, the same procedure has been done with the Lorentz model: finding the Sellmeier transformation, iterating the transformed equation in order to remove the highest time derivative by the nonlinear term, expressing the solution as the inverse Fourier transform of 4 independent modes, finding the relations between them and calculating the initial condition for the amplitude equation. The Sellmeier transformation now produces a 4-th order differential equation in time, so we get a 4×4 nonlinear system of ODEs.

To validate the amplitude equation for the Lorentz model of dispersion, we chose two different set of parameters a, b, c. These two sets correspond to two physical models where the material resonance occurs at different wavelengths of light. The first choice of parameters represents a material with ultraviolet resonance (wavelength is of order of tens of nanometers). The frequency of the initial pulse was placed to the left of the resonance. In the second set of parameters we chose the material to have infrared resonance (wavelength is of order of tens of micrometers) and the frequency of the initial pulse is at the right of the resonance. In both examples, the overlap between the MMS solution and the high precision solution of the model equation is satisfactory both in their spatial and spectral domains. The investigation of the stability for both sets of parameters resulted into well-posedness for the amplitude equation in the ultraviolet case and ill-posedness for the infrared case. The instability in the latter case did not pose any danger as the numerical tests showed.

This paper contains successful derivation of the MMS solution to Maxwell's equation (2.3.10) and validation of its numerical accuracy. We introduced the Sellmeier transformation which helped to remove the pseudo-differential operator from the Maxwell's equation. The MMS produced a linear amplitude equation which made it much faster to solve compared to solving the original Maxwell's equation. We found that the amplitude equation can be ill-posed depending on the parameters of the susceptibility, but a careful examination showed that it does not represent a problem due to the nature of the instability and the location of the initial condition. The amplitude equation was validated through three numerical tests using two different dispersion models (susceptibilities). We found the amplitude equation to be an accurate approximation of the exact solution in each test. However, countless other numerical tests revealed some limitations of the MMS solution. In all cases, the main source of failure was the choice of parameters in the susceptibility and the initial frequency ω_0. It affected the Taylor expansion of $\hat{\chi}(\omega)$ around ω_0 where the ordering of the terms became compromised. Also, when one was near resonance, the MMS solution did not perform as good as expected, partly because the magnitude of the nonlinear terms exceeded the order ε^2 in the numerical solution. Further investigations of these problems are definitely required.

References

[1] Maria Göppert-Mayer. Über elementarakte mit zwei quantensprüngen. *Annalen der Physik*, 401(3):273–294, 1931.

[2] W. Kaiser and C. G. B. Garrett. Two-photon excitation in caf_2: eu^{2+}. *Phys. Rev. Lett.*, 7:229–231, Sep 1961.

[3] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich. Generation of optical harmonics. *Phys. Rev. Lett.*, 7:118–119, Aug 1961.

[4] T. H. Maiman. Stimulated optical radiation in ruby. *Nature*, 187(4736):493–494, Aug 1960.

[5] R. J. Collins, D. F. Nelson, A. L. Schawlow, W. Bond, C. G. B. Garrett, and W. Kaiser. Coherence, narrowing, directionality, and relaxation oscillations in the light emission from ruby. *Phys. Rev. Lett.*, 5:303–305, Oct 1960.

[6] Simon Mahler, Chene Tradonsky, Ronen Chriki, Asher A. Friesem, and Nir Davidson. Coupling of laser arrays with intracavity elements in the far-field. *OSA Continuum*, 2(6):2077–2084, Jun 2019.

[7] Michael Koziol. Infinera and windstream beam 800 gigabits per second through a single optical fiber, 2020.

[8] G. P. Puccioni, A. Poggi, W. Gadomski, J. R. Tredicce, and F. T. Arecchi. Measurement of the formation and evolution of a strange attractor in a laser. *Phys. Rev. Lett.*, 55:339–342, Jul 1985.

[9] Hendrik Antoon Lorentz. Versuch einer theorie der electrischen und optischen erscheinungen in bewegten körpern. *Cambridge University Press*, 2013.

[10] jr. N.N. Bogolyubov. Perturbation theory. *Encyclopedia of Mathematics.*, 2011.

[11] The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 114(767):243–265, March 1927.

[12] F. J. Dyson. Divergence of perturbation theory in quantum electrodynamics. *Phys. Rev.*, 85:631–632, Feb 1952.

[13] J. D. Cole J. Kevorkian. Multiple scale and singular perturbation methods. *Springer*, 1996.

[14] E. J. Hinch. Perturbation methods. *Cambridge University Press*, 2012.

[15] Ali H. Nayfeh. Perturbation methods. *John Wiley & Sons*, 1973.

[16] Bhimsen K. Shivamoggi. Perturbation methods for differential equations. *Birkhäuser*, 2003.

[17] James A. Murdock. Perturbations theory and methods. *John Wiley & Sons*, 1991.

[18] Henri Poincare. Les methodes nouvelles de la mecanique celeste. 1892.

[19] Alan C. Newell and Jerome V. Moloney. Nonlinear optics. *Addison Wesley*, 1992.

[20] Ali H. Nayfeh. Order reduction of retarded nonlinear systems – the method of multiple scales versus center-manifold reduction. *Nonlinear Dynamics*, 51(4):483–500, Mar 2008.

[21] Per Kristen Jakobsen. Topics in applied mathematics and nonlinear waves, 2019.

[22] J. Cole and J. Kevorkian. Uniformly valid asymptotic approximations for certain non-linear differential equations. 1963.

[23] Maurice L. Rasmussen. Uniformly valid approximations for non-linear oscillations with small damping. *International Journal of Non-Linear Mechanics*, 5(4):687–696, 1970.

[24] R.T. Davis and K.T. Alfriend. Solutions to van der pol's equation using a perturbation method. *International Journal of Non-Linear Mechanics*, 2(2):153–162, 1967.

[25] Edward L. Reiss. On multivariable asymptotic expansions. *SIAM Review*, 13(2):189–196, 1971.

[26] G.E Kuzmak. Asymptotic solutions of nonlinear second order differential equations with variable coefficients. *Journal of Applied Mathematics and Mechanics*, 23(3):730–744, 1959.

[27] N. D. Fowkes. A singular perturbation method. part i. *Quarterly of Applied Mathematics*, 26(1):57–69, 1968.

[28] R.V Ramnath and G Sandri. A generalized multiple scales approach to a class of linear differential equations. *Journal of Mathematical Analysis and Applications*, 28(2):339–364, 1969.

[29] Hung Cheng and Tai Tsun Wu. An aging spring. *Studies in Applied Mathematics*, 49(2):183–185, 1970.

[30] P. Noerdlinger and V. Petrosian. The effect of cosmological expansion on self-gravitating ensembles of particles. *The Astrophysical Journal*, 168:1, 1971.

[31] J. Kevorkian. Passage through resonance for a one-dimensional oscillator with slowly varying frequency. *SIAM Journal on Applied Mathematics*, 20(3):364–373, 1971.

[32] J. W. Searl. Expansions for singular perturbations. *IMA Journal of Applied Mathematics*, 8(2):131–138, Oct 1971.

[33] K. K. Tam. On the asymptotic solution of the orr–sommerfeld equation by the method of multiple-scales. *Journal of Fluid Mechanics*, 34:145–158, 1968.

[34] Ali Hasan Nayfeh. A comparison of three perturbation methods for earth-moon-spaceship problem. *AIAA Journal*, 3(9):1682–1687, 1965.

[35] Lu Ting and Sherwood Brofman. On take-off from circular orbit by small thrust. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 44(10-11):417–428, 1964.

[36] Yun Y. Shi and Martin C. Eckstein. Ascent or descent from satellite orbit by low thrust. *AIAA Journal*, 4(12):2203–2209, 1966.

[37] J. Kevorkian. The two variable expansion procedure for the approximate solution of certain non-linear differential equations. *American Mathematical Society*, 1966.

[38] K. T. Alfriend and R. H. Rand. Stability of the triangular points in the elliptic restricted problem of three bodies. *AIAA Journal*, 7(6):1024–1028, 1969.

[39] M. Eckstein, Y. Shi, and J. Kevorkian. Use of the energy integral to evaluate higher-order terms in the time history of satellite motion. *Astronomical Journal*, 71:301, June 1966.

[40] Yun-Yuan Shi and Martin C. Eckstein. Application of Singular Perturbation Methods to Resonance Problems. *Astronomical Journal*, 73:275, May 1968.

[41] Ali Hasan Nayfeh and William S. Saric. An analysis of asymmetric rolling bodies with nonlinear aerodynamics. *AIAA Journal*, 10(8):1004–1011, 1972.

[42] Ali Hasan Nayfeh. A multiple time scaling analysis of re-entry roll dynamics. *AIAA Journal*, 7(11):2155–2157, 1969.

[43] Edward L. Reiss and Bernard J. Matkowsky. Nonlinear dynamic buckling of a compressed elastic column. *Quarterly of Applied Mathematics*, 29(2):245–260, 1971.

[44] Michael P. Mortell. Traveling load on a cylindrical shell. *The Journal of the Acoustical Society of America*, 44(6):1664–1670, 1968.

[45] Luigi Morino. A perturbation method for treating nonlinear panel flutter problems. *AIAA Journal*, 7(3):405–411, 1969.

[46] Robert T. Wingate and R. T. Davis. Perturbation solution of a hyperbolic equation governing longitudinal wave propagation in certain nonuniform bars. *The Journal of the Acoustical Society of America*, 47(5B):1334–1337, 1970.

[47] J. C. Luke. A perturbation method for nonlinear dispersive wave problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 292(1430):403–412, 1966.

[48] M. J. Ablowitz and D. J. Benney. The evolution of multi-phase modes for nonlinear dispersive waves. *Studies in Applied Mathematics*, 49(3):225–238, 1970.

[49] Robert W. Boyd. Nonlinear optics. *Elsevier, AP Academic Press*, 2020.

[50] J. M. Brown, P. Jakobsen, A. Bahl, J. V. Moloney, and M. Kolesik. On the convergence of quantum resonant-state expansion. *Journal of Mathematical Physics*, 57(3):032105, 2016.

[51] Boris A. Malomed. Chapter 2 - variational methods in nonlinear fiber optics and related fields. volume 43 of *Progress in Optics*, pages 71–193. Elsevier, 2002.

[52] Thomas Brabec and Ferenc Krausz. Nonlinear optical pulse propagation in the single-cycle regime. *Phys. Rev. Lett.*, 78:3282–3285, Apr 1997.

[53] Alex A. Zozulya and Scott A. Diddams. Dynamics of self-focused femtosecond laser pulses in the near and far fields. *Opt. Express*, 4(9):336–343, Apr 1999.

[54] A. Couairon, E. Brambilla, T. Corti, D. Majus, O. de J. Ramírez-Góngora, and M. Kolesik. Practitioner's guide to laser pulse propagation models and simulation. *The European Physical Journal Special Topics*, 199(1):5–76, Nov 2011.

[55] M. Kolesik, J. V. Moloney, and M. Mlejnek. Unidirectional optical pulse propagation equation. *Phys. Rev. Lett.*, 89:283902, Dec 2002.

[56] M. Kolesik and J. V. Moloney. Nonlinear optical pulse propagation simulation: From maxwell's to unidirectional equations. *Phys. Rev. E*, 70:036604, Sep 2004.

[57] P. Kinsler. Limits of the unidirectional pulse propagation approximation. *J. Opt. Soc. Am. B*, 24(9):2363–2368, Sep 2007.

[58] Per Jakobsen. Bidirectional pulse propagation equation for extreme nonlinear optics. *Physica Scripta*, 89(9):095502, aug 2014.

[59] Mingsian R. Bai. Application of bem (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries. *The Journal of the Acoustical Society of America*, 92(1):533–549, 1992.

[60] Michael C. Harris, Jonathan D. Blotter, and Scott D. Sommerfeldt. Obtaining the complex pressure field at the hologram surface for use in near-field acoustical holography when pressure and in-plane velocities are measured. *The Journal of the Acoustical Society of America*, 119(2):808–816, 2006.

[61] Finn Jacobsen and Virginie Jaud. Statistically optimized near field acoustic holography using an array of pressure-velocity probes. *The Journal of the Acoustical Society of America*, 121(3):1550–1558, 2007.

[62] Yei-Chin Chao. An implicit least-square method for the inverse problem of acoustic radiation. *The Journal of the Acoustical Society of America*, 81(5):1288–1292, 1987.

[63] In-Youl Jeon and Jeong-Guon Ih. On the holographic reconstruction of vibroacoustic fields using equivalent sources and inverse boundary element method. *The Journal of the Acoustical Society of America*, 118(6):3473–3482, 2005.

[64] Hua Lee and Douglas P. Sullivan. Fundamental limitation of resolution enhancement by wave-field extrapolation. *The Journal of the Acoustical Society of America*, 84(2):611–617, 1988.

[65] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.

[66] Steven Johnson. Notes on perfectly matched layers (pmls). 01 2007.

[67] Steffen Marburg. Computational acoustics of noise propagation in fluids: finite and boundary element methods. *Springer*, 2008.

[68] J. J. Thomson. On electrical oscillations and the effects produced by the motion of an electrified sphere. *Proceedings of the London Mathematical Society*, s1-15(1):197–219, 1883.

[69] Michael Cross and Henry Greenside. Pattern formation and dynamics in nonequilibrium systems. *Cambridge University Press*, 2009.

3 Paper 1

Convergence and completeness for square-well Stark resonant state expansions

Journal of Mathematical Physics, 59(11):113501, 2018

JOURNAL OF MATHEMATICAL PHYSICS **59**, 113501 (2018)

Convergence and completeness for square-well Stark resonant state expansions

David Juhasz,[1] Miro Kolesik,[2] and Per Kristen Jakobsen[1]
[1]*Department of Mathematics and Statistics, The Arctic University of Norway, 9019 Tromsø, Norway*
[2]*College of Optical Sciences, University of Arizona, 1630 East University Boulevard, Tucson, Arizona 85721, USA*

(Received 1 June 2018; accepted 16 October 2018; published online 2 November 2018)

In this paper, we investigate the completeness of the Stark resonant states for a particle in a square-well potential. We find that the resonant state expansions for target functions converge inside the potential well and that the existence of this convergence does not depend on the depth of the potential well, V_0. By analyzing the asymptotic form of the terms in these expansions, we prove some results on the relation between smoothness of target functions and the asymptotic rate of convergence of the corresponding resonant state expansion and show that the asymptotic rate of convergence is also independent of V_0, but the absolute size terms in the series asymptotically goes as V_0^{-1}. *Published by AIP Publishing.* https://doi.org/10.1063/1.5042523

I. INTRODUCTION

Decaying quantum states were first introduced in 1928 by Gamow[1,2] and independently by Gurney and Condon[3,4] in the context of nuclear physics, to describe long-lived wave functions of particles that eventually "escape" from a confining but unstable potential. This was however not the first time decaying eigenstates were used in physics. As early as 1884, Thomson[5] used them to describe decay phenomena in electromagnetism. Characterization of the unstable states using the absence of incoming waves was first introduced by Siegert[6] in the context of the nuclear scattering matrix. The Siegert characterization was taken up by Peierls,[7] Couteur,[8] and Humblet[9] and developed into a powerful tool in nuclear scattering theory. The wave functions satisfying the Siegert outgoing-wave conditions became known as resonant states and their properties have been investigated for many years (for example, Refs. 10–19).

The fact that resonant states decay exponentially in time implies that it is more likely to find the released particle far from the nucleus than closer to it since it is more likely to have been released by an earlier time than a later one. The resonant states are thus not normalizable, at least not in the standard sense introduced by von Neumann when he gave a proper mathematical foundation[20] for a subset of the Dirac formalization of quantum mechanics[21] in terms of Hilbert spaces for the states and Hermitian operators for the observables. Subsequently, after many intervening years of steady mathematical progress involving many people, an extended mathematical foundation for the full Dirac formalization, involving rigged Hilbert spaces, was developed by Gelfand.[22] It turned out that the context of rigged Hilbert spaces was flexible enough to also include resonant states even if such states were not part of Dirac's original formalization of quantum mechanics. However, the theory of rigged Hilbert spaces is perhaps a bit heavy on the mathematical side, and therefore, several other extensions of the von Neumann foundation of quantum mechanics designed to accommodate resonant states has been developed over the years.[23–28] Over and above these mathematical developments, a key issue involving resonant states has been their physical interpretation. The fact that they cannot be normalized in the Hilbert space setting means that the Born interpretation of quantum states in terms of probability theory fails. However, it has been shown that a state with infinite norm can be given a probabilistic interpretation by considering an expanding sphere in which the probability density (and the norm) remains conserved.[29] Be this as it may, it has always been recognized that such states contain useful physical information. For example, the temporal decay of the resonant state

corresponding to a bound state in the zero field limit is related to the life time of that state under the influence of the field.

The mathematical peculiarities were perhaps perceived less as an obstacle in electromagnetics where the counterparts of quantum resonances are the so-called leaky modes. These are configurations of the electromagnetic fields inside unstable resonators and waveguides that can "survive" for long times, but eventually radiate their energy away. They have been studied with keen interest in resonator cavities,[30,31] optical waveguides,[32] and photonic[33] and plasmonic[34,35] structures and are often utilized for numerical simulations.

However, despite their long history of utility (e.g., Refs. 36–39) in various fields, the unstable states, both quantum and electromagnetic, have not yet been fully understood. On the quantum side, the lack of a general theory for non-self adjoint operators is challenging and is the reason why non-Hermitian systems are mostly investigated on a case-by-case basis (see, e.g., Ref. 25). Ours is precisely such a study of a concrete quantum system.

Our particular motivation for investigating this problem comes from our long time involvement in the problem of high intensity optical pulse propagation.[40,41] For such high intensity fields, it is very challenging to come up with a material response theory that is reasonably accurate and also reasonably fast to evaluate. This last requirement is of paramount importance for long distance propagation[42] and rules out any scheme involving a direct integration of the Schrödinger equation[43] because of the large spectrum of space and time scales involved. There are roughly three orders of magnitude between the oscillation time scale for the electrons in the atom and the optical time scale. This difference, which is a problem for any direct integration scheme, also presents an opportunity for simplification; if one can find a complete set of resonant states for an atom in a constant field, then the solutions for a variable field can be expanded, to a good approximation, in the same set of resonant states by simply making the parameter representing the field strength time dependent. This is an adiabatic assumption that becomes more accurate with the larger the gap between electronic time scale and the optical time scale becomes.

The purpose of this paper is to demonstrate the completeness of the resonant eigenstates for a quantum particle in a square-well potential that is exposed to a homogeneous external electric field and to investigate the convergence properties of the corresponding resonant-state expansion. The external field, even a weak one, has a profound effect on the energetic spectrum of the system; As soon as the field is switched on, all discrete-energy eigenstates dissolve into the energy-continuum which fills up the whole real axis.[44] At the same time, resonances appear in the complex energy plane, and it is these decaying states our work is concerned with. The results presented in this work generalize and extend the findings in our previous study,[45] where the convergence of a resonant state expansion was investigated for the case of a zero-range Dirac-delta potential.[46–48] In particular, the more realistic system investigated here allows us to make conjectures concerning a wide family of one-dimensional quantum systems.

The paper is organized as follows. In Sec. II, we set up the problem by introducing the resonant states for a square-well potential and the locations of their energy eigenvalues in the complex plane; these results are known from the literature.[49] In an attempt to make the paper more readable, we decided to omit most of the technical but rather standard calculation details in favor of demonstrating the validity of important intermediate results with the help of high-precision numerical tests. In particular, we have made extensive numerical investigations into the convergence of the resonant state expansions and in Sec. III we show some of the results of these investigations. They indicate strongly that the resonant state expansions converge point-wise to the left (i.e., against the pull of the field) of the well and inside the well, but diverge to the right of the well, assuming here that the external field points to the right. This result is consistent with what was obtained for the case of a Dirac delta potential in Ref. 45. In Sec. IV, we prove that what the numerical evidence indicates is indeed true. Resonant state expansions are shown to converge point-wise to the left of the well and inside the well and diverge to the right of the well for all reasonable initial data for the Schrödinger equation. In Sec. V, we investigate the asymptotic form of the terms in the resonant state expansion and make precise statements about the rate of convergence and how this rate relates to the smoothness of the function that is being expanded. We also present high-precision numerical calculations to verify the correctness of our asymptotic formulas and the statements on convergence rates. One surprising

conclusion that came out of this investigation is that the depth of the potential well does not play much of a role; both the question of convergence of the resonant state expansion and its asymptotic rate of convergence are independent of the depth of the well. We finish the paper with Sec. VI where we briefly discuss on what has been achieved and where we also spell out the aforementioned conjecture in some more detail.

II. STARK RESONANT STATES FOR A SQUARE WELL

Let us consider the following Hamiltonian:

$$H = -\frac{1}{2}\partial_{xx} + V(x) - \varepsilon x, \tag{1}$$

where ε is the strength of the external field and where we without loss of generality assume that $\varepsilon > 0$, which corresponds to the external electric field pulling the electron to the right. The atomic potential, $V(x)$, is modeled by a square well of width $2d$ and depth V_0

$$V(x) = \begin{cases} -V_0, & |x| < d, \\ 0, & |x| > d. \end{cases}$$

According to Siegert's characterization, Stark resonant states for the square well are wave functions of the form

$$\psi(x,t) = \psi_\omega(x) e^{-i\omega t}, \tag{2}$$

where $\psi_\omega(x)$ are solutions to the equation

$$H\psi_\omega = \omega\psi, \tag{3}$$

which satisfy the boundary conditions

$$\begin{aligned} &\psi_\omega(x) \to 0 \ \text{ when } x \to -\infty, \\ &\psi_\omega(x) \text{ and } \psi'_\omega(x) \text{ are continuous at } x = -d, d, \\ &\psi_\omega(x) \text{ is a purely outgoing wave at } \ x = \infty. \end{aligned} \tag{4}$$

The resonant states can be expressed in terms of Airy functions in the form

$$\psi_p(x) \equiv \psi_{\omega_p}(x) = \begin{cases} a_1 \mathrm{Ai}(y_1(x,\omega_p)) & x < -d, \\ a_2 \mathrm{Ai}(y_2(x,\omega_p)) + a_3 \mathrm{Bi}(y_2(x,\omega_p)), & -d < x < d, \\ a_4 \mathrm{Ci}^+(y_1(x,\omega_p)) & d < x < \infty, \end{cases} \tag{5}$$

with the notation $\mathrm{Ci}^+ = \mathrm{Bi} + i\mathrm{Ai}$ representing an Airy combination that asymptotically behaves as an outgoing wave. The functions $y_1(x, \omega)$ and $y_2(x, \omega)$ that parametrize the arguments of the Airy functions above are given by

$$\begin{aligned} y_1(x,\omega) &= -2(2\varepsilon)^{-\frac{2}{3}}(\varepsilon x + \omega), \\ y_2(x,\omega) &= -2(2\varepsilon)^{-\frac{2}{3}}(\varepsilon x + V_0 + \omega). \end{aligned}$$

The *resonant eigenvalues* ω_p are obtained as solvability conditions for the coefficients a_i ensuring that the wave function and its derivative are both continuous at the edges of the potential well, at $x = \pm d$. There is a countable set of such solutions in the complex plane, determined by the equation

$$\det M(\omega_p) = 0, \tag{6}$$

where $M(\omega)$ is a certain 4×4 matrix. The explicit expression for the determinant of $M(\omega)$ is

$$\det M(\omega) = (A_0A'_1 - A'_0A_1)(B_2C'_3 - B'_2C_3) - (A_0B'_1 - A'_0B_1)(A_2C'_3 - A'_2C_3),$$

where we have defined

$$\begin{aligned} &A_0 = \mathrm{Ai}\left(\mu\left(-d + \frac{\omega}{\varepsilon}\right)\right), A_1 = \mathrm{Ai}\left(\mu\left(-d + \frac{\omega + V_0}{\varepsilon}\right)\right), B_1 = \mathrm{Bi}\left(\mu\left(-d + \frac{\omega + V_0}{\varepsilon}\right)\right), \\ &A_2 = \mathrm{Ai}\left(\mu\left(d + \frac{\omega + V_0}{\varepsilon}\right)\right), B_2 = \mathrm{Bi}\left(\mu\left(d + \frac{\omega + V_0}{\varepsilon}\right)\right), C_3 = \mathrm{Ci}^+\left(\mu\left(d + \frac{\omega}{\varepsilon}\right)\right) \end{aligned}$$

with $\mu = -(2\varepsilon)^{\frac{1}{3}}$.

The resonant eigenvalues lie in the lower complex half plane which means that the corresponding resonant-state wave functions (2) are decaying in time, as they should. Figure 1 shows the zero contours of the real and imaginary parts of the equation. The points where they cross are the zero points of the determinant and thus are the resonant eigenvalues. There exist two infinite families of resonant eigenvalues. The family on the right side of the imaginary axis, called as the A-series in this paper, has zeros located close to the positive real axis and correspond to longer living states, while the family to the left of the imaginary axis has eigenvalues located along the ray $\arg(z) = -\frac{2\pi}{3}$ and corresponds to fast decaying states. We call this second family as the C-series. Figure 1 also shows a finite family of resonant states lying close to the negative real axis that corresponds to the bound state eigenvalues for the square well in the limit when the external field approach zero. We call these states as perturbed bound states. This structural division of the resonant states into a finite number of perturbed bound states and the two infinite A-series and C-series is also seen for the short range delta potential in Ref. 45, is probably generic, and should be expected for more general potentials also.

Our focus in this paper is to investigate to what extent the resonant states we have found can be used to expand initial conditions for the Schrödinger equation. If any given initial condition can be expanded in resonant states

$$f(x) = \sum_p \frac{\left(f, \psi_p\right)}{\left(\psi_p, \psi_p\right)} \psi_p(x), \tag{7}$$

the solution to the time dependent Schrödinger equation for the Hamiltonian operator (1), with initial data $\psi(x, 0) = f(x)$, is given by

$$\psi(x, t) = \sum_p \frac{\left(f, \psi_p\right)}{\left(\psi_p, \psi_p\right)} \psi_p(x) e^{-i\omega_p t}. \tag{8}$$

This is an exact solution for a static external field corresponding to a fixed value for ε in (1). For the case of an external field that varies slowly on the atomic time scale, which we take to be atto-seconds, the expansion (8), where now the resonant states and their complex eigenvalues vary in time through their dependence on $\varepsilon = \varepsilon(t)$, is a good approximation to the exact solution.

The eigenvalue problem (3) and (4) is not self-adjoint and as a consequence the eigenvalues displayed in Fig. 1 are complex and the resonant states (5) are not normalizable on the real line. This is, as discussed in the Introduction, to be expected on physical grounds.

In order to achieve the normalization for the resonant states, we rely on the technique used in Ref. 50. In this technique, one introduces a complex contour $\mathcal{L}$ on which the resonant states are

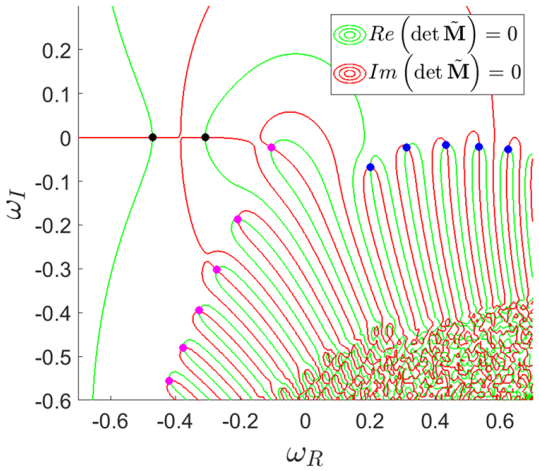


FIG. 1. Contour plot of det $M(\omega_R + i\omega_I)$. The green and red lines are the zero contours of the real and imaginary part of det $M(\omega_R + i\omega_I)$. The parameters chosen for this illustration were $\varepsilon = 0.03$, $V_0 = 0.5$, $d = 4$. The eigenvalues from the A-series are the blue dots, those from the C-series are purple, and the perturbed bound states are in black.

evaluated. The contour we will be using has the form

$$\mathcal{L} = z(x) = \begin{cases} x, & x < x_c, \\ x_c + e^{i\theta}(x - x_c) & x > x_c \end{cases} \tag{9}$$

which serves its purpose for any chosen parameters $x_c > 0$ and $0 < \theta \leq \frac{\pi}{2}$. In the following, we will consider $\theta = \frac{\pi}{2}$ and $x_c > d$ to the right of the well. One can see from the form of the resonant states that they decay on the negative axis, while on the positive part, where the outgoing wave dominates the wave function shape, they show exponential growth. It turns out that for any $0 < \theta \leq \frac{\pi}{2}$, the outgoing part of the wave function decays along the contour and integration along the contour can be used to normalize the states. The normalization is achieved by replacing the usual Hermitian inner product for complex valued functions on the real line by another bilinear product for complex valued functions defined on the contour $\mathcal{L}$; this is the product denoted by $(\cdot, \cdot)$ in (7) and (8).

More generally, for any complex contour C and analytic functions Φ and Ψ, we can define a bilinear complex valued product (Φ, Ψ) by the formula

$$(\Phi, \Psi) = \int_C dz \Phi(z) \overline{\Psi(z)},$$

where $\overline{\Psi}(z) = \overline{\Psi(\bar{z})}$ is the natural generalization of a complex conjugate preserving the class of analytic functions. For the particular choice of contour (9) with $\theta = \frac{\pi}{2}$, we have the explicit expression

$$(\varphi, \psi) = \int_{-\infty}^{x_c} dx \varphi(x)\psi(x) + i \int_{x_c}^{\infty} dx \varphi(x)\psi(x), \tag{10}$$

where now $\varphi(x) = \Phi(z(x))$ and $\psi(x) = \overline{\Psi(z(x))}$ are smooth functions defined on the real axes that are singular at the point $x = x_c$. At the point x_c, we have

$$\partial_x \varphi|^a = i \partial_x \varphi|_a,$$
$$\partial_x \psi|^a = i \partial_x \psi|_a,$$

where for any function $f(x)$ we have introduced the notation $f|^{x_c} = \lim_{x \to x_c^+}$ and $f|_{x_c} = \lim_{x \to x_c^-} f(x)$. This set of functions clearly forms a vector space over the complex numbers, but it is an unusual complex vector space in several ways. Taking the complex conjugate of vectors in V brings us out of the space, the product, $(\cdot, \cdot)$, is not positive, (φ, φ) is in general a complex number, and this complex valued product is symmetric $(\varphi, \psi) = (\psi, \varphi)$. However, the space V equipped with the product (10) is the natural setting for working with resonant states; not only does it contain the resonant states as vectors but also the Hamiltonian (1) is self-adjoint on this space

$$(H\varphi, \psi) = (\varphi, H\psi).$$

This means that resonant states corresponding to different eigenvalues are orthogonal and this fact leads to the expansion (1). For self-adjoint operators in Hilbert space, one has a completeness theorem. We are not aware of the existence of such a result for the space V. In order to prove such a result for the space V, one has first to construct an appropriate topology and then use this topology as a starting point for a completeness theorem. This kind of investigation, which certainly is worthwhile to pursue mathematically, will however at best to lead to some kind of convergence that is weaker than the pointwise convergence which is our goal to investigate in the current paper. We will therefore not pursue these mathematical issues here.

Note that the relevance of this kind of mathematical structure for describing expansion into decaying states was noted already in 1938 by Kapur and Peierls while studying the dispersion formulas for nuclear reactions.[51]

III. NUMERICAL RESULTS

We have done extensive numerical tests of the resonant state expansions (7). For particular choices of functions, $f(x)$, and particular square wells, the expansion coefficients were calculated

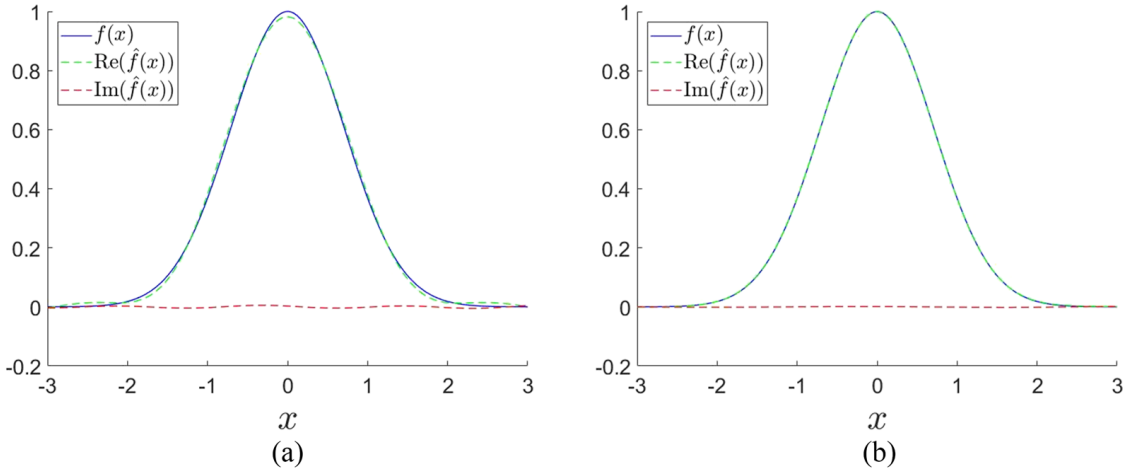


FIG. 2. The test function $f(x) = e^{-x^2}$ and the corresponding expansion $\hat{f}(x)$ using resonant states from A-series. The parameters are $V_0 = 2$, $d = 14$, $\varepsilon = 0.03$. Note that the function to be expanded in this and the following figure is well localized inside the potential well. (a) 100 terms used in the expansion. (b) 1000 terms used in the expansion.

using resonant states from the set of perturbed bound states and the A-series. The sample results displayed in Figs. 2 and 3 show what we find is generic behaviour. For this illustration, we used a Gaussian function and a Gaussian wave packet, and computed from 20 to 1000 terms taken from the set of perturbed bound states and from the A-series in the resonant state expansion thereby defining a function $\hat{f}(x) = \sum_n c_n \psi_{\omega_n}(x)$, which could possibly be different from our original f. We calculate numerically, the complex energy eigenvalues, ω_p, which are solutions of (6) and the coefficients $a_1 \cdots a_4$, ensuring the continuity of the resonant states at the points $x = -d, d$. From these pictures and many like them, our conjecture is that resonant state expansions based on the set of perturbed bound states and the A-series will converge point-wise for all reasonable initial data for the Schrödinger equation. The convergence appears to be fairly slow compared to the convergence of a regular Fourier series for the same functions. The slow convergence is, in particular, well illustrated by Fig. 3. The correspondence between the convergence and the pictures in Fig. 3 lies in the fact that if we increase the number of terms in the expansion, the difference between the two functions decreases.

In the remainder of this paper, we will prove that our conjecture is true and also investigate in detail how the rate of convergence depends on all parameters in the problem.

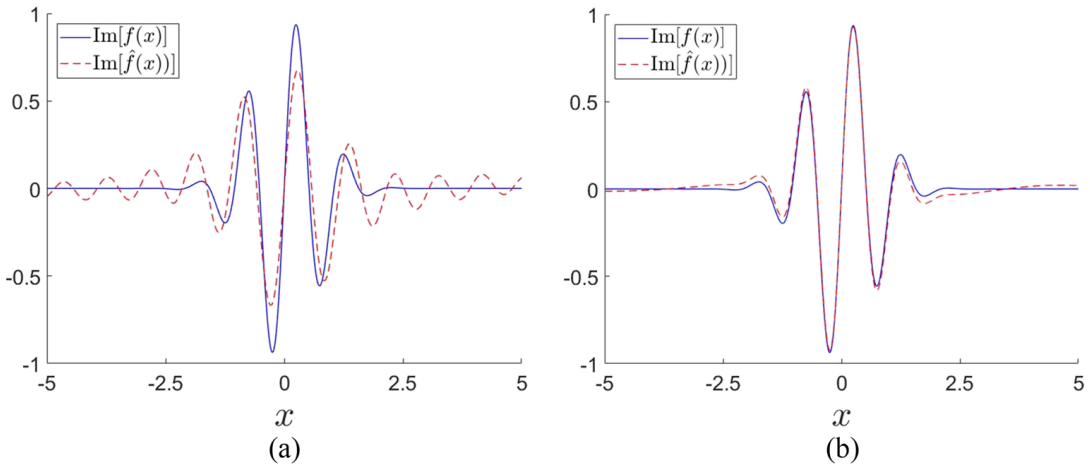


FIG. 3. The test function $f(x) = e^{-x^2} e^{i6x}$ and the corresponding expansion $\hat{f}(x)$ using resonant states from A-series. The parameters are $V_0 = 2$, $d = 14$, $\varepsilon = 0.03$. (a) 1000 terms used in the expansion. (b) 10086 terms used in the expansion.

IV. RESONANT STATES EXPANSION

The numerical tests in Sec. III suggest that the resonant state expansion (7) converges inside the well, although the rate of convergence is rather slow. In this chapter, we present a proof that confirms the pointwise convergence. Let us begin by noting that the Stark Hamiltonian (1) has no bound states. The continuous spectrum is found by imposing scattering boundary conditions on the Stark Hamiltonian where we have both incoming and outgoing waves at positive infinity and decaying outgoing waves at negative infinity.

The Stark Hamiltonian is an unbounded self-adjoint operator and as such it has an associated resolution of the identity.[52] In general, the spectral resolution contains an integration over point spectra and also absolute continuous and singularly continuous spectral components. For the Stark Hamiltonian in this paper, it has been proven[53–55] that both the point spectrum and the singularly continuous spectrum are empty and that the absolutely continuous spectrum is equal to the whole real line. Thus we have

$$\int_{-\infty}^{\infty} \psi_\omega(x)\psi_\omega(x')\mathrm{d}\omega = \delta(x-x'), \tag{11}$$

where the complex conjugate in (11) is missing because in our case the scattering states are real. For our particular Stark Hamiltonian (1), we find that the scattering states are given by

$$\psi_\omega(x) = \chi \begin{cases} \frac{2}{\pi^2 |\det \mathbf{M}(\omega)|}\mathrm{Ai}\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right), & x<-d, \\ \frac{2}{\pi |\det \mathbf{M}(\omega)|}\Big[(B_1'A_0 - B_1A_0')\mathrm{Ai}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right) & \\ \quad +(A_1A_0' - A_1'A_0)\mathrm{Bi}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right)\Big] & -d<x<d, \\ i\left(\frac{\overline{\det \mathbf{M}}(\omega)}{\det \mathbf{M}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^+\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right) & \\ -i\left(\frac{\det \mathbf{M}(\omega)}{\overline{\det \mathbf{M}}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^-\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right) & d<x, \end{cases}$$

where $\mu = -(2\varepsilon)^{\frac{1}{3}}$, $\mathrm{Ci}^\pm = \mathrm{Bi}(x) \pm i\mathrm{Ai}(x)$ and where $\overline{\det M}(\omega) = \overline{\det M(\overline{\omega})}$. The normalization constant $\chi = 2^{-\frac{2}{3}}\varepsilon^{-\frac{1}{6}}$ ensures that the multiplier of the delta functions on the right-hand side of (11) is one.

Since Ci^+ and Ci^- represent outgoing and incoming waves at positive infinity, there is a natural split of the scattering states in outgoing, $\psi_\omega^+(x)$, and incoming, $\psi_\omega^-(x)$, parts. These are

$$\psi_\omega^+(x) = \chi \begin{cases} i\left(\frac{\overline{\det \mathbf{M}}(\omega)}{\det \mathbf{M}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^+\left(\mu\left(d+\frac{\omega}{\varepsilon}\right)\right)\frac{\mathrm{Ai}\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right)}{\pi p(\omega)} & x<-d, \\ i\left(\frac{\overline{\det \mathbf{M}}(\omega)}{\det \mathbf{M}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^+\left(\mu\left(d+\frac{\omega}{\varepsilon}\right)\right)\frac{1}{p(\omega)} & \\ \Big[(B_1'A_0 - B_1A_0')\mathrm{Ai}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right) & \\ +(A_1A_0' - A_1'A_0)\mathrm{Bi}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right)\Big], & -d<x<d, \\ i\left(\frac{\overline{\det \mathbf{M}}(\omega)}{\det \mathbf{M}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^+\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right) & d<x, \end{cases} \tag{12}$$

$$\psi_\omega^-(x) = \chi \begin{cases} -i\left(\frac{\det \mathbf{M}(\omega)}{\overline{\det \mathbf{M}}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^-\left(\mu\left(d+\frac{\omega}{\varepsilon}\right)\right)\frac{\mathrm{Ai}\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right)}{\pi p(\omega)} & x<-d, \\ -i\left(\frac{\det \mathbf{M}(\omega)}{\overline{\det \mathbf{M}}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^-\left(\mu\left(d+\frac{\omega}{\varepsilon}\right)\right)\frac{1}{p(\omega)} & \\ \Big[(B_1'A_0 - B_1A_0')\mathrm{Ai}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right) & \\ +(A_1A_0' - A_1'A_0)\mathrm{Bi}\left(\mu\left(x+\frac{\omega+V_0}{\varepsilon}\right)\right)\Big] & -d<x<d, \\ -i\left(\frac{\det \mathbf{M}(\omega)}{\overline{\det \mathbf{M}}(\omega)}\right)^{\frac{1}{2}}\mathrm{Ci}^-\left(\mu\left(x+\frac{\omega}{\varepsilon}\right)\right) & d<x, \end{cases} \tag{13}$$

where we have defined $p(\omega) = (B_1'A_0 - B_1A_0')A_2 + (A_1A_0' - A_1'A_0)B_2$. Observe that by construction, we have

$$\psi_\omega(x) = \psi_\omega^+(x) + \psi_\omega^-(x).$$

We can use the completeness to split any functions in the span of $\{\psi_\omega\}$ into outgoing and incoming parts

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} \mathrm{d}s\, \delta(x-s) f(s) \\ &= \int_{-\infty}^{\infty} \mathrm{d}s \int_{-\infty}^{\infty} \mathrm{d}\omega\, \psi_\omega(x)\psi_\omega(s) \int_{-\infty}^{\infty} \mathrm{d}\omega'\, a(\omega')\psi_{\omega'}(s) \\ &= \int_{-\infty}^{\infty} \mathrm{d}\omega'\, a(\omega') \int_{-\infty}^{\infty} \mathrm{d}\omega\, \psi_\omega(x) \int_{-\infty}^{\infty} \mathrm{d}s\, \psi_\omega(s)\psi_{\omega'}(s), \end{aligned}$$

where $a(\omega') = \int_{-\infty}^{\infty} \mathrm{d}x\, \psi_{\omega'}(x) f(x)$ is the energy representation of $f(x)$. We now split the scattering state $\psi_{\omega'}(s)$ into outgoing and incoming parts using (12) and (13) and find that $f(x) = f^+(x) + f^-(x)$, where

$$f^\pm(x) = \int_{-\infty}^{\infty} \mathrm{d}\omega'\, a(\omega') \int_{-\infty}^{\infty} \mathrm{d}\omega\, \psi_\omega(x) \int_{-\infty}^{\infty} \mathrm{d}s\, \psi_\omega(s)\psi_{\omega'}^\pm(s).$$

Using special rules for anti-derivatives of Airy functions[56] and generalizing the approach used in Ref. 45, we get the following expressions for the outgoing and incoming parts of $f(x)$:

$$f^\pm(x) = \frac{1}{2} \int_{-\infty}^{\infty} \mathrm{d}\omega'\, a(\omega') \left[\mp i \frac{|\det \mathbf{M}(\omega')|}{p(\omega')} \right] \int_{-\infty}^{\infty} \mathrm{d}\omega\, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' \mp i\xi}. \tag{14}$$

Note that the inner integral in (14) has been regularized using a parameter ξ that will be removed at the end of our calculations by letting it approach zero from above.

We will now focus on the outgoing part of $f(x)$; the incoming part of $f(x)$ is treated in an entirely similar manner. For the outgoing part of $f(x)$, we observe that the integrand in the inner integral in the expression (14) defining $f^+(x)$

$$P_\xi(\omega') = \int_{-\infty}^{\infty} \mathrm{d}\omega\, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi} \tag{15}$$

has poles at the point $\omega = \omega' + i\xi$ in the upper half plane and at the zeroes for the functions $\det M(\omega)$ and $\overline{\det M}(\omega)$ in the lower and upper half plane.

We now rewrite the quantity $P_\xi(\omega')$ using the residue theorem on the contour Γ_R which is closed in the lower half plane. This contour is depicted in Fig. 4. The only poles of the integrand inside this contour are the zeroes of $\det M(\omega)$ and we thus have

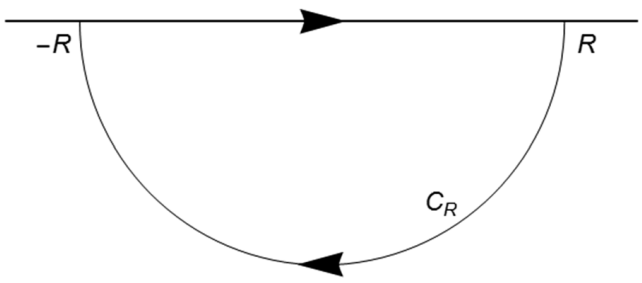


FIG. 4. Closed contour Γ_R in the lower complex frequency half plane, ω.

$$P_\xi^R(\omega') \equiv \int_{-R}^{R} \mathrm{d}\omega\, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi} \tag{16}$$
$$= \frac{2\pi}{i} \sum_j \mathrm{Res}\left(\psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi}, \omega_j\right) - \int_{C_R} \mathrm{d}\omega\, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi}.$$

Observe that P_ξ^R converges to P_ξ in (15) as $R \to \infty$. From (16), it is evident that if the integral over the contour C_R vanishes in the limit when R approaches infinity

$$\lim_{R\to\infty} \int_{C_R} \mathrm{d}\omega\, \psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi} = 0, \tag{17}$$

then the $P_\xi(\omega')$ will be a sum of residues

$$P_\xi(\omega') = \frac{2\pi}{i} \sum_j \mathrm{Res}\left(\psi_\omega(x) \frac{\psi_\omega(d)}{\omega - \omega' - i\xi}, \omega_j\right),$$

and from this it is then a simple matter to show that the regulator ξ can be safely removed and that formula (14) reduces to a resonant state expansion for the outgoing part of f

$$f^+(x) = \sum_j c_j \psi_j(x), \tag{18}$$

which is what we wanted to prove.

Thus the proof of convergence for the resonant state expansion for any given x is reduced to showing that the limit (17) holds for that x. In order to investigate this limit, we parameterize the circular arc C_R in the lower half plane using $\omega = Re^{i\theta}$ and use standard formulas for the Airy functions[57] in the large argument limit to derive asymptotic formulas for the scattering states ψ_ω in the limit of large R. Because of the well known Stokes phenomenon, the resulting asymptotic expressions for the scattering states are different for the two angular sectors $-\frac{2\pi}{3} < \theta < 0$ and $-\pi < \theta < -\frac{2\pi}{3}$. We find that in the second sector, the scattering state $\psi_\omega(x)$ decays exponentially for all x and thus this angular sector gives no contribution to the limit (17). In the first angular sector, we have the following asymptotic expression for the scattering state:

$$\psi_\omega(x) = \chi \begin{cases} \frac{2(\kappa R)^{\frac{1}{4}}}{i\pi^{\frac{1}{2}}\sigma^{\frac{1}{2}}} e^{-i\beta\varrho R^{\frac{1}{2}}} e^{\varpi x R^{\frac{1}{2}}} & x < -d, \\ \frac{2(\kappa R)^{\frac{1}{4}}}{i\pi^{\frac{1}{2}}\sigma^{\frac{1}{2}}} e^{-i(\beta+\sigma)\varrho R^{\frac{1}{2}}} e^{\varpi R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}\right)} & -d < x < d, \\ -\frac{\sigma^{\frac{1}{2}}}{2(\kappa R)^{\frac{3}{4}}\pi^{\frac{1}{2}}} e^{i\varrho R^{\frac{1}{2}}(\sigma-3\beta)} e^{-\varpi R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}\right)} & \\ i\frac{2(\kappa R)^{\frac{1}{4}}}{\pi^{\frac{1}{2}}\sigma^{\frac{1}{2}}} e^{-i\beta\varrho R^{\frac{1}{2}}} e^{\varpi x R^{\frac{1}{2}}} - i\frac{\sigma^{\frac{1}{2}}}{2\pi^{\frac{1}{2}}(\kappa R)^{\frac{3}{4}}} e^{i\beta\varrho R^{\frac{1}{2}}} e^{-\varpi x R^{\frac{1}{2}}} & d < x, \end{cases}$$

where ϖ, ϱ are complex numbers depending on θ and $\mu = (2\varepsilon)^{-\frac{2}{3}}$, $\beta = 2\mu\varepsilon d$, $\sigma = 2\mu V_0$, and $\kappa = 2\mu e^{i\theta}$. Let us start analyzing the various regions in x. For $x < -d$, we find that the limit is zero and that for $x > d$ it does not exist. Thus the resonant state expansion converges for $x < -d$, but not for $x > d$. This is consistent with what was found for the Dirac delta potential in Ref. 45 if we imagine approaching the Dirac delta function using a sequence of square wells where V_0 approach minus infinity, while d approach zero.

For x inside the square well, where $-d < x < d$, the asymptotic expression for the integrand in (17) is found to be

$$\begin{aligned}\psi_\omega(x)\frac{\psi_\omega(d^+)}{\omega-\omega'\mp i\xi} \approx -\frac{\chi^2}{\omega-\omega'\mp i\xi}\Bigg(&\frac{1}{\pi(\kappa R)^{\frac{1}{2}}}\\ &e^{-i\sigma\varrho_r R^{\frac{1}{2}}}e^{(2\mu)^{\frac{1}{2}}\left|\sin\left(\frac{1}{2}\theta\right)\right|R^{\frac{1}{2}}\left[\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}-d\right)-\sigma\right]}e^{i\varpi_i R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}-d\right)}\\ &-\frac{\sigma}{8\pi(\kappa R)^{\frac{3}{2}}}\\ &e^{i\varrho_r R^{\frac{1}{2}}(\sigma-2\beta)}e^{(2\mu)^{\frac{1}{2}}\left|\sin\left(\frac{1}{2}\theta\right)\right|R^{\frac{1}{2}}\left[-\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}+d\right)+(\sigma-2\beta)\right]}e^{-i\varpi_i R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}+d\right)}\\ &-\frac{4(\kappa R)^{\frac{1}{2}}}{\pi\sigma}\\ &e^{-i(2\beta+\sigma)\varrho_r R^{\frac{1}{2}}}e^{(2\mu)^{\frac{1}{2}}\left|\sin\left(\frac{1}{2}\theta\right)\right|R^{\frac{1}{2}}\left[\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}+d\right)-(2\beta+\sigma)\right]}e^{i\varpi_i R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}+d\right)}\\ &+\frac{1}{\pi(\kappa R)^{\frac{1}{2}}}\\ &e^{i\varrho_r R^{\frac{1}{2}}(\sigma-4\beta)}e^{(2\mu)^{\frac{1}{2}}\left|\sin\left(\frac{1}{2}\theta\right)\right|R^{\frac{1}{2}}\left[-\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}-d\right)+(\sigma-4\beta)\right]}e^{-i\varpi_i R^{\frac{1}{2}}\left(x+\frac{V_0}{\varepsilon}-d\right)}\Bigg).\end{aligned} \tag{19}$$

If this expression decays exponentially in the limit $R\to\infty$, the resonant state expansion will converge inside the square well. In (19), we have four different terms that need to be checked separately. The first two exponentials in (19) decay if

$$\begin{aligned}&\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}-d\right)-\sigma<0, && -\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}+d\right)+(\sigma-2\beta)<0,\\ &\Downarrow && \Downarrow\\ &d>x, && -3d<x.\end{aligned} \tag{20}$$

These conditions are satisfied inside the well, $-d < x < d$.

The second two parts of (19) converge to zero if

$$\begin{aligned}&\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}+d\right)-(2\beta+\sigma)<0, && -\varepsilon 2\mu\left(x+\frac{V_0}{\varepsilon}-d\right)+(\sigma-4\beta)<0,\\ &\Downarrow && \Downarrow\\ &d>x, && -3d<x,\end{aligned} \tag{21}$$

which are also satisfied inside the well. Note that the depth of the well V_0 disappears from the inequalities (20) and (21). Thus we reach the surprising conclusion that the resonant state expansion converges for all x inside the square well *independently* of the depth of the well.

V. RATE OF CONVERGENCE

In this section, we directly investigate the rate of convergence of the resonant state expansion (18) found to converge for $-\infty < x < d$ in Sec. IV. As we have discussed in Sec. II of this paper, the energy eigenvalues can be categorized into three groups. The first is the finite set of perturbed bound states, the second set is the A-series, which are located in the fourth quadrant in the complex plane, and the third is the C-series located in the third quadrant. The perturbed set of bound states is finite and thus does not contribute to the rate of convergence, but the A-series and the C-series are both infinite and they do contribute to the rate of convergence. First, we look at the part of the sum corresponding to the eigenvalues from the A-series and after that briefly sum up the results for the C-series. In order to do this, we need an asymptotic expression for resonant state eigenvalues from the A-series. In principle, we find this expression by first finding the leading order contribution of the determinant function, here denoted by $\det M^\infty(\omega)$, for $|\omega|\to\infty$ and then finding the leading order

expression for the roots of the equation $\det M^{\infty}(\omega) = 0$ that define the A-series. The final formula for the resonant eigenvalues ω_p is simple and is reproduced below in (22), but even though nothing beyond the standard asymptotic formulas for Airy functions are required for this derivation, the details are technical and will not be presented (this remark applies to most of the explicit formulas derived in this section). The formula for ω_p is

$$\omega_p \approx \gamma^{-1}\left(\frac{3\pi}{2}\xi_p\right)^{\frac{2}{3}}, \tag{22}$$

where

$$\xi_p = p - \frac{i}{3\pi}\ln(p) + i\rho + \frac{i}{2\pi}\ln\left(\left|\sin\left(c_2 p^{\frac{1}{3}}\right)\right|\right) - \frac{1}{2\pi}h(p), \tag{23}$$

where $p \gg 1$ is the index of the eigenvalues. The parameters appearing in (22) and (23) are given by

$$\begin{aligned}
\gamma &= 2(2\varepsilon)^{-\frac{2}{3}},\\
c_1 &= \frac{V_0\gamma}{2^{\frac{1}{3}}3^{\frac{2}{3}}\pi^{\frac{8}{3}}},\\
c_2 &= 2d(3\pi\varepsilon)^{\frac{1}{3}},\\
\rho &= \frac{\ln\left(\pi^2 c_1\right)}{2\pi},\\
h(p) &= \arg\left(\sin\left(c_2 p^{\frac{1}{3}}\right)\right).
\end{aligned}$$

In Fig. 5, we compare the asymptotic formula for the resonant eigenvalues with an arbitrary precision numerical calculation of the eigenvalues. As we can see, our asymptotic formula is highly accurate.

We will now use the asymptotic formula for the location of the resonant eigenvalues of the A-series to get an estimate of the rate of convergence of the resonant state expansion for functions f whose support are inside the square well. The rate of expansion in general depends on x, but here we will focus on the point $x = 0$. The resonant state expansion for f then becomes the following numerical series:

$$\sum_p \frac{b_p}{N_p},$$

where

$$\begin{aligned}
b_p &= \psi_{\omega_p}(0)\int_{-d}^{d} f(x)\psi_{\omega_p}(x)\mathrm{d}x,\\
N_p &= \int_{-d}^{d} \psi_{\omega_p}(x)^2\mathrm{d}x.
\end{aligned}$$

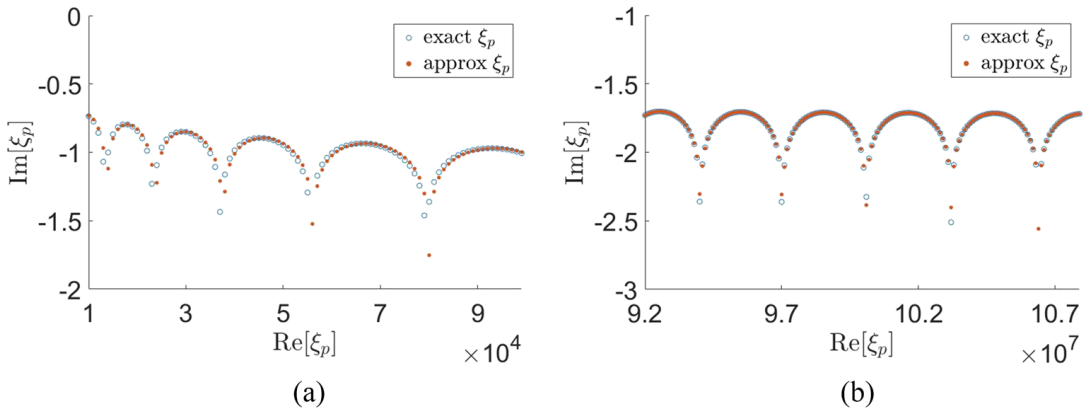


FIG. 5. Comparison between a high precision numerical calculation of the pole positions and the asymptotic formula (22) for two different ranges of the pole index.

Using the asymptotic formula for the resonant eigenvalues from the A-series (22), we get the following asymptotic expressions for b_p and N_p:

$$\begin{aligned} b_p &\approx \pi^{-1}\left(\frac{3\pi}{2}\right)^{-\frac{1}{3}} p^{-\frac{1}{3}}\{a_3\cos(g_1(p)+\frac{\pi}{4})+a_2\sin(g_1(p)+\frac{\pi}{4})\} \\ &\quad \int_{-d}^{d} dx f(x)\{a_3\cos(g_1(p)+xg_2(p)+\frac{\pi}{4})+a_2\sin(g_1(p)+xg_2(p)+\frac{\pi}{4})\}, \\ N_p &\approx \frac{a_1^2}{\gamma\varepsilon}\pi^{-1}\left(\frac{3\pi p}{2}\right)^{\frac{1}{3}}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} g_1(p) &= \frac{2}{3}\mu^{\frac{3}{2}}\xi_p + \mu^{\frac{1}{2}}p^{\frac{1}{3}}\gamma V_0, \\ g_2(p) &= \mu^{\frac{1}{2}}\gamma\varepsilon p^{\frac{1}{3}}. \end{aligned}$$

Recall that $(a_1, \ldots, a_4)$ is a vector that spans the null space of the matrix $M_p = M(\omega_p)$. In general, this vector also depends on p. However, in the asymptotic range, when $p \gg 1$ the matrix M_p simplifies in such a way that the null-space vector can be chosen to be independent of p.

In Fig. 6, where we have chosen $f(x)$ to be a Gaussian function, we see how accurate the asymptotic expressions are. They are showing two sets of points, the absolute value of the numerically calculated value of the term, and the value we get from the asymptotic formulas. For smaller values of p, the approximation seems to fluctuate around the exact values and does not seem to catch up on the most extreme swings and breaks. However, as we move forward to higher values, the asymptotic formula eventually catches all the wild swings of the terms in the series.

We will now derive some analytic estimates for the rate of convergence using the asymptotic expressions (24) for b_p and N_p. First we note that for the special case of a function that is constant equal to one inside the well and zero outside, we can find an analytic expression for b_p. Let us call this

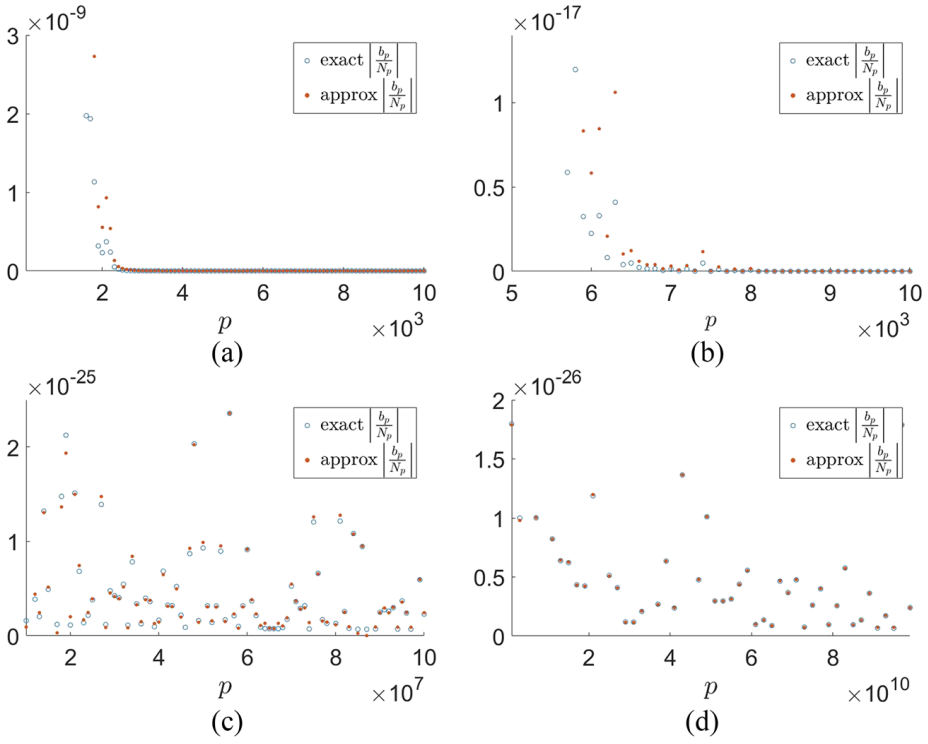


FIG. 6. The absolute value of exact and asymptotic ratios $\left|\frac{b_p}{N_p}\right|$ for different ranges of the eigenvalue index p. In these pictures, we used $V_0 = 2$, $d = 14$, $\varepsilon = 0.03$, $x_c = 20$, and $\theta = \frac{\pi}{2}$.

quantity as b_p^0. The details of the calculation are relegated to the Appendix. The resulting expression is

$$\left|b_p^0\right| \approx \frac{\alpha}{V_0\left|\cos\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|},$$

where α is a numerical factor that does not depend on p or V_0. Now let $f(x)$ be a function that is zero at $x = \pm d$ and is n times continuously differentiable. From formula (24), it is clear that each time the trigonometric sum under the integral sign is integrated, a factor of $g_2(p)$ appears in the denominator. It is also clear that any number of integrations will preserve the form of the trigonometric term in the integrand up to sign after an even number of integrations. Thus, using integration by parts, $n = 2m$ times will give us

$$\left|\frac{b_p}{N_p}\right| \lesssim p^{-\frac{n+1}{3}} \frac{\beta M_n}{V_0\left|\cos\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|},$$

where β is another numerical factor not depending on p or V_0 and $\left|f^{(n)}(x)\right| \le M_n$ for x inside the well. This formula shows explicitly how the size of the terms in the resonant state series for a function depends on the smoothness of that function and the depth of the well. Note that the depth of the well cancels if we divide two consecutive terms to get the rate of convergence, and we can conclude that the absolute size of the terms depends on the depth of the well as V_0^{-1}, but that the rate of convergence is independent of V_0.

So far we have only considered convergence of the part of the resonant state expansion that comes from the A-series. However, detailed investigations of the terms in the series coming from the C-series have shown that they are decaying exponentially in the resonant state index, p, and thus have no influence on the question of convergence for the series as a whole. These investigations, which follow the same approach as for the A-series, will not be presented here.

VI. CONCLUSION

We have investigated the completeness of the Stark resonant states in a system with a square-well potential and a homogeneous external field. Our conclusion is that when the field pulls the particle to the right, the resonant state expansions converge pointwise to the left of the well as well as inside the well. Interestingly, we have found that the existence of convergence is independent of the depth of the potential well, V_0. In other words, no matter how shallow the well might be, there is a convergent resonant-state expansion.

We have also derived formulas that show how the *rate of convergence* depends on the smoothness of the function being expanded. These formulas indicate that the rate of convergence is also independent of the depth of the well, but the absolute size of the terms grows like V_0^{-1}. Thus, for any given target accuracy, a smaller potential depth means that more terms have to be included in the series.

Taking into account the similar nature of the convergence results for the Dirac delta potential[45] and for the square potential treated in this paper, we conjecture a similar result for a general potential of compact support. Such a potential can be approximated by a finite set of conjoined square wells, as illustrated in Fig. 7, and we expect that our approach can be generalized to this setting using a transfer-matrix technique. In the limit when the set of conjoined square wells approach the smooth potential, there will inevitably be square wells that are arbitrarily shallow but we expect, based on the result derived in this paper, that this will not destroy the convergence. In fact, this same result leads us to conjecture pointwise convergence even for non-compact potentials.

Finally let us note that there are issues that invite further investigations. Our resonant-state expansion is constructed separately for the incoming and outgoing portions of the given wave function, and there is certain similarity here with the Green's function approach in which the imposed boundary conditions can select the outgoing waves. Naturally, what constitutes the outgoing and incoming parts of the given function in the region far from the origin (in the direction of the field) is given and unique, and corresponds to the decomposition into Ci^+ and Ci^-. However, the in- and out-split is not

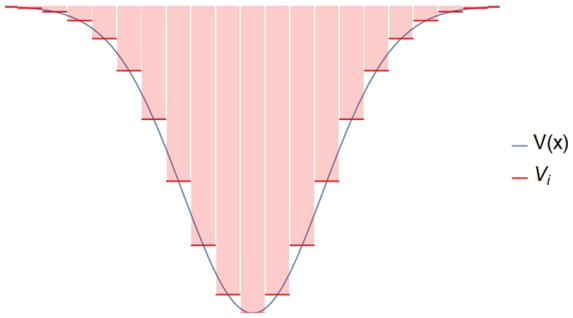


FIG. 7. A continuous potential $V(x)$ with a suitable discretization V_i.

necessarily unique inside the potential well. It is an open question of how such a degree of freedom affects the resulting expansion.

ACKNOWLEDGMENTS

The authors are thankful for support from the Department of mathematics and statistics at the Arctic University of Norway, from the Arizona Center for Mathematical Sciences at the University of Arizona, and for the support from the Air Force Office for Scientific Research under Grant No. FA9550-16-1-0088. Miro Kolesik was supported by the Air Force Office for Scientific Research under Grant No. FA9550-18-1-0183.

APPENDIX: EVALUATION OF b_p^0

The integral in (24) is computed as

$$\begin{aligned}
&\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\int_{-d}^{d}dx f(x)\{a_3\cos(g_1(p)+xg_2(p)+\frac{\pi}{4})+a_2\sin(g_1(p)+xg_2(p)+\frac{\pi}{4})\}\\
&=\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\frac{1}{g_2(p)}\int_{g_1(p)-dg_2(p)+\frac{\pi}{4}}^{g_1(p)+dg_2(p)+\frac{\pi}{4}}[a_2\sin(y)+a_3\cos(y)]\mathrm{d}y\\
&=\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\frac{1}{g_2(p)}[a_3\sin(y)-a_2\cos(y)]\Big|_{y=g_1(p)-dg_2(p)+\frac{\pi}{4}}^{y=g_1(p)+dg_2(p)+\frac{\pi}{4}}\\
&=\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\frac{1}{g_2(p)}\Big[a_3\sin\Big(g_1(p)+dg_2(p)+\frac{\pi}{4}\Big)-a_2\cos\Big(g_1(p)+dg_2(p)+\frac{\pi}{4}\Big)\\
&-a_3\sin\Big(g_1(p)-dg_2(p)+\frac{\pi}{4}\Big)+a_2\cos\Big(g_1(p)-dg_2(p)+\frac{\pi}{4}\Big)\Big]\\
&=\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\frac{2\sin(dg_2(p))}{g_2(p)}\Big[a_3\cos\Big(g_1(p)+\frac{\pi}{4}\Big)+a_2\sin\Big(g_1(p)+\frac{\pi}{4}\Big)\Big].
\end{aligned}\tag{A1}$$

Using (A1) and the asymptotic forms of the Airy functions, the coefficients b_p in (24) become

$$\begin{aligned}
b_p^0&\approx\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\Big[a_2\sin\Big(g_1(p)+\frac{\pi}{4}\Big)+a_3\cos\Big(g_1(p)+\frac{\pi}{4}\Big)\Big]\\
&\quad\pi^{-\frac{1}{2}}\left(\frac{3\pi}{2}\right)^{-\frac{1}{6}}p^{-\frac{1}{6}}\frac{2\sin(ag_2(p))}{g_2(p)}\Big[a_3\cos\Big(g_1(p)+\frac{\pi}{4}\Big)+a_2\sin\Big(g_1(p)+\frac{\pi}{4}\Big)\Big]\\
&=\pi^{-1}\left(\frac{3\pi}{2}\right)^{-\frac{1}{3}}p^{-\frac{1}{3}}\frac{2\sin(ag_2(p))}{g_2(p)}\Big[a_2\sin\Big(g_1(p)+\frac{\pi}{4}\Big)+a_3\cos\Big(g_1(p)+\frac{\pi}{4}\Big)\Big]^2.
\end{aligned}\tag{A2}$$

We take a further look at the trigonometric terms. After writing them in exponential forms, we find that one of the two exponentials can be neglected in the limit $p \gg 1$. Therefore we get

$$\sin\left(g_1(p)+\frac{\pi}{4}\right) \approx \frac{1}{2i}(-1)^p p^{\frac{1}{3}}\left(\pi^2 c_1\right)^{-\frac{1}{2}}\left|\sin\left(c_3 p^{\frac{1}{3}}\right)\right|^{-\frac{1}{2}} e^{i\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)},$$
$$\cos\left(g_1(p)+\frac{\pi}{4}\right) \approx \frac{1}{2}(-1)^p p^{\frac{1}{3}}\left(\pi^2 c_1\right)^{-\frac{1}{2}}\left|\sin\left(c_3 p^{\frac{1}{3}}\right)\right|^{-\frac{1}{2}} e^{i\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)}. \tag{A3}$$

Using (A3), we rewrite (A2) as

$$\begin{aligned}
b_p^0 \approx & \pi^{-1}\left(\frac{3\pi}{2}\right)^{-\frac{1}{3}} p^{-\frac{1}{3}} \frac{2\sin\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}{g_2(p)}\left[a_2\frac{1}{2i}(-1)^p p^{\frac{1}{3}}\left(\pi^2 c_1\right)^{-\frac{1}{2}}\left|\sin\left(c_3 p^{\frac{1}{3}}\right)\right|^{-\frac{1}{2}}\right. \\
& e^{i\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)} \\
& \left. + a_3\frac{1}{2}(-1)^p p^{\frac{1}{3}}\left(\pi^2 c_1\right)^{-\frac{1}{2}}\left|\sin\left(c_3 p^{\frac{1}{3}}\right)\right|^{-\frac{1}{2}} e^{i\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)}\right]^2 \\
= & \pi^{-1}\left(\frac{3\pi}{2}\right)^{-\frac{1}{3}} p^{-\frac{1}{3}} \frac{2\sin\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}{\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma\varepsilon} \\
& \left(\frac{1}{2}(-1)^p p^{\frac{1}{3}}\left(\pi^2 c_1\right)^{-\frac{1}{2}}\left|\sin\left(c_3 p^{\frac{1}{3}}\right)\right|^{-\frac{1}{2}} e^{i\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)}\right)^2\left(\frac{a_2}{i}+a_3\right)^2 \\
= & \left(\frac{3\pi}{2}\right)^{-\frac{1}{3}} p^{-\frac{1}{3}} \frac{\sin\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}{2(3\pi\varepsilon p)^{\frac{1}{3}}} p^{\frac{2}{3}}\left(\pi^3 c_1\right)^{-1}\left|\sin\left(2d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|^{-1} \\
& e^{i2\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)}\left(\frac{a_2}{i}+a_3\right)^2 \\
= & \left(\frac{3\pi}{2}\right)^{-\frac{1}{3}} \frac{\sin\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}{2(3\pi\varepsilon)^{\frac{1}{3}}}\left(\pi^3 c_1\right)^{-1}\left|\sin\left(2d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|^{-1} \\
& e^{i2\left(\mu^{\frac{1}{2}} p^{\frac{1}{3}}\gamma V_0 - \frac{1}{2}h(p)+\frac{\pi}{4}\right)}\left(\frac{a_2}{i}+a_3\right)^2.
\end{aligned}$$

Thus

$$\begin{aligned}
\left|b_p^0\right| &\approx \frac{\alpha}{V_0}\left|\frac{\sin\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}{\sin\left(2d(3\pi\varepsilon p)^{\frac{1}{3}}\right)}\right| \\
&= \frac{\alpha}{V_0\left|\cos\left(d(3\pi\varepsilon p)^{\frac{1}{3}}\right)\right|},
\end{aligned}$$

where

$$\alpha = \frac{\varepsilon^{\frac{1}{3}}}{2^{\frac{5}{3}}\pi}|(a_3 - i\,a_2)|^2.$$

[1] G. A. Gamow, "Zur quantentheorie des atomkernes," Z. Phys. **51**, 204–212 (1928).
[2] G. A. Gamow, "Zur quantentheorie des atomzertrummerung," Z. Phys. **52**, 510–515 (1928).
[3] R. W. Gurney and E. U. Condon, "Quantum mechanics and radioactive disintegration," Phys. Rev. **33**, 127–140 (1929).
[4] R. W. Gurney and E. U. Condon, "Quantum mechanics and radioactive disintegration," Nature **122**, 439 (1928).
[5] J. J. Thomson, "On electrical oscillations and the effects produces by the motion of an electric sphere," Proc. London Math. Soc. **s1-15**, 197–218 (1884).
[6] J. A. J. F. Siegert, "On the derivation of the dispersion formula for nuclear reactions," Phys. Rev. **56**, 750–752 (1939).
[7] R. E. Peierls, "Complex eigenvalues in scattering theory," Proc. R. Soc. London **253**, 16–36 (1959).
[8] K. J. Le Couteur, "The structure of a non-relativistic s-matrix," Proc. R. Soc. London **256**, 115–127 (1960).
[9] J. Humblet, "Theory of nuclear reactions," Nucl. Phys. **26**, 529–578 (1961).
[10] T. Goto, "On the unstable states in quantum field theory," Prog. Theor. Phys. **21**(1), 1–17 (1959).
[11] P. Lin, "Completeness relations and the resonant state expansions," Phys. Rev. C **47**, 1903 (1993).

[12] G. Garcia-Calderon, "An expansion of continuum wave functions in terms of resonant states," Nucl. Phys. A **261**, 130 (1976).
[13] V. N. Ostrovsky, O. I. Tolstikhin, and H. Nakamura, "Siegert pseudo state formulation of scattering theory: One-channel case," Phys. Rev. A **58**, 2077 (1998).
[14] O. I. Tolstikhin, "Siegert-state expansion for nonstationary systems: Coupled equations in the one-channel case," Phys. Rev. A **73**, 062705 (2006).
[15] O. I. Tolstikhin, "Siegert-state expansion for nonstationary systems. IV. Three-dimensional case," Phys. Rev. A **77**, 032712 (2008).
[16] T. Berggren, "On the use of resonant states in eigenfunction expansions of scattering and reaction amplitudes," Nucl. Phys. A **109**, 265–287 (1968).
[17] R. M. More, "Theory of decaying states," Phys. Rev. A **4**(5), 1782–1790 (1971).
[18] G. Garcia-Calderon and R. Peierls, "Resonant states and their uses," Nucl. Phys. A **265**, 443–460 (1976).
[19] T. Bergren and P. Lin, "Resonant state expansion of the resolvent," Phys. Rev. C **47**, 768 (1993).
[20] J. von Neumann, *The Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1932).
[21] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford at the Clarendon Press, 1930).
[22] I. M. Gelfand and N. Y. Vilenkin, *Generalized Functions* (Academic Press, 1964), Vol. 4.
[23] D. C. Brody, "Boirthogonal quantum mechanics," J. Phys. A: Math. Theor. **47**, 035305 (2013).
[24] N. Moiseyev, *Non-Hermitian Quantum Mechanics* (Cambridge University Press, 2011).
[25] G. Garcia-Calderon, A. Mattar, and J. Villavicencio, "Hermitian and non-Hermitian formulations of the time evolution of quantum decay," Phys. Scr. **T151**, 014076 (2012).
[26] N. H. K. Sasada and G. Ordonez, "Resonant spectrum analysis of the conductance of an open quantum system and three types of fano parameter," Jpn. Phys. Soc. **80**, 104707 (2011).
[27] Gonzalo Ordonez and N. Hatano, "The arrow of time in quantum systems and dynamical breaking of the resonance-anti-resonance symmetry," J. Phys. A: Math. Theor. **50**, 405304 (2017).
[28] N. Hatano and G. Ordonez, "Time-reversal symmetric resolution of unit without background integrals in open quantum systems," J. Math. Phys. **55**, 122106 (2014).
[29] N. Hatano, "Probabilistic interpretation of resonant states," Pramana **73**, 553 (2009).
[30] J. Rosenkrantz de Lasson, P. T. Kristensen, J. Mørk, and N. Gregersen, "Roundtrip matrix method for calculating the leaky resonant modes of open nanophotonic structures," J. Opt. Soc. Am. A **31**(10), 2142–2151 (2014).
[31] P. T. Kristensen and S. Hughes, "Modes and mode volumes of leaky optical cavities and plasmonic nanoresonators," ACS Photonics **1**(1), 2–10 (2014).
[32] E. F. Franchimon, K. R. Hiremath, R. Stoffer, and M. Hammer, "Interaction of whispering gallery modes in integrated optical microring or microdisk circuits: Hybrid coupled mode theory model," J. Opt. Soc. Am. B **30**(4), 1048–1057 (2013).
[33] A. Settimi, S. Severini, and B. J. Hoenders, "Quasi-normal-modes description of transmission properties for photonic bandgap structures," J. Opt. Soc. Am. B **26**(4), 876–891 (2009).
[34] R.-C. Ge, J. F. Young, and S. Hughes, "Quasi-normal mode approach to the local-field problem in quantum optics," Optica **2**(3), 246–249 (2015).
[35] F. Yang, H. Liu, H. Jia, and Y. Zhong, "Analytical description of quasi-normal mode in resonant plasmonic nano cavities," J. Opt. **18**(3), 035003 (2016).
[36] O. I. Tolstikhin, V. N. Ostrovsky, and H. Nakamura, "Siegert pseudo-states as a universal tool: Resonances, *S* matrix, green function," Phys. Rev. Lett. **79**, 2026–2029 (1997).
[37] *Resonances the Unifying Route Towards the Formulation of Dynamical Processes Foundations and Applications in Nuclear, Atomic and Molecular Physics*, Lecture Notes in Physics, edited by E. Brändas and N. Elander (Erkki Brändas, 1987).
[38] R. de la Madrid, G. García-Calderón, and J. Gonzalo Muga, "Resonant expansions in quantum mechanics," Czech. J. Phys. **55**, 1141 (2005).
[39] N. Moiseyev, "Quantum theory of resonances: Calculating energies, widths and cross-sections by complex scaling," Phys. Rep. **302**(5-6), 212–293 (1998).
[40] A. Teleki, P. Jakobsen, J. V. Moloney, M. Kolesik, J. M. Brown, and E. M. Wright, "Metastable electronic states and nonlinear response for high-intensity optical pulses," Optica **1**(5), 323 (2014).
[41] A. Bahl, J. K. Wahlstrand, S. Zahedpour, H. M. Milchberg, and M. Kolesik, "Nonlinear optical polarization response and plasma generation in noble gases: Comparison of metastable-electronic-state-approach models to experiments," Phys. Rev. A **96**, 043867 (2017).
[42] A. Couairon, E. Brambilla, T. Corti, D. Majus, O. J. Ramirez-Gongora, and M. Kolesik, "Practitioner's guide to laser pulse propagation models and simulation," Eur. Phys. J.: Spec. Top. **199**(1), 5–76 (2011).
[43] E. Lorin, S. Chelkowski, and A. Bandrauk, "Maxwell-Schrödinger-plasma (MASP) model for laser-molecule interactions: Towards an understanding of filamentation with intense ultrashort pulses," Physica D **241**(12), 1059–1071 (2012).
[44] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon, *Schrödinger Operators* (Springer-Verlag, 1987).
[45] J. M. Brown, P. Jakobsen, A. Bahl, J. V. Moloney, and M. Kolesik, "On the convergence of quantum resonant-state expansion," J. Math. Phys. **57**, 032105 (2016).
[46] W. Elberfeld and M. Kleber, "Tunneling from an ultrathin quantum well in a strong electrostatic field: A comparison of different methods," Z. Phys. B: Condens. Matter **73**(1), 23–32 (1988).
[47] S. Geltman, "Ionisation dynamics of a model atom in an electrostatic field," J. Phys. B: At. Mol. Phys. **11**(19), 3323–3337 (1978).
[48] A. Teleki, E. M. Wright, and M. Kolesik, "Microscopic model for the higher-order nonlinearity in optical filaments," Phys. Rev. A **82**, 065801 (2010).
[49] A. Emmanouilidou, "Stark and field-born resonances of an open square well in a static external electric field," J. Chem. Phys. **122**, 194101 (2005).
[50] B. Gyarmati and T. Vertse, "On the normalization of gamow functions," Nucl. Phys. A **160**, 523–528 (1971).

[51] P. L. Kapur and R. Peierls, "The dispersion formula for nuclear reactions," Proc. R. Soc. London **166**, 277–295 (1939).
[52] N. Dunford and J. T. Schwartz, *Linear Operators. Part II. Spectral Theory* (Wiley, 1963), Theorem 3 p. 1192.
[53] I. W. Herbst, "Unitary equivalence of Stark Hamiltonians," Math. Z. **155**, 55–70 (1977).
[54] J. E. Avron and I. W. Herbst, "Spectral and scattering theory of Schrodinger operators related to the Stark effect," Commun. Math. Phys. **52**, 239–254 (1977).
[55] B. Simon, "Phase space analysis of simple scattering systems: Extensions of some work of Enss," Duke Math. J. **46**(1), 119–168 (1979).
[56] O. Valee and M. Soares, *Airy Functions and Applications to Physics* (Imperial College Press, 2004).
[57] NIST Digital Library of Mathematical Functions, http://dlmf.nist.gov/, Release 1.0.18 of March 3, 27, edited by F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, and B. V. Saunders.

4 Paper 2

Constructing a partially transparent computational boundary for UPPE using leaky modes

Journal of Mathematical Physics, 60(08):083505, 2019

Journal of Mathematical Physics ARTICLE scitation.org/journal/jmp

Constructing a partially transparent computational boundary for UPPE using leaky modes

Cite as: J. Math. Phys. 60, 083505 (2019); doi: 10.1063/1.5099193
Submitted: 8 April 2019 • Accepted: 25 July 2019 •
Published Online: 19 August 2019

David Juhasz and Per Kristen Jakobsen

AFFILIATIONS

Department of Mathematics and Statistics, the Arctic University of Norway, 9019 Tromsø, Norway

ABSTRACT

In this paper, we introduce a method for creating a transparent computational boundary for the simulation of unidirectional propagation of optical beams and pulses using leaky modes. The key element of the method is the introduction of an artificial-index material outside a chosen computational domain and utilization of the quasi-normal modes associated with such artificial structure. The method is tested on the free space propagation of TE electromagnetic waves. By choosing the material to have appropriate optical properties, one can greatly reduce the reflection at the computational boundary. In contrast to the well-known approach based on a perfectly matched layer, our method is especially well suited for spectral propagators.

I. INTRODUCTION

Treatment of domain boundaries in numerical simulations, especially in the solution of partial differential equations, presents a long-standing problem. While powerful methods have been developed for certain situations, they often introduce significant additional complexity and computational overhead. The perfectly matched layer (PML)[1] approach stands as a prime example of methods that work extremely well in situations where a transparent boundary is meant to mimic a connection of the given "computational box" to an infinite outside space. Indeed, PML-based methods are routinely employed for wave-propagation simulation, for example, in finite-difference Maxwell solvers[2] and in beam-propagation simulation.[3]

Nevertheless, there are applications for which good boundary treatments are still lacking. For example, in extreme nonlinear optics, characterized by high intensity, few cycle pulses, which, through their interaction with material degrees of freedom, display very broad and complex spatiotemporal spectra, *spectral beam, and pulse propagators*[4] are the preferred methods of choice. Unfortunately, they do not mesh well with the boundary treatments developed for the finite-difference solvers such as PML.

While spectral propagators applied to pulses and/or beams shine in many situations that are next to impossible to handle with finite-difference approaches, the boundary treatment can be a significant problem. For example, long-distance propagation of highly nonlinear optical pulses[5] is often connected with light-matter interactions that send significant energy propagating toward the boundaries of computational domains where it must be "absorbed" as if propagating into infinite space. In connection to spectral-based numerical simulation of beam and pulse propagation, this is a difficult problem that we aim to address in this work.

The method we put forward can be understood as an extension of an approximation that is sometimes used to simulate beam and pulse propagation in leaky waveguides[6,7] such as hollow-core fibers or capillaries.[8] In such a context, the propagating modes are approximated[9,10] by real parts of the true leaky modes for the given waveguide,[11] while their propagation constants are redefined by the inclusion of the imaginary parts that reflect the propagation loss of a leaky mode. Such an approach can be interpreted as a first-order perturbation theory where eigenvalues are corrected, while the wave functions are kept unchanged. Needless to say, this only works when the physics dictates that the propagation is dominated by a relatively small number of modes that have small propagation losses.

We propose to utilize the true leaky modes, without approximations, as the basis for both the numerical representation of the optical field and for the realization of transparent boundary conditions. We introduce an artificial structure outside of the given computational domain in order to introduce an infinite set of leaky modes, and construct an expansion of an arbitrary beam profile. While we present the treatment for a fixed frequency, the generalization to pulsed waveform is straightforward.

Leaky modes have had a long history in the field of electromagnetics. They were used already as early as in 1884 by Thomson[12] in his study of decay phenomena in electromagnetics. Since then, they have been of enduring interest in electromagnetics, for resonator cavities,[13,14] optical waveguides,[15] photonic,[16] and plasmonic[17,18] structures, and are often used for numerical simulations, which is also what we propose to do in the current paper. Leaky modes are decaying eigenstates and as such have played an important role in quantum theory from its very inception until today. In this setting, they describe unstable states. Such states were first defined in terms of the absence of incoming waves by Siegert[19] for the nuclear scattering matrix. Siegert's definition of unstable states was taken up by Peierls,[20] Le Couteur,[21] and Humblet[22] and by them refined into an important tool for nuclear scattering theory. The wave functions satisfying the Siegert outgoing-wave conditions are known as resonant states, and their properties have been of interest for many years.[23–28]

As is evident from the previous paragraph, leaky modes and unstable states have a long history and have been, and are, of great utility[29–32] in various fields. However, the fact that they are decaying eigenstates means that the corresponding eigenvalue problems are not self-adjoint. Consequently, the matter of projecting general field configurations into sum of leaky modes or resonant states, and the question of completeness of the resulting expansions, are not backed up by any general theory, like for the self-adjoint case. In fact, the leaky modes and resonant states are invariably growing exponentially in space, and thus cannot be placed in some well-known inner product spaces. The lack of a general theory for non-self-adjoint operators is challenging, and it means that questions of projection and completeness have to be handled in a case by case basis. In this paper, we will introduce a projection method for leaky modes based on a naturally occurring complex non-Hermitian inner product, but will not present a convergence proof for our leaky mode expansions.

The paper is organized as follows: In Sec. II, we introduce the model which we will use to access the feasibility of our proposed approach to setting up a partially reflective boundary for spectral pulse propagators. We are most familiar with the particular spectral pulse propagator UPPE,[33] but our approach is applicable to any of the spectral pulse propagators that are commonly used in optics today. The model describes the propagation of TE electromagnetic waves in a homogeneous medium that we for convenience assume is a vacuum. We then proceed to set up and solve the eigenvalue problem for the complex transverse wave numbers that define the leaky modes. In this section, we also derive a very accurate explicit asymptotic formula for the location of leaky mode wave numbers in the complex plane. In Sec. III, we introduce the leaky modes and show that, by using the technique, well-know from the study of resonant states in quantum theory, of shifting them over to a complex spatial contour outside the transverse computational domain, the leaky modes can be identified with vectors in a vector space of functions on the real line that is endowed with a complex non-Hermitian inner product.[34,35] The leaky modes are orthogonal with respect to this product, and we can thus write down generalized Fourier series for any given function based on the orthogonal leaky modes and this non-Hermitian inner product. This solves the projection problem for our leaky modes.

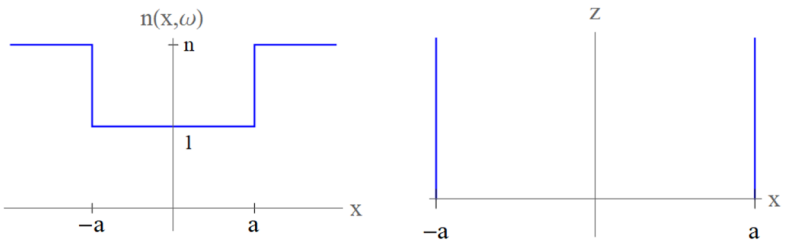
We have done extensive numerical experiments using our leaky mode expansions, and in Sec. IV, we present some examples and the conclusions we draw from these examples, with regards to their suitability for representing initial data for the spectral propagator. We argue that the leaky mode expansions converge pointwise for all sufficiently smooth functions in our space, but that they do not always converge to the function used to generate the expansion. The pointwise convergence only becomes problematic in the limit when the index, of the artificial material introduced outside the computational domain, tends toward the same value as the index inside the domain. One would expect that problems with the leaky mode expansions would appear in this limit since if the limit is reached, there is no index difference between the inside and the outside of the computational domain and leaky modes cease to exist. However, in order to minimize the reflection from the boundary of the computational domain, we want to choose the difference between the inside index and the outside, artificial index, as small as possible. It thus becomes a trade-off between making it small in order to minimize reflections and not making it so small that the leaky mode expansions stops giving a good representation of the functions used to generate the expansions. At the end of Sec. IV, we argue, using a dimensionless quantity that appear from our theory, that an acceptable trade-off can be made.

In this paper, we do not present a proof that the leaky mode expansion converge to the function used to generate the series. The chief reason for this is that we believe that they never really do converge point wise to the function used to generate them. This is what our numerical results from Sec. IV indicated. In Sec. V, we present analytical arguments that points to the same conclusion. However, the conjectured lack of pointwise convergence to the desired function does not make the leaky mode expansions useless from a more practical point of view. This is what we argue in Sec. IV, where we use a certain dimensionless quantity to specify what we mean by a practical point of view in this context.

II. THE MODEL

We will assume that the spatial inhomogeneity of the refractive index takes the form of a straight channel oriented along the z-axis of our coordinate system, of uniform width $2a$ in the transverse direction, which is oriented along the x-axis of our coordinate system. The geometry of the channel is illustrated in Fig. 1. Consistent with the geometry, we assume that the electromagnetic field is transverse electric. Thus, we have

$$\begin{aligned}\mathbf{E}(\mathbf{r},\omega) &= (0, e(x,z,\omega), 0),\\ \mathbf{P}(\mathbf{r},\omega) &= (0, p(x,z,\omega), 0).\end{aligned} \tag{1}$$



(a) Refractive index $n(x, \omega)$ (b) Geometry of the channel

FIG. 1. Geometry of the refractive index channel.

Here, we use the sign convention for the inverse Fourier transform with respect to time that is standard in optics

$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} d\omega\, \mathbf{E}(\mathbf{r}, \omega)\, e^{-i\omega t}. \tag{2}$$

Using Maxwell's equations with no free charges or currents, we find that $e(x, z, \omega)$ is a solution to the following model equation:

$$\partial_{zz} e(x, z, \omega) + \partial_{xx} e(x, z, \omega) + \left(\frac{\omega}{c}\right)^2 (1 + \chi(x, \omega)) e(x, z, \omega) = p(x, z, \omega), \tag{3}$$

$$\Downarrow \text{ linearization}$$

$$\partial_{zz} e(x, z, \omega) + \partial_{xx} e(x, z, \omega) + \left(\frac{\omega}{c}\right)^2 n^2(x, \omega) e(x, z, \omega) = 0. \tag{4}$$

In addition to the model Eq. (3), the electric field $e(x, z, \omega)$ must satisfy the conditions

$$e(\pm a_-, z, \omega) = e(\pm a_+, z, \omega), \tag{5}$$

$$\partial_x e(\pm a_-, z, \omega) = \partial_x e(\pm a_+, z, \omega), \tag{6}$$

which follows from the electromagnetic interface conditions for transverse electric fields at $x = \pm a$.

The goal is now to find leaky modes for the linearized equation. These modes can then be used to write down a modified spectral propagator, like UPPE, for the model Eq. (3), where a leaky mode transform takes the place of the usual transverse Fourier transform. The rest of the paper is focused on constructing the leaky modes and evaluating for which transverse field configurations they form a suitable basis.

III. LEAKY MODES

Leaky modes are solutions to the linearized model Eq. (4) that are propagating in the direction of the positive z-axis, satisfy the electromagnetic interface conditions (5) and (6), and are outgoing at positive and negative infinity.

Such functions must be of the form

$$\begin{aligned} e(x, z, \omega) &= D e^{i\beta z} e^{i\xi x}, & x > a, \\ e(x, z, \omega) &= e^{i\beta_0 z}\left(B e^{i\xi_0 x} + C e^{-i\xi_0 x}\right), & -a < x < a, \\ e(x, z, \omega) &= A e^{i\beta z} e^{-i\xi x}, & x < -a, \end{aligned} \tag{7}$$

where β_0, ξ_0 and β, ξ are the propagation constants and transverse wave numbers inside and outside the channel, respectively. The propagation constants are determined by the transverse wave numbers by the identities

$$\beta = \left(\left(\frac{\omega}{c}\right)^2 n^2 - \xi^2\right)^{\frac{1}{2}}, \quad \beta_0 = \left(\left(\frac{\omega}{c}\right)^2 - \xi_0^2\right)^{\frac{1}{2}}. \tag{8}$$

From the physical point of view, the modes represents electromagnetic disturbances that propagate in the direction of the positive z-axis while they are partially reflected and transmitted at the lateral boundaries defining the index channel. This is illustrated in Fig. 2.

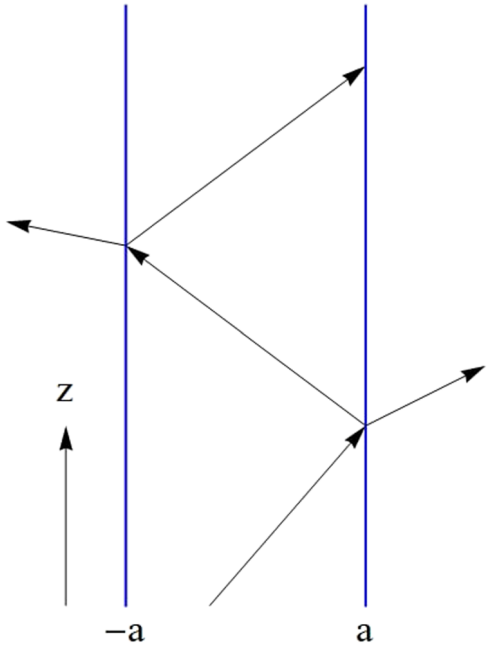


FIG. 2. Plane-wave propagation.

In order for the functions (7) to satisfy the electromagnetic boundary conditions (6), and thus be leaky modes, the two propagation constants β and β_0 must be equal, which is only true if the following identity holds:

$$\xi^2 = \left(\frac{\omega}{c}\right)^2 (n^2 - 1) + \xi_0^2. \tag{9}$$

This is Snell's law. In addition, the following linear algebraic system:

$$\begin{pmatrix} e^{ia\xi} & -e^{-ia\xi_0} & -e^{ia\xi_0} & 0 \\ -ie^{ia\xi}\xi & -ie^{-ia\xi_0}\xi_0 & ie^{ia\xi_0}\xi_0 & 0 \\ 0 & e^{ia\xi_0} & e^{-ia\xi_0} & -e^{ia\xi} \\ 0 & ie^{ia\xi_0}\xi_0 & -ie^{-ia\xi_0}\xi_0 & -ie^{ia\xi}\xi \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{10}$$

must have a unique solution. This can only happen if the determinant of the matrix defining the system is zero. One can show that the determinant is zero if and only if the transverse wavenumber satisfy following equation:

$$\tan(2a\xi_0) + i\frac{2\xi\xi_0}{\xi^2 + \xi_0^2} = 0. \tag{11}$$

Equation (11), together with Snell's law (9), will determine the dispersion law pertaining to each separate leaky mode.

Note that the system of (9)–(11) has two symmetries connecting solutions. If we denote solutions using the notation $\{\{\xi, \xi_0\}, (A, B, C, D)\}$, the two symmetries are of the form

$$\{\{\xi, \xi_0\}, (A, B, C, D)\} \to \{\{\xi, -\xi_0\}, (A, C, B, D)\}, \tag{12}$$

$$\{\{\xi, \xi_0\}, (A, B, C, D)\} \to \left\{\left\{-\xi^*, -\xi_0^*\right\}, \left(A^*, B^*, C^*, D^*\right)\right\}. \tag{13}$$

A. Dispersion laws

In this section, we will design asymptotic formulas for all solutions ξ, ξ_0 to Eqs. (9) and (11), and thus determine all modes for the system and their respective dispersion laws.

Let us start by observing that $\xi_0 = 0$ is a solution to Eq. (11) and that the corresponding solution vector to the linear system (9) is given by

$$\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}. \tag{14}$$

However, we observe that if we insert $\xi_0 = 0$ and the vector (14) into the formula for the modes (7), we find that the corresponding mode is identically zero. Thus, the solution $\xi_0 = 0$ only gives us a trivial mode which can be disregarded when we use the modes for expanding electric field configurations.

Observe that because of the symmetries (12) and (13), it is sufficient to consider the case when ξ_0 is in the second quadrant. Any solution in one of the other quadrants can be generated from a solution in the second quadrant by using the symmetries. In the second quadrant, we can split the system (9) and (11) into two separate systems depending on which square root we take when Eq. (9) is used to express ξ as a function of ξ_0,

$$\tan(2a\xi_0) = -i\frac{2\xi_0\sqrt{\alpha+\xi_0^2}}{\xi^2+\xi_0^2}, \qquad \xi = \sqrt{\alpha+\xi_0^2}, \tag{15}$$

$$\tan(2a\xi_0) = i\frac{2\xi_0\sqrt{\alpha+\xi_0^2}}{\xi^2+\xi_0^2}, \qquad \xi = -\sqrt{\alpha+\xi_0^2}, \tag{16}$$

where we have defined $\alpha = (\omega/c)^2(n^2-1)$. In Fig. 3, we display the solutions of the first of the two systems, (15). In the figure, the solutions are defined by the intersection of the zero contours for the real and imaginary part of the equation for ξ_0 in (15). There clearly exists an infinite set of solutions, each one corresponding to a distinct mode with its associated dispersion law. A similar plot for the second of the two systems, (16), gives convincing numerical evidence that it has no solutions in the second quadrant and thus this system does not give us any additional modes in the second quadrant.

It is evident that for most solutions displayed in Fig. 3, the real part strongly dominate the imaginary part. This fact can be used to find an asymptotic formula for the solutions to Eq. (11).

Assuming that $|\xi_0| \gg \sqrt{\alpha}$, Eq. (11) can be approximated by

$$\tan(2a\xi_0) = -i\frac{2\xi_0|\xi_0|\sqrt{1+\frac{\alpha}{\xi_0^2}}}{2\xi_0^2\left(1+\frac{\alpha}{2\xi_0^2}\right)} = i\sqrt{1+\frac{\alpha}{\xi_0^2}}\left(\frac{1}{1+\frac{\alpha}{2\xi_0^2}}\right) \approx i\left(1-\frac{\alpha^2}{8x^4}\right) \tag{17}$$

because $|\xi_0| \equiv \sqrt{\xi_0^2}$ is equal to $-\xi_0$ when ξ_0 is in the second quadrant.

Judging from the locations of the zeros in Fig. 3, most of them will be found in regions of the complex plane, where $|\mathrm{Re}[\xi_0]| \gg |\mathrm{Im}[\xi_0]|$. We therefore write $\xi_0 = x + iy$, where $|x| \gg |y|$ and use this to simplify Eq. (15) in the following way:

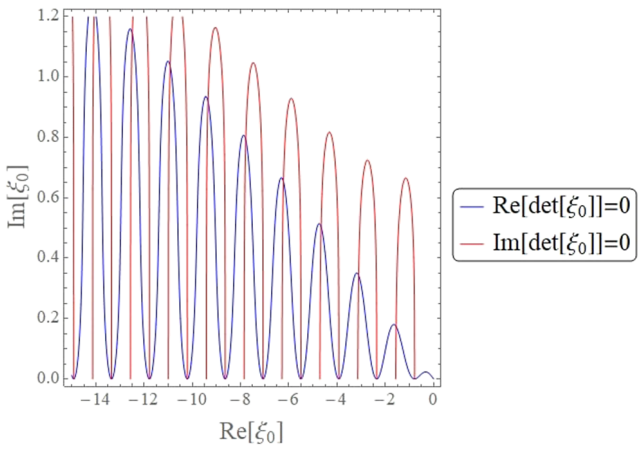


FIG. 3. Zero contours of the determinant in the second quadrant. Parameter values used in this plot were $a = 1$, $n = 1 + 10^{-12}$ and infrared light at wavelength 1 μm.

$$\frac{1-r^2}{r^2+2r\cos\theta+1}-i\frac{2r\sin\theta}{r^2+2r\cos\theta+1}\approx\left(1-\frac{\alpha^2}{8x^4}\right), \tag{18}$$

where $r=\exp(-4ay)$ and $\theta=4ax$. After careful analysis, we find the following formula:

$$\xi_{0p}=-\frac{p\pi}{2a}+\frac{i}{4a}\mathrm{Log}\left[\frac{16\left(-\frac{p\pi}{2a}+\frac{i}{4a}\ln\left[\frac{p^4\pi^4}{\alpha^2a^4}\right]\right)^4}{\alpha^2}\right] \tag{19}$$

which is an asymptotic approximation to the solutions of (9) and (11) in the 2^{st} quadrant. Using the symmetries (12) and (13), similar formulas for modes residing in the other three quadrants can be obtained.

Writing the formula defining ξ in terms of ξ_0 from (15) in the form $\xi=\xi'+i\xi''$, it is evident that if ξ_0 is in the second quadrant, then $\xi'>0$ and $\xi''<0$. Using the formula (7) and our convention for the inverse Fourier transform (2), we can conclude that the modes in the second quadrant, determined by formula (19), are outgoing and exponentially growing in the transverse direction. From the formula for the propagation constants (8), it is also evident that they are decaying in the propagation direction. These are thus leaky modes. In a similar way, the modes determined by formula (19) for the 4-th are also outgoing and decaying in the propagation direction, and thus, are also leaky modes. We find however that the modes determined by formulas for 1-st and 3-rd are incoming and growing in the propagation direction. These modes are thus not leaky modes, but gaining modes.

Even if we assumed p being large, the formulas for all four quadrants give surprisingly good results, even when p is of order 1. However, it is exactly in this region where the formulas can break down. Observe that the inner logarithm in formula (19) must be positive in order to stay in the same quadrant. Therefore, these formula becomes invalid if

$$\frac{p^4\pi^4}{\alpha^2a^4}\lesssim 1\Rightarrow p\lesssim\frac{a\sqrt{\alpha}}{\pi}. \tag{20}$$

Let us assume that $\sqrt{\alpha}\gg|\xi_{0p}|$. Applying this assumption to the equation for ξ_0 from (15), which determine the leaky modes in the 2-nd quadrant, gives us

$$\begin{aligned}\tan(2a\xi_0)=-i2\frac{\xi_0\sqrt{\alpha\left(1+\xi_0^2/\alpha\right)^2}}{\alpha\left(1+2\xi_0^2/\alpha\right)}&\approx-i\frac{2\xi_0}{\sqrt{\alpha}},\\ &\Downarrow\\ \frac{2r\sin\theta}{r^2+2r\cos\theta+1}+i\frac{1-r^2}{r^2+2r\cos\theta+1}&\approx-i\frac{2\xi_0}{\sqrt{\alpha}},\end{aligned} \tag{21}$$

where $\xi_0=x+iy$ and $r=\exp(-4ay)$, $\theta=4ax$. Using the same approach as before, we find that

$$\xi_{0p}=-\frac{\pi p}{2a}+i\frac{1}{4a}\mathrm{Log}\left[\frac{\sqrt{\alpha}-2\xi_0}{\sqrt{\alpha}+2\xi_0}\right]\approx-\frac{\pi p}{2a}+i\frac{1}{4a}\mathrm{Log}\left[1-\frac{4\xi_0}{\sqrt{\alpha}}\right]\approx-\frac{\pi p}{2a}-i\frac{\xi_0}{a\sqrt{\alpha}}. \tag{22}$$

Under the assumed condition $\sqrt{\alpha}\gg|\xi_{0p}|$, the second term in Eq. (22) is a small correction to the first term. This allows us to look at Eq. (22) as a recursion formula for the solution ξ_{0p}. Starting from the leading term $\xi_{0p}^0=-\pi p/(2a)$, we obtain from the first iteration

$$\xi_{0p}\approx-\frac{\pi p}{2a}+i\frac{1}{a\sqrt{\alpha}}\frac{\pi p}{2a}=-\frac{\pi p}{2a}+i\frac{\pi p}{2a^2\sqrt{\alpha}}. \tag{23}$$

We have thus obtained a different asymptotic formula for the solutions ξ_{0p} in the 2-nd quadrant, a formula where we know that the imaginary part is a small correction to the real part, $\pi p/(2a)\gg\pi p/(2a^2\sqrt{\alpha})$, or equivalently, $p\ll a\sqrt{\alpha}$. This condition implies that condition (20) holds. Thus, we can conclude that the asymptotic formula (23) holds exactly when the asymptotic formula (19) breaks down. Formulas similar to (23) can be derived for the other quadrants.

Some of the leaky modes are paraxial, whereas others are not. In order to be more precise about which modes are paraxial, note that the propagation vector for the light beam is of the form (ξ_0,β_0). This is clear from Eq. (7), and it allows us to calculate the propagation angle of the beam with respect to the z-axis. This angle is

$$\theta_p=\tan^{-1}\left(\frac{\mathrm{Re}[\xi_{0p}]}{\mathrm{Re}[\beta(\xi_{0p})]}\right). \tag{24}$$

Clearly, for each ξ_{0p}, we get a different angle. In order for a mode to be paraxial, the angle θ_p must be small, and this holds only if $\mathrm{Re}[\xi_{0p}] \ll \mathrm{Re}[\beta(\xi_{0p})]$, or in other words

$$p \ll \frac{a\sqrt{2}\omega}{\pi c}, \tag{25}$$

where we have used the fact to leading order $\xi_{0p} \approx -\pi p/(2a)$. Formula (25) determines which leaky modes are paraxial.

We now investigate if there are zeros in parts of the complex ξ_0-plane that are not covered by the asymptotic formulas we have found so far. We will focus on the second quadrant, the other quadrants can be treated in a similar way with corresponding results. These investigations are necessary because the exponential smallness of the equation determining ξ_0, in the part of the second quadrant well away from the real axis, makes a direct numerical search for solutions, like the one in Fig. 3, very challenging.

Let us first look for zeros in the part of the second quadrant where $y = \mathrm{Im}[\xi_0]$ is much larger than $x = \mathrm{Re}[\xi_0]$ and $\mathrm{Im}[\xi_0] \gg 1 \gg \alpha$. Under these conditions on ξ_0, we have $\tan(2a\xi_0) \approx i$ and the equation for ξ_0 in (15) can be simplified using Taylor series with the solution $y = \sqrt{\alpha}$. We thus end up with a solution $\xi_0 = i\sqrt{\alpha}$ that contradicting the assumptions imposed on ξ_0. Hence, no zeros can exist in this part of the second quadrant.

Let us next look at the region where $y = \mathrm{Im}[\xi_0] \sim \mathrm{Re}[\xi_0] = x$ and $y, x \gg 1 \gg \alpha$. At this point, observe that the left-hand side of Eq. (11) is not actually equal to the determinant of the system(10). Imposing Eq. (11) only implies that the determinant is zero. If we rather equates the full determinant of (10) to zero, we get

$$e^{4ia\xi_0} = \frac{(\xi_0 + \xi)^2}{(\xi_0 - \xi)^2}. \tag{26}$$

The above equation can be simplified using the assumptions $|\xi_0| \gg \alpha$. Writing $\xi_0 = -x + iy$, where $x, y > 0$, we get

$$\text{Re:} \quad x^4 - 6x^2y^2 + y^4 = 0, \tag{27}$$

$$\text{Im:} \quad x^3y - xy^3 = 0. \tag{28}$$

The only possible solutions to Eq. (27) are $x = y(\sqrt{2} - 1), x = y(\sqrt{2} + 1)$. Substituting these solutions into (28), and solving for y, we find in both cases t $y = 0$. This contradict our assumptions, and thus, there are no solutions in this region of the second quadrant either. We have now covered all possible regions of the second quadrant and thus conclude that there are no other zeros of the determinant, and thus leaky modes, than the ones we have already found and that is covered by our asymptotic formulas.

There is however one remaining issue related to the zeros, and thus leaky modes, that needs to be discussed. As we have already noted, the asymptotic formulas for the zeros, which, by design, are expected to be accurate only in the limit when the index p is very large, in fact works surprisingly well even for p as small as 2. However, the very first zero, the one corresponding to $p = 1$, is never very accurate. The first zero also behave differently when the parameter α is varied. Recall that the value of this parameter is proportional to the size of the index step defining the channel where the waves will be propagating. We are interested in minimizing reflections from the edges of the channel and therefore would want to make the index step, and hence the parameter α, as small as possible. When we let alpha decrease, we observe that all the zeros in the second quadrant, except the first, move slowly up, and even more slowly toward the imaginary axis. This behavior is to be expected from of the logarithmic dependence of the imaginary part of the zero on the parameter α. The first zero approaches the imaginary axis at a fast rate when α is decreased, and for a finite value of $\alpha = \alpha_c$, it simply vanishes. For $\alpha < \alpha_c$, we observe that for one value of the index p, the formula (19) indicate the presence of a double zero. The index for which this occurs increase when α keeps decreasing towards zero. These double zeros are however spurious; careful numerical investigations show that there are no double zeros. However, this abrupt change in the prediction derived from formula (19), when α vary smoothly, alerted us to the possibility that the root cause to why our formula predicted both the vanishing of the first zero and the existence of double zeros is the crossing of a branch cut. Observe that the argument of the logarithm in formula (19) is an expression with complex values, so there is indeed a branch cut implied by the formula and thus the argument crossing this branch cut when α vary smoothly is a real possibility.

Formula (19) can be written in the form

$$\xi_{0p} = -\frac{p\pi}{2a} + \frac{i}{4a}\mathrm{Log}[z], \quad z = \frac{16\left(-\frac{p\pi}{2a} + \frac{i}{4a}\ln\left[\frac{p^4\pi^4}{\alpha^2a^4}\right]\right)^4}{\alpha^2}. \tag{29}$$

We use the standard branch of the logarithm in our calculations, and thus, there is a branch cut along the negative real axis. The real part of ξ_{0p} is negative, so we will have a crossing of the branch cut whenever the imaginary part of z vanish. Expanding the polynomial expression defining z in Eq. (29), and taking the imaginary part, we find that there is a crossing of the branch cut whenever α crosses the value

$$\alpha_p = \frac{p^2\pi^2}{a^2}\exp(-p\pi). \tag{30}$$

Further numerical investigations show that $\alpha_c = \alpha_1$ and that α_p for $p > 1$ correspond to the values of α where formula (19) predicts a double zero for the value of the index equal to p. The impact of the disappearing of the first zero on our leaky mode expansions will be discussed later, at the end of Sec. IV.

B. Mode shapes, normalization, and projection

For $\{\xi, \xi_0\}$, solving Eqs. (9) and (11) with ξ_0 in the second quadrant and ξ in the fourth quadrant, this is the case specified in (15), we have a leaky mode whose formula which, according to (7), is given by

$$u_p^-(x) = \begin{cases} De^{i\xi_p x}, & x > a \\ Be^{i\xi_{0p}x} + Ce^{-i\xi_{0p}x}, & -a < x < a, \\ Ae^{-i\xi_p x}, & x < -a \end{cases} \quad \xi_p = \left(\alpha + (\xi_{0p})^2\right)^{1/2}. \tag{31}$$

Using the symmetries (12) and (13), we can conclude that there is a corresponding incoming, gaining mode, in the first quadrant whose formula is given by

$$u_p^+(x) = \begin{cases} D^* e^{i\xi_p^* x}, & x > a \\ B^* e^{-i\xi_{0p}^* x} + C^* e^{i\xi_{0p}^* x}, & -a < x < a, \\ A^* e^{-i\xi_p^* x}, & x < -a \end{cases} \quad \xi_p = -(\alpha + \xi_{0p}^{*2})^{1/2}. \tag{32}$$

Observe that we have $(u_p^+)^* = u_p^-$. In Fig. 4, we see an outgoing mode corresponding to the index $p = 20$. The mode is evidently exponentially growing in x. This holds true for all modes, both incoming and outgoing.

Since the modes are exponentially growing in x, they are clearly not normalizable. We can make the modes normalizable by analytically continuing them into a complexified spatial domain and restricting the analytically continued modes to carefully chosen complex contours. The contours will be different depending on whether the modes are incoming or outgoing. The contours we will be using are of the form

$$z^+(x) = \begin{cases} a - i(x-a), & x > a \\ x, & |x| < a, \\ -a - i(x+a), & x < -a \end{cases} \quad z^-(x) = \begin{cases} a + i(x-a), & x < a \\ x, & |x| < a, \\ -a + i(x+a), & x < -a \end{cases} \tag{33}$$

where z^+ is used for the incoming modes and z^- is used for the outgoing modes.

Evaluating Eqs. (31) and (32) on these contours, we find that they exponentially decay in both directions on the real axis. Define functions $\psi_p^+(x)$ and $\psi_p^-(x)$ on the positive real axis as

$$\psi_p^+(x) = u_p^+\left(z^+(x)\right), \tag{34}$$

$$\psi_p^-(x) = u_p^-\left(z^-(x)\right). \tag{35}$$

The formulas for these functions are

$$\psi_p^-(x) = \begin{cases} De^{i\xi_p a}e^{-\xi_p(x-a)}, & x > a \\ Be^{i\xi_{0p}x} + Ce^{-i\xi_{0p}x}, & |x| < a, \\ Ae^{i\xi_p a}e^{+\xi_p(x+a)}, & x < -a \end{cases} \quad \xi_p = (\alpha + \xi_{0p}^2)^{1/2}, \tag{36}$$

$$\psi_p^+(x) = \begin{cases} D^* e^{i\xi_p^* a}e^{\xi_p(x-a)}, & x > a \\ B^* e^{-i\xi_{0p}^* x} + C^* e^{i\xi_{0p}^* x}, & |x| < a, \\ A^* e^{i\xi_p^* a}e^{-\xi_p(x+a)}, & x < -a \end{cases} \quad \xi_p = -\left(\alpha + \left(\xi_{0p}^*\right)^2\right)^{1/2}. \tag{37}$$

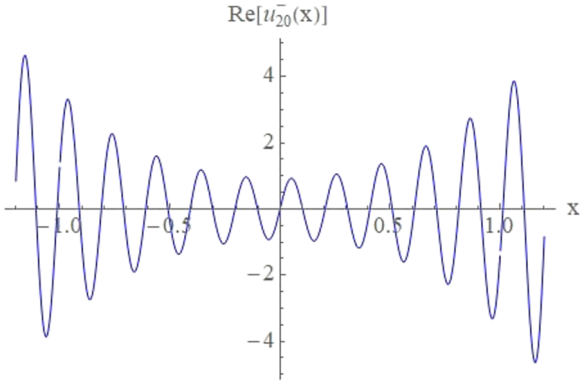


FIG. 4. Outgoing mode corresponding to the index $p = 20$. Parameters used in this plot were $a = 1$, $n = 1 + 10^{-12}$ and infrared light at wavelength 1 μm.

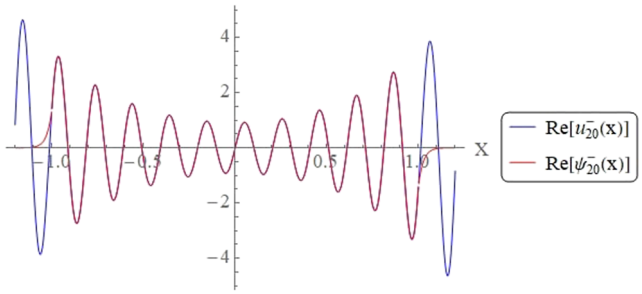


FIG. 5. Outgoing mode and its complexified version corresponding to the index $p = 20$. Parameters used in this plot were $a = 1$, $n = 1 + 10^{-12}$ and infrared light at wavelength 1 μm.

Note that for these complexified modes, we also have the relation $\left(\psi_p^+(x)\right)^* = \psi_p^-(x)$. Figure 5 shows the functions $u_{20}^-(x)$ and $\psi_{20}^-(x)$ in the same picture. We observe that $\psi_{20}^-(x)$ decay exponentially outside the channel, which is confined to the interval $[-a, a]$. Also note that the complexified modes are not continuously differentiable at the points $x = \pm a$. This is because we restricted the analytically continued modes to a contour that is singular at $x = \pm a$. We made this choice in order to get fastest possible decay of the complexified modes and the simplest possible expressions for certain key differential operators acting on the modes.

Using the analyticity of the complexified modes at the two points $z = \pm a$ and the formulas for the two singular contours (33), it is easy to verify that the following boundary conditions holds for $\psi_p^+(x)$ and $\psi_p^-(x)$ at the two points $x = \pm a$:

$$\begin{aligned}
\psi_p^+(\pm a^-, \omega) &= \psi_p^+(\pm a^+, \omega), & \psi_p^-(\pm a^-, \omega) &= \psi_p^-(\pm a^+, \omega),\\
\partial_x\psi_p^+(-a^-, \omega) &= -i\partial_x\psi_p^+(-a^+, \omega), & \partial_x\psi_p^-(-a^-, \omega) &= i\partial_x\psi_p^-(-a^+, \omega),\\
\partial_x\psi_p^+(a^-, \omega) &= i\partial_x\psi_p^+(a^+, \omega), & \partial_x\psi_p^-(a^-, \omega) &= -i\partial_x\psi_p^-(a^+, \omega).
\end{aligned} \tag{38}$$

This fact tells us that complexified modes ψ^+, ψ^- belong to two different spaces of functions, V^+ and V^-. Here, V^- is the space of smooth functions on real line which satisfies the boundary conditions for $\psi_p^-(x)$ (38), and similarly for V^+. We evidently have

$$\left\{\psi_p^+(x)\right\}_{p=1}^{\infty} \subset V^+, \qquad \left\{\psi_p^-(x)\right\}_{p=1}^{\infty} \subset V^-. \tag{39}$$

It is easy to verify that the complexified modes are in fact eigenfunctions to the differential operator

$$\mathcal{L}_x = \begin{cases} \partial_{xx} + \left(\frac{\omega}{c}\right)^2, & |x| < a \\ -\partial_{xx} + \left(\frac{\omega}{c}\right)^2(n^2 - 1), & |x| > a \end{cases}. \tag{40}$$

We have

$$\mathcal{L}_x\psi^-(x) = \lambda_p\psi_p^-(x), \quad \lambda_p = \left(\left(\frac{\omega}{c}\right)^2 - \xi_{0p}^2\right)^{\frac{1}{2}}, \tag{41}$$

$$\mathcal{L}_x\psi^+(x) = \mu_p\psi_p^+(x), \quad \mu_p = \lambda_p^*. \tag{42}$$

In order use the complexified modes as a tool for expanding functions in V^-, functions that are in the span of $\left\{\psi_p^-(x)\right\}_{p=1}^{\infty}$, we need an inner product on the space. Furthermore, with respect to this inner product the leaky modes must be normalizable and orthogonal. Orthogonality would be assured if the operator $\mathcal{L}_x$, defined in (40), is self-adjoint with respect to the chosen inner product. This, however, seems like an impossible task since we know that the eigenvalues λ_p, defined in (42), are in fact complex.

Nevertheless, an inner product that satisfy all the requirements can be constructed. In order to do this, note that for any contour $\mathcal{C}$ in the complex plane, we can define a complex values scalar product on the space of functions analytic in an open set containing the contour

$$(\Phi, \Psi) = \int_{\mathcal{C}} \Phi(z)\overline{\Psi}(z)\mathrm{d}z \in \mathbb{C}, \tag{43}$$

where $\overline{\Psi}(z)$ is an analytic function defined by $\overline{\Psi}(z) = \Psi^*\left(z^*\right)$. Applying this definition of scalar product of analytic functions to the contour z^-, we get the following complex valued scalar product on the space V^-, defined for any pair of function $\psi, \phi \in V^-$ by the expression

$$(\psi,\phi)^- = i\int_{-\infty}^{-a}\psi(x)\phi(x)\mathrm{d}x + \int_{-a}^{a}\psi(x)\phi(x)\mathrm{d}x + i\int_{a}^{\infty}\psi(x)\phi(x)\mathrm{d}x. \tag{44}$$

It is now straight forward to show that the differential operator $\mathcal{L}_x$ is self-adjoint with respect to the inner product (44) on the space V^-. The orthogonality of the leaky modes can be readily verified.

Any function in $f \in V^-$ which is in the span of the leaky modes $\{\psi_p^-(x)\}_{p=1}^{\infty}$ can now be expanded in terms of a generalized Fourier series of the form

$$f(x) \in V^- \Rightarrow f(x) = \sum_{p=1}^{\infty}\frac{\left(f(x),\psi_p^-\right)^-}{\left(\psi_p^-,\psi_p^-\right)^-}\psi_p^-(x). \tag{45}$$

In a similar way, an inner product can be introduced on the space of gaining modes V^+.

Observe that the boundary conditions (38) implies that $\phi \in V^- \Leftrightarrow \psi^* \in V^+$. Thus, the complex conjugate maps between these two spaces. In a similar way, the complex conjugate maps between the inner products on the two spaces

$$(\psi,\phi)^{-*} = (\psi^*,\phi^*)^+. \tag{46}$$

The spaces of leaky modes V^- and gaining modes V^+ are not only linear spaces but also complex algebras. This holds because products of functions preserve the boundary conditions at $x = \pm a$. For any pair of functions in $\psi, \phi \in V^-$, we have, for example, $\partial_x(\psi\phi)\ (a^-) = -i\partial_x(\psi\phi)\ (a^+)$. Thus, we can conclude that $(\psi\phi) \in V^-$.

IV. NUMERICAL RESULTS

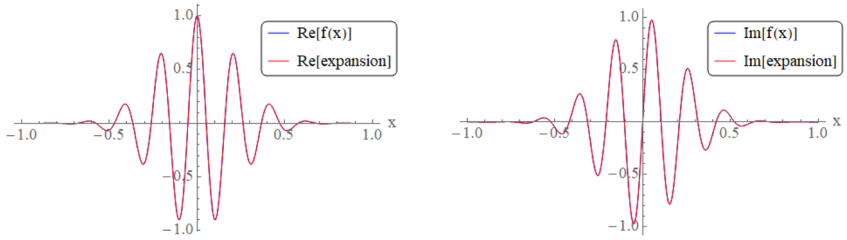
In this paper, we will not present a formal proof specifying precisely which space of functions are in the span of the set of leaky modes, and thus, for which space of functions the expansions (45) converge point wise. We will however present some arguments in Sec. V that addresses the question of convergence of the leaky mode expansions (45).

In this section, we present some numerical tests of the leaky mode expansions that will indicate strongly that they are indeed useful for the optical beam propagation context, we have designed them for. In order for the leaky mode expansions to be useful for modelling (semi) transparent computational boundaries for UPPE, and other spectral pulse propagators, there are two conditions that must be met.

First, physically reasonable initial data must be in the span of the leaky modes. Second, products of functions in the span must also be in the span. In Fig. 6, we display an expansion of a Gaussian wave packet using only 30 terms in the leaky mode expansion (45). In these plots, the refractive index outside the slab is $n = 1 + 10^{-12}$. As we can see, the Gaussian wave packet and its leaky mode expansion are indistinguishable.

Second, since the spectral pulse propagator must be able to handle nonlinear interactions, products of functions in the span must also be in the span (Fig. 7). In order to investigate this, we expanded squares of the leaky modes, namely, $\left(\psi_{10}^-(x)\right)^2$ and $\left(\psi_{50}^-(x)\right)^2$. The results are clearly very satisfying. Note that for the second mode, we needed more terms in the leaky mode expansion because of its highly oscillatory nature.

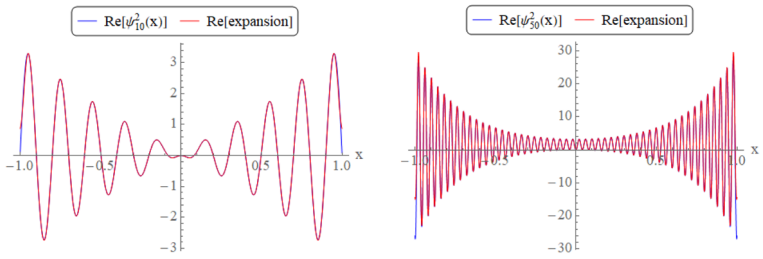
In this paper, we are not going to implement our leaky mode expansions in a fully nonlinear spectral pulse propagation algorithm. Before this can be done, more work has to be put into ensuring the accuracy and efficiency of the transformation from a function to its leaky mode expansion and back again. Here, we will show a linear propagation example, where we compare the approach using the leaky modes, to one using regular Fourier modes, which corresponds to imposing perfectly reflecting boundary condition at $x = \pm a$. Both are compared to the exact, infinite domain solution, which, for any z, can be approximated arbitrarily well by using a regular Fourier series on a much larger transverse domain. To appreciate how well our leaky modes expansion does, we demonstrate a numerical experiment where model a CW Gaussian beam propagation using Fourier expansion for finite as well as for infinite domain and compare it to the leaky modes expansion.



(a) Real part of a gaussian wavepacket and its leaky mode expansion.

(b) Imaginary part of a gaussian wavepacket and its leaky mode expansion.

FIG. 6. A Gaussian wavepacket $f(x) = \exp(-mx^2)\exp(ikx)$, where $m = 10$, $k = 30$ and its leaky mode expansion using 30 outgoing terms. The parameters used in this expansion were $a = 1$, $n = 1 + 10^{-12}$ and infrared light at wavelength 1 μm.



(a) Real part of the square of outgoing mode corresponding to $p = 10$ and its leaky mode expansion using 60 outgoing terms.

(b) Real part of the square of outgoing mode corresponding to $p = 50$ and its leaky mode expansion using 200 outgoing terms.

FIG. 7. Two squares of outgoing modes corresponding to $p = 10$ and $p = 50$ and their leaky mode expansion using 60, respectively, 200 outgoing terms. The parameters used in this expansion were $a = 1$, $n = 1 + 10^{-11}$ and infrared light at wavelength 1 μm.

As indicated earlier, we also express the infinite domain solution using Fourier modes, now on a larger domain, say $10a$. The exact solution and this numerical solution will not deviate until the diffracting Gaussian hit the boundary of the extended domain.

Figure 8 depicts the solutions for all three approaches. The width of the domain for the finite Fourier method as well as the slab for the leaky modes is $a = 1/2$ and the refractive index outside the slab is $n = 1 + 10^{-12}$ in the optical regime of infrared light. Approximately at the point $z \approx 4 \times 10^4$, the wave hits the boundary of the slab and the finite Fourier solution starts to deviate from the other two solutions. In Fig. 8(a), we see that the leaky modes solution and infinite domain Fourier overlap reasonably well. Propagating the wave further in the slab, we observe that around $z \approx 4 \times 10^5$ the leaky modes solution starts to deviate from the infinite domain Fourier solution. Up to this moment, the leaky modes and infinite Fourier solutions were very close to each other. We can therefore say that with the leaky modes method, we were able to propagate the wave approximately 10 times longer than with the finite Fourier method. The reason why the leaky modes eventually collapsed is, that the slab is not perfectly transparent. In other words, the difference in the refractive indices for the slab and the outside domain is nonzero. This leads to reflections that gradually build up as the wave propagates in z causing it to interfere with itself.

To be able to propagate the wave using leaky modes even further, we could lower the index n even more, to, say, $n = 1 + 10^{-15}$. However, it turns out that here we come across some serious issues. Let us first look at an expansions for a Gaussian function using $n = 1 + 10^{-15}$. Looking at Fig. 9, we see that the width of the domain is the same as in the propagation example; however, we made refractive index n outside the slab closer to 1. The number of terms used in this expansion was 200. The badness of this expansion suggests that one should use perhaps more terms to make it better. But the truth is, the expansion does not change with more terms. Thus, the numerics indicate that the leaky mode expansion for the Gaussian does converge pointwise, but unfortunately to some other function than the target Gaussian. There are two possible explanations for what happens here. The first is that the series actually diverges, but so slowly that we cannot detect it numerically. The second is that the series does converge pointwise, but not to the function used to generate it.

Extensive numerical investigations, using very high numerical precision, leads us to conjecture that it is the second explanation that is correct. In fact, we suspect that the leaky mode *never* converge pointwise to the function used to generate it. We will look more into these issues in Sec. V using asymptotic methods. Here, we just note that even though we very likely do not have pointwise convergence for the leaky mode expansion, the expansion is nevertheless useful for the task it was designed for. We find that the deviation between a function and its

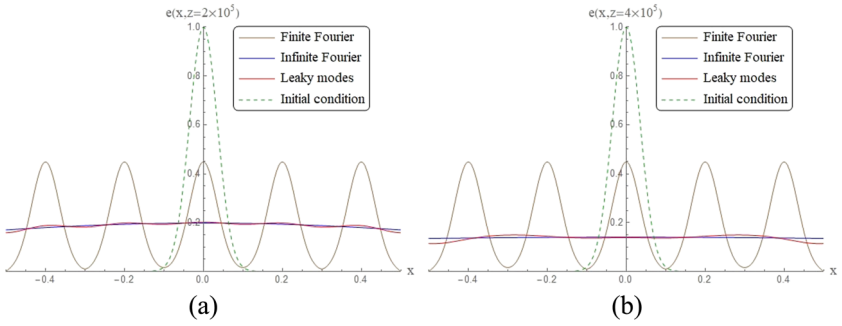


FIG. 8. Comparing different solutions to Eq. (4) using the Fourier method in a finite and infinite domain and leaky modes. For the Fourier method in a finite domain and leaky modes, 100 terms in the expansions were used while for the Fourier method in an infinite domain we used 300 terms. Parameters used in these plots were $a = 1/2$, $n = 1 + 10^{-12}$ and infrared light at wavelength 1 μm.

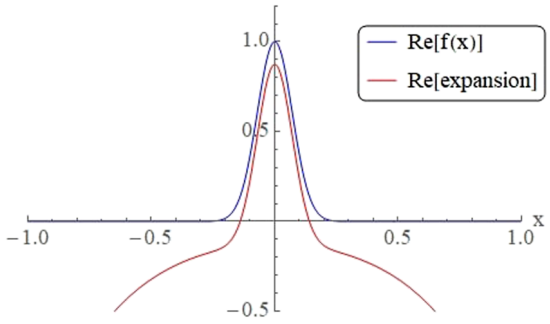


FIG. 9. The test function in this figure is $f(x) = \exp(-(10x)^2)$. Parameters used in this expansion were $a = 1$, $n = 1 + 10^{-15}$ and infrared light at wavelength 1 μm.

leaky mode expansion is only noticeable when the dimensionless number

$$\eta = a^2\left(\frac{\omega}{c}\right)^2(n^2 - 1) \tag{47}$$

is not too small. For the series to give a, practically speaking, faithful representation of functions, we need at least $\eta \gtrsim 0.8$. We find that for an infrared light at wavelength 1 μm ($\omega/c \approx 6.283 \times 10^6$), we have a good representation of Gaussian initial data if

$$\begin{aligned} a &= 10^{-1}\ \text{m}, \quad n \geq 1 + 10^{-10}, \\ a &= 10^{-2}\ \text{m}, \quad n \geq 1 + 10^{-8}, \\ a &= 10^{-3}\ \text{m}, \quad n \geq 1 + 10^{-6}, \\ a &= 10^{-4}\ \text{m}, \quad n \geq 1 + 10^{-4}. \end{aligned}$$

An important requirement for using the leaky mode expansion is that the main part of the pulse, where the bulk of the nonlinear interactions takes place, is well inside the domain $[-a, a]$. The choices for the transverse width a of the domain in the above list are chosen because they corresponds to actual dimensions used in high energy, long distance, propagation of optical pulses in air, using the UPPE code developed at the Center for Mathematical Sciences at the University of Arizona.

In Sec. III, we mentioned a phenomenon that occurs when one varies the value of α. In particular, if α becomes less than (30) for $p = 1$, the first zero disappears. In other words, we loosen the first eigenfunction completely, an eigenfunction which determine the first term in the leaky mode expansion. Since one usually expects that the first terms in the expansion are the most important ones for generic functions, the loss off the first eigenfunction is ominous. We conjecture that this loss, at least partly, explains why the leaky mode expansion loses its ability to accurately represent important boundary data like a Gaussian, when the parameter α become small enough.

In support of this conjecture, note that the eigenfunctions $\psi_p^-(x)$ alternate between being odd and even functions depending on the index p. Before the disappearing of the first zero, $\psi_1^-(x)$ is an even function. Let us denote (30) for $p = 1$ as α^*. Then, for $\alpha \gtrsim \alpha^*$ the expansion is a good representation of the Gaussian, which is even. However, for $\alpha < \alpha^*$, the expansion goes bad because we have lost the first term in the sum. Now we understand why the expansion goes wrong. Because the most important first term in the expansion is an odd function trying to

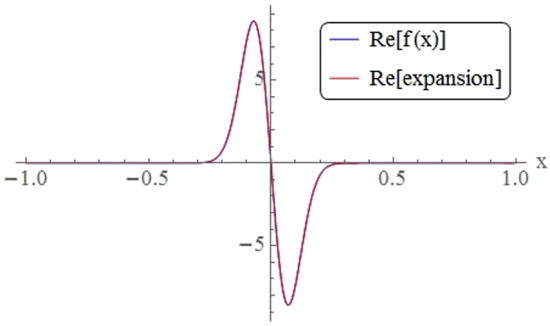


FIG. 10. The test function in this figure is $f(x) = \exp(-(10x)^2)\,(-200x)$. Parameters used in this expansion were $a = 1$, $n = 1 + 10^{-24}$, and infrared light at wavelength 1 μm.

represent an even Gaussian. With this in mind, let us expand an odd function instead of the even Gaussian for $\alpha < \alpha^*$. Let us, for example, use derivative of the Gaussian. Indeed, as can be seen in Fig. 10, the leaky mode expansion represent the odd functions much better than the even Gaussian. While the Gaussian was badly represented by its leaky mode expansion for $n = 1 + 10^{-15}$, for the derivative of the Gaussian, which is an odd function, we have a very precise leaky mode expansion, even for an index step as small as $n = 1 + 10^{-24}$.

V. ASYMPTOTIC SERIES

In Sec. IV, we conjectured that for small index steps, the leaky mode expansions does converge, but not to the functions used to generate the expansion. We used high precision numerical calculations to support this conjecture. In this section, we will lend additional support to the conjecture by proving that in the limit of small index step, the leaky mode expansion does indeed converge point wise, but to the wrong function. The asymptotic regime we are exploring are more here conveniently defined in terms of the parameter α. The requirement of the analysis in this section is that

$$\sqrt{\alpha} \ll |\xi_{0p}|. \tag{48}$$

The validity of this inequality is what we in this section mean by the asymptotic limit. Note that (48) is in fact also the requirement for the formula (19) to be an accurate approximation to the locating of the leaky mode wave numbers, ξ_{0p}. Recall that the leaky mode expansion for a function $f(x)$ is given by

$$f(x) = \sum_{p=1}^{\infty} \frac{\left(f(x), \psi^-_{\xi_{0p}}(x)\right)^-}{\left(\psi^-_{\xi_{0p}}(x), \psi^-_{\xi_{0p}}(x)\right)^-} \psi^-_{\xi_{0p}}(x). \tag{49}$$

We will be interested in finding an asymptotic approximation to the terms in this sum for two sample functions. However, before we proceed to the actual terms for our sample functions, we first need to know, how the coefficients A, B, C, D in Eq. (37) depend on the index p in the asymptotic limit. First of all, we realize that the vector $(A, B, C, D)^T$ is the null space and thus eigenvector belonging to the eigenvalue 0. To compute this eigenvector we can proceed, as we would normally do, by row-reducing the matrix Eq. (10), which gives us the matrix whose last row is of the form $(0, 0, 0, \eta(\xi_0))$, where $\eta(\xi_0)$ contains the determinant of $\mathbf{M}$. Thus, we get all zeros in the last row of the matrix, if $\xi_0 = \xi_{0p}$. Using this simplification, it is easy to find a basis for the one dimensional null space in the form

$$\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} \frac{2\xi_0 e^{2ia\xi_0}}{\xi\left(1-e^{4ia\xi_0}\right)+\xi_0\left(1+e^{4ia\xi_0}\right)} \\ -\frac{(\xi-\xi_0)e^{ia(\xi+3\xi_0)}}{\xi\left(1-e^{4ia\xi_0}\right)+\xi_0\left(1+e^{4ia\xi_0}\right)} \\ \frac{(\xi+\xi_0)e^{ia(\xi+\xi_0)}}{\xi\left(1-e^{4ia\xi_0}\right)+\xi_0\left(1+e^{4ia\xi_0}\right)} \\ 1 \end{pmatrix}. \tag{50}$$

Using the asymptotic expression (19), for the location of the zeros of the determinant $\xi_0 = \xi_{0p}$, we find the asymptotic formula for the term $\exp(\pm ia\xi_{0p})$ grows linearly in p, while $\exp(+ia\xi_{0p})$ decays. In the asymptotic limit, we can approximate (50) by the expression

$$\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} \approx \begin{pmatrix} (-1)^{p+1} \\ (-1)^{p+1} \\ 1 \\ 1 \end{pmatrix}. \tag{51}$$

For any given function $f(x)$, we can write the leaky mode expansion in the form

$$f(x) = \sum_p \frac{b_p(x)}{N_p} = \sum_p c_p(x), \tag{52}$$

where according to Eq. (44)

$$b_p(x) = \int_{-a}^{a} f(x)\psi^-_p(x)\mathrm{d}x\psi^-_p(x), \tag{53}$$

$$N_p = \left(i\int_{-\infty}^{-a} + \int_{-a}^{a} + i\int_{a}^{\infty}\right)\left(\psi^-_p(x)\right)^2\mathrm{d}x. \tag{54}$$

We now turn our attention to the normalization terms N_p in Eq. (54). In the limit of small α since the zeros ξ_{0p} in Eq. (49) are from the 2-nd quadrant, we can assume $\xi_p \approx -\xi_{0p}$. Using this assumption, and doing the integrals in (54) exactly, we obtain

$$N_p \approx \frac{k}{\xi_{0p}}\left(\frac{4aBC\xi_{0p}}{k} + \frac{\exp(2ia\xi_{0p})}{2i}\right) \approx 4aBC \approx 4a(-1)^{p+1}, \tag{55}$$

where $k = A^2 + D^2 = B^2 + C^2$. In the last step, we used the fact that $\exp(2ia\xi_{0p})$ decays as p^{-2}.

Let us next make a general statement about the decay rate, as a function of the index p, of the projection (53) of a given function $f(x)$ onto the leaky mode $\psi_p^-(x)$. Let us assume that $f(x)$ is a function that is zero at $x = \pm a$ and is n-times continuously differentiable. It is clear that each time we perform integration by parts in (53), i.e., differentiating $f(x)$ and integrating $\psi_p^-(x)$, we get an extra factor $i\xi_{0p}$ in the denominator. After n consecutive integrations by parts, we get

$$b_p(x) = \psi_p^-(x)\left(\frac{i}{\xi_{0p}}\right)^n \int_{-d}^{d} f^{(n)}(x)\left(\psi_p^-(x)\right)^{\star(n)} dx, \tag{56}$$

where

$$\left(\psi_p^-(x)\right)^{\star(n)} = \begin{cases} B\exp(i\xi_{0p}x) + C\exp(-i\xi_{0p}x) & n \text{ is even} \\ B\exp(i\xi_{0p}x) - C\exp(-i\xi_{0p}x) & n \text{ is odd} \end{cases}. \tag{57}$$

The term $1/\xi_{0p}^n$ can be approximated as $\approx(-2a)^n/(p\pi)^n$, and so the asymptotic expression for Eq. (56) becomes

$$b_p(x) \approx \psi_p^-(x)\left(\frac{-2ai}{p\pi}\right)^n \int_{-d}^{d} f^{(n)}(x)\left(\psi_p^-(x)\right)^{\star(n)} dx. \tag{58}$$

Next, we will find asymptotic approximations to (58) for our two chosen sample functions. This will give us an asymptotic approximation to the terms of the leaky mode expansion for the two sample functions.

As our first sample, we choose the following triangle function:

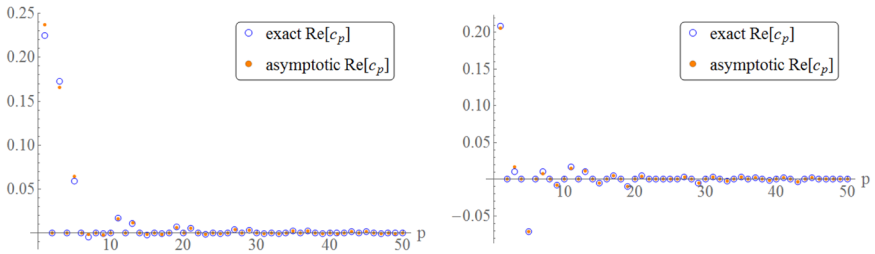
$$f(x) = \begin{cases} x + d & -d < x < 0 \\ -x + d & 0 < x < d \end{cases}, \tag{59}$$

whose derivative is 1 for $-d < x < 0$ and -1 for $0 < x < d$. In this case, the approximative coefficients $c_p(x) = b_p(x)/N_p$ after one integration by parts are

$$c_p(x) \approx -\frac{4a^2}{(p\pi)^2}\frac{((-1)^p - 1)}{4a(-1)^{p+1}}\left(-1 + \exp\left(\frac{i\pi dp}{2a}\right)\left(\frac{p\pi}{a\sqrt{\alpha}}\right)^{d/a}\right)^2$$
$$\exp\left(-\frac{i\pi p(d+x)}{2a}\right)\left((-1)^p - \exp\left(\frac{i\pi px}{a}\right)\left(\frac{p\pi}{a\sqrt{\alpha}}\right)^{\frac{2x}{a}}\right)\left(\frac{p\pi}{a\sqrt{\alpha}}\right)^{-\frac{d+x}{a}}. \tag{60}$$

In Figs. 11 and 12, we compare the asymptotic expressions for the terms in the leaky mode expansion with the exact terms calculated using high precision numerics. As we can see, there is a remarkable agreement between the values predicted by the asymptotic formulas and the exact values, even for small values of the mode index p. Numerically, the terms appear to approach zero fairly quickly, indicating the series itself converge.

In order to see if, and for which values of x the series converge or diverge, we write the terms in the series (60) in the form of sum of terms proportional to z^p/p^n, for some complex number z and n being one of the following expressions: $2 \pm x/a$, $2 \pm (d + x)/a$ and $2 \pm (d - x)a$. The sum over p of each term can be expressed using the polylogarithm function $Li(n, z)$, which is defined by the expression



(a) $x = 0$ (b) $x = 1/3$

FIG. 11. Comparing the real part of the exact values of $c_p(x)$ with their asymptotic forms. The parameters used in these plots were $a = 1$, $n = 1 + 10^{-13}$, $d = 1/2$ and infrared light at wavelength 1 μm. The test function for these coefficients is Eq. (59).

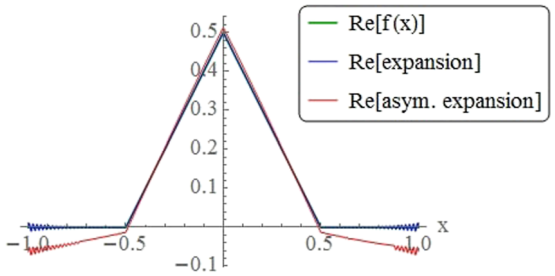


FIG. 12. Comparing the original test function Eq. (59) with its leaky modes expansion and the asymptotic leaky modes expansion. Parameters used in this expansion were $a = 1$, $n = 1 + 10^{-13}$, $d = \frac{1}{2}$, and infrared light at wavelength 1 μm.

$$Li(n,z) = \sum_{p=1}^{\infty} \frac{z^p}{p^n}. \tag{61}$$

If n is strictly larger than 1, the series defining the polylogarithm converge absolutely. After analyzing the various inequalities we find that we get absolute convergence of the leaky mode expansion for the triangle function only if $-a + d < x < a - d$. For the case in Fig. 12, this region is $-1/2 < x < 1/2$. However, we also have a convergence in the entire channel. In the region outside the region of absolute convergence, we also have convergence. The convergence here is ensured by cancellations among terms spiraling toward the origin in the complex plane. In the region of the channel outside the domain $-a + d < x < a - d$, the amplitude of the terms does not decay fast enough ensure absolute convergence, and the cancellation among the spiraling terms are needed for convergence. The resulting convergence is evidently only conditional.

As our second sample function, we pick a Gaussian wave packet

$$f(x) = \exp(-(mx)^2)\exp(ikx), \tag{62}$$

for some real numbers $m, k > 0$. The asymptotic terms $c_p(x)$ for this case are

$$c_p(x) \approx \frac{1}{4a(-1)^{p+1}} \psi_p^-(x) \int_{-a}^{a} \exp(-(mx)^2)\exp(ikx)(B\exp(i\xi_{0p}x) + C\exp(-i\xi_{0p}x)\mathrm{d}x)\mathrm{d}x. \tag{63}$$

Let the parameters m, k be such that $f(x)$ has its support well inside the slab and $f(\pm a) \approx 0$. Then, we can evaluate the integral analytically as

$$c_p(x) \approx \frac{\sqrt{\pi}}{4ma(-1)^{p+1}} \left(B\exp\left(-\frac{(k+\xi_{0p})^2}{4m^2} \right) + C\exp\left(-\frac{(k-\xi_{0p})^2}{4m^2} \right) \right)$$
$$\left((-1)^{p+1} \exp\left(-i\frac{xp\pi}{2a} \right) \left(\frac{p\pi}{a\sqrt{\alpha}} \right)^{-x/a} + \exp\left(i\frac{xp\pi}{2a} \right) \left(\frac{p\pi}{a\sqrt{\alpha}} \right)^{x/a} \right), \tag{64}$$

where ξ_{0p} is defined in (19). The terms $c_p(x)$ in this case decay exponentially and thus ensure that the leaky mode expansion converge for all x in the channel. From Fig. 13 we see that the leaky mode expansion and the exact numerical expansion both are very close to the original Gaussian wave packet for all x in the channel.

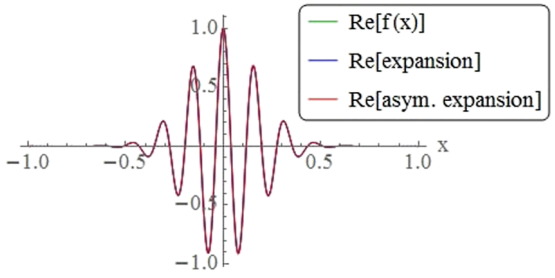


FIG. 13. Comparing the original test function Eq. (62) with its leaky modes expansion and the asymptotic leaky modes expansion. Parameters used in this expansion were $a = 1$, $n = 1 + 10^{-14}$, $m = 4$, $k = 40$ and infrared light at wavelength 1 μm.

We have seen that the expansion we introduced in Eq. (49) can represent a function very well as long as the parameter values are chosen such that $\eta = a^2\alpha \gtrsim 0.8$. These bounds are sufficient for all practical purposes. However, looking at the expansion under such circumstances where the value of α is small enough, we see that the expansion is a very bad representation of the target function. Although we cannot state that we know the reason for this, we have done some preliminary investigations that points to a likely explanation.

Recall that we do have completeness for the scattering modes. Formally this is expressed by the identity

$$\int_{-\infty}^{\infty} \varphi_{\xi_0}(x)\varphi^*_{\xi_0}(x')\mathrm{d}\xi_0 = \delta(x-x'). \tag{65}$$

Here, $\varphi_{\xi_0}(x)$, can be any linear combination of scattering modes. The usual way to get from the completeness for scattering modes to the completeness for the leaky modes is to analytically extend the scattering modes into the complex frequency space and then use the Cauchy theorem. This allows us to write it as a discrete sum of residues evaluated at the poles which are ξ_{0j}. Thus the scattering states get converted into resonant leaky modes at these points leaving us with a sum similar to the one in (49).

In order to be more precise about this, we introduce an integration contour $\mathcal{C}$ in Fig. 14 that contains the zeros ξ_{0j} in the second and fourth quadrant. In the scattering states, the continuity coefficients contain the determinant of the matrix (10), which contains the variable $\xi = \sqrt{\alpha + \xi_0^2}$. This is a complex square root that has a branch cut on the negative real axis. Figure 14 depicts one possible complex contour. We indicated the branch points, where the branch cut begins.

Integrating the integrand in (65) over the contour $\mathcal{C}$, multiplying by $f(x)$ and integrating over the real axis, we get

$$2\pi i \sum_{p=0}^{\infty} \mathrm{Res}\left(\varphi_{\xi_0}(x)\varphi^*_{\xi_0}(x'), \xi_{0p}\right)\left(f(x), \psi^-_{\xi_{0p}}(x)\right)^- = \sum_{p=1}^{\infty} \frac{\left(f(x), \psi^-_{\xi_{0p}}(x)\right)^-}{\left(\psi^-_{\xi_{0p}}(x), \psi^-_{\xi_{0p}}(x)\right)^-}\psi^-_{\xi_{0p}}(x). \tag{66}$$

With the usual boundary conditions for the scattering states we end up with a system where we have one free parameter $a^+(\xi_{0p})$. The expressions on the two sides of (66) are identical only if the free parameter $a^+(\xi_{0p})$ is chosen to be

$$a^+(\xi_{0p}) = \sqrt{\frac{\det{}' \mathbf{M}(\xi_{0p})}{2\pi i\left(\psi^-_{\xi_{0p}}(x), \psi^-_{\xi_{0p}}(x)\right)^- \overline{\det \mathbf{M}}(\xi_{0p})}}. \tag{67}$$

However, this choice for $a^+(\xi_{0p})$ is not an analytic function because both functions $\det{}' \mathbf{M}(\xi_{0p})$ and $\left(\psi^-_{\xi_{0p}}(x), \psi^-_{\xi_{0p}}(x)\right)^-$ are zero inside the integration contour. Each of these two families of countably many zeros give rise to equally many branch cuts. The parametric formulas for these branch cuts are possible to find but while applying the Cauchy theorem we must now include terms representing integrals around all these additional branch cuts. Thus, what we get from the Cauchy theorem is that any function with compact support inside that channel

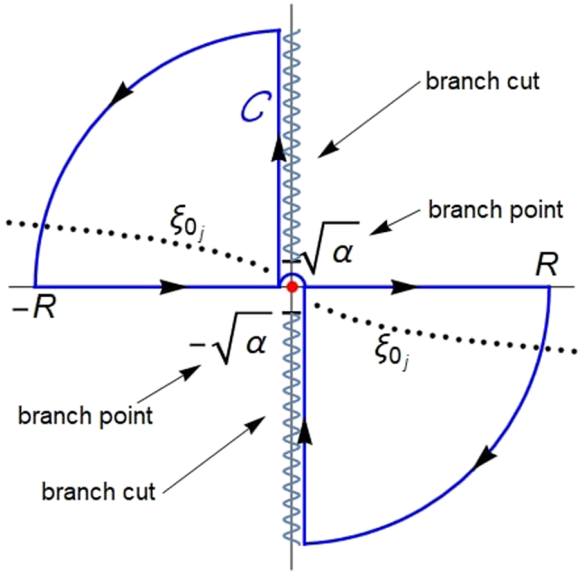


FIG. 14. Complex integration contour $\mathcal{C}$.

is equal to its leaky mode expansion, plus additional terms that includes integrals along the branch cuts on and off the imaginary axis as described above. What we know is that, unless alpha is smaller than the critical value α^*, which we introduced in Sec. IV, the function is well represented by the leaky mode expansion alone. This means that the contribution from all the other terms for such values of alpha are negligible. For smaller values of alpha the contributions from the rest of the terms are not negligible and as $\eta = a^2\alpha$ approaches zero, these terms will come to dominate. For such values of α the leaky mode expansion still converge, but it does not converge to the function used to construct the expansion. By deriving asymptotic formulas for the all the terms defined by integrals around branch cuts, in the limit when η approaches zero, one could compare their sizes and identify the dominant ones. If, say, one term dominate, then this term could be added to the leaky mode expansion resulting in an expansion that represents the function to be expanded in a much better way than the leaky mode expansion is able to do on its own. We believe that the asymptotic expressions for the terms could be found, but there might not be a dominant term, and even if there is, extending the leaky mode expansion by adding this term might easily make the expansion too hard to use for practical calculations.

VI. CONCLUSION

In this paper, we have presented an new approach to minimizing the reflections from finite computational boundaries for wave equations formulated in the spectral domain. This approach is based on representing the field in the transverse spatial direction using leaky mode expansions supported by an artificial index channel. We have shown that at the linear level, our approach makes it possible propagate the waves much further than what is possible if a regular Fourier expansion is used. The leaky modes are not reflection less at the boundary, and eventually the small but finite reflections build up, and the computed solutions starts to deviate from the infinite domain solutions. This reflection can be minimized by reducing the index step, but at the price of getting a progressively worse representation of the solution to the wave equation. In Secs. IV and V, we have argued, using both numerical and analytical approaches, that a practically useful trade-off can be made between minimizing reflections from the boundary and maximizing the accuracy of the representation of solutions of the wave equations using leaky modes.

We have illustrated our approach using the case of a TE electromagnetic wave in vacuum, but the approach can clearly be generalized to much more general wave propagation problems than this. In the optical context, the obvious next step would be to consider waves with cylinder symmetry.

An important issue that we have not discussed in this paper is how to compute the transformation from fields to leaky mode amplitudes and back in an accurate, stable, and efficient way. In order for the method described in this paper to become part of the toolbox of computational optics, this problem has to be addressed.

ACKNOWLEDGMENTS

The authors are thankful for support from the Department of mathematics and statistics at the Arctic University of Norway, from members of the Arizona Center for Mathematical Sciences at the University of Arizona, in particular, Miro Kolesik, and for the support from the Air Force Office for Scientific Research under Grant No. FA9550-19-1-0032.

REFERENCES

[1] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**(2), 185–200 (1994).
[2] W. C. Chew and W. H. Weedon, "A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates," Microwave Opt. Technol. Lett. **7**, 599–604 (1994).
[3] W. P. Huang, C. L. Xu, W. Lui, and K. Yokoyama, "The perfectly matched layer (PML) boundary condition for the beam propagation method," IEEE Photonics Technol. Lett. **8**(5), 649–651 (1996).
[4] A. Couairon, E. Brambilla, T. Corti, D. Majus, O. de J. Ramírez-Góngora, and M. Kolesik, "Practitioner's guide to laser pulse propagation models and simulation," Eur. Phys. J. Spec. Top. **199**(1), 5–76 (2011).
[5] K. Schuh, P. Panagiotopoulos, M. Kolesik, S. W. Koch, and J. V. Moloney, "Multi-terawatt 10 μm pulse atmospheric delivery over multiple Rayleigh ranges," Opt. Lett. **42**(19), 3722–3725 (2017).
[6] J. Chen, A. Suda, E. J. Takahashi, M. Nurhuda, and K. Midorikawa, "Compression of intense ultrashort laser pulses in a gas-filled planar waveguide," Opt. Lett. **33**(24), 2992–2994 (2008).
[7] S. Chen, A. Jarnac, A. Houard, Y. Liu, C. L. Arnold, B. Zhou, B. Forestier, B. Prade, and A. Mysyrowicz, "Compression of high-energy ultrashort laser pulses through an argon-filled tapered planar waveguide," J. Opt. Soc. Am. B **28**(5), 1009–1012 (2011).
[8] T. Popmintchev, M.-C. Chen, D. Popmintchev, P. Arpin, S. Brown, S. Ališauskas, G. Andriukaitis, T. Balčiunas, O. D. Mücke, A. Pugzlys, A. Baltuška, B. Shim, S. E. Schrauth, A. Gaeta, C. Hernández-García, L. Plaja, A. Becker, A. Jaron-Becker, M. M. Murnane, and H. C. Kapteyn, "Bright coherent ultrahigh harmonics in the keV x-ray regime from mid-infrared femtosecond lasers," Science **336**(6086), 1287–1291 (2012).
[9] C. L. Arnold, S. Akturk, M. Franco, A. Couairon, and A. Mysyrowicz, "Compression of ultrashort laser pulses in planar hollow waveguides: A stability analysis," Opt. Express **17**(13), 11122–11129 (2009).
[10] C. L. Arnold, B. Zhou, S. Akturk, S. Chen, A. Couairon, and A. Mysyrowicz, "Pulse compression with planar hollow waveguides: A pathway towards relativistic intensity with table-top lasers," New J. Phys. **12**(7), 073015 (2010).
[11] E. A. J. Marcatili and R. A. Schmeltzer, "Hollow metallic and dielectric waveguides for long distance optical transmission and lasers," Bell Syst. Techn. J. **43**(4), 1783–1809 (1964).

[12]J. J. Thomson, "On electrical oscillations and the effects produces by the motion of an electric sphere," Proc. London Math. Soc. **15**, 197–218 (1884).
[13]J. R. de Lasson, P. T. Kristensen, J. Mørk, and N. Gregersen, "Roundtrip matrix method for calculating the leaky resonant modes of open nanophotonic structures," J. Opt. Soc. Am. A **31**(10), 2142–2151 (2014).
[14]P. T. Kristensen and S. Hughes, "Modes and mode volumes of leaky optical cavities and plasmonic nanoresonators," ACS Photonics **1**(1), 2–10 (2014).
[15]E. F. Franchimon, K. R. Hiremath, R. Stoffer, and M. Hammer, "Interaction of whispering gallery modes in integrated optical microring or microdisk circuits: Hybrid coupled mode theory model," J. Opt. Soc. Am. B **30**(4), 1048–1057 (2013).
[16]A. Settimi, S. Severini, and B. J. Hoenders, "Quasi-normal-modes description of transmission properties for photonic bandgap structures," J. Opt. Soc. Am. B **26**(4), 876–891 (2009).
[17]R.-C. Ge, J. F. Young, and S. Hughes, "Quasi-normal mode approach to the local-field problem in quantum optics," Optica **2**(3), 246–249 (2015).
[18]F. Yang, H. Liu, H. Jia, and Y. Zhong, "Analytical description of quasi-normal mode in resonant plasmonic nano cavities," J. Opt. **18**(3), 035003 (2016).
[19]J. A. J. F. Siegert, "On the derivation of the dispersion formula for nuclear reactions," Phys. Rev. **56**, 750–752 (1939).
[20]R. E. Peierls, "Complex eigenvalues in scattering theory," Proc. R. Soc. London **253**, 16–36 (1959).
[21]K. J. Le Couteur, "The structure of a non-relativistic s-matrix," Proc. R. Soc. London **256**, 115–127 (1960).
[22]J. Humblet, "Theory of nuclear reactions," Nucl. Pys. **26**, 529–578 (1961).
[23]T. Goto, "On the unstable states in quantum field theory," Prog. Theor. Phys. **21**(1), 1–17 (1959).
[24]P. Lin, "Completeness relations and the resonant state expansions," Phys. Rev. C **47**, 1903 (1993).
[25]G. Garcia-Calderon, "An expansion of continuum wave functions in terms of resonant states," Nucl. Phys. A **261**, 130–140 (1976).
[26]O. I. Tolstikhin, V. N. Ostrovsky, and H. Nakamura, "Siegert pseudo state formulation of scattering theory: One-channel case," Phys. Rev. A **58**, 2077–2096 (1998).
[27]O. I. Tolstikhin, "Siegert-state expansion for nonstationary systems: Coupled equations in the one-channel case," Phys. Rev. A **73**, 062705 (2006).
[28]O. I. Tolstikhin, "Siegert-state expansion for nonstationary systems. IV. Three-dimensional case," Phys. Rev. A **77**, 032712 (2008).
[29]O. I. Tolstikhin, V. N. Ostrovsky, and H. Nakamura, "Siegert pseudo-states as a universal tool: Resonances, *S* matrix, green function," Phys. Rev. Lett. **79**, 2026–2029 (1997).
[30]*Resonances the Unifying Route towards the Formulation of Dynamical Processes Foundations and Applications in Nuclear, Atomic and Molecular Physics*, Lecture Notes in Physics, edited by E. Brändas and N. Elander (Erkki Brändas, 1987).
[31]R. de la Madrid, G. García-Calderón, and J. G. Muga, "Resonant expansions in quantum mechanics," Czech J. Phys. **55**, 1141 (2005).
[32]N. Moiseyev, "Quantum theory of resonances: Calculating energies, widths and cross-sections by complex scaling," Phys. Rep. **302**(5-6), 212–293 (1998).
[33]M. Kolesik, J. V. Moloney, and M. Mlejnek, "Unidirectional optical pulse propagation equation," Phys. Rev. Lett. **89**(28), 283902 (2002).
[34]D. Juhasz, M. Kolesik, and P. K. Jakobsen, "Convergence and completeness for square-well Stark resonant state expansions," J. Math. Phys. **59**, 113501 (2018).
[35]J. M. Brown, P. Jakobsen, A. Bahl, J. V. Moloney, and M. Kolesik, "On the convergence of quantum resonant-state expansion," J. Math. Phys. **57**, 032105 (2016).

5 Paper 3

Modelling pulse propagation in complex index materials using the method of multiple scales

Submitted to *Physica Scripta* on 20th of April 2021. Currently in process of revision (as of July 2021, manuscript ID: PHYSSCR-114255).

Modelling pulse propagation in complex index materials using the method of multiple scales

David Juhasz* and Per Kristen Jakobsen*

*Department of Mathematics and Statistics, the Arctic University of Norway, 9019 Tromsø, Norway

July 5, 2021

Abstract

In this paper we study pulse propagation in complex refractive index materials, modelled by a dispersive wave equation, using the method of multiple scales (MMS), and perform several numerical tests to investigate its accuracy. We assume a complex valued refracting index in a noncentrosymmetric medium with a Kerr response. The key feature of our MMS solution is the linearity of the amplitude equation and the complex nature of the mode-frequency. The MMS is tested as an initial value problem using three different dispersion models. Depending on the parameters of the problem, the amplitude equation can be both well- or ill-posed. Despite the ill-posedness, the MMS solution remains a valid approximation of the solution to the original nonlinear model.

1 Introduction

For mathematical models of waves propagating in material media, the phenomenon of *dispersion* is frequently caused by the requirement of causality.

For example, for the case of a light pulse propagating through an isotropic and homogeneous media, the propagating pulse induces a local dipole density, $\mathbf{P}$, which for the simplest cases takes the form

$$\mathbf{P}(\mathbf{x},t) = \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t' \chi(t-t')\mathbf{E}(\mathbf{x},t'), \tag{1}$$

meaning that the polarization at a time t only depends on the electric field at times previous to t. This memory effect, which is the embodiment of causality, is in optics called *temporal dispersion*, or just dispersion. In this paper, in order to make our discussion specific, we will focus on this particular context in our work, but our methods and results apply quite widely to dispersive wave propagation.

The presence of dispersion evidently spells trouble for the integration of the governing equations for the waves. In general, they cannot be solved as an initial value problem.

This phenomenon is of course well known in the theory of wave propagation, and in optics in particular, and various more or less ingenious methods has been invented for getting around the problem.

In optics one frequently tries to get around this problem by solving Maxwell's equations for optical pulse propagation as a boundary value problem, rather as an initial value problem.

In fact, one could argue that the boundary value problem is more closely aligned with the way experiments are done, than the initial value problem is. Waves in a material slab are launched by shining a source laser at the interface of the slab, and therefore one could say that the incoming field at the boundary of the slab is fully controlled by using the laser. Thus, we have the data necessary for solving a boundary value problem for optical pulse propagation.

This is the basic, and in fact the only premise, which underlie the Unidirectional Pulse Propagation Equation(UPPE) [1][2][3] approach to optical pulse propagation. Of course, if there is significant back scatter of light from the interface and/or the material, the boundary condition is not fully controlled, and UPPE, and also other, less general, boundary solvers, which all rely on unidirectionality, are in trouble. For such cases, one can still solve Maxwell's equations for optical pulse propagation as a boundary value problem by using a more general approach than UPPE. This approach is called the Bidirectional Pulse Propagation Equation(BPPE) [4][5].

The BPPE approach is an exact method in the sense that no solutions has been lost when transitioning from Maxwell to BPPE. It is however also a purely numerical method and does not give any analytical insight into the pulse propagation problem. The UPPE is also an exact method, but only if one restricts to solutions of Maxwell that satisfy the condition of unidirectionality, and it is also a purely numerical method.

A much older approach to optical pulse propagation solves Maxwell's equations as an initial value problem by restricting to a class of solutions that are spectrally narrow, also called narrow band solutions. This restriction makes it possible to derive equations, in general called *amplitude equations*, that, for a limited time, give a good approximation to Maxwell's equations, for solutions that are spectrally narrow. These amplitude equations can be solved as initial value problems. The systematic approach for deriving these equations is the method of multiple scales(MMS) [6][7]. In optics, the best known such amplitude equation is the nonlinear Schrödinger equation(NLSE). Another well known amplitude equation, which is particularly useful for driven optical systems like a laser, is the Complex Ginzburg-Landau equation [8].

From a purely numerical point of view, the great thing about amplitude equations is that they are much faster to solve numerically than the original Maxwell's equations. The root cause of this is that for the narrow spectrum solutions represented by the amplitude equations, the fast frequency at the center of the spectrum, the so-called carrier wave, needs not be temporally resolved, only deviations from the center frequency needs to be resolved, and the range of these deviations is by assumption small.

In addition to being fast to solve numerically as an initial value problem, amplitude equations tend to have a universal form, at least to leading order, and this universal form, quite frequently, makes the equation amenable to analytical investigations. In the best cases, a complete analytical solution can be found. This is true for the NLSE equation [9]. However, this analytic solvability is not robust. If we want to extend the amplitude equation beyond the leading order, which we must, if we want an equation that approximates the narrow spectrum solutions to Maxwell for a longer interval of time, the analytic solvability is typically lost, but the fast numerical solvability is not. So, one can say that the important feature of amplitude equations in optics, is that they form the basis for a fast numerical approach for obtaining certain types of solutions to Maxwell's equations.

Apart from the use of amplitude equations, there are also other analytic approaches to simplifying PDE's by restricting to a subset of their full solution space. Two such alternative approaches are the use of center manifolds [10][11] and the renormalization group [12][13].

The aim of this paper is two-fold.

Firstly, we want to go beyond the use of stationary modes as a basis for the MMS expansion, and consider the more general case of modes that are decaying, usually because of loss of some sort in the material, or gain for that matter. In this paper we will focus on the situation of lossy materials.

The case of weak losses can treated by using stationary modes, assuming that the decaying terms are small compared to the leading part of Maxwell's equations. For situations where the loss is too large to be included as a perturbation, one must use a MMS based on decaying modes. This is a situation that occur if one is investigating optical pulse propagation close to a material resonance. It is also the case if one want to derive amplitude equations in near-zero index situations. In fact, this last case is the major motivation for the work we do in this paper. Two interesting features of the resulting amplitude equations, features that we explore in detail later in this paper, is that the amplitude equations we derive are necessarily *linear* and that they are frequently *ill posed* . The interesting thing is that they, nevertheless, accurately represent both the linear and the nonlinear dynamics of the narrow band solutions of Maxwell's equations, for time intervals for which the amplitude equations according to the MMS procedure should approximate such solutions well.

Secondly, we want to investigate the validity of the amplitude equations, as a fast numerical scheme for narrow band solutions to Maxwell's equations, by comparing the numerical solution of the amplitude equation to the corresponding numerical solution of Maxwell's equations. Since we cannot solve Maxwell's equation numerically as an initial value problem, this validity check has an obvious problem that needs to be handled. Handling this problem is the second major focus of this paper. It relies on the fact that in optics, one almost always approximates the linear dispersion of materials using Sellmeier equations [14].

2 A model wave equation, including nonlinearity and general temporal dispersion

The basic model equation we will use to illustrate our methods, and for which we will state our main conclusions, is the simplest nontrivial wave equation from nonlinear optics. It is scalar, and includes general linear dispersion and a Kerr, cubic nonlinear material response. Our methods and conclusions apply much more widely than this, but for the sake of clarity and because numerical methods play an important role in this paper, it is necessary to work within a specific class of equations.

In order to put our model equation into a real physical context, we start this section by deriving the equation from Maxwell's equations under some reasonable physical assumptions on the material response.

Maxwell's equations for a situation where there are no free charges or currents, are given by

$$\begin{aligned} \partial_t \mathbf{B} + \nabla \times \mathbf{E} &= 0, \\ \partial_t \mathbf{D} - \nabla \times \mathbf{H} &= 0, \\ \nabla \cdot \mathbf{D} &= 0, \\ \nabla \cdot \mathbf{B} &= 0. \end{aligned} \tag{2.1}$$

Most materials show no magnetic response at optical frequencies, thus we assume that

$$\begin{aligned} \mathbf{H} &= \frac{1}{\mu_0}\mathbf{B}, \\ \mathbf{D} &= \varepsilon_0 \mathbf{E} + \mathbf{P}. \end{aligned} \tag{2.2}$$

The polarization is in general a sum of a term that is linear in $\mathbf{E}$ and ones that are nonlinear in $\mathbf{E}$. We thus have

$$\mathbf{P} = \mathbf{P}_L + \mathbf{P}_{NL}. \tag{2.3}$$

We assume that the linear material response is isotropic, homogeneous and causal

$$\mathbf{P}_L(\mathbf{x},t) = \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t' \chi(t-t')\mathbf{E}(\mathbf{x},t'). \tag{2.4}$$

In the explicit calculations that we do in this paper, we will assume that the nonlinear polarization is restricted to the Kerr effect. Thus we will assume that

$$\mathbf{P}_{NL} = \varepsilon_0 \eta \mathbf{E}\cdot\mathbf{E}\mathbf{E}, \tag{2.5}$$

where η is the Kerr coefficient. This is a choice we make just to be specific, the applicability of our methods, and the validity of our conclusions, derived in this paper, in no way depend on this particular choice for the nonlinear response.

Inserting (2.2)-(2.5) into (2.1), we can rewrite Maxwell's equations into the form

$$\begin{aligned}
\partial_t \mathbf{B} + \nabla\times\mathbf{E} &= 0,\\
\partial_t \mathbf{E} - c^2\nabla\times\mathbf{B} + \sqrt{2\pi}\partial_t\hat{\chi}(i\partial_t)\mathbf{E} &= -\frac{1}{\varepsilon_0}\partial_t\mathbf{P}_{NL},\\
\nabla\cdot\left(\mathbf{E} + \sqrt{2\pi}\hat{\chi}(i\partial_t)\mathbf{E}\right) &= -\frac{1}{\varepsilon_0}\nabla\cdot\mathbf{P}_{NL},\\
\nabla\cdot\mathbf{B} &= 0,
\end{aligned} \tag{2.6}$$

where we have used an alternative form of (2.4) which is derived in Appendix A. The factor $\sqrt{2\pi}$ is a consequence of our conventions for the Fourier transform.

We will now restrict ourselves to solutions of the form

$$\mathbf{E}(z,t) = E(z,t)\mathbf{e}_y, \tag{2.7}$$

$$\mathbf{B}(z,t) = B_1(z,t)\mathbf{e}_x + B_2(z,t)\mathbf{e}_z, \tag{2.8}$$

$$\mathbf{P}_{NL}(z,t) = P_{NL}(z,t)\mathbf{e}_y,$$

which are called *transverse electric waves*(TE). For this simplified case, Maxwell's equations take the form

$$\begin{aligned}
\partial_t B_1 - \partial_z E &= 0,\\
\partial_t B_2 &= 0,\\
\partial_t E - c^2\partial_z B_1 + \sqrt{2\pi}\partial_t\hat{\chi}(i\partial_t)E &= -\frac{1}{\varepsilon_0}\partial_t P_{NL},\\
\partial_z B_2 &= 0,
\end{aligned} \tag{2.9}$$

where for the assumed Kerr effect we have

$$P_{NL} = \varepsilon_0 \eta E^3. \tag{2.10}$$

By taking cross derivatives it is easy to eliminate the magnetic field components and arrive at the equation

$$\partial_{tt}E - c^2\partial_{zz}E + \sqrt{2\pi}\partial_{tt}\hat{\chi}(i\partial_t)E = -\frac{1}{\varepsilon_0}\partial_{tt}P_{NL}, \tag{2.11}$$

which is the basic model equation we will be using in the rest of this paper.

2.1 Scaling of the model equation

If the aim is to solve Maxwell's equations numerically, in some specific physical context, it is not really necessary to scale the equation, and frequently this is not done, even if one can argue that it could still be a useful thing to do. However, if one is going to derive an amplitude equation for the same physical situation, it might not be essential to scale the equation, but it certainly is extremely useful. After all, the essence of MMS is the ordering of terms in certain expansions according to size, and ensuring that this ordering, according to size, persists, up to some time of our choosing.

Equation (2.11) will be the staring point for our multiple scale approach. Let us start by picking scales Z_0, T_0 and E_0, for space, time and electric field amplitude, so that we have

$$\begin{aligned} z &= Z_0 z', \\ t &= T_0 t', \\ E &= E_0 E', \end{aligned} \tag{2.12}$$

where the primed symbols are the scaled quantities. With these choices of scales the model equation (2.11) takes the form

$$\partial_{t't'} E' - \frac{c^2 T_0^2}{Z_0^2} \partial_{z'z'} E' + \sqrt{2\pi} \partial_{t't'} \hat{\chi}' \left(i\partial_{t'}\right) E' = -\eta E_0^2 \partial_{t't'} E'^3, \tag{2.13}$$

where $\hat{\chi}'\left(i\partial_{t'}\right) = \hat{\chi}\left(i\partial_{t'} \frac{1}{T_0}\right)$. We are at this point free to choose the time scale T_0, so let us choose it so that the factor before the z-derivative becomes one

$$T_0 = \frac{Z_0}{c}. \tag{2.14}$$

We next set the scale for the electric field to be the initial peak electric field amplitude. Thus

$$E_0 = \max_z |E(z,0)|. \tag{2.15}$$

With this we can write the model equation in the form

$$\partial_{tt} E - \partial_{zz} E + \sqrt{2\pi} \partial_{tt} \hat{\chi}(i\partial_t) E = -\varepsilon^2 \partial_{tt} E^3, \tag{2.16}$$

where we now have dropped primes on all quantities, since from this point on, only scaled quantities will appear. In this equation we have introduced the dimensionless parameter $\varepsilon = \sqrt{\eta} E_0$. Typically, the Kerr parameter is fixed for any given material, whereas E_0 is at our disposal to vary over many orders of magnitude, depending on the strength of the laser used to generate the initial electric field. This means that ε can be made to vary over many orders of magnitude, but for realistic field intensities it is always smaller than one, usually much smaller than one. ε is the small perturbation parameter which we need for the MMS expansion.

In this paper, Fourier transforms and plane waves play a prominent role, and whenever that is the case, it is convenient to pick the units for wave number and frequency in such a way that the phase of plane waves, and the Fourier transform, and its inverse, retain the same symbolic form in scaled and unscaled quantities. It is easy to verify that this is the case if we make the choice

$$K_0 = \frac{1}{Z_0}, \tag{2.17}$$

$$\Omega_0 = \frac{1}{T_0}. \tag{2.18}$$

Thus we measure frequency in the well known unit Hertz, or cycles per unit time. Wavelength is in a similar way measured in periods per unit length.

Note that (2.14), which fixes the time scale in terms of the length scale, also, because of (2.18), fix the frequency scale in terms of the wave number scale

$$\Omega_0 = cK_0. \tag{2.19}$$

Thus, the only scale that remains to pick is the one for wave number. The initial field is in a lab situation generated using a laser. These days most optical labs have lasers that produce pulses of femto second duration [15], and labs with more specialized equipment can produce pulses down to atto second durations [16]. Such short pulses have a very broad spectrum and their dynamics are therefore hard to approximate using amplitude equations, which require narrow band pulses. There are versions of MMS that can handle such broad band pulses, but only in the weak dispersion limit. In this paper we apply MMS in a form that is tailored to the opposite limit of strong dispersion. Thus, in this paper we must assume that the initial field has a narrow wave number spectrum centred on a wave length determined by the lasing wave length of the laser generating the initial pulse. What the word "narrow" in the previous sentence means, will be clarified in the MMS expansion in the next section.

In this paper we are focused on validating amplitude equations derived using decaying modes in the vicinity of a material resonance, and it thus makes sense to pick the scale, K_0, for wave number, in such a way that the position of the resonance is centred on scales wavelengths that are of order one. The initial laser pulse will then have a narrow band spectrum centred on a scaled wave length that is also of order one. We will in the rest of the paper assume that this has been done, and from now on only refer to scaled quantities, both in the model equation and in the specific material models that we will introduce in the sections to come.

2.2 Decaying mode, amplitude equation, for the model equation

We proceed with the multiple scale method by introducing the expansions

$$\begin{aligned}
\partial_t &= \partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots, \\
\partial_z &= \partial_{z_0} + \varepsilon\partial_{z_1} + \varepsilon^2\partial_{z_2} + \ldots, \\
e &= e_0 + \varepsilon e_1 + \varepsilon^2 e_2 + \ldots,
\end{aligned} \tag{2.20}$$

where the connection between the multiple scale field amplitude e and the electric field amplitude E is given by

$$E(z,t) = e(z_0, t_0, z_1, t_1, \ldots)\big|_{t_j=\varepsilon^j t, z_j=\varepsilon^j z}. \tag{2.21}$$

The function $\hat{\chi}(i\partial_t)$ is expressed as a Taylor series as follows.

$$\begin{aligned}
\hat{\chi}(i\partial_t) &= \hat{\chi}(i(\partial_{t_0} + \varepsilon\partial_{t_1} + \varepsilon^2\partial_{t_2} + \ldots)) = \hat{\chi}(i\partial_{t_0}) + \hat{\chi}'(i\partial_{t_0})\left(\varepsilon i\partial_{t_1} + \varepsilon^2 i\partial_{t_2} + \ldots\right) \\
&\quad + \frac{\hat{\chi}''(i\partial_{t_0})}{2}\left(\varepsilon i\partial_{t_1} + \varepsilon^2 i\partial_{t_2} + \ldots\right)^2 \\
&= \hat{\chi}(i\partial_{t_0}) + \varepsilon i\partial_{t_1}\hat{\chi}'(i\partial_{t_0}) + \varepsilon^2\left(i\partial_{t_2}\hat{\chi}'(i\partial_{t_0}) - \frac{1}{2}\hat{\chi}''(i\partial_{t_0})\partial_{t_1 t_1} + \ldots\right) + \ldots.
\end{aligned} \tag{2.22}$$

We now insert (2.20), (2.21) and (2.22) into (2.16) and expand everything in sight. This gives us the following perturbation hierarchy

$$\varepsilon^0: \quad \partial_{t_0t_0}e_0 - \partial_{z_0z_0}e_0 + \sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}(i\partial_{t_0})e_0 = 0, \tag{2.23}$$

$$\begin{aligned}\varepsilon^1: \quad \partial_{t_0t_0}e_1 - \partial_{z_0z_0}e_1 + \sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}(i\partial_{t_0})e_1 &= -2\partial_{t_0t_1}e_0 + 2\partial_{z_0z_1}e_0 - 2\sqrt{2\pi}\partial_{t_0}\hat{\chi}(i\partial_{t_0})\partial_{t_1}e_0 \\ &\quad - i\sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}'(i\partial_{t_0})\partial_{t_1}e_0,\end{aligned} \tag{2.24}$$

$$\begin{aligned}\varepsilon^2: \quad \partial_{t_0t_0}e_2 - \partial_{z_0z_0}e_2 + \sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}(i\partial_{t_0})e_2 &= -\partial_{t_1t_1}e_0 - 2\partial_{t_0t_2}e_0 + \partial_{z_1z_1}e_0 + 2\partial_{z_0z_2}e_0 \\ &\quad - \sqrt{2\pi}\hat{\chi}(i\partial_{t_0})\partial_{t_1t_1}e_0 - 2\sqrt{2\pi}\partial_{t_0}\hat{\chi}(i\partial_{t_0})\partial_{t_2}e_0 \\ &\quad - 2i\sqrt{2\pi}\partial_{t_0}\hat{\chi}'(i\partial_{t_0})\partial_{t_1t_1}e_0 - i\sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}'(i\partial_{t_0})\partial_{t_2}e_0 \\ &\quad + \frac{1}{2}\sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}''(i\partial_{t_0})\partial_{t_1t_1}e_0 - \eta E_0^2\partial_{t_0t_0}e_0^3 \\ &\quad - i\sqrt{2\pi}\partial_{t_0t_0t_1}\hat{\chi}'(i\partial_{t_0})e_1 - 2\sqrt{2\pi}\partial_{t_0t_1}\hat{\chi}(i\partial_{t_0})e_1 \\ &\quad - 2\partial_{t_0t_1}e_1 + 2\partial_{z_0z_1}e_1.\end{aligned} \tag{2.25}$$

For the order ε^0 equation we choose the wave packet solution

$$e_0(z_0, t_0, z_1, t_1, \ldots) = A_0(z_1, t_1, \ldots)e^{i\theta_0} + (*), \tag{2.26}$$

where

$$\theta_0 = kz_0 - \omega t_0, \tag{2.27}$$

and where $\omega = \omega(k)$ is a *complex* solution to the dispersion relation

$$\omega^2 n^2(\omega) = k^2. \tag{2.28}$$

Here, the complex refractive index, $n(\omega)$, is defined by

$$n^2(\omega) = 1 + \sqrt{2\pi}\hat{\chi}(\omega). \tag{2.29}$$

Observe that our multiple scale expansion is based on a complex, decaying mode, not a complex, stationary mode, which is usual when one applies MMS far from any resonances of the material. We will see here and also in later sections that this fact will change many aspects of the resulting amplitude equations.

We must now calculate the right-hand side of the order ε^1 equation. Inserting (2.26) into (2.24), we get

$$\begin{aligned}\partial_{t_0t_0}e_1 - \partial_{z_0z_0}e_1 + \sqrt{2\pi}\partial_{t_0t_0}\hat{\chi}(i\partial_{t_0})e_1 &= \Big(2i\omega\partial_{t_1}A_0 + 2ik\partial_{z_1}A_0 + 2i\omega\sqrt{2\pi}\hat{\chi}(\omega)\partial_{t_1}A_0 \\ &\quad + i\omega^2\sqrt{2\pi}\hat{\chi}'(\omega)\partial_{t_1}A_0\Big)e^{i\theta_0} + (*).\end{aligned} \tag{2.30}$$

In order to remove secular terms we must postulate that

$$2ik\partial_{z_1}A_0 + i\left(2\omega + 2\omega\sqrt{2\pi}\hat{\chi}(\omega) + \omega^2\sqrt{2\pi}\hat{\chi}'(\omega)\right)\partial_{t_1}A_0 = 0. \tag{2.31}$$

Observe that from the dispersion relation (2.28) we have

$$\begin{aligned}\omega^2(1 + \sqrt{2\pi}\hat{\chi}(\omega)) &= k^2, \\ &\Downarrow \\ \omega'(k)\left(2\omega + 2\omega\sqrt{2\pi}\hat{\chi}(\omega) + \sqrt{2\pi}\omega^2\hat{\chi}'(\omega)\right) &= 2k.\end{aligned} \tag{2.32}$$

Thus (2.31) can be written in the form

$$\partial_{t_1} A_0 + \omega'(k)\partial_{z_1} A_0 = 0. \tag{2.33}$$

For the case of stationary modes, the quantity $\omega'(k)$ is real and by definition equal to the group velocity for an initial light pulse with a narrow spectrum centred on the wave number k.

The order ε^1 equation simplifies into

$$\partial_{t_0 t_0} e_1 - \partial_{z_0 z_0} e_1 + \sqrt{2\pi}\partial_{t_0 t_0}\hat{\chi}(i\partial_{t_0})e_1 = 0. \tag{2.34}$$

At this point we face a choice; which solution should we pick for this equation? The equation is homogeneous, and thus does not have any nontrivial particular solution, like the equation for e_2, at the next order, do.

Here we pick the simplest possible solution

$$e_1 = 0, \tag{2.35}$$

for (2.34). The discussion of why we make this pick here, and what the consequences would be to make another less trivial choice, is best postponed until after we complete the derivation of the amplitude equation at order ε^2.

We now must compute the right-hand side of the order ε^2 equation. Inserting (2.26) and (2.35) into the right-hand side of the order ε^2 equation we get

$$\begin{aligned}
\partial_{t_0 t_0} e_2 - \partial_{z_0 z_0} e_2 + \sqrt{2\pi}\partial_{t_0 t_0}\hat{\chi}(i\partial_{t_0})e_2 = (&-\partial_{t_1 t_1} A_0 + 2i\omega\partial_{t_2} A_0 + \partial_{z_1 z_1} A_0 + 2ik\partial_{z_2} A_0 \\
&- \sqrt{2\pi}\hat{\chi}(\omega)\partial_{t_1 t_1} A_0 + i2\omega\sqrt{2\pi}\hat{\chi}(\omega)\partial_{t_2} A_0 \\
&- 2\omega\sqrt{2\pi}\hat{\chi}'(\omega)\partial_{t_1 t_1} A_0 + i\omega^2\sqrt{2\pi}\hat{\chi}'(\omega)\partial_{t_2} A_0 \\
&- \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)\partial_{t_1 t_1} A_0\Big) e^{i\theta_0} + 9\omega^2 A_0^3 e^{i3\theta_0} \\
&+ 3(\omega_i - 2\omega)^2 |A_0|^2 A_0 e^{i\theta_0} e^{2t_0\omega_i} + (*),
\end{aligned} \tag{2.36}$$

where $\omega_i = \operatorname{Im}\omega$. At this point it is worth observing that none of the nonlinear terms in (2.36) are secular. In addition to the usual nonsecular term $e^{i3\theta_0}$, we also have the term $e^{i\theta_0}e^{2t_0\omega_i}$, which would have been included into the secular terms, and thus, in the final amplitude equation, if it wasn't for the complex nature of ω. This makes MMS based on decaying modes essentially different from the case of stationary modes.

Continuing the calculation, we observe that in order to remove secular terms from (2.36), we must postulate that

$$\begin{aligned}
&- \partial_{t_1 t_1} A_0 + 2i\omega\partial_{t_2} A_0 + \partial_{z_1 z_1} A_0 + 2ik\partial_{z_2} A_0 - \sqrt{2\pi}\hat{\chi}(\omega)\partial_{t_1 t_1} A_0 + i2\omega\sqrt{2\pi}\hat{\chi}(\omega)\partial_{t_2} A_0 \\
&- 2\omega\sqrt{2\pi}\hat{\chi}'(\omega)\partial_{t_1 t_1} A_0 + i\omega^2\sqrt{2\pi}\hat{\chi}'(\omega)\partial_{t_2} A_0 - \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)\partial_{t_1 t_1} A_0 = 0, \\
&\Updownarrow \\
&\partial_{t_1 t_1} A_0 \left(-1 - \sqrt{2\pi}\hat{\chi}(\omega) - 2\omega\sqrt{2\pi}\hat{\chi}'(\omega) - \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)\right) \\
&+ \partial_{t_2} A_0 \left(2i\omega + i2\omega\sqrt{2\pi}\hat{\chi}(\omega) + i\omega^2\sqrt{2\pi}\hat{\chi}'(\omega)\right) + \partial_{z_1 z_1} A_0 + 2ik\partial_{z_2} A_0 = 0.
\end{aligned} \tag{2.37}$$

The factor multiplying the term $\partial_{t_2} A_0$ can be expressed using (2.32), and equation (2.37) therefore simplifies as follows

$$\partial_{t_1 t_1} A_0 \left(-1 - \sqrt{2\pi}\hat{\chi}(\omega) - 2\omega\sqrt{2\pi}\hat{\chi}'(\omega) - \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega) \right) + \partial_{t_2} A_0 \frac{2ik}{\omega'(k)} + \partial_{z_1 z_1} A_0 + 2ik\partial_{z_2} A_0 = 0,$$

$$\Downarrow$$

$$\partial_{t_2} A_0 + \omega'(k)\partial_{z_2} A_0 - i\beta\partial_{z_1 z_1} A_0 + i\alpha\partial_{t_1 t_1} A_0 = 0, \tag{2.38}$$

where

$$\alpha = \omega'(k)\frac{n^2(\omega) + 2\omega\sqrt{2\pi}\hat{\chi}'(\omega) + \frac{1}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)}{2k}, \tag{2.39}$$

$$\beta = \frac{\omega'(k)}{2k}. \tag{2.40}$$

By removing the secular terms from equation (2.36), the order ε^2 equation turns into

$$\partial_{t_0 t_0} e_2 - \partial_{z_0 z_0} e_2 + \sqrt{2\pi}\partial_{t_0 t_0}\hat{\chi}(i\partial_{t_0}) e_2 = 9\omega^2 A_0^3 e^{i3\theta_0} + 3(\omega + 2i\omega_i)^2 |A_0|^2 A_0 e^{i\theta_0} e^{2t_0\omega_i} + (*). \tag{2.41}$$

This equation is not homogeneous, and we chose at this point to solve for e_2, using only a particular solution. One such particular solution is evidently

$$e_2(z_0, t_0, \ldots) = c_1 A_0^3 e^{i3\theta_0} + c_2 |A_0|^2 A_0 e^{i\theta_0} e^{2t_0\omega_i} + (*), \tag{2.42}$$

where

$$c_1 = \frac{\omega^2}{k^2 - \omega^2 n^2(3\omega)} = \frac{1}{n^2(\omega) - n^2(3\omega)}, \tag{2.43}$$

$$c_2 = \frac{3(\omega + 2i\omega_i)^2}{k^2 - (1 + \hat{\chi}(\omega + i2\omega_i))(\omega + 2i\omega_i)^2} \tag{2.44}$$

Defining an amplitude $A(z, t)$ by

$$A(z, t) = A_0(z_1, t_1, \ldots)\big|_{t_j = \varepsilon^j t, z_j = \varepsilon^j z}, \tag{2.45}$$

and proceeding in the usual way, using (2.33) and (2.38) we finally get the following amplitude equation

$$\partial_t A + \omega'(k)\partial_z A - i\beta\partial_{zz} A + i\alpha\partial_{tt} A = 0. \tag{2.46}$$

The amplitude A is related to the electric field amplitude E through the formula

$$E(z, t) = A(z, t)e^{i(kz - \omega t)} + \varepsilon^2 (c_1 A^3(z, t)e^{i3(kz - \omega t)} + c_2 |A(z, t)|^2 A(z, t) e^{i(kz - \omega t)} e^{2t\omega_i}) + (*). \tag{2.47}$$

By design, for (2.46), (2.47) to be an approximate solution to (2.16), we must have

$$\beta\partial_{zz} A \sim \alpha\partial_{tt} A \sim \mathcal{O}(\varepsilon^2), \quad \partial_t A \sim \partial_z A \sim \mathcal{O}(\varepsilon), \tag{2.48}$$

where we recall that ε is a number much smaller than 1. Given these circumstances, we observe that

$$\begin{aligned} \partial_t A &= -\omega'(k)\partial_z A \sim \mathcal{O}(\varepsilon), \\ &\Downarrow \\ \partial_{tt} A &= \omega'(k)^2 \partial_{zz} A \sim \mathcal{O}(\varepsilon^2), \end{aligned} \tag{2.49}$$

and thus, the second order in time, amplitude equation (2.46), is asymptotically equivalent to the more convenient, first order in time, equation

$$\partial_t A + \omega'(k)\partial_z A - i\left(\beta - \alpha\left(\omega'(k)\right)^2\right)\partial_{zz} A = 0. \tag{2.50}$$

This amplitude equation, together with relation (2.47), are the two key elements defining a fast numerical scheme for narrow band solutions to (2.16).

Note that this amplitude equation is a linear equation. This is very different from the nonlinear Schrödinger equation, which is the leading order amplitude equation for our model equation far from any material resonances. The solution to the amplitude equation (2.50), can be an accurate approximation for our nonlinear model equation, despite the linearity of the amplitude equation, because the expression (2.47) that connects the amplitude A to the electric field amplitude E, is nonlinear. Note that if we want to have an amplitude equation which is a good approximation to the model equation beyond a time of order ε^{-2}, we have to extend the MMS procedure to higher order in ε. It is evident from what we have said about secular terms in the paragraph following equation (2.1), that these extended amplitude equations will all be linear, no matter to which order the MMS procedure is extended.

Since the amplitude equation is linear, the solution space, S_A, is of course a linear space. The relation (2.47) amounts to a map, M, from S_A into the solution space, S_E, of the exact equation (2.16). The map is certainly not surjective, and neither is it injective. The lack of injectivity means that the map cannot be used to induce a nonlinear superposition principle on its image set, $M(S_E) \subset S_E$, using the usual pullback/pushforward approach.

In order to get the approximate solution to our model equation defined by the amplitude equation (2.50), and the relation (2.47), we decided to pick particular solutions at order ε and ε^2, at both orders disregarding the general solution to the homogeneous equation. The consequence of adding a solution to the homogeneous equation, in the form of wave packets, at one or both orders, would be to add one or two new independent amplitudes to the problem. Each of these amplitudes would, through the removal of secular terms at order ε, ε^2 and ε^3, satisfy their own amplitude equations. Both these extra amplitude equations would also be linear, and uncoupled from each other and the one for the amplitude A. The relation defining the electric field in terms of the amplitudes would now be much more complicated and involve sums of products of all three amplitudes.

The deciding factor for whether we include these extra amplitudes, or not, is what kind of solutions of the model equation we are trying to approximate. This comes down to which kind of initial conditions for the model equation we are able to represent faithfully using our amplitude equation.

Any choice of a narrow band initial amplitude for the amplitude equation (2.50), will lead to a narrow band wave packet solution of the model equation, with dispersive properties determined by the choice of a complex solution $\omega = \omega(k)$ to the dispersion relation (2.28), at order ε^0 of our MMS expansion. This is certainly a valid choice of initial condition, if our aim is to validate the amplitude equation (2.50) together with its corresponding defining relation (2.47) for E.

However, such a choice of initial condition for E is awkward from a physical point of view. Usually the initial field is determined by a laser, whose output, in most cases, can be approximated spectrally by a narrow Gaussian centred at the operating wave length of the laser. The initial condition for

E, defined above using a narrow band initial condition for A, consists of two Gaussians centred at k and $3k$, and with a special relationship between the amplitudes and phases of the two Gaussians.

From the relation (2.47), it is evident that in order for E to be a Gaussian centred on some wave number k, the amplitude A must be a Gaussian centred on $k = 0$, and we have to introduce a solution to the homogeneous solution at order ε^2 of the form

$$e_2(z_0, t_0, z_1, t_1, \ldots) = B_0(z_1, t_1, \ldots)e^{i(3kz_0 - \omega(3k)t_0)} + (*), \tag{2.51}$$

leading, in the end, to a relation determining the electric field in terms of the amplitudes A and

$$B(z,t) = B_0(z_1, t_1, \ldots)\big|_{t_j=\varepsilon^j t, z_j=\varepsilon^j z},$$

of the form

$$\begin{aligned} E(z,t) = A(z,t)e^{i(kz-\omega t)} + \varepsilon^2((B + c_1A^3(z,t))e^{i3(kz-\omega t)} \\ +c_2|A(z,t)|^2A(z,t)e^{i(kz-\omega t)}e^{2t\omega_i}) + (*). \end{aligned} \tag{2.52}$$

By fixing the initial value of the amplitude B to be

$$B(z,0) = -c_1A^3(z,0)), \tag{2.53}$$

we ensure that the initial spectrum for E is a Gaussian, whose center and width, is determined by the initial spectrum for the amplitude A.

The initial value for the amplitude A is found by solving

$$E(z,0) = A(z,0)e^{ikz} + \varepsilon^2 c_2|A(z,0)|^2A(z,0)e^{ikz} + (*), \tag{2.54}$$

for A using for example Newton's method.

Since our aim is to get an approximation to the solution of the model equation to order ε^2 for times $t \leq \varepsilon^{-2}$, we need to remove the secular terms generated by the amplitude B, at order ε^3, since such terms would grow linearly and potentially disturb the spectrum at $3k$, already for times of order ε^{-1}.

From detailed calculations done while deriving the amplitude equation (2.50), it is not hard to see that the resulting amplitude equation for the amplitude B must take the form

$$\partial_t B + \omega'(3k)\partial_z B = 0. \tag{2.55}$$

For the validity tests we will discussing later in the paper, we will always introduce such an extra amplitude B, whenever it is necessary for faithfully representing a Gaussian initial spectrum for E.

3 Testing the validity of the amplitude equations

The first goal of this paper was to use the MMS approach to derive the amplitude equation close to a material resonance for nonlinear wave equations with arbitrary, but of course causal, material response. Our main model equation and motivation comes from the field of nonlinear optics, but our methods and results can evidently be applied much wider than this.

Our second goal, which is the focus in this section and the rest of the paper, is to test the validity of the derived amplitude equations using high precision numerical simulations. If one is far from a material resonance, the dispersion relation of the material can to a good approximation be modelled by a polynomial function of the frequency, usually a second order polynomial is sufficient like in the

Lorentz model [17]. In this situation, a direct simulation of the model equation can fairly easily be achieved, and the validity of the amplitude equation can thus be tested. However, even in this situation, numerical validation is not all that common. One reason for this is that the narrow band solutions, which are well approximated by the amplitude equations, are also the type of solutions to Maxwell that are most challenging to simulate numerically. This is because such solutions, which basically are wave packets, have a slowly varying, and consequently very wide, envelope, and at the same time contain a very large number of oscillations under the envelope. Thus one need both a large computational domain and a very high resolution of that domain. This makes for a large number discretization points and thus long running times.

When one is close to a material resonance, low order polynomial approximations do not work as well as when one is far from a resonance. In this situation different and more complicated approximations must be used. This typically turns the model equation, which is a differential equation in time, into an equation that is a pseudo-differential equation in time. Solving such a thing as an initial value problem is not an easy proposition.

A frequently used class of such, more general approximations, are the rational functions. This type of approximations are in particular very much used for approximating the electric susceptibility in optics. For this situation, one can exactly transform the model equation into a equation that is a differential equation in time, and whose initial value problem can be solved numerically. The transformation is based on what is called the Sellmeier formulas in optics, and it thus makes sense to denote the associated transformation for the *Sellmeier transformation*.

3.1 The Sellmeier Transformation

The Sellmeier formulas in optics are approximations of the electric susceptibility in terms of sums of rational functions of the simple type

$$R(\omega) = \frac{1}{a\omega^2 + b\omega + c}, \tag{3.1.1}$$

where a, b and c are complex constants. Causality, and questions of loss or gain in the material, put restrictions on the constants that we leave aside for now. Later, when we do our numerical simulation in order to validate the amplitude equations, these restrictions come to the fore. Finite sums of functions of the type (3.1.1) will produce general rational functions. Thus the Sellmeier formulas are simply rational approximations to the electric susceptibility

$$\hat{\chi}(\omega) = \frac{P(\omega)}{Q(\omega)}. \tag{3.1.2}$$

In order to describe the Sellmeier transformation, observe first that (2.16) can be written in the form $\mathcal{L}(E, E^3) = 0$, where $\mathcal{L}$ is a suitable operator that produces our equation (2.16) and which includes the integral operator $\hat{\chi}(i\partial_t)$. Secondly, observe that the Fourier transform of our model equation can be rewritten into the form

$$\begin{aligned} \hat{\mathcal{L}}\left(E, E^3\right) &= 0, \\ &\Updownarrow \\ \frac{1}{Q(\omega)}\left[Q(\omega)\hat{\mathcal{L}}\left(E, E^3\right)\right] &= 0, \\ &\Updownarrow \\ \frac{1}{Q(\omega)}\,\hat{\hat{\mathcal{L}}}\left(E, E^3\right) &= 0, \end{aligned} \tag{3.1.3}$$

where $\hat{\tilde{\mathcal{L}}} = Q(\omega)\hat{\mathcal{L}}$.

Assuming the form of the susceptibility to be described by the very simplest Sellmeier formula (3.1.1), $\hat{\chi}(\omega) = R(\omega)$, the operator $\tilde{\mathcal{L}}$ is the differential operator

$$\tilde{\mathcal{L}} = \left(c + ib\partial_t - a\partial_t^2\right) \circ \mathcal{L}. \tag{3.1.4}$$

The equation $\tilde{\mathcal{L}}\left(E, E^3\right) = 0$ is the Sellmeier transformation of the original equation $\mathcal{L}(E, E^3) = 0$.

It is evident from (3.1.3) that any solution to the differential equation

$$\tilde{\mathcal{L}}\left(E, E^3\right) = 0, \tag{3.1.5}$$

is also a solution to the original pseudo differential equation (2.16). If S_M denotes the space of solutions for this original model equation and S_D denote the space of solutions for the Sellmeier transformed equation (3.1.5), we thus have $S_D \subset S_M$. We are not going to do a detailed mathematical analysis of situations where the Sellmeier transformation breaks down, these are situations where $Q(\omega)$ pass through zero, or is close to zero on a region of positive measure. This situation is very rarely realized for ordinary materials.

Our idea is now to restrict the numerical validation to solutions in the smaller solution space S_D. This of course only make sense if the MMS procedure applied to the Sellmeier transformed equation (3.1.5) produces the exact same amplitude equation as the one we got earlier, starting from the original pseudo differential equation (2.16). From our short discussion of the relation between S_M and S_D in the previous paragraph, we certainly expect to get the same amplitude equation, but we still feel that it is prudent to directly verify that this is the case for each of the two explicit examples of Sellmeier formulas discussed in the following sections.

3.2 A toy model for dispersion

In this first validation calculation for our amplitude equation (2.50) and associated reconstructed electric field amplitude (2.47) we chose a material response function of the simple form

$$\chi(t) = \begin{cases} ue^{-vt}, & t > 0 \\ 0, & t < 0 \end{cases}, \tag{3.2.1}$$

for some real positive constants u and v. The corresponding electric susceptibility, which is the Fourier transform of (3.2.1), is given by the formula

$$\hat{\chi}(\omega) = \frac{u}{\sqrt{2\pi}} \frac{1}{v - i\omega}. \tag{3.2.2}$$

Defining parameters $\gamma = \frac{v}{u}, a = \frac{1}{u}$, the formula for the susceptibility can be written in the more convenient form

$$\hat{\chi}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{\gamma - ia\omega}. \tag{3.2.3}$$

We are not aware of any material response that in the optical regime is described by this electric susceptibility, but it does satisfy the, all important Kramer-Kronig relations, and thus describe a causal response. For us, this choice is a device for testing the amplitude equation in the simplest setting possible. Later in the paper we will investigate the validity of the amplitude equation when the electric susceptibility describes a very common atomic model, the Lorentz oscillator. For this case, the validation is conceptually the same as for the simple model above, but it is technically much harder.

3.2.1 The Sellmeier transformation

The model for electric susceptibility described above, leads to a Sellmeier transformed equation that, in our scheme of things, is as simple as possible to handle numerically. Following the procedure described in the previous section, we find that the Sellmeier transform of our model equation is

$$a\partial_{ttt}E + (\gamma+1)\partial_{tt}E - \gamma\partial_{zz}E - a\partial_{zzt}E + \varepsilon^2\gamma\partial_{tt}E^3 + \varepsilon^2 a\partial_{ttt}E^3 = 0. \tag{3.2.4}$$

This equation is, as expected, a differential equation and not a pseudo differential equation. It can be solved numerically using, for example a pseudo spectral method, which is what we do in this paper. As noted earlier, for the validation test to make sense, we must ensure that the MMS procedure applied to (3.2.4) gives us the same amplitude equation (2.50), which we got from the original model equation (2.16). We do this is Appendix B and observe there that the derived amplitude equation is indeed the same as the original one (2.50).

3.2.2 Numerical results

Actually solving the Sellmeier transformed equation (3.2.4), is awkward, because it is not solved explicitly with respect to the highest derivative. However, the fact that ε is small means that, by iteration, we can approximate (3.2.4) by an equation that *is* solved explicitly with respect to highest derivative. We achieve this by expanding out the offending term $a\partial_{ttt}E^3$ as follows

$$a\partial_{ttt}E^3 = a\left(6\left(\partial_t E\right)^3 + 18E\partial_t E\partial_{tt}E + 3E^2\partial_{ttt}E\right), \tag{3.2.5}$$

and substituting for $\partial_{ttt}E$ using (3.2.4).

Dropping terms of order ε^4, which we must do in order to be consistent with our order ε^2 MMS expansion, we get the following explicit equation

$$\begin{aligned} &a\partial_{ttt}E + (\gamma+1)\partial_{tt}E - \gamma\partial_{zz}E - a\partial_{zzt}E + \eta E_0^2\gamma\left(6E\left(\partial_t E\right)^2 + 3E^2\partial_{tt}E\right) \\ &+ \varepsilon^2 a\left[6\left(\partial_t E\right)^3 + 18E\partial_t E\partial_{tt}E + \frac{3E^2}{a}\left(\gamma\partial_{zz}E + a\partial_{zzt}E - (\gamma+1)\partial_{tt}E\right)\right] = 0. \end{aligned} \tag{3.2.6}$$

Since the explicit equation (3.2.6) agrees with the exact implicit equation (3.2.4) to order ε^2, the amplitude equation to order ε^2 for these two equations must be the same. Thus we can use (3.2.6) to validate our original amplitude equation (2.50).

Recall that the amplitude equation is derived using one specific mode of the linearization. Thus, to do the validation we must select one of these modes. Given this mode, all the parameters in the amplitude equation are determined in terms of the parameters of (3.2.6). The detailed expressions are derived in Appendix B. This linear mode, which is an explicit function of z and t, can now be used to determine all the three initial conditions needed to solve (3.2.6) numerically, and thus to complete the validation.

The equation (3.2.6) has three independent modes. The solution to the linear part of (3.2.6) can be expressed as inverse Fourier integrals of all three modes

$$E(z,t) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty}\mathrm{d}kA_1(k)e^{i(kz-\omega_1(k)t)} + \int_{-\infty}^{\infty}\mathrm{d}kA_2(k)e^{i(kz-\omega_2(k)t)} + \int_{-\infty}^{\infty}\mathrm{d}kA_3(k)e^{i(kz-\omega_3(k)t)}\right). \tag{3.2.7}$$

The electric field $E(z,t)$ must be real so (3.2.7) must be equal to its complex conjugate

$$E^*(z,t) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty} \mathrm{d}k A_1^*(-k)e^{i(kz+\omega_1^*(-k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_2^*(-k)e^{i(kz+\omega_2^*(-k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_3^*(-k)e^{i(kz+\omega_3^*(-k))t)}\right). \tag{3.2.8}$$

Let us first assume that $k > 0$. From the form of the dispersion relation (B.6)

$$p(k,\omega) = a\omega^3 + i\omega(\gamma+1) - ak^2\omega - i\gamma k^2, \tag{3.2.9}$$

we observe that if $\omega(k)$ is a solution to $p(k,\omega) = 0$, then we have

$$\begin{aligned} p(k,-\omega^*) &= -a(\omega^*)^3 - i\omega^*(\gamma+1) + ak^2\omega^* - i\gamma k^2 \\ &= -\left[a\omega^3 + i\omega(\gamma+1) - ak^2\omega - i\gamma k^2\right]^* = 0. \end{aligned} \tag{3.2.10}$$

This implies that if $\omega(k)$ is a solution, then $-\omega^*(k)$ is a solution as well. Thus, assuming that the three solutions are distinct, we can number them in such a way that

$$\omega_1^*(k) = -\omega_2(k), \tag{3.2.11}$$

$$\omega_3^*(k) = -\omega_3(k). \tag{3.2.12}$$

Next observe that the dispersion relation (B.6) is even in k. Using this fact, we can number the solutions for negative k in such a way that $\omega_{1,2,3}(k) = \omega_{1,2,3}(-k)$. Using these relations, together with (3.2.11) and (3.2.12), formula (3.2.8) turns to

$$E^*(z,t) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty} \mathrm{d}k A_1^*(-k)e^{i(kz-\omega_2(k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_2^*(-k)e^{i(kz-\omega_1(k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_3^*(-k)e^{i(kz-\omega_3(k))t)}\right). \tag{3.2.13}$$

For (3.2.7) and (3.2.13) to be the same, the following relations between the amplitudes $A_{1,2,3}(k)$ must hold

$$\begin{aligned} A_1(k) &= A_2^*(-k), \\ A_3(k) &= A_3^*(-k). \end{aligned} \tag{3.2.14}$$

From these relations, we see that the amplitudes for $k < 0$ are determined from their values for $k > 0$ and vice versa.

While deriving the asymptotic solution in Appendix B, we assumed the solution to the 0-th order equation in the perturbation hierarchy (B.1) to be one of the three possible modes. The amplitude equation (2.50) is the correct equation only for narrow band solutions consisting of one such mode, centred around wave number k_0 which we, without loss of generality, can assume is positive. To be specific, assume that this mode is $A_1(k)$. The narrow band property implies that $A_1(k) = 0$ for $k < 0$, and from this the relations (3.2.14) implies that we can consistently choose $A_3(k) = 0$ for all k and $A_2(k) = 0$ for $k > 0$.

Using these assumptions and (3.2.14), we get from (3.2.13)

$$\begin{aligned} E(z,t) &= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty} \mathrm{d}k A_1^*(-k)e^{i(kz-\omega_2(k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_2^*(-k)e^{i(kz-\omega_1(k)t)}\right) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{d}k A_1(k)e^{i(kz-\omega_1(k)t)} + \int_{-\infty}^{\infty} \mathrm{d}k A_1^*(k)e^{-i(kz-\omega_1^*(k)t)} \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{d}k \underbrace{\left[A_1(k)e^{-i\omega(k)t} + A_1^*(-k)e^{i\omega^*(-k)t}\right]}_{\hat{E}(k,t)} e^{ikz} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{d}k \hat{E}(k,t)e^{ikz}, \end{aligned} \tag{3.2.15}$$

where we let $\omega_1(k) = \omega(k)$ to match the parameters in our asymptotic solution (2.47) with (3.2.15).

We now turn back to equation (3.2.6). Upon transforming this PDE into its spectral domain k, we get a third order ODE with a nonlinear right-hand side. We want to solve this ODE as an initial value problem, and therefore need three initial conditions, here denoted by $f(k), g(k)$ and $h(k)$. These three initial conditions we obtain from (3.2.15)

$$\begin{aligned}\hat{f}(k) &= \hat{E}(k,0),\\ \hat{g}(k) &= \partial_t \hat{E}(k,0),\\ \hat{h}(k) &= \partial_{tt}\hat{E}(k,0).\end{aligned} \tag{3.2.16}$$

These conditions can be also expressed in terms of the amplitude $A_1(k)$ using (3.2.15)

$$\hat{f}(k) = A_1(k) + A_1^*(-k), \tag{3.2.17}$$

$$\hat{g}(k) = -i\omega(k)A_1(k) + i\omega^*(-k)A_1^*(-k), \tag{3.2.18}$$

$$\hat{h}(k) = -\omega^2(k)A_1(k) - \left(\omega^*(-k)\right)^2 A_1^*(-k). \tag{3.2.19}$$

The only thing left to do now, is to compute the initial condition for the amplitude equation (2.50) in terms of our chosen amplitude $A_1(k)$.

As discussed at the end of section 2.2, the initial condition for E, which is natural from a physical point of view, is one whose spectrum is a narrow Gaussian. This is taken care of by letting the spectral amplitude $A_1(k)$ be a Gaussian centered at some wave number k_0

$$A_1(k) = \begin{cases} De^{-\delta(k-k_0)^2}, & k > 0 \\ 0, & k < 0 \end{cases}, \tag{3.2.20}$$

where $D, \delta > 0$.

As also discussed at the end of section 2.3, we might have to introduce an extra amplitude at order ε^2 in order to faithfully represent the initial condition on E in terms of an initial condition for the amplitude equation.

The way to determine if any extra amplitude has to be introduced at order ε^2, is to assume the opposite. Given this, the relation between the amplitude A and the electric field E is determined by (2.47).

We now take the inverse Fourier transform of equation (3.2.17) and equate it to the right-hand side of (2.47) evaluated at $t = 0$. Matching separately the first part and the second part, which is the complex conjugates of the first part, on both sides, we get

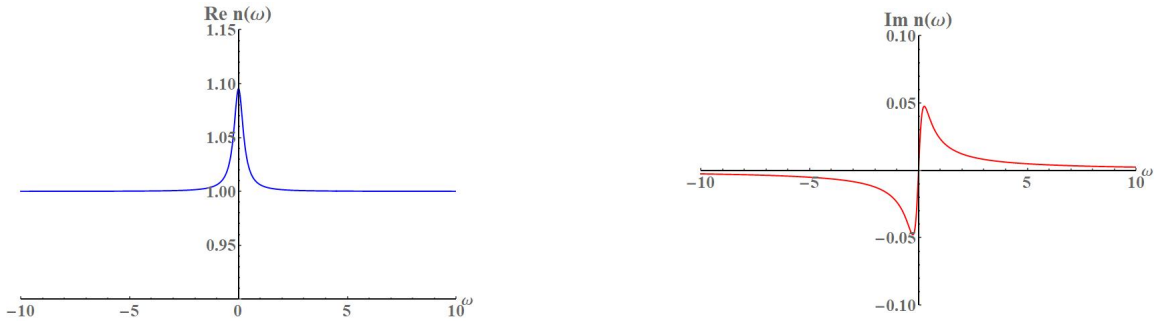
$$\begin{aligned}\mathcal{F}^{-1}\left\{A_1(k)\right\} &= A(z,0)e^{ik_0 z} + c_1\eta E_0^2 A^3(z,0)e^{i3k_0 z} + c_2\eta E_0^2|A(z,0)|^2 A^*(z,0)e^{ik_0 z},\\ \mathcal{F}^{-1}\left\{A_1^*(-k)\right\} &= A^*(z,0)e^{-ik_0 z} + c_1^*\eta E_0^2\left(A^*\right)^3(z,0)e^{-i3k_0 z} + c_2^*\eta E_0^2|A(z,0)|^2 A(z,0)e^{-ik_0 z},\end{aligned} \tag{3.2.21}$$

which is a nonlinear system of algebraic equations, consisting of two equations and two unknowns $A(z,0)$ and $A^*(z,0)$. This system can easily be solved numerically, for example using Newtons method.

Note that the parameter δ controls the width of $A_1(k)$. Given that the amplitude $A(z,t)$ should be slowly varying in z, $\partial_z A \sim \mathcal{O}(\varepsilon)$, the parameter δ should be chosen such that $\delta \sim 1/\varepsilon$.

In order to do a numerical comparison, we fix the refractive index by choosing the parameter values $\gamma = 5$ and $a = 20$. In figure 1 we see the real and imaginary part of the resulting refractive index of our material.

At this point, the only remaining parameter to set is the, all important, nonlinearity parameter ε. In all our derivation, we assumed that this parameter was small, only then will there be separation of scales, which is the key assumption underlying our derivations of amplitude equations using MMS.



(a) (b)

Figure 1: Refractive index $n(\omega)$ for the toy model after scaling. The parameters in this figure are $\gamma = 5$ and $a = 20$.

Usually, one is deriving amplitude equations in order to gain some analytical insight into the wave propagation problem at hand, and the actual numerical value of ε does not need to be fixed. Here however, where we are doing a numerical validation of the amplitude equations, the situation is different, here we need to give a specific value to the nonlinearity parameter. And, with respect to this, we face a tradeoff.

On the one hand, we should pick a value for ε that is as small as possible, in order for the separation of scales to be as large as possible. Only for such values can we expect our amplitude equations to accurately approximate the original model equation.

On the other hand, a very small value for the nonlinearity parameter means that we have to propagate the waves for very long distances for the nonlinearity in the model equation to influence the spectrum of the waves. Very long distances mean very long times which translates into very long running time for the simulations of the model equation.

In this paper we choose the value for the nonlinearity parameter to be $\varepsilon = 10^{-1}$. This certainly does not seem to be a very small number, but as we will see, even for a nonlinearity parameter as large as this, our amplitude equation does actually approximate the exact model equation well. This is just another example of what one could call the phenomenon of the *unreasonable accuracy of asymptotic methods*.

As it turns out, our chosen value for the nonlinearity parameter is also physically reasonable. For example, for a light pulse in a visible part of the spectrum, at 586 nm, propagating through Argon gas at atmospheric pressure, of an intensity equal to half the critical ionization threshold, the value for the nonlinearity parameter is $\varepsilon = 0.13$ [18].

Let us now point our attention to calculating the correct initial condition for A. In order for the amplitude equation (2.50) to be valid, the amplitude A must be spectrally narrow. In fact, the whole MMS expansion is based on the assumption that the width of the spectrum of A is of order epsilon.

The spectrum for A which we find assuming that no new amplitude is required at order ε^2 is displayed in figure 2. The parameters used to calculate the initial condition for A were $a = 20$ and $\gamma = 5$.

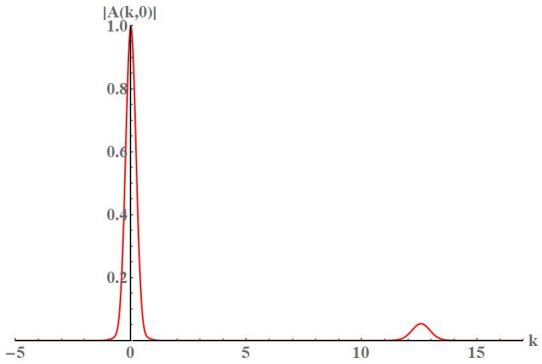


Figure 2: The Fourier transform of the initial conditions $A(k,0)$ obtained with Newtons iterative method from the equations (3.2.21).

Evidently, the width of this spectrum is not of order ε. The main part of the spectrum is narrow, but there is an additional peak in the spectrum centred at $2k_0$ whose distance from the origin is of order one, not order ε. Thus, assuming that no new amplitude is needed at order ε^2 leads to a contradiction.

From relation (2.47) we observe that the peak in the spectrum for A corresponds to a peak at $3k_0$ for E. Furthermore, from figure 2 we observe that the height of the peak is of order ε^2. From these two observations it is evident that the peak in the spectrum for A can be taken into account by introducing an extra amplitude B in the MMS expansion at order ε^2. As we outlined at the end of section 2.2, the amplitude B will come equipped with its own linear amplitude equation, decoupled from the one for A.

The initial values for A and B are now found by solving equation (2.54) for A, using the approach from (3.2.21), and then using the identity (2.53) to determine the initial value for B in terms of the one for A.

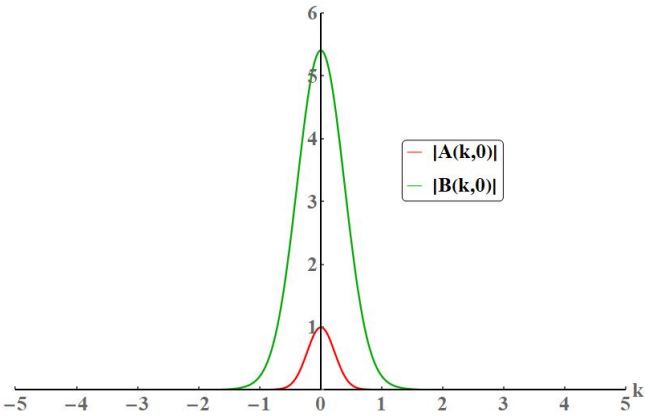


Figure 3: The Fourier transform of the initial conditions $A(k,0)$ and $B(k,0)$ obtained with Newtons iterative method from the equations (3.2.21) and (2.53).

In figure 3 we display an example of the initial condition $A(k,0)$. In this example we use the value $k_0 = 2\pi$. Note that this means that our initial pulse spectrum is located to the right of the frequency defining the material resonance, in a region of anomalous dispersion, which means that the real part of the refractive index decrease for increasing frequency.

Now we have everything we need in order to compare the numerical solution of the model equation (3.2.6) to it's corresponding amplitude equation (2.50).

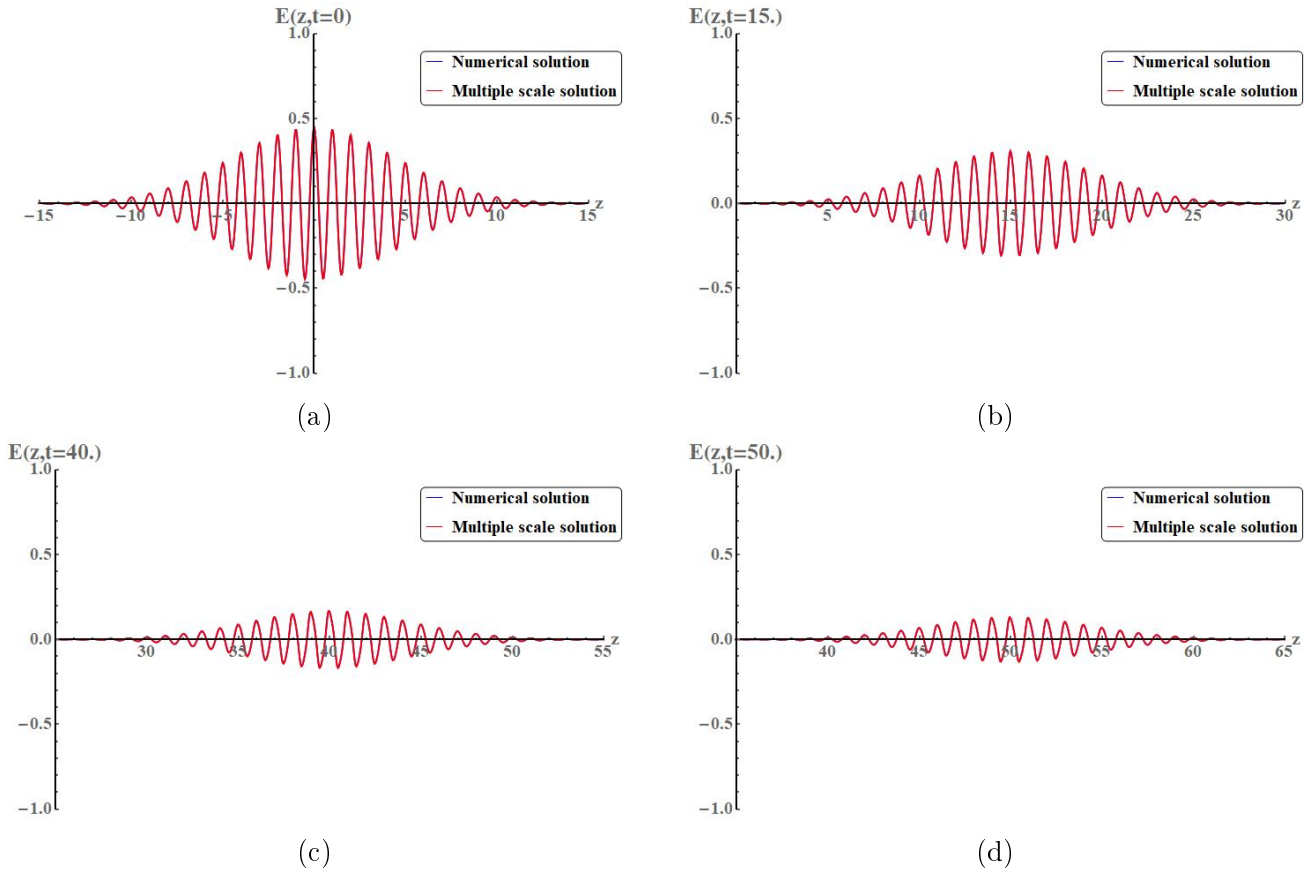


Figure 4: The solution of (3.2.6) using a system of first order ODE's (blue graph) compared with solution (2.47) using the amplitude $A(z,t)$ computed from (2.50) (red graph) for the toy model dispersion.

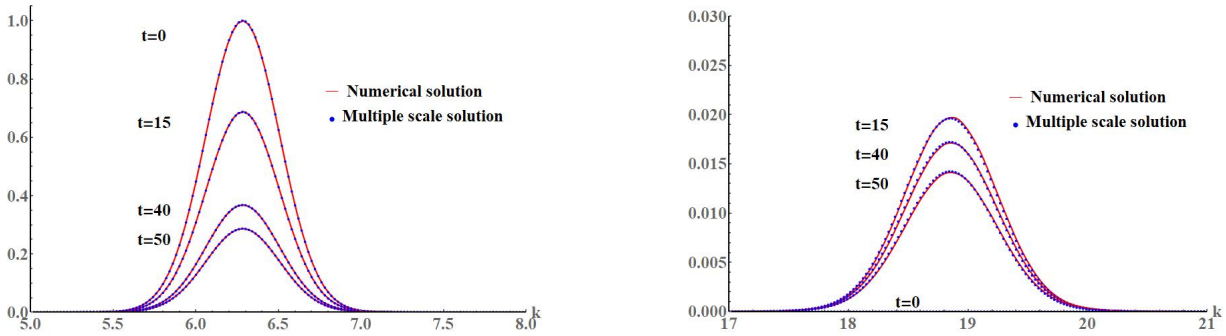
Figure 4 depicts these two solutions at different times. We can see that they are indeed very close. For the values of the parameters used in this comparison, the complex frequency of the chosen mode is given by $\omega_0 = \omega(k_0) = 6.28 - i2.5 \times 10^{-2}$.

In order to get more insight into the accuracy, let us compare the wave number spectrum of these two solutions. This is what is displayed in figure 5, which includes both the major bumps in the k spectrum. The larger one has its support at the frequency $k = k_0 = 2\pi$ as expected. This represents the linear part of the solution. The smaller one is located around the frequency $k = 3k_0 = 6\pi$. It makes sense that it sits precisely at $3k_0$, since the Kerr nonlinearity has the form E^3, which means that when inserting a plane wave for E into E^3, the wave numbers get multiplied by a factor of 3.

From figure 5b it does appear that there is a small deviation between the exact numerical solution and the solution derived from the amplitude equation.This deviation is however one or more orders of magnitude smaller than ε^2, and such deviations are to be expected for an MMS expansion truncated at order ε^2.

Based on the experience from testing the MMS solution, it became clear that in order to stay in the correct asymptotic regime, where the MMS solution is valid, the parameter values fixing the problem are subject to certain constraints.

First of all, the constants α and β occurring in the amplitude equation, defined in (2.39), (2.40), include the derivatives of the susceptibility $\hat{\chi}(\omega)$. These derivatives come from the Taylor expansion of this function. We expect this Taylor series to converge, so by assumption, these derivatives must not break the order of the preceding terms in the expansion. This assumption depends mainly on the parameters γ and a in (3.2.3) and the frequency ω around which the Taylor series is expanded.



(a) (b)

Figure 5: The spectrum of the solutions of (3.2.6) using a system of first order ODE's (blue graph) compared with the spectrum of the solution (2.47) using the amplitude $A(z,t)$ computed from (2.50) (red graph) for the toy model dispersion.

Preserving the order of these terms is closely tied to the possibility of making the assumptions in (2.48) and consequently arriving at the amplitude equation. The assumption made in (2.48) are related also to the choice of the parameter δ in the initial condition (3.2.20). The amplitude $A(z,t)$ is by assumption slowly varying in z compared to the exponential $e^{ik_0 z}$. Therefore the initial condition (3.2.20) should be fast varying in k and the parameter δ needs to be chosen accordingly, for example $\delta = 1/\varepsilon$.

3.2.3 Stability and well-posedness

From the numerical solution of the amplitude equation in the previous section, and its ability to accurately approximate the exact, narrow band, solutions for Maxwell, one might think that our amplitude equation is just fine. However, nothing could be further from the truth. In this section we will show that the amplitude equation is in fact ill posed as a PDE.

Since the amplitude equation is linear, this fact can be easily proven by merely calculating the growth curve for the equation. We find this curve by inserting $A(z,t) = e^{\lambda(k)t}e^{ikz}$ into the equation (2.50), cancelling the common factor $e^{\lambda(k)t}e^{ikz}$, and extracting the real part of the resulting algebraic equation.

$$\begin{aligned} \lambda + ik\omega'(k_0) - i\left(\beta - \alpha\left(\omega'(k_0)\right)^2\right)(ik)^2 = 0, \\ \Updownarrow \\ \lambda(k) = -ik\omega'(k_0) - ik^2\left(\beta - \alpha\left(\omega'(k_0)\right)^2\right), \\ \Updownarrow \\ \text{Re}\,\lambda(k) = a_1 k^2 + a_2 k, \end{aligned} \tag{3.2.22}$$

where a_1 and a_2 are real parameters that depend on the complex parameters α, β and $\omega'(k)$ is the derivative of the dispersion relation (2.28). This function is a parabola which pass through the origin, and is displayed in figure 6, using the parameter values from our numerical test.

From figure 6 it is evident that the amplitude equation is ill posed by definition; arbitrary high spatial frequencies will grow exponentially in time, with no upper bound for the growth rate.

Being ill posed is usually, for good reasons, thought of as ringing the death knell for any proposed mathematical model. But, still, in the situation investigated in the previous section, the ill posed amplitude equation is an excellent tool for simulating narrow band solutions to Maxwell's equation.

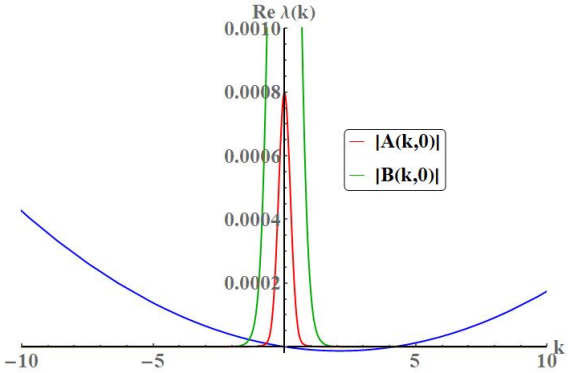


Figure 6: Stability of the amplitude equation (2.50) for the toy model dispersion. The red and the green graphs are representing the amplitudes $|A(k,0)|$ and $|B(k,0)|$, respectively. These function were scaled to fit the figure. The parameter values in this figure are $a_1 = 3 \times 10^{-6}$ and $a_2 = -1.26 \times 10^{-5}$.

We regard this as another example of the way asymptotic methods make use of, and give meaning to, otherwise meaningless mathematical expressions. Anyone that is a user of asymptotic methods can not help noticing this fact. The classical case of this is Eulers example, where the useless power series $\sum_n (-1)^n n!\, x^n$, with zero radius of convergence, is, nevertheless, an accurate approximation to a certain exponential integral. Similarly, the expansion for the energy of the anharmonic quantum oscillator is an asymptotic series with zero radius of convergence [19]. In fact, essentially all series used in quantum theory, in both the particle and the field incarnations of the theory, is known to, or believed to, have zero radius of convergence [20][21]. But, nevertheless, their usefulness is indisputable, their predictions give some of the most accurate correspondences between theory and experiment in all of science.

The reason why the ill posed amplitude equation nevertheless is an accurate numerical model for the narrow band solutions of Maxwell, for which it was designed, is simple. From the form of the stability curve (3.2.22) it is clear that it passes through 0 for every parameter. Combining this with that fact that the initial condition for A will always be centered at $k = 0$, we can conclude that in the case of ill-posedness, the growth rate around $k = 0$ will not be large. For any solution that satisfies the assumptions used to derive the amplitude equation, the fastest growing wave number component in the spectrum of the amplitude $A(z,t)$, will, during the time for which the amplitude equation is valid, $t \leq \varepsilon^{-2}$, not grow large enough to affect the solution to order ε^2 or greater.

3.3 Lorentzian model of dispersion

In the previous section we used the simplest possible rational approximation to the electric susceptibility. As we noted there, this model is, as far as we know, not a realistic physical model, even though it is causal and thus does satisfies the optical Kramer-Kronig relations. In this subsection we consider the next simplest one, which is of the form

$$\hat{\chi}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{\omega_p^2}{\omega_r^2 - \omega^2 - i\gamma\omega}. \tag{3.3.1}$$

This dispersion is a realistic physical model of the electric susceptibility. It can be derived from a purely classical model of an atom consisting of a positive charge, representing the nucleus, together with all the electrons, save one [22]. The remaining electron is singled out by being the one that resonantly responds to an imposed oscillatory electric field. Since the single electron is much lighter

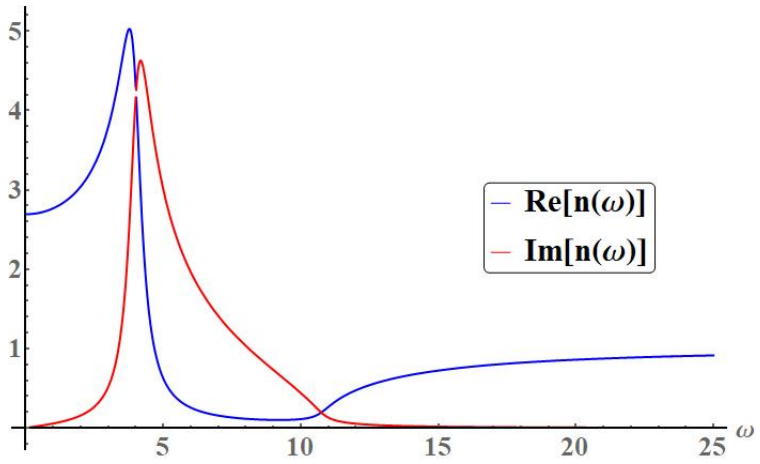


Figure 7: Refractive index $n(\omega)$ for the Lorentzian model of dispersion (3.3.1). The parameters in this figure are $a = -0.01, b = -7 \times 10^{-3}$ and $c = 0.16$.

than the rest of the atom, which after all contains the nucleus, we are in effect describing a simple oscillator, which is a harmonic oscillator, unless the field is strong enough to pull the electron too far from the atom. The model is called the Lorentz oscillator, and have been applied to a vast range of materials in the gaseous, liquid and solid phase. The parameters ω_r, ω_p and γ are interpreted as the resonance frequency of the oscillator, the plasma frequency and the absorptive loss. The factor of 2π is, as noted before, a consequence of our Fourier transform conventions, which are introduced in Appendix A. The Lorentz oscillator model is causal and thus satisfies the Kramer-Kronig relations.

We now repeat the calculations from the previous section for the Lorentz oscillator. We start by finding a differential equation $\tilde{\mathcal{L}}(E, E^3) = 0$, corresponding to our model pseudo differential equation (2.16), using the Sellmeier transformation. After that, we use MMS to verify that we get the same amplitude equation for the Sellmeier transformed equation as the one we got from the original equation (2.16). We then move on to doing a numerical comparison of the accuracy of the amplitude equation with respect to the Sellmeier transformed equation. As noted earlier in this paper, this amounts to a direct comparison of the numerical accuracy of the amplitude equation with respect to the original model equation (2.16), which is the goal of this paper. In our calculations we rewrite the function (3.3.1) in a, for us more convenient, form

$$\hat{\chi}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{\frac{\omega_r^2}{\omega_p^2} - \frac{1}{\omega_p^2}\omega^2 - i\frac{\gamma}{\omega_p^2}\omega} = \frac{1}{\sqrt{2\pi}} \frac{1}{a\omega^2 + ib\omega + c}, \tag{3.3.2}$$

where we have defined $a = -1/\omega_p^2, b = -\gamma/\omega_p^2$ and $c = \omega_r^2/\omega_p^2$. In our case, the refractive index is defined as $n^2(\omega) = 1 + \sqrt{2\pi}\hat{\chi}(\omega)$. In Figure (7) we display an example of such a refractive index, corresponding to the numerical values $a = -0.01, b = -7 \times 10^{-3}$ and $c = 0.16$.

In the next subsections we derive the equation (3.1.3) for the Lorentzian model of dispersion and obtain the MMS solution for it. Next we test the MMS solution using two choices of parameters for the Lorentz model.

3.3.1 The Sellmeier transformation

The Lorentzian model of dispersion leads to the following Sellmeier transformed equation

$$(1+c)\partial_{tt}E - b\partial_{ttt}E - a\partial_{tttt}E + (b\partial_{zzt} - c\partial_{zz} + a\partial_{zztt})E + \varepsilon^2(c\partial_{tt} - b\partial_{ttt} - a\partial_{tttt})E^3 = 0. \tag{3.3.3}$$

Again, the derived equation (3.3.3) does not include any pseudo differential operator. We have verified that the amplitude equation for (3.3.3) matches the one for (2.16).

3.3.2 Numerical results

We proceed by testing the results based on the amplitude equation (2.50) and a solution to (3.3.3) for the Lorentz model of dispersion. Two tests will be conducted, each of them with a different choice of parameters a, b, c for the Lorentz model.

The Sellmeier transformed equation (3.3.3) is implicit in its highest time derivative, but can be approximated by an equation explicit in its highest time derivative. Following the same procedure as in chapter 3.2.2 we have

$$\begin{aligned}
&(1+c)\partial_{tt}E - b\partial_{ttt}E - a\partial_{tttt}E + (b\partial_{zzt} - c\partial_{zz} + a\partial_{zztt})E \\
&= \varepsilon^2 \left[-c\left(6E\left(\partial_t E\right)^2 + 3E^2\partial_t^{(2)}E\right) + b\left(6\left(\partial_t E\right)^3 + 18E\partial_t E\partial_t^{(2)}E + 3E^2\partial_t^{(3)}E\right)\right. \\
&\left. + a\left(3\partial_t^{(4)}EE^2 + 18E\left(\partial_t^{(2)}E\right)^2 + 24\left(\partial_t^{(3)}E\right)E\partial_t E + 36\left(\partial_t E\right)^2\partial_t^{(2)}E\right)\right],
\end{aligned} \tag{3.3.4}$$

where $\partial_t^{(4)}E$ is expressed as

$$\partial_t^{(4)}E = \frac{1}{a}\left((1+c)\partial_{tt}E - b\partial_{ttt}E + (b\partial_{zzt} - c\partial_{zz} + a\partial_{zztt})E\right). \tag{3.3.5}$$

As before, the amplitude equation (2.50) is an equation for one of the four independent modes equation (3.3.4) has. In order to test the MMS solution we need to choose one mode for which we have the amplitude equation (2.50). In order to arrive at the point where we can chose the mode, we proceed like we did in subsection 3.2.2. The electric field $E(z,t)$ is expressed as the inverse Fourier transform of the sum of all 4 modes and is then equated with its complex conjugate. In order to proceed, we need to establish relations between the mode frequencies found by solving the corresponding dispersion equation.

We observe that

$$\begin{aligned}
p(k,\omega) &= a\omega^4 + ib\omega^3 + \omega^2(c+1) - k^2\left(a\omega^2 + ib\omega + c\right) = 0 \\
&\Downarrow \\
p(k,-\omega^*) &= a(\omega^*)^4 - ib(\omega^*)^3 + (\omega^*)^2(c+1) - k^2\left(a(\omega^*)^2 - ib\omega^* + c\right) = 0 \\
&= \left[a\omega^4 + ib\omega^3 + \omega^2(c+1) - k^2\left(a\omega^2 + ib\omega + c\right)\right]^* = 0.
\end{aligned} \tag{3.3.6}$$

Thus like in subsection 3.2.2, solution space of $p(k,\omega) = 0$ is the same as for $p(k,-\omega^*) = 0$. We can therefore conclude that if $\omega(k)$ is a solution, then $-\omega^*(k)$ is as well. Assuming that all solutions are distinct, we can enumerate the four the solutions in such a way that the following relations between $\omega_{1,2,3,4}(k)$ hold:

$$\omega_1^*(-k) = -\omega_2(k), \tag{3.3.7}$$
$$\omega_3^*(-k) = -\omega_4(k). \tag{3.3.8}$$

The equation $p(k,\omega) = 0$ is also even in k, and thus we can enumerate the four solutions corresponding to negative k in such a way that $\omega_i(-k) = \omega_j(k)$, for some i and j. Using this fact, together with the relations (3.3.7), (3.3.8), the realness of the electric field implies the following relations between the amplitudes $A_{1,2,3,4}(k)$:

$$A_1(k) = A_2^*(-k), \tag{3.3.9}$$
$$A_3(k) = A_4^*(-k). \tag{3.3.10}$$

The amplitudes for negative argument are defined from their values for $k > 0$. As before, we want the electric field consisting of only one amplitude, for which we obtained the amplitude equation (2.50), therefore we set $A_2(k) = A_3(k) = A_4(k) = 0$ for $k > 0$. This implies $A_{3,4} = 0, \forall k$. We thus arrive at

$$E(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k \underbrace{\left[A_1(k)e^{-i\omega(k)t} + A_1^*(-k)e^{i\omega^*(-k)t}\right]}_{\hat{E}(k,t)} e^{ikz} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}k \hat{E}(k,t)e^{ikz}. \tag{3.3.11}$$

The amplitude $A_1(k)$ can be chosen arbitrary and it will define the initial condition $E(z,0)$.

The equation (3.3.4) is solved as a 4-th order ODE in the spectral domain k. For this to work, we need four initial conditions, denoted by $\hat{f}_j(k)$ for $j = 0, 1, 2, 3$. These conditions are obtained from (3.3.11) and can be expressed in terms of the given amplitude $A_1(k)$ as

$$\partial_t^{(j)} \hat{E}(k,0) = \hat{f}_j(k) = (-i\omega(k))^j A_1(k) + (i\omega^*(-k))^j A_1^*(-k). \tag{3.3.12}$$

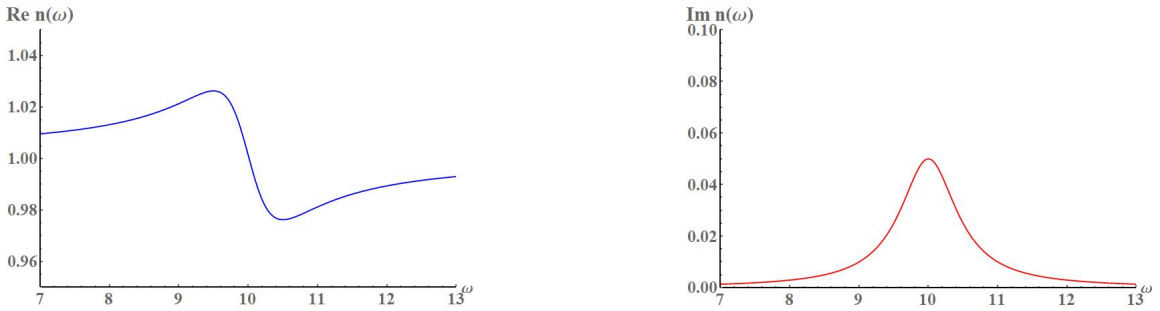
By choosing $A_1(k), k > 0$, all four initial conditions are defined. From the relation (2.47) between $E(z,t)$ and $A(z,t)$, we then calculate the initial condition for the amplitude equation (2.50) in exactly the same way as in (3.2.21). The same procedure also applies in terms of the extra mode $B(z,t)$ whether or not it should be included. As we will see, this extra mode will be included in the following numerical tests.

We proceed to the implementation part and choose the initial amplitude $A_1(k)$ to be the same as for the toy model. A Gaussian centered at the wave number k_0.

$$A_1(k) = \begin{cases} De^{-\delta(k-k_0)^2}, & k > 0 \\ 0, & k < 0 \end{cases}, \tag{3.3.13}$$

for some $D, \delta > 0$. We will use different values for k_0 in the next two numerical tests.

Lorentz test for ultraviolet resonance The model in chapter 3.2.2 resulted in an ill posed amplitude equation that turned out to give us a correct solution within the asymptotic regime for our perturbation scheme. In this test we picked the parameters in such a way that it results in a well posed amplitude equation.



(a) (b)

Figure 8: Refractive index $n(\omega)$ for Lorentz test for ultraviolet resonance. The parameters in this figure are $a = -1, b = -1$ and $c = 100$.

In figure 8 we see the real and imaginary part of the refractive index. The choice of parameters in this figure are

$$\begin{aligned} a &= -1, \\ b &= -1, \\ c &= 100. \end{aligned} \tag{3.3.14}$$

The parameter for the nonlinear term is again chosen to be $\varepsilon = 10^{-1}$.

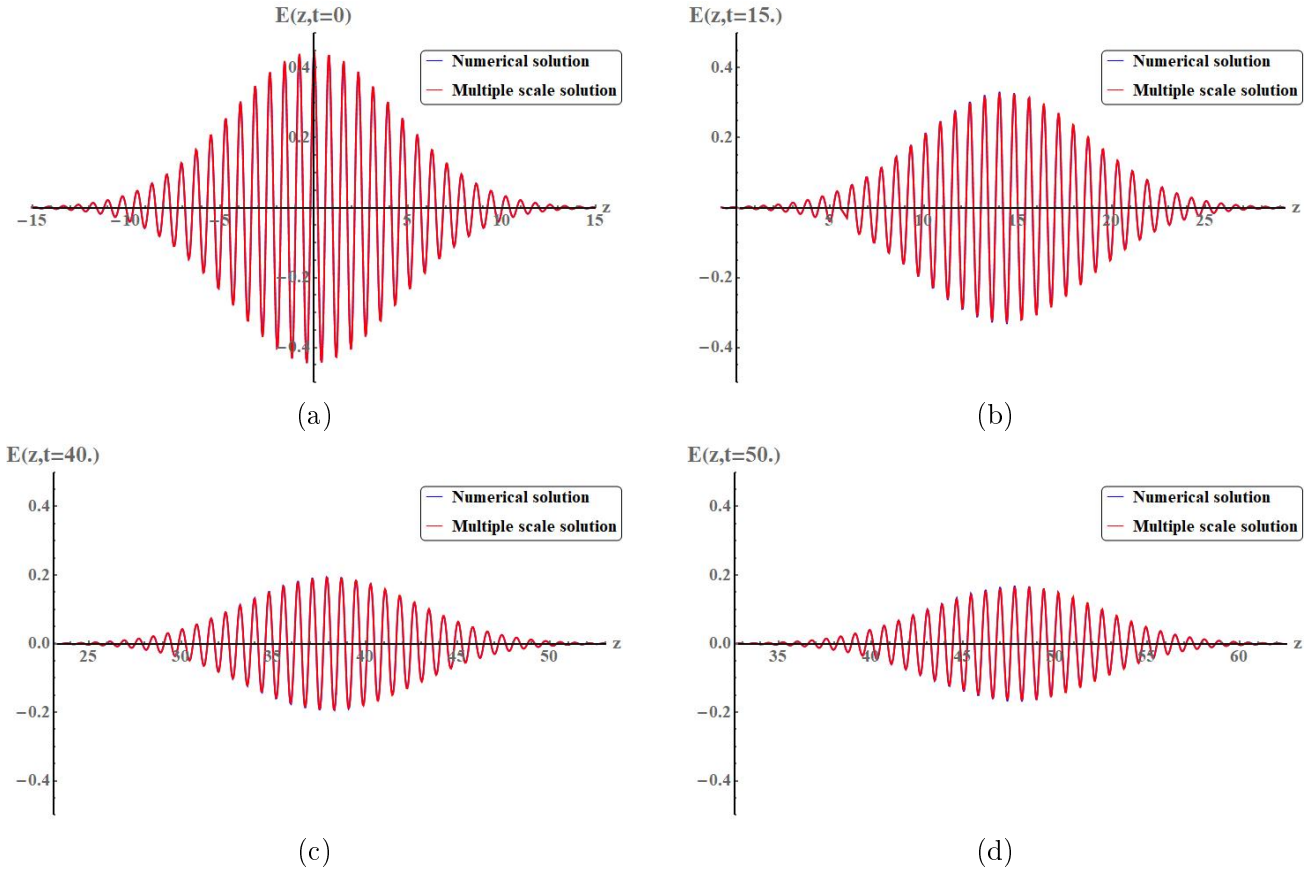


Figure 9: Lorentz test for ultraviolet resonance. The solution of (3.3.4) using a system of first order ODE's (blue function) compared with the MMS solution (2.47) (red function).

When comparing the two solutions from both equations (3.3.4) and the MMS solution in figure 9, we see that they overlap sufficiently at all the presented times. However, comparing the solutions in the spectral domain can reveal the more subtle differences.

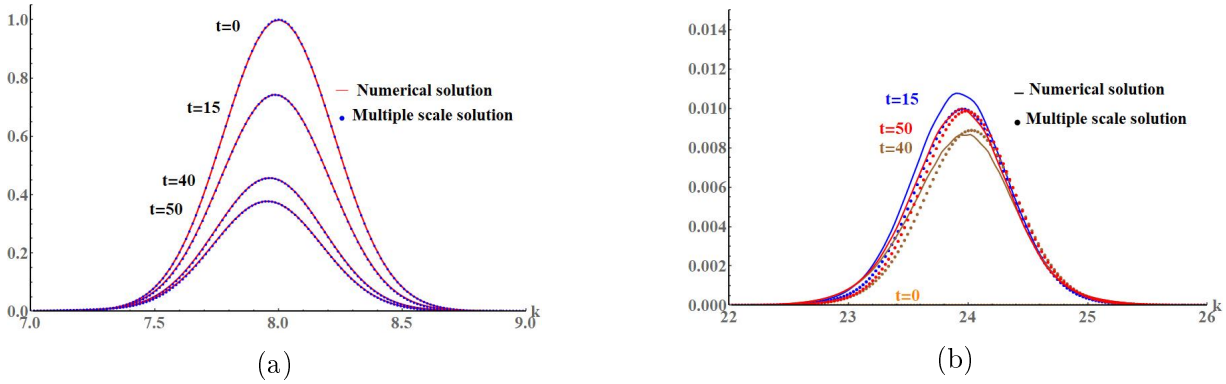


Figure 10: Lorentz test for ultraviolet resonance. A close-up for the spectrum in k of the solutions of (3.3.4) (blue function) compared with the spectrum of the MMS solution (2.47) (red function).

On figure 10 we see that they overlap quite nicely. The parameters used in figures 9 and 10 are

$k_0 = 8, \omega_0 = \omega(k_0) = 7.9 - i1.99 \times 10^{-2}$. The part of the spectrum we are interested in is the smaller gaussian bump centered at $3k_0 = 24$, where the nonlinearity manifests. In order to see the details how well these two functions overlap, we can look at the figure 10 for both the gaussian bumps. Figure 10b is rather convincing us of the accuracy of the MMS solution. On figure 10a is the main gaussian that represents the linear part of the solution. As we can see, the MMS solution is indeed accurate up to the order of ε^2 for both the toy model and the Lorentz model of dispersion. In the next paragraph we run one more numerical test with a different set of parameters.

Lorentz test for infrared resonance Let us now choose the parameters a, b, c for the Lorentz test for infrared resonance. In this model we are using the following parameters in the refracting index:

$$\begin{aligned} a &= -0.25, \\ b &= -10, \\ c &= 1. \end{aligned} \tag{3.3.15}$$

We can see the plot of this refraction index in figure (11). We also notice that the graph of this function is somewhat unusual. The resonance for this function is much closer to 0 than for the index in figure 8. However, the resonance area in the non-scaled version of figure 11 is located at the frequencies that are around 1.5×10^{14} (with the scaling factor $\Omega_0 = 1.5 \times 10^{15}$) which is in the infrared range.

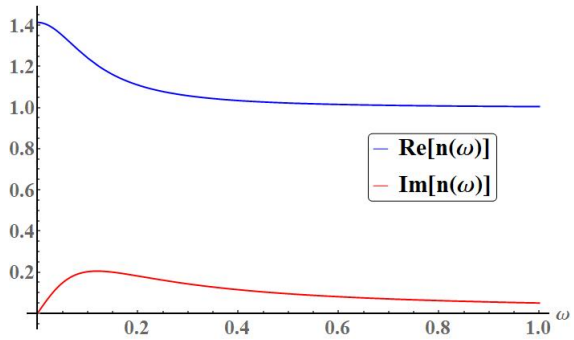
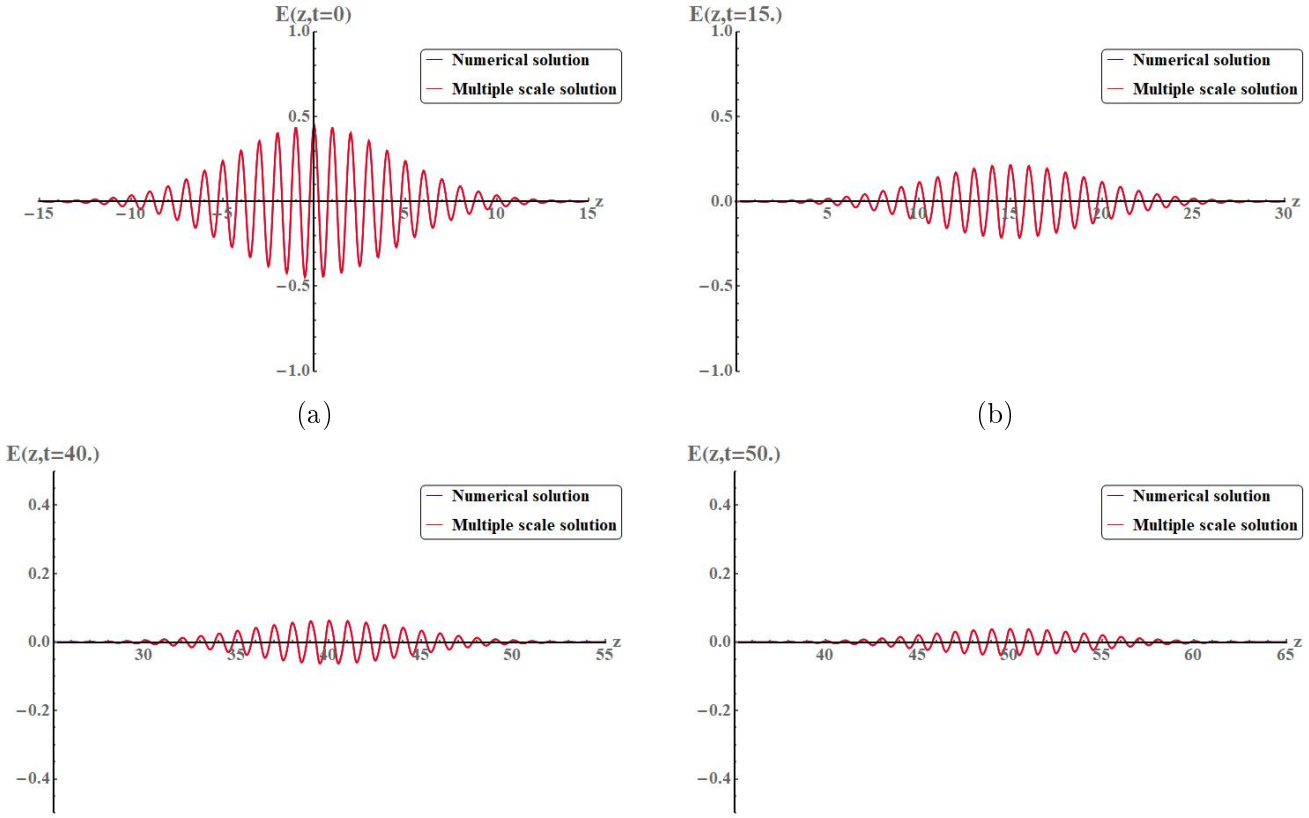


Figure 11: Refractive index $n(\omega)$ for Lorentz test for infrared resonance. The parameters in this figure are $a = -0.25, b = -10$ and $c = 1$.

The initial condition for the amplitude remains the same as in (3.3.13) as well as the parameters δ, D, η and E_0. The refractive index parameters give us the scaled frequency of the pulse obtained from the dispersion relation $\omega_0 = \omega(k_0) = 6.29 - i4.91 \times 10^{-2}$.

As seen from the results in figures (12) and (13) we can conclude that the MMS solution proved itself and it is sufficiently accurate compared to the high precision numerical solution up to the error of order $\varepsilon^2 = 10^{-2}$. The parameters used in figures 12 and 13 are $k_0 = 2\pi, \omega_0 = \omega(k_0) = 6.29 - i4.91 \times 10^{-2}$.

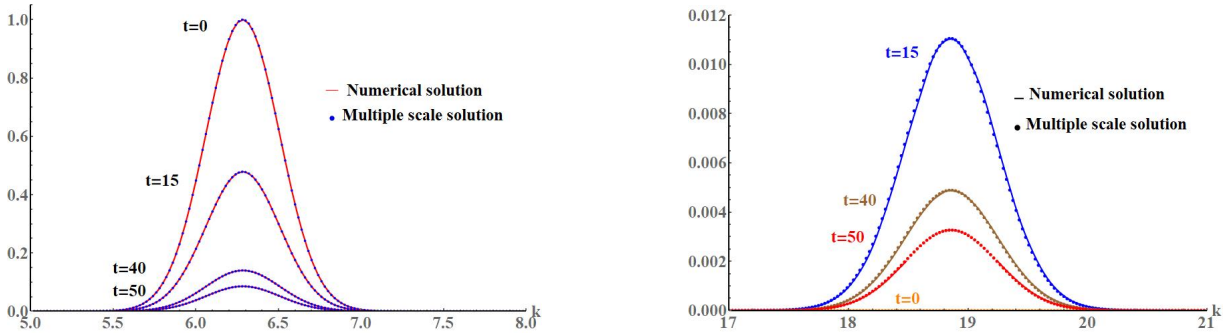
In the next subsection we investigate the stability of the amplitude equation for both the Lorentz tests for ultraviolet and infrared resonance.



(a) (b)

(c) (d)

Figure 12: Lorentz test for infrared resonance. The solution of (3.3.4) using a system of first order ODE's (blue function) compared with the MMS solution (2.47) (red function).

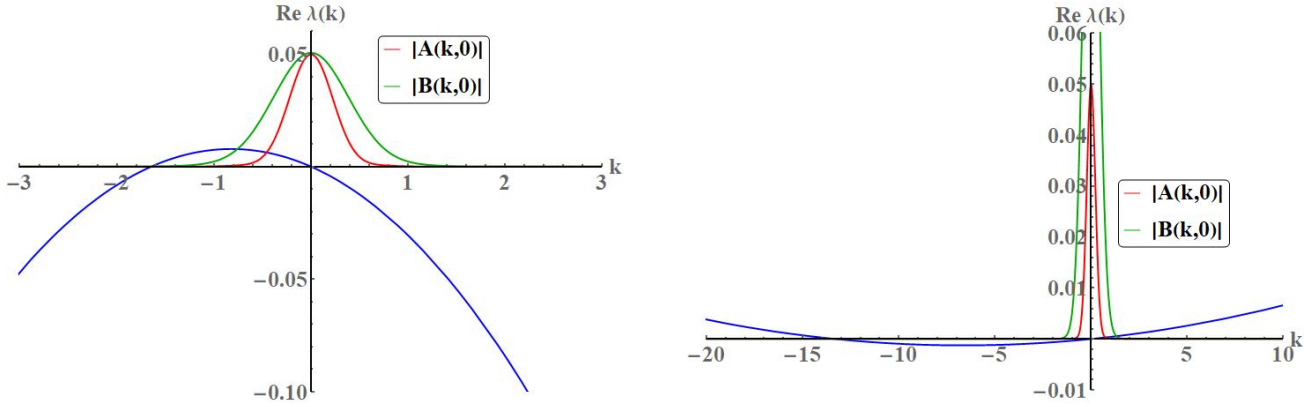


(a) (b)

Figure 13: Lorentz test for infrared resonance. A close-up for the spectrum in k of the solutions of (3.3.4) (blue function) compared with the spectrum of the MMS solution (2.47) (red function).

3.3.3 Stability and well-posedness

In this subsection we look at the stability of the amplitude equation (2.50) with the parameters used in Lorentz tests for ultraviolet and infrared resonance. The formula for the stability is the same as in (3.2.22). On figure (14a) we can see the graph of the function Re $\lambda(k)$ for both tests. As mentioned earlier in Lorentz test for ultraviolet resonance, we chose the parameters for this test such that it



(a) Lorentz test for ultraviolet resonance.

(b) Lorentz test for infrared resonance.

Figure 14: Stability curve for the amplitude equation (2.50). The parameters in 14a are $a_1 = -1.15 \times 10^{-2}$ and $a_2 = -1.9 \times 10^{-2}$ and in 14b are $a_1 = 2.83 \times 10^{-5}$ and $a_2 = 3.77 \times 10^{-4}$.

results in a well posed amplitude equation. Figure 14a confirms this. For the Lorentz test for infrared resonance we got ill posedness again. For both cases there is a small range of k where the amplitude is stable (in the ill posed case) or unstable (in the well posed case). Figure 14a tells us that we do not have to worry about the exponential growth for Lorentz model for ultraviolet resonance whereas the opposite is true for Lorentz model for infrared resonance 14b.

4 Conclusion

In this paper we have successfully derived the MMS solution to Maxwell's equation (2.16) and demonstrated its numerical accuracy. During the process we introduced the Sellmeier transformation that helped us express Maxwell's equation without any pseudodifferential operators. We showed that the obtained MMS solutions provide a good approximation to the solution of the nonlinear Maxwell's equation (2.16) up to order ε^2. The key features of our MMS solutions are the linearity of the amplitude equation and the complex nature of the modes. The linearity made the amplitude equation analytically solvable in a much faster way than the original Maxwell's equation which is one of the main advantages of MMS. On the other hand, in some cases, the amplitude equation turned out to be ill posed. However it does not represent a problem because of the nature of the stability curve and the location of the initial condition.

We did three numerical tests; the first corresponds to a toy model of dispersion, and the other two to the, more physical, Lorentz model of dispersion. The MMS solution performed very well in all three cases. For the two Lorentz cases, we chose the frequency of the initial pulse to be in front of and behind the resonance, where the first case gave us a well posed amplitude equation while the second one turned out to be an ill posed one.

The key idea behind MMS is to maintain the ordering in asymptotic expansions. This ordering also applies to the Taylor expansion of the susceptibility (2.22). It is clear, that this ordering depends on the parameters in $\hat{\chi}(\omega)$ itself, but also on the frequency ω_0 around which we are expanding. In our countless numerical tests we have tried many different combinations of the parameters. We observed that this ordering failed for certain values for the parameters. Especially when one wants to be near or at resonance. The derivatives of $\hat{\chi}(\omega)$ also appear in the constant α, which is required to be of a certain order since it appears in the ε^2 part of the amplitude equation. Alternatively, different methods may be used to expand the susceptibility, for example rational function expansion.

Another problem with our approach that also relates to the values of parameters in $\hat{\chi}(\omega)$, is the growth of the nonlinear term in the Maxwell's equation. As expected, if the height of the spectral peak at $3k_0$ exceeds the order of ε^2 in the numerical solution, our MMS solution fails to maintain its accuracy. This problem, however, occurred less often than the aforementioned violation of ordering in the $\hat{\chi}(\omega)$ expansion.

Further investigations are needed to clarify the issues discussed in the last two paragraphs. Resolving them, by perhaps modifying the way MMS is used in these kind problems, could lead to an MMS solution preserving its accuracy for an even broader set of parameter values.

Data Availability

The data that supports the findings of this study are available within the article (and its supplementary material).

Acknowledgements

The authors are thankful for support from the department of mathematics and statistics at the Arctic University of Norway, from the Arizona Center for Mathematical Sciences at the University of Arizona and for the support from the Air Force Office for Scientific Research under grant # FA9550-16-1-0088.

Appendices

Appendix A

In this appendix we introduce a more convenient representation of the linear polarization (2.4). First, let us mention that our convention for Fourier transform is

$$\begin{aligned} F(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}x f(t) e^{-i\omega t}, \\ f(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}x F(\omega) e^{i\omega t}. \end{aligned} \tag{A.1}$$

Then using (A.1) we have according to (2.4)

$$\begin{aligned}
\mathbf{P}_L &= \varepsilon_0 \int_{-\infty}^{t} \mathrm{d}t' \chi(t-t')\mathbf{E}(\mathbf{x},t'), \\
&= \varepsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\omega \hat{\chi}(\omega)\hat{\mathbf{E}}(\mathbf{x},\omega)e^{-i\omega t}, \\
&= \varepsilon_0 \int_{-\infty}^{\infty} \mathrm{d}\omega \left(\sum_{n=0}^{\infty} \frac{\hat{\chi}(0)}{n!}\omega^n \right) \hat{\mathbf{E}}(\mathbf{x},\omega)e^{-i\omega t}, \\
&= \varepsilon_0 \sum_{n=0}^{\infty} \frac{\hat{\chi}(0)}{n!} \left(\int_{-\infty}^{\infty} \mathrm{d}\omega \omega^n \hat{\mathbf{E}}(\mathbf{x},\omega)e^{-i\omega t} \right), \\
&= \varepsilon_0 \sum_{n=0}^{\infty} \frac{\hat{\chi}(0)}{n!} \left(\int_{-\infty}^{\infty} \mathrm{d}\omega \left(i\partial_t\right)^n \hat{\mathbf{E}}(\mathbf{x},\omega)e^{-i\omega t} \right), \\
&= \varepsilon_0 \sum_{n=0}^{\infty} \frac{\hat{\chi}(0)}{n!} \left(i\partial_t\right)^n \left(\int_{-\infty}^{\infty} \mathrm{d}\omega \left(i\partial_t\right)^n \hat{\mathbf{E}}(\mathbf{x},\omega)e^{-i\omega t} \right), \\
&= \varepsilon_0 \sqrt{2\pi} \hat{\chi}\left(i\partial_t\right) \mathbf{E}(\mathbf{x},t),
\end{aligned} \tag{A.2}$$

where $\hat{\chi}(\omega)$ is the Fourier transform of $\chi(t)$.

Appendix B

In this appendix we derive the amplitude equation and the MMS solution to (3.2.4). The multiple scales were introduced in (2.20). Using these on the equation (3.2.4), we get the following perturbation hierarchy

$$\varepsilon^0 : \quad a\partial_{t_0t_0t_0}e_0 + (\gamma+1)\partial_{t_0t_0}e_0 - a\partial_{z_0z_0t_0}e_0 - \gamma\partial_{z_0z_0}e_0 = 0, \tag{B.1}$$

$$\begin{aligned}
\varepsilon^1 : \quad & a\partial_{t_0t_0t_0}e_1 + (\gamma+1)\partial_{t_0t_0}e_1 - a\partial_{z_0z_0t_0}e_1 - \gamma\partial_{z_0z_0}e_1 = \\
& \left(-2\partial_{t_0t_1} - 3a\partial_{t_0t_0t_1} + a\partial_{z_0z_0t_1} + 2a\partial_{t_0z_0z_1} - 2\gamma\partial_{t_0t_1} + 2\gamma\partial_{z_0z_1}\right)e_0,
\end{aligned} \tag{B.2}$$

$$\begin{aligned}
\varepsilon^2 : \quad & a\partial_{t_0t_0t_0}e_2 + (\gamma+1)\partial_{t_0t_0}e_2 - a\partial_{z_0z_0t_0}e_2 - \gamma\partial_{z_0z_0}e_2 = \\
& \left(-3a\partial_{t_0t_0t_2} - 3a\partial_{t_0t_1t_1} + 2a\partial_{t_0z_0z_2} + a\partial_{t_0z_1z_1} + 2a\partial_{z_0z_1t_1} + a\partial_{z_0z_0t_2}\right. \\
& -2\partial_{t_0t_2}(\gamma+1) - \partial_{t_1t_1}(\gamma+1) + 2\gamma\partial_{z_0z_2} + \gamma\partial_{z_1z_1}\big)e_0 + \left(-3a\partial_{t_0t_0t_1} + 2a\partial_{t_0z_0z_1}\right. \\
& \left. +a\partial_{z_0z_0t_1} - 2\partial_{t_0t_1}(\gamma+1) + 2\gamma\partial_{z_0z_1}\right)e_1 - \left(a\partial_{t_0t_0t_0} + \gamma\partial_{t_0t_0}\right)e_0^3.
\end{aligned} \tag{B.3}$$

For the ε^0 order equation we choose the wave packet solution

$$e_0(z_0,t_0,z_1,t_1,\ldots) = A_0(z_1,t_1,\ldots)e^{i\theta_0} + (*), \tag{B.4}$$

where

$$\theta_0 = kz_0 - \omega t_0, \tag{B.5}$$

and where $\omega = \omega(k)$ is a complex number and a solution to the dispersion relation

$$a\omega^3 + i\omega^2(\gamma+1) - ak^2\omega - i\gamma k^2 = 0. \tag{B.6}$$

Note that k is the scaled initial spatial frequency and is used to obtain the complex frequency ω.

We now proceed to the ε order equation. Inserting (B.4) into (B.2) we get

$$\begin{aligned}&a\partial_{t_0t_0t_0}e_1+(\gamma+1)\partial_{t_0t_0}e_1-a\partial_{z_0z_0t_0}e_1-\gamma\partial_{z_0z_0}e_1=\\&\left(2i\omega\partial_{t_1}(\gamma+1)A_0+3a\omega^2\partial_{t_1}A_0-ak^2\partial_{t_1}A_0+2a\omega k\partial_{z_1}A_0+2ik\gamma\partial_{z_1}A_0\right)e^{i\theta_0}+(*).\end{aligned}\tag{B.7}$$

In order to remove secular terms we postulate that

$$\begin{aligned}2i\omega\partial_{t_1}(\gamma+1)A_0+3a\omega^2\partial_{t_1}A_0-ak^2\partial_{t_1}A_0+2a\omega k\partial_{z_1}A_0+2ik\gamma\partial_{z_1}A_0=0,\\\partial_{t_1}A_0\left(2i\omega(\gamma+1)+3a\omega^2-ak^2\right)+\partial_{z_1}A_0\left(2ak\omega+2i\gamma k\right)=0.\end{aligned}\tag{B.8}$$

Observe that by differentiating the dispersion relation (B.6) with respect to k we get

$$\begin{aligned}a\omega^3+i\omega^2(\gamma+1)-ak^2\omega-i\gamma k^2=0,\\\Downarrow\\3a\omega^2\omega'(k)+i2\omega\omega'(k)(\gamma+1)-a2k\omega-ak^2\omega'(k)-i2\gamma k=0,\\3a\omega^2+i2\omega(\gamma+1)-ak^2=\frac{a2k\omega+i2\gamma k}{\omega'(k)},\end{aligned}\tag{B.9}$$

Using (B.9), the equation (B.8) can be written in the form

$$\partial_{t_1}A_0+\omega'(k)\partial_{z_1}A_0=0.\tag{B.10}$$

The ε order equation (B.7) now simplifies into

$$a\partial_{t_0t_0t_0}e_1+(\gamma+1)\partial_{t_0t_0}e_1-a\partial_{z_0z_0t_0}e_1-\gamma\partial_{z_0z_0}e_1=0.\tag{B.11}$$

We choose the special zero solution to (B.11)

$$e_1=0.\tag{B.12}$$

We now compute the right-hand side of the order ε^2 equation. Inserting (B.12) into the right-hand side of the order ε^2 equation (B.3) we get

$$\begin{aligned}a\partial_{t_0t_0t_0}e_2+(\gamma+1)\partial_{t_0t_0}e_2-a\partial_{z_0z_0t_0}e_2-\gamma\partial_{z_0z_0}e_2=\big(&3a\omega^2\partial_{t_2}A_0+3ia\omega\partial_{t_1t_1}A_0+2ak\omega\partial_{z_2}A_0\\&-ia\omega\partial_{z_1z_1}A_0+2iak\partial_{t_1z_1}A_0-ak^2\partial_{t_2}A_0\\&+2i\omega(\gamma+1)\partial_{t_2}A_0-(\gamma+1)\partial_{t_1t_1}A_0\\&+2i\gamma k\partial_{z_2}A_0+\gamma n_0^2\partial_{z_1z_1}A_0\big)e^{i\theta_0}-NST+(*),\end{aligned}\tag{B.13}$$

where

$$\begin{aligned}NST=a27i\omega^3A_0^3e^{3i\theta_0}+A_0^2A_0^*e^{i\theta_0}e^{2t_0\omega_i}\left(3ia\omega^3-18a\omega^2\omega_i-36ia\omega\omega_i^2+24a\omega_i^3\right)\\-9\gamma\omega^2A_0^2e^{3i\theta_0}+3\gamma A_0^2A_0^*e^{i\theta_0}e^{2t_0\omega_i}\left(2\omega_i-i\omega\right)^2,\end{aligned}\tag{B.14}$$

are the non-secular terms and where $\omega_i=\operatorname{Im}\omega$. In order to remove secular terms we postulate that

$$\begin{aligned}&3a\omega^2\partial_{t_2}A_0+3ia\omega\partial_{t_1t_1}A_0+2ak\omega\partial_{z_2}A_0-ia\omega\partial_{z_1z_1}A_0+2iak\partial_{t_1z_1}A_0-ak^2\partial_{t_2}A_0\\&+2i\omega(\gamma+1)\partial_{t_2}A_0-(\gamma+1)\partial_{t_1t_1}A_0+2i\gamma k\partial_{z_2}A_0+\gamma\partial_{z_1z_1}A_0=0,\\&\Downarrow\\&\partial_{t_2}A_0\left(3a\omega^2-ak^2+2i\omega(\gamma+1)\right)+\partial_{t_1t_1}A_0\left(3ia\omega-(\gamma+1)\right)+\partial_{z_2}A_0\left(2ak\omega+2i\gamma k\right)\\&+2iak\partial_{t_1z_1}A_0+\partial_{z_1z_1}A_0\left(\gamma-ia\omega\right)=0.\end{aligned}\tag{B.15}$$

From the equation (B.8) we express the $\partial_{t_1 z_1}$ derivative in terms of $\partial_{t_1 t_1}$ in the following way

$$\begin{aligned}
\partial_{t_1} A_0 \left(2i\omega(\gamma+1) + 3a\omega^2 - ak^2\right) &= -\partial_{z_1} A_0 \left(2ak\omega + 2i\gamma k\right) \qquad /\partial_{t_1}, \\
\partial_{t_1 t_1} A_0 \frac{ak^2 - 2i\omega(\gamma+1) - 3a\omega^2}{2ak\omega + 2i\gamma k} &= \partial_{t_1 z_1} A_0.
\end{aligned} \tag{B.16}$$

Substituting (B.16) back to (B.15) and using the relation (B.9) by then term $\partial_{t_2} A_0$ we get

$$\begin{aligned}
&\partial_{t_2} A_0 \frac{a2k\omega + i2\gamma k}{\omega'(k)} + \partial_{t_1 t_1} A_0 \left(3ia\omega - (\gamma+1)\right) + \partial_{z_2} A_0 \left(2ak\omega + 2i\gamma k\right) \\
&+ 2iak \frac{ak^2 - 2i\omega(\gamma+1) - 3a\omega^2}{2ak\omega + 2i\gamma k} \partial_{t_1 t_1} A_0 + \partial_{z_1 z_1} A_0 \left(\gamma - ia\omega\right) = 0, \\
&\qquad\qquad \Downarrow \\
&\partial_{t_2} A_0 + \omega'(k) \partial_{z_2} A_0 + \frac{\omega'(k)}{a2k\omega + i2\gamma k} \left(3ia\omega - (\gamma+1) + 2iak \frac{ak^2 - 2i\omega(\gamma+1) - 3a\omega^2}{2ak\omega + 2i\gamma k}\right) \partial_{t_1 t_1} A_0 \\
&- i \frac{\omega'(k)}{2k} \partial_{z_1 z_1} A_0 = 0.
\end{aligned} \tag{B.17}$$

Let us deal with the factor by the term $\partial_{t_1 t_1} A_0$ separately.

$$\begin{aligned}
&\frac{\omega'(k)}{a2k\omega + i2\gamma k} \left(3ia\omega - (\gamma+1) + 2iak \frac{ak^2 - 2i\omega(\gamma+1) - 3a\omega^2}{2ak\omega + 2i\gamma k}\right) \\
&= \frac{i\omega'(k)}{2k} \left(\frac{\gamma + 1 - 3ia\omega}{\gamma - ia\omega} + \frac{a\left(2i\omega(\gamma+1) + 3a\omega^2 - ak^2\right)}{(\gamma - ia\omega)^2}\right) \\
&= \frac{i\omega'(k)}{2k} \frac{\left(\gamma + 1 - 3ia\omega\right)\left(\gamma - ia\omega\right) + a2i\omega(\gamma+1) + 3a^2\omega^2 - a^2k^2}{(\gamma - ia\omega)^2} \\
&= \frac{i\omega'(k)}{2k} \frac{\gamma^2 + \gamma - 2ia\omega\gamma + ia\omega - a^2k^2}{(\gamma - ia\omega)^2}.
\end{aligned} \tag{B.18}$$

Now we turn once again for help to the dispersion relation (B.6) and find

$$\begin{aligned}
a\omega^3 + i\omega^2(\gamma+1) - ak^2\omega - i\gamma k^2 &= 0, \\
&\Downarrow \\
i\omega^2(\gamma - ia\omega) + i\omega^2 - ik^2(\gamma - ia\omega) &= 0, \\
k^2 &= \frac{\omega^2}{\gamma - ia\omega} + \omega^2.
\end{aligned} \tag{B.19}$$

Inserting (B.19) back into (B.18), we obtain

$$\begin{aligned}
\frac{i\omega'(k)}{2k} \frac{\gamma^2 + \gamma - 2ia\omega\gamma + ia\omega - a^2k^2}{(\gamma - ia\omega)^2} &= \frac{i\omega'(k)}{2k} \frac{\gamma^2 + \gamma - 2ia\omega\gamma + ia\omega - a^2\frac{\omega^2}{\gamma - ia\omega} - a^2\omega^2}{(\gamma - ia\omega)^2} \\
&= \frac{i\omega'(k)}{2k} \frac{(\gamma - ia\omega)^2 + \gamma + ia\omega - a^2\frac{\omega^2}{\gamma - ia\omega}}{(\gamma - ia\omega)^2} \\
&= \frac{i\omega'(k)}{2k} \frac{(\gamma - ia\omega)^2 + \frac{(\gamma + ia\omega)(\gamma - ia\omega) - a^2\omega^2}{\gamma - ia\omega}}{(\gamma - ia\omega)^2} \\
&= \frac{i\omega'(k)}{2k} \left(1 + \frac{\frac{\gamma^2 + a^2\omega^2 - a^2\omega^2}{\gamma - ia\omega}}{(\gamma - ia\omega)^2}\right) = \frac{i\omega'(k)}{2k} \left(1 + \frac{\gamma^2}{(\gamma - ia\omega)^3}\right).
\end{aligned} \tag{B.20}$$

Using this result in (B.17), our final amplitude equation becomes

$$\partial_{t_2}A_0+\omega'(k)\partial_{z_2}A_0+\frac{i\omega'(k)}{2k}\left(1+\frac{\gamma^2}{(\gamma-ia\omega)^3}\right)\partial_{t_1t_1}A_0-i\frac{\omega'(k)}{2k}\partial_{z_1z_1}A_0=0. \tag{B.21}$$

By removing the secular terms from the equation (B.13), the order ε^2 equation turns to

$$\begin{aligned}&a\partial_{t_0t_0t_0}e_2+(\gamma+1)\partial_{t_0t_0}e_2-a\partial_{z_0z_0t_0}e_2-\gamma\partial_{z_0z_0}e_2=-\left(a27i\omega^3A_0^3e^{3i\theta_0}\right.\\&+A_0^2A_0^*e^{i\theta_0}e^{2t_0\omega_i}\left(3ia\omega^3-18a\omega^2\omega_i-36ia\omega\omega_i^2+24a\omega_i^3\right)-9\gamma\omega^2A_0^32e^{3i\theta_0}\\&\left.+3\gamma A_0^2A_0^*e^{i\theta_0}e^{2t_0\omega_i}\left(2\omega_i-i\omega\right)^2\right)+(*),\end{aligned} \tag{B.22}$$

which we solve for e_2 taking only the particular solution.

$$e_2(z_0,t_0,\ldots)=c_1A_0^3e^{i3\theta_0}+c_2|A_0|^2A_0e^{i\theta_0}e^{2\omega_i}+(*), \tag{B.23}$$

where

$$\begin{aligned}c_1&=\frac{9\omega^2(\gamma-3ia\omega)}{-9\omega^2(\gamma-3ia\omega)+9k^2(\gamma-3ia\omega)-9\omega^2}=\frac{1}{-1+n^2(\omega)-1/(\gamma-i3a\omega)}\\&=\frac{1}{n^2(\omega)-n^2(3\omega)},\end{aligned} \tag{B.24}$$

$$\begin{aligned}c_2&=\frac{-\left(3a(2\omega_i-i\omega)^3+3\gamma(2\omega_i-i\omega)^2\right)}{a(2\omega_i-2i\omega)^3+k^2(\gamma+a(2\omega_i-i\omega))+(2\omega_i-i\omega)^2(\gamma+1)}\\&=\frac{-3(2\omega_i-i\omega)^2}{k^2+(2\omega_i-i\omega)^2\left(1+1/(\gamma-ia(\omega+2i\omega_i))\right)}\\&=\frac{3(\omega+2i\omega_i)^2}{k^2-(\omega+2i\omega_i)^2\left(1+1/(\gamma-ia(\omega+2i\omega_i))\right)}\\&=\frac{3(\omega+2i\omega_i)^2}{k^2-\left(1+\hat{\chi}(\omega+i2\omega_i)\right)\left(\omega+2i\omega_i\right)^2},\end{aligned} \tag{B.25}$$

with $\omega_i=\operatorname{Im}\omega$.

Defining as before the amplitude as in (2.45) and proceeding the usual way using (B.10) and (B.21) we get the amplitude equation

$$\partial_tA+\omega'(k)\partial_zA-i\beta\partial_{zz}A+i\alpha\partial_{tt}A=0, \tag{B.26}$$

where

$$\alpha=\frac{\omega'(k)}{2k}\left(1+\frac{\gamma^2}{(\gamma-ia\omega)^3}\right), \tag{B.27}$$

$$\beta=\frac{\omega'(k)}{2k}. \tag{B.28}$$

The overall approximate solution to (3.2.4) is then

$$E(z,t)=A(z,t)e^{i(kz-\omega t)}+c_1\varepsilon^2A^3(z,t)e^{i3(kz-\omega t)}+c_2\varepsilon^2|A(z,t)|^2A(z,t)e^{i(kz-\omega t)}e^{2t\omega_i}+(*), \tag{B.29}$$

where c_1,c_2 are defined in (B.24) and (B.25). To verify that α,β in (2.46) and (B.26) are the same, we look at (2.39) and get

$$\begin{aligned}&\omega'(k)\frac{n^2(\omega)+2\omega a\sqrt{2\pi}\hat{\chi}'(\omega)+\frac{a^2}{2}\omega^2\sqrt{2\pi}\hat{\chi}''(\omega)}{2k}=\frac{\omega'(k)}{2k}\left(1+\frac{1}{\gamma-ia\omega}+2\omega a\frac{i}{(\gamma-ia\omega)^2}\right.\\&\left.+\frac{a^2}{2}\omega^2\frac{-2}{(\gamma-ia\omega)^3}\right)=\frac{\omega'(k)}{2k}\left(\frac{(\gamma-ia\omega)^3+(\gamma-ia\omega)^2+2\omega ai(\gamma-ia\omega)-a^2\omega^2}{(\gamma-ia\omega)^3}\right)\\&=\frac{\omega'(k)}{2k}\left(\frac{(\gamma-ia\omega)^3+\gamma^2}{(\gamma-ia\omega)^3}\right)=\frac{\omega'(k)}{2k}\left(1+\frac{\gamma^2}{(\gamma-ia\omega)^3}\right),\end{aligned} \tag{B.30}$$

which is the same as (B.27).

Using the same argument as in (2.48) we can simplify the amplitude equation (B.26) into

$$\partial_t A + \omega'(k)\partial_z A - i\partial_{zz} A \left(\beta - \alpha \left(\omega'(k)\right)^2\right) = 0. \tag{B.31}$$

References

[1] M. Kolesik, J. V. Moloney, and M. Mlejnek. Unidirectional optical pulse propagation equation. *Phys. Rev. Lett.*, 89:283902, Dec 2002.

[2] M. Kolesik, P. Jakobsen, and J. V. Moloney. Quantifying the limits of unidirectional ultrashort optical pulse propagation. *Phys. Rev. A*, 86:035801, Sep 2012.

[3] M. Kolesik and J. V. Moloney. Nonlinear optical pulse propagation simulation: From maxwell's to unidirectional equations. *Phys. Rev. E*, 70:036604, Sep 2004.

[4] A. Hofstrand, P. Jakobsen, and J. V. Moloney. Bidirectional shooting method for extreme nonlinear optics. *Phys. Rev. A*, 100:053818, Nov 2019.

[5] Per Jakobsen. Bidirectional pulse propagation equation for extreme nonlinear optics. *Physica Scripta*, 89(9):095502, aug 2014.

[6] Alan C. Newell and Jerome V. Moloney. *Nonlinear optics*. Addison Wesley, 1992.

[7] Per Kristen Jakobsen. Topics in applied mathematics and nonlinear waves. *arXiv.org*, 2019.

[8] Igor S. Aranson and Lorenz Kramer. The world of the complex ginzburg-landau equation. *Rev. Mod. Phys.*, 74:99–143, Feb 2002.

[9] V. Zakharov and A. Shabat. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Journal of Experimental and Theoretical Physics*, 34:62–69, 1970.

[10] Al Kelley. The stable, center-stable, center, center-unstable, unstable manifolds. *Journal of Differential Equations*, 3(4):546–570, 1967.

[11] A. J. Roberts. Macroscale, slowly varying, models emerge from the microscale dynamics. *IMA Journal of Applied Mathematics*, 80(5):1492–1518, Oct 2015.

[12] Kenneth G Wilson. Renormalization group methods. *Advances in Mathematics*, 16(2):170–186, 1975.

[13] Lin-Yuan Chen, Nigel Goldenfeld, and Y. Oono. Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory. *Phys. Rev. E*, 54:376–394, Jul 1996.

[14] W. Sellmeier. Ueber die durch die aetherschwingungen erregten mitschwingungen der körpertheilchen und deren rückwirkung auf die ersteren, besonders zur erklärung der dispersion und ihrer anomalien. *Annalen der Physik*, 223(11):386–403, 1872.

[15] Rafael R. Gattass and Eric Mazur. Femtosecond laser micromachining in transparent materials. *Nature Photonics*, 2(4):219–225, Apr 2008.

[16] P. B. Corkum and Ferenc Krausz. Attosecond science. *Nature Physics*, 3(6):381–387, Jun 2007.

[17] Hendrik Antoon Lorentz. Versuch einer theorie der electrischen und optischen erscheinungen in bewegten körpern. *Cambridge University Press*, 2013.

[18] M. Mlejnek, E. M. Wright, and J. V. Moloney. Femtosecond pulse propagation in argon: A pressure dependence study. *Phys. Rev. E*, 58:4903–4910, Oct 1998.

[19] Carl M. Bender and Tai Tsun Wu. Anharmonic oscillator. *Phys. Rev.*, 184:1231–1260, Aug 1969.

[20] F. J. Dyson. Divergence of perturbation theory in quantum electrodynamics. *Phys. Rev.*, 85:631–632, Feb 1952.

[21] Michael Cross and Henry Greenside. Pattern formation and dynamics in nonequilibrium systems. *Cambridge University Press*, 2009.

[22] Masud Mansuripur. Field, force, energy and momentum in classical electrodynamics. *Bentham Science Publishers*, 09 2011.