TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

WKB-based methods for the solution of highly oscillatory differential equations

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Anton Arnold
E101 – Institut für Analysis und Scientific Computing, TU Wien

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Jannis Körner, MSc
Matrikelnummer: 12038056

Diese Dissertation haben begutachtet:

1. **Univ.Prof. Dr. Anton Arnold**
   Institut für Analysis und Scientific Computing, TU Wien
2. **Prof. Dr. Claudia Negulescu**
   Institut de Mathématiques de Toulouse, Université Paul Sabatier
3. **Ao.Univ.-Prof. Dr. Ewa B. Weinmüller**
   Institut für Analysis und Scientific Computing, TU Wien

Wien, am 08. März 2024

Kurzfassung

Diese Arbeit widmet sich der effizienten numerischen Lösung von gewöhnlichen Differentialgleichungen (GDG), welche hochoszillatorische Lösungen aufweisen. Das Modell, welches hier von Interesse ist, ist die eindimensionale stationäre (d.h. zeitunabhängige) Schrödingergleichung im hochoszillatorischen Bereich. Standard Verfahren für GDG zur Lösung dieser Gleichung sind sehr ineffizient, da sie sehr kleine Gittergrößen verwenden müssen, um die schnellen Oszillationen genau aufzulösen. Anstelle dessen entwickeln und analysieren wir hier numerische Verfahren, welche, aufbauend auf WKB-Theorie (benannt nach den Physikern Wentzel, Kramers und Brillouin), auf a priori Informationen über das asymptotische Verhalten der Lösung basieren.

Die Arbeit ist in drei Teile gegliedert, wobei in jedem eine neue WKB-basierte numerische Methode zur Lösung der hochoszillatorischen Schrödingergleichung präsentiert wird.

Im ersten Teil dieser Arbeit erweitern wir ein bereits existierendes WKB-basiertes Einschrittverfahren zweiter Ordnung um eine adaptive Schrittweitensteuerung und einen automatischen Mechanismus zum Wechseln numerischer Methoden. Diese Erweiterung erlaubt es dem Algorithmus, zwischen der WKB-basierten Methode für oszillatorische Bereiche und einem Runge-Kutta-Verfahren für glattere (d.h. weniger oszillatorische) Bereiche zu wechseln, was insgesamt zu einer Effizienzsteigerung führt. Durch den Vergleich mit einer ähnlichen Strategie aus der Literatur zeigen wir, dass unser neuer Ansatz bezüglich der Genauigkeit und Effizienz überlegen ist.

Im zweiten Teil entwickeln wir eine Erweiterung (mit höherer Ordnung) des WKB-basierten Einschrittverfahrens, welches im ersten Teil dieser Arbeit verwendet wurde. Das Verfahren beruht auf einer WKB-basierten Transformation der Schrödingergleichung in eine weniger oszillatorische Gleichung, welche numerisch auf einem groben Gitter gelöst werden kann. Durch die Herleitung hinreichend genauer Quadraturformeln für mehrere oszillatorische Integrale, welche in der Picard-Iteration der Lösung des transformierten Problems auftreten, erhalten wir ein Einschrittverfahren dritter Ordnung in Bezug auf die Schrittweite. Die Genauigkeit und Effizienz des neuen Verfahrens werden anhand mehrerer numerischer Beispiele demonstriert.

Im letzten Teil dieser Arbeit implementieren wir direkt eine WKB-Approximation beliebiger Ordnung als Näherungslösung für die Schrödingergleichung. Unsere Fehleranalyse stützt sich hauptsächlich auf die Annahme, dass der Koeffizient in der Gleichung analytisch ist. Wir leiten Fehlerabschätzungen her, welche explizit bezüglich des kleinen Parameters in der GDG sowie der gewählten Trunkierungsordnung für die zugrunde liegende asymptotische WKB-Reihe ist. Diese WKB-Reihe wird insbesondere im Hinblick auf die optimale Anzahl von Termen analysiert, um den resultierenden Approximationsfehler zu minimieren. Unsere Untersuchung zeigt, dass die optimale Anzahl an Termen umgekehrt proportional zum in der GDG auftretenden kleinen Parameter ist, was einen zugehörigen minimalen Fehler liefert, welcher exponentiell klein in Bezug auf diesen Parameter ist.

Abstract

This thesis is dedicated to the efficient numerical treatment of ordinary differential equations (ODEs) exhibiting highly oscillatory solutions. The model of interest here is the one-dimensional stationary (i.e. time-independent) Schrödinger equation in the highly oscillatory regime. Standard ODE methods become highly inefficient when solving this equation, as they have to use very small grid sizes in order to resolve the rapid oscillations accurately. Instead, we develop and analyze numerical methods which utilize a priori information on the asymptotic behavior of the solution, relying on WKB theory (named after the physicists Wentzel, Kramers and Brillouin).

The thesis is divided into three parts, each corresponding to a novel WKB-based numerical approach for solving the highly oscillatory Schrödinger equation.

In the first part of this thesis, we enhance an existing WKB-based second order one-step method with an adaptive step size controller and an automated methods switching. This extension allows the algorithm to switch between the WKB-based method for oscillatory regions and a standard Runge-Kutta method for smoother (i.e. less oscillatory) regions, leading to an overall increase in efficiency. By comparing with a similar strategy from existing literature, we find that our novel approach outperforms in terms of both accuracy and efficiency.

In the second part, we develop a higher order extension to the WKB-based one-step method employed in the first part of this thesis. The method relies on a WKB-based transformation of the Schrödinger equation into a smoother equation, which can be solved numerically on a coarse grid. By establishing sufficiently accurate quadrature formulas for several oscillatory integrals encountered in the Picard approximation of the solution of this transformed problem, we obtain a one-step method that is third order with respect to the step size. The accuracy and efficiency of the novel method are demonstrated through several numerical examples.

In the final part of this thesis, we implement directly an arbitrary order WKB approximation as an approximate solution to the Schrödinger equation. Our error analysis relies mainly on the assumption that the coefficient in the equation is analytic. We derive error estimates explicitly in terms of the small parameter present in the ODE, as well as of the chosen truncation order for the underlying asymptotic WKB series. This WKB series is then analyzed particularly with regard to the optimal number of terms required to minimize the resulting approximation error. Our investigation reveals that the optimal number of terms is inversely proportional to the small parameter in the ODE, yielding a corresponding minimal error which is exponentially small with respect to this parameter.

Acknowledgement

First and foremost, I would like to thank my advisor Prof. Dr. Anton Arnold for choosing me as a member of his research group four years ago, and giving me the opportunity to start my doctoral studies at TU Vienna. I am also very grateful for his professional advice; I always felt like he was asking the "right" questions during our mathematical discussions, which have helped me to move forward with my topic.

I am also very grateful to Prof. Dr. Christian Klein for giving me the opportunity of a short research visit in Dijon.

I further thank Prof. Dr. Markus Melenk, Prof. Dr. Christian Klein, and Kirian Döpfner for fruitful mathematical discussions and research collaboration.

Special thanks to Prof. Dr. Markus Melenk for contacting me after one of my conference talks, providing me with a very helpful mathematical idea.

I would also like to express my gratitude to all the nice people at TU Vienna for an encouraging work environment. In particular, thanks to Christoph and Lorenza for several (mathematical) conversations.

Lastly, a special thanks to my parents, who always supported me throughout my academic years.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 08. März 2024

Jannis Körner

Contents

1 Introduction

This thesis is devoted to the efficient numerical treatment of a class of ordinary differential equations (ODEs) that have highly oscillatory solutions. Highly oscillatory phenomena appear in a wide range of applications, including plasma physics, electromagnetic and acoustic scattering, Hamiltonian dynamics, inflationary cosmology, and quantum mechanics.

We focus here on problems that can be described by the one-dimensional stationary (i.e. time-independent) Schrödinger equation:

$$\varepsilon^2 \varphi''(x) + a(x)\varphi(x) = 0\,, \quad x \in I \subseteq \mathbb{R}\,. \tag{1.0.1}$$

Here, $0 < \varepsilon \ll 1$ is a very small parameter, I is a real interval, and the coefficient function $a : I \to \mathbb{R}$ is assumed to be sufficiently smooth. Furthermore, we assume that a is bounded away from zero, i.e. $a(x) \geq a_0 > 0$, which means that the solutions of equation (1.0.1) are highly oscillatory.

To comprehend the oscillatory nature of the (possibly complex-valued) solution φ of (1.0.1), one may initially consider the case where $a(x) \equiv a_0 > 0$ is just a positive constant. The general solution is then given by $\varphi(x) = \alpha \sin\left(\frac{\sqrt{a_0}}{\varepsilon}x\right) + \beta \cos\left(\frac{\sqrt{a_0}}{\varepsilon}x\right)$, where $\alpha, \beta \in \mathbb{C}$ are arbitrary constants. Consequently, for a small parameter $0 < \varepsilon \ll 1$, the solution is highly oscillatory, with an amplitude of order $\varphi = \mathcal{O}(1)$, and $\varphi' = \mathcal{O}(\varepsilon^{-1})$, as $\varepsilon \to 0$. These properties persist even with a non-constant function $a(x) \geq a_0 > 0$. In fact, it is known that the (local) wave length λ of a solution φ of (1.0.1) is given by $\lambda(x) = (2\pi\varepsilon)/\sqrt{a(x)}$. As a consequence, the solution exhibits rapid oscillations for a small parameter ε, especially in the semi-classical limit $\varepsilon \to 0$.

The Schrödinger equation (1.0.1) is particularly important within the context of quantum mechanical problems, e.g., for the simulation of electron transport in semiconductor devices [MJK13, Neg05, SHMS98]. In these applications, φ represents the Schrödinger wave function, and $a(x) := E - V(x)$ is related to a prescribed electrostatic potential V, where $E \in \mathbb{R}$ denotes the injection energy of an electron of mass m. The small parameter $\varepsilon := \frac{\hbar}{\sqrt{2m}}$ is then proportional to the (reduced) Planck constant $\hbar$. We note that there are numerous additional applications of equation (1.0.1), e.g., in plasma physics [CS58, Lew68] and cosmology [MS03, Win05].

Given the highly oscillatory behavior of a solution φ of (1.0.1), employing standard ODE methods (e.g. Runge-Kutta methods) can be very costly and inefficient. Indeed, this is because one has to use a very small grid size h in order to accurately resolve every oscillation of the solution. This efficiency issue, which we shall discuss in more detail in the next subsection, is a particular problem in applications, where the equation (1.0.1) has to be solved many times in parallel, e.g., in the modeling of semiconductor devices [AAN11]. This underlines the substantial demand for efficient numerical methods that are suitable for solving problems corresponding to equation (1.0.1).

1.1 The main problem when using standard ODE methods

In this subsection we will address the inefficiency problem encountered when using standard ODE methods (e.g. Runge-Kutta methods) for solving the highly oscillatory Schrödinger equation (1.0.1). For simplicity, we will consider here only explicit one-step methods.

Consider for any initial value $\mathbf{Y}_0 \in \mathbb{C}^m$, $m \in \mathbb{N}$ and a given interval $I := [\xi, \eta] \subset \mathbb{R}$ the initial value problem (IVP)

$$\begin{cases} \mathbf{Y}'(x) = \mathbf{F}(x, \mathbf{Y}(x)) \,, & x \in I \,, \\ \mathbf{Y}(\xi) = \mathbf{Y}_0 \,, \end{cases} \tag{1.1.1}$$

with $\mathbf{F}$ being sufficiently smooth and Lipschitz continuous w.r.t. its second argument. Further, we consider a discretization $\{x_0, \dots, x_M\}$ of the interval I, denote with $h_n := x_{n+1} - x_n$, $n = 0, \dots, M-1$ the corresponding step sizes, and set $h_{max} := \max_{0 \le n \le M-1} h_n$ as the maximum step size. Let

$$\mathbf{Y}_{n+1} := \mathbf{Y}_n + h_n \mathbf{\Phi}(x_n, \mathbf{Y}_n, h_n) \,, \quad n = 0, \dots, M-1 \,, \tag{1.1.2}$$

be a general explicit one-step method with incremental function $\mathbf{\Phi}$. It is well-known that such a method is convergent with order $p \in \mathbb{N}$, if it satisfies the consistency estimate

$$\forall\, x \in [\xi, \eta)\, \forall\, h \in (0, \eta - x] : \quad \|\mathbf{Y}(x+h) - [\mathbf{Y}(x) + h\mathbf{\Phi}(x, \mathbf{Y}(x), h)]\| \le C_{cons}\, h^{p+1}$$

as well as the stability estimate

$$\forall\, x \in [\xi, \eta)\, \forall\, h \in (0, \eta - x]\, \forall\, \mathbf{Y}, \widetilde{\mathbf{Y}} \in \mathbb{C}^m : \quad \|\mathbf{\Phi}(x, \mathbf{Y}, h) - \mathbf{\Phi}(x, \widetilde{\mathbf{Y}}, h)\| \le C_{stab}\, \|\mathbf{Y} - \widetilde{\mathbf{Y}}\| \,,$$

with some constants $C_{cons}, C_{stab} > 0$ (e.g., see [Gea71]). Here, $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{C}^m$. More precisely, the global error can then be estimated as follows:

$$\max_{n=0,\dots,M} \|\mathbf{Y}_n - \mathbf{Y}(x_n)\| \le \frac{C_{cons}}{C_{stab}} \left(\exp(C_{stab}(\eta - \xi)) - 1\right) h_{max}^p \,. \tag{1.1.3}$$

Estimate (1.1.3) with its explicit dependence on the consistency and stability constants C_{cons} and C_{stab} reveals the main problem that lies behind the use of a standard method for solving the highly oscillatory equation (1.0.1). Indeed, although the global error of a numerical method might asymptotically behave like $\mathcal{O}(h^p)$, as $h \to 0$, it is also of great importance that the values of the consistency and stability constants C_{cons} and C_{stab} are moderate, since for very large values of C_{cons} and C_{stab} one is forced to use very small step sizes h_{max} in order to reduce the r.h.s. of (1.1.3).

If we aim to solve the highly oscillatory Schrödinger equation (1.0.1) (augmented with some initial values) using a standard ODE method, the conventional approach of doing so is to first introduce the notation $\mathbf{Y}(x) := (\varphi(x), \varphi'(x))^T$ and then rewrite the second order ODE (1.0.1) into the following first order system:

$$\mathbf{Y}'(x) = \begin{pmatrix} 0 & 1 \\ -a(x)/\varepsilon^2 & 0 \end{pmatrix} \mathbf{Y}(x) \,. \tag{1.1.4}$$

The problem with this approach is that the system matrix on the r.h.s. of (1.1.4) is of order $\mathcal{O}(\varepsilon^{-2})$, which, depending on the chosen numerical method, typically results in very large values for the (ε-dependent) constants C_{cons} and C_{stab}. To illustrate this, let us consider as an example the simplest method one can think of, namely, the explicit Euler method, specified by the incremental function $\mathbf{\Phi}^{Euler}(x, \mathbf{Y}, h) := \mathbf{F}(x, \mathbf{Y}(x))$. For this method, we shall now determine the (ε-dependent) consistency and stability constants C_{cons} and C_{stab}. First, Taylor's theorem with the remainder in Lagrange form yields that

$$\|\mathbf{Y}(x+h) - [\mathbf{Y}(x) + h\mathbf{F}(x, \mathbf{Y}(x))]\| = \frac{\|\mathbf{Y}''(\tau)\|}{2} h^2 \leq \frac{\|\mathbf{Y}''\|_{L^\infty(I)}}{2} h^2 \,, \tag{1.1.5}$$

for some appropriate $\tau \in [x, x+h]$. To illustrate the dependence of the r.h.s. of (1.1.5) on the parameter ε, we use the ODE (1.1.4) and the fact that $\varphi' = \mathcal{O}(\varepsilon^{-1})$, to estimate

$$\|\mathbf{Y}''\|_{L^\infty(I)} \leq C\varepsilon^{-3} \,, \tag{1.1.6}$$

where $\varepsilon \leq \varepsilon_0$ for some sufficiently small ε_0. Thus, we conclude that $C_{cons} = \mathcal{O}(\varepsilon^{-3})$. Moreover, the stability constant C_{stab} is simply given by the Lipschitz constant of $\mathbf{F}$. In fact, we see from (1.1.4) that $C_{stab} = \mathcal{O}(\varepsilon^{-2})$. According to (1.1.3), the global error for the explicit Euler method can thus be estimated by

$$\max_{n=0,\dots,M} \|\mathbf{Y}_n^{Euler} - \mathbf{Y}(x_n)\| \leq C_1 \varepsilon^{-1} \left(\exp(C_2 \varepsilon^{-2}) - 1\right) h_{max} \,, \tag{1.1.7}$$

where $\varepsilon \leq \varepsilon_0$ and $C_1, C_2 > 0$ are constants independent of ε and h_{max}. Hence, we observe that the total (ε-dependent) constant factor on the r.h.s. of (1.1.7) blows up for small parameters ε. While we have chosen a straightforward example here, it is worth noting that this problem arises similarly when employing other standard ODE methods.

The downside of having an ε-dependent large constant in the global error estimate (1.1.3) can also be demonstrated through a simple numerical experiment. To this end, we solve the first order system (1.1.4) on the interval $[0, 1]$ for the coefficient function $a(x) = (x+1)^2$ and the initial value $\mathbf{Y}(0) = (1, 0)^T$. For this, we use the well-known Dormand-Prince method [DP80], which is an explicit Runge-Kutta method with convergence order $p = 5$.[1] In Figure 1.1.1 and Figure 1.1.2 we show results for the choices $\varepsilon = 2^{-4}$ and $\varepsilon = 2^{-6}$, respectively. For both figures we use the same number of grid points. On the left of each figure we plot the real part of the exact solution as a solid line as well as the numerical approximations, indicated as dots. According to the left plot of Figure 1.1.1, the numerical solution matches well with the exact solution. This can also be observed from the right plot of Figure 1.1.1, which shows the corresponding global absolute error. Indeed, the error remains permanently less than $5 \cdot 10^{-4}$. Thus, there are enough grid points for the numerical method to accurately resolve every oscillation. By contrast, as seen in the left plot of Figure 1.1.2 for $\varepsilon = 2^{-6}$, the numerical approximations become unacceptably inaccurate near the end of the interval. Indeed, the right plot of Figure 1.1.2 reveals that the corresponding global absolute errors accumulate up to 10^0 at the end of the interval.

[1] Actually, the Dormand-Prince method [DP80] is an embedded method that provides both a fourth and a fifth order numerical solution. Here, we use only the fifth order one.

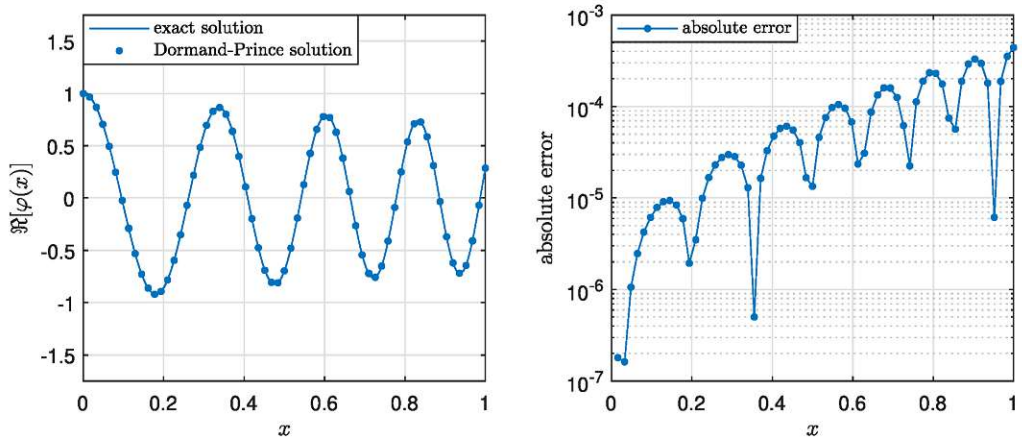

Figure 1.1.1: Numerical results for the Dormand-Prince method ($p = 5$) when solving the system (1.1.4) on the interval $[0, 1]$ for the coefficient function $a(x) = (x + 1)^2$, an initial value $\mathbf{Y}(0) = (1, 0)^T$, and a parameter $\varepsilon = 2^{-4}$. Left: Real part of the exact solution (solid lines) and numerical approximations (dots). Right: The corresponding global absolute error.

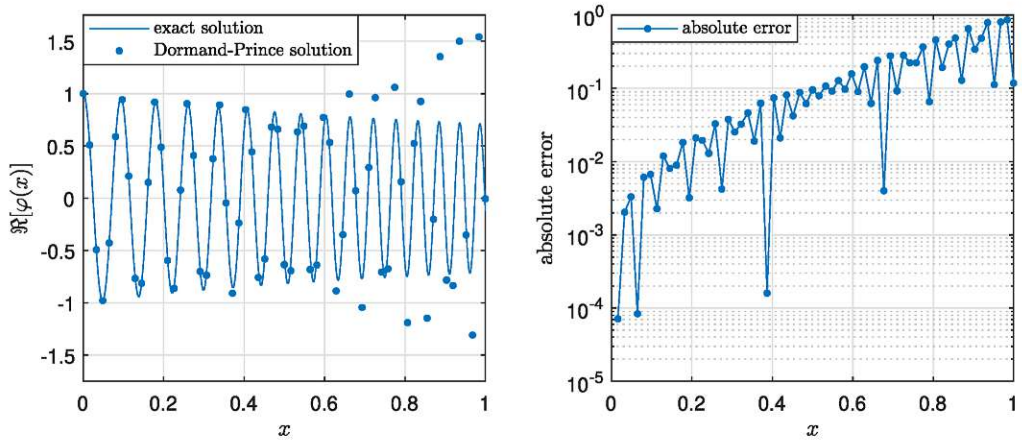

Figure 1.1.2: Numerical results for the Dormand-Prince method ($p = 5$) when solving the system (1.1.4) on the interval $[0, 1]$ for the coefficient function $a(x) = (x + 1)^2$, an initial value $\mathbf{Y}(0) = (1, 0)^T$, and a parameter $\varepsilon = 2^{-6}$. Left: Real part of the exact solution (solid lines) and numerical approximations (dots). Right: The corresponding global absolute error.

Hence, the number of grid points is too small for the numerical method to resolve every oscillation accurately.

The above discussion makes clear why it is of great interest to develop numerical methods with consistency and stability constants that are independent of ε. Indeed, such methods are referred to as *uniformly correct* w.r.t. ε. Even more desirable are methods that are *asymptotically correct* w.r.t. ε, i.e., the global error tends to zero for $\varepsilon \to 0$, even for a fixed grid size. In other words, the more oscillatory the solution φ of (1.0.1) becomes, the more accurate is the numerical solution of an asymptotically correct method. This is a feature shared by the methods presented in this thesis, provided that certain additional assumptions are satisfied.

1.2 WKB theory: utilizing a priori information on the solution

The methods presented in this thesis all rely on WKB theory, which is named after the physicists Gregor Wentzel, Hendrik Anthony Kramers, and Léon Brillouin (e.g., see [LL85]). This theory provides a method for approximating the solution of a linear differential equation in which the highest derivative is scaled by a small parameter ε. The basic idea is that the rapid oscillatory behavior of such a solution is typically characterized by an exponential function. More precisely, the technique relies on an asymptotic ansatz for the solution of a given differential equation. In this way one is able to derive asymptotic WKB approximations of the solution, which provide a priori insights into the asymptotic behavior of the solution, as $\varepsilon \to 0$. This a priori information can then be utilized to develop numerical methods that are uniformly or even asymptotically correct w.r.t. ε.

The first step in order to derive an asymptotic WKB approximation for the solution φ of the highly oscillatory Schrödinger equation (1.0.1) consists of making the WKB-ansatz

$$\varphi(x) \sim \exp\left(\frac{1}{\varepsilon} S(x)\right), \quad \varepsilon \to 0\,. \tag{1.2.1}$$

Here, S is a complex-valued function that contains information about the phase and the amplitude of the solution φ. It is then convenient to express S as an asymptotic expansion w.r.t. ε:

$$S(x) \sim \sum_{n=0}^{\infty} \varepsilon^n S_n(x)\,, \quad \varepsilon \to 0\,; \quad S_n(x) \in \mathbb{C}\,. \tag{1.2.2}$$

This asymptotic series is typically divergent (as usual for asymptotic series) and must therefore be truncated in order to obtain an approximate solution.

By substituting (1.2.1)-(1.2.2) into the Schrödinger equation (1.0.1), one formally obtains

$$\left(\sum_{n=0}^{\infty} \varepsilon^n S_n'(x)\right)^2 + \sum_{n=0}^{\infty} \varepsilon^{n+1} S_n''(x) + a(x) = 0. \tag{1.2.3}$$

A comparison of different ε-powers then yields the following well-known recurrence relation for the functions S_n' (e.g., see [BO99, §10]):

$$(S_0^{\pm})' = \pm \mathrm{i}\sqrt{a}\,, \tag{1.2.4}$$

$$(S_1^{\pm})' = -\frac{(S_0^{\pm})''}{2(S_0^{\pm})'} = -\frac{a'}{4a} = -\frac{1}{4}(\ln(a))'\,, \tag{1.2.5}$$

$$(S_n^{\pm})' = -\frac{1}{2(S_0^{\pm})'}\left(\sum_{j=1}^{n-1} (S_j^{\pm})'(S_{n-j}^{\pm})' + (S_{n-1}^{\pm})''\right), \quad n \geq 2\,. \tag{1.2.6}$$

Thus, there are two sequences $(S_n^{\pm})_{n\in\mathbb{N}_0}$ of functions which lead to approximate solutions of the Schrödinger equation (1.0.1), by analogy to the two fundamental solutions of the second order ODE (1.0.1). The functions of these sequences are all unique up to an additive

integration constant. The general WKB approximation for (1.0.1) is then defined as the linear combination

$$\varphi \approx \varphi_N^{WKB} := \alpha_{N,\varepsilon} \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^-\right) + \beta_{N,\varepsilon} \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^+\right), \tag{1.2.7}$$

where $\alpha_{N,\varepsilon}, \beta_{N,\varepsilon} \in \mathbb{C}$ are complex constants, which can be determined by initial or boundary conditions, in general. The WKB approximation (1.2.7) is the basis for all methods presented in this thesis.

In fact, the methods presented in Chapter 2 and Chapter 3 both rely on a second order WKB approximation, which is obtained by truncating the sums in (1.2.7) after $N = 2$:

$$\varphi_2^{WKB}(x) = \alpha_{2,\varepsilon} \frac{\exp\left(-\frac{\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)\right)}{a(x)^{1/4}} + \beta_{2,\varepsilon} \frac{\exp\left(\frac{\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)\right)}{a(x)^{1/4}}, \tag{1.2.8}$$

where ϕ^ε denotes the phase of this WKB approximation and is given by

$$\phi^\varepsilon(x) := \frac{1}{\mathrm{i}} \int_\xi^x \left((S_0^+)'(y) + \varepsilon^2 (S_2^+)'(y)\right) \mathrm{d}y = \int_\xi^x \left(\sqrt{a(y)} - \varepsilon^2 b(y)\right) \mathrm{d}y, \tag{1.2.9}$$

with

$$b(y) := -\frac{1}{2a(y)^{1/4}} \left(a(y)^{-1/4}\right)''. \tag{1.2.10}$$

More precisely, the basis for both methods is an analytical pre-processing of the highly oscillatory ODE (1.0.1), by using the phase (1.2.9) to transform (1.0.1) into a smoother (i.e. less oscillatory) problem. For more details about the specific form of this transformation, we refer to Chapter 2 or Chapter 3. The resulting problem is a first order ODE system for a new variable $Z(x) = (z_1(x), z_2(x))^T$ and has the form

$$Z'(x) = \varepsilon \mathbf{N}^\varepsilon(x) Z(x), \quad x \in I \subseteq \mathbb{R}, \tag{1.2.11}$$

where $\mathbf{N}^\varepsilon$ is a 2×2 matrix of order $\mathcal{O}(1)$, as $\varepsilon \to 0$. It is worth noting that a solution Z of (1.2.11) is still highly oscillatory. However, these oscillations have an amplitude only of the order $\mathcal{O}(\varepsilon^2)$, in contrast to the $\mathcal{O}(1)$ oscillations of a solution φ of the original ODE (1.0.1). Hence, the dominant oscillations are eliminated and the resulting problem can be solved numerically on a coarse grid. This strategy was developed in [AAN11] and used there to construct efficient and accurate one-step methods of first and second order w.r.t. the step size h. In fact, even for a fixed step size h, their methods both yield global errors of order $\mathcal{O}(\varepsilon^3)$, as $\varepsilon \to 0$, provided that the phase integral (1.2.9) can be computed exactly. This underlines again the benefit of employing an asymptotically correct method (w.r.t. ε) for (1.0.1): as the solution of (1.0.1) becomes more and more oscillatory (for smaller values of ε), the numerical solution becomes increasingly accurate.

As stated in [BO99, §10.2], in order for the WKB-ansatz (1.2.1)-(1.2.2) to be valid on the entire interval I, it is necessary that for any $n \in \mathbb{N}_0$ the asymptotic relation

$$\varepsilon^n S_{n+1}(x) = o(\varepsilon^{n-1} S_n(x)), \quad \varepsilon \to 0, \tag{1.2.12}$$

must hold uniformly in $x \in I$. This validity condition for the WKB approximation is clearly violated if $a(x)$ vanishes on I, as $|S_1^{\pm}(x)|$ blows up in this case, according to (1.2.5). In fact, the zeros of a have a particular meaning within a quantum mechanical context. Then, equation (1.0.1) with $a(x) = E - V(x)$ describes a quantum mechanical particle in a potential $V(x)$, where E is the total energy of the particle. A zero of $a(x)$ thus corresponds to a point where the potential energy equals the total energy. At such points the motion of the particle stops and changes to the opposite direction. For this reason, a zero of $a(x)$ is referred to as a *turning point*. Turning points also often characterize the connection between oscillatory regions ($a(x) > 0$) and so-called *evanescent regions* ($a(x) < 0$). In evanescent regions, the solution φ of (1.0.1) does not exhibit oscillatory behavior. Instead, one observes (essentially) an ε-dependent exponential growth or decay. Given that one of our main assumptions is $a(x) \geq a_0 > 0$, we focus here only on the oscillatory region, excluding turning points and the associated validity problems of the WKB approximation.

1.3 Adaptive step size control and automated methods switching

Even when excluding turning points, and hence the technical problems they imply, WKB-based methods can still have efficiency and accuracy problems in regions where the coefficient function $a(x)$, although non-zero, is very small. Indeed, as the solution becomes less oscillatory in the neighbourhood of a turning point, a WKB-based method loses its advantage and might even become increasingly inaccurate. Therefore, one possible remedy is to switch in such regions to a standard ODE method (e.g., a Runge-Kutta method). Such an approach is taken in Chapter 2, where we implement an automated switching mechanism for the $\mathcal{O}(h^2)$ one-step method from [AAN11]. Inspired by the strategy in [HLH16], this switching mechanism is based on and combined with an adaptive step size control, which aims to compute the numerical solution more accurately and efficiently, when compared to using just a uniform grid. In the following, we shall briefly explain this approach as well as formulate our main goals and questions; for more details, see Chapter 2.

The basis for the step size control in Chapter 2 is an estimator for the local truncation error:

$$\mathrm{est}_n := \|Y_n^{(k)} - Y_n^{(k+1)}\|_\infty \,. \tag{1.3.1}$$

Here, $Y_n^{(k)} = (\varphi_n^{(k)}, \varphi_n'^{(k)})^T$ and $Y_n^{(k+1)} = (\varphi_n^{(k+1)}, \varphi_n'^{(k+1)})^T$ represent two numerical solutions of order k and $k+1$ to approximate the exact solution $Y(x_n) = (\varphi(x_n), \varphi'(x_n))^T$ of an IVP corresponding to (1.0.1), x_n denoting a grid point. Suppose that these approximations were computed using the (trial) step size $h_{n,trial} := x_n - x_{n-1}$. The step size is then varied by $h_{new} := \theta_n \cdot h_{n,trial}$, where the multiplicative factor θ_n is given by

$$\theta_n := \max\left(0.5, \min\left(2, 0.9\left(\frac{\mathrm{ATol} + \mathrm{RTol}\cdot\|Y_n^{(k+1)}\|_\infty}{\mathrm{est}_n}\right)^{\frac{1}{k+1}}\right)\right) . \tag{1.3.2}$$

Here, ATol and RTol are prescribed absolute and relative error tolerances, the values 0.5 and 2 are design parameters that limit the ratio of two consecutive step sizes from below

and above, and 0.9 is a common safety factor for increasing the probability of the next step to be accepted. Note that for $\|Y_n^{(k+1)}\|_\infty \to 0$ the ATol term in the numerator in (1.3.2) dominates, whereas for large values of $\|Y_n^{(k+1)}\|_\infty$ the RTol term dominates. Hence, the factor (1.3.2) leads to a step size control that adequately incorporates a gradual switch-over between absolute and relative errors, depending on the behavior of the solution; for more details, see Chapter 2. The step size control is then based on an acceptance criterion that aims to maintain the local error in each step as close as possible to the given error tolerances: If the fraction in (1.3.2) containing est_n is greater than or equal to 1, the nth step is accepted and the step size is defined as $h_n := h_{n,trial}$. The trial step size for the next step is then defined as $h_{n+1,trial} := h_{new}$. However, if this fraction is smaller than 1, the nth step is rejected and a reattempt is done with the smaller (trial) step size $h_{n,trial}$ by updating its value as $h_{n,trial} \to h_{new}$.

The above explained adaptive step size control is also the basis for the dynamical switching mechanism: Consider two pairs of numerical methods, one of order $k^{(1)}$ and $k^{(1)}+1$, and the other of order $k^{(2)}$ and $k^{(2)}+1$. Based on the adaptive step size control above, one computes in each step the multiplicative factors $\theta^{(1)}$ and $\theta^{(2)}$ and checks the acceptance criterion for each $(i) \in \{(1),(2)\}$. The switching mechanism then intervenes by selecting the acceptable pair of methods that yields the larger value of $\theta_n^{(i)}$, $i = 1, 2$, hence yielding the larger proposed new step size. More precisely, we define

$$\Theta_n := \begin{cases} \theta_n^{(1)}, & \text{(1) accepted, (2) rejected} \\ \theta_n^{(2)}, & \text{(1) rejected, (2) accepted} \\ \max\left(\theta_n^{(1)}, \theta_n^{(2)}\right), & \text{otherwise}. \end{cases} \tag{1.3.3}$$

Thus, the switching mechanism favors the pair of methods with the smaller error estimator, discounted by its respective order $k^{(i)}$. Hence, the error estimator is not only used to determine the next step size, but also to decide between the two methods (or more precisely, the two pairs of methods). If at least one pair of methods was accepted, the algorithm sets $h_n := h_{n,trial}$ as the step size and $h_{n+1,trial} := \Theta_n \cdot h_{n,trial}$ as the trial step size for the next step. Otherwise, a reattempt is made with the smaller trial step size, updating its value as $h_{n,trial} \to \Theta_n \cdot h_{n,trial}$.

In Chapter 2, this adaptive step size control with the automated switching mechanism is implemented for the WKB-based $\mathcal{O}(h)$ and $\mathcal{O}(h^2)$ methods from [AAN11], combined with a Runge-Kutta pair of order $\mathcal{O}(h^4)$ and $\mathcal{O}(h^5)$ as the standard ODE methods. In [HLH16] the authors proposed a similar coupling of a WKB-based method with a Runge-Kutta method, called the Runge-Kutta-WKB (RKWKB) method. However, the WKB-based method they use in this coupling is only of order $\mathcal{O}(h)$. Further, the dynamic switching mechanism the authors use is different from the one sketched above. The main difference is that, in each step, they simply choose the method with the smaller error estimator, without incorporating the respective order of the method. We note, however, that the work [HLH16] was the main motivation for the investigations in Chapter 2. A similar approach can also be found in their more recent paper [AHLH20], where the authors again use a WKB-based method of order $\mathcal{O}(h)$, but the algorithm is also suitable for equations involving first derivative terms of the form $\gamma(x)\varphi'(x)$.

In particular, we focus here on the following goals and questions:

- One main goal is the comparison of numerical results obtained with our approach, i.e., the WKB-based method from [AAN11] coupled to a Runge-Kutta method, with the RKWKB method proposed in [HLH16]. For both approaches we use the same Runge-Kutta scheme as the standard ODE method. Which approach is more efficient?

- Is the error estimator (1.3.1) reliable? That is, is the error estimator in good agreement with the actual local truncation error?

- Does our switching mechanism ensure well defined switching points between the WKB-based method (for oscillatory regions) and the standard ODE method (for smoother regions)?

- The coupling of two methods (or more precisely, two pairs of methods) incurs additional computational costs, as in each step both methods have to be applied. Hence we raise the question: Do the two selected methods complement each other well enough so that the overall efficiency of the algorithm is increased?

For the two numerical examples presented in Chapter 2, we find that our approach outperforms the RKWKB method from [HLH16] significantly, as our algorithm yields smaller global errors (particularly for small ε-values and small prescribed error tolerances) for the same CPU time. The numerical tests reveal that the efficiency difference is mostly due to the higher order (w.r.t. the step size h) WKB-based method used in our approach; the different step size controls and switching mechanisms (here vs. [HLH16, AHLH20]) play a less important role. This motivates the development of an even higher order WKB-based method, which will be our main goal in Chapter 3.

1.4 Construction of third order WKB-based one-step schemes

Motivated by the efficiency and accuracy of the second order one-step scheme (w.r.t. the step size h) from [AAN11], particularly when implemented with an adaptive step size control as in Chapter 2, we develop in Chapter 3 an extension of this method to a third order scheme. Indeed, especially when prescribing very low error tolerances for the step size controller, a higher numerical order of the method can be highly beneficial, as lower tolerances typically force the algorithm to choose smaller step sizes. Further, such a third order extension of [AAN11] should of course have at least the same asymptotic accuracy as the second order scheme, i.e. $\mathcal{O}(\varepsilon^3)$ as $\varepsilon \to 0$. In this subsection we will sketch the steps needed for this extension, and formulate our main goals and questions; for more details, we refer the reader to Chapter 3.

The development of a third order scheme can be realized similarly to the second order scheme from [AAN11]. To this end, we consider the Schrödinger equation (1.0.1) on an real interval $I = [\xi, \eta]$, augmented with the initial values

$$\varphi(\xi) = \varphi_0\,, \quad \varepsilon\varphi'(\xi) = \varphi_1\,, \tag{1.4.1}$$

where $\varphi_0, \varphi_1 \in \mathbb{C}$ are complex constants. Employing the WKB-based transformation from [AAN11], which exploits the phase (1.2.9) of the second order WKB approximation (1.2.8), one obtains for a new variable $Z(x) = (z_1(x), z_2(x))^T$ the transformed problem (for more details, see Chapter 3)

$$\begin{cases} Z'(x) = \varepsilon \mathbf{N}^\varepsilon(x) Z(x)\,, & x \in I = [\xi, \eta]\,, \\ Z(\xi) = Z_0\,. \end{cases} \tag{1.4.2}$$

Here, the (Hermitian) matrix $\mathbf{N}^\varepsilon(x)$ is bounded independently of ε and has only off-diagonal non-zero entries:

$$N^\varepsilon_{1,2}(x) = b(x)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)}\,, \quad N^\varepsilon_{2,1}(x) = b(x)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)}\,, \tag{1.4.3}$$

with ϕ^ε and b defined in (1.2.9)-(1.2.10). For the construction of a third order method we then consider a discretization $\{x_0, \dots, x_M\}$ of the interval I and the third order Picard approximation of the solution to the IVP (1.4.2). This Picard approximation was already derived in [AAN11] and has the form

$$Z(x_{n+1}) = \left[\mathbf{I} + \varepsilon \mathbf{M}^\varepsilon_1(x_{n+1}; x_n) + \varepsilon^2 \mathbf{M}^\varepsilon_2(x_{n+1}; x_n) + \varepsilon^3 \mathbf{M}^\varepsilon_3(x_{n+1}; x_n)\right] Z(x_n) + r_n\,, \tag{1.4.4}$$

where $\mathbf{I}$ denotes the 2×2 identity matrix, the matrices $\mathbf{M}^\varepsilon_p$, $p = 1, 2, 3$ are given by

$$\mathbf{M}^\varepsilon_p(x_{n+1}; x_n) = \int_{x_n}^{x_{n+1}} \mathbf{N}^\varepsilon(y) \mathbf{M}^\varepsilon_{p-1}(y; x_n)\,\mathrm{d}y\,, \quad \mathbf{M}^\varepsilon_0 = \mathbf{I}\,, \tag{1.4.5}$$

and the remainder r_n satisfies the estimate (see (3.3.2) in Chapter 3)

$$\|r_n\|_\infty \leq C \varepsilon^4 h^3 \min(\varepsilon, h)\,, \tag{1.4.6}$$

where $C > 0$ is a constant independent of ε and h. Given the highly oscillatory entries (1.4.3) of the matrix $\mathbf{N}^\varepsilon$, it is evident that the matrices $\mathbf{M}^\varepsilon_p$, $p = 1, 2, 3$ involve (iterated) oscillatory integrals and can hence not be computed exactly in general. Therefore, in order to derive one-step methods from the Picard approximation (1.4.4), which are asymptotically correct w.r.t. ε, it is necessary to design sufficiently accurate quadratures for these integrals. To this end, we will use techniques such as the *asymptotic method* from [INO06] and the *shifted asymptotic method* from [AAN11]. Then, after suitable quadratures for the integrals $\mathbf{M}^\varepsilon_p$ have been established, we are able to define a numerical scheme for solving the IVP (1.4.2), yielding approximations Z_n of $Z(x_n)$, $n = 0, \dots, M$. Finally, we obtain numerical solutions φ_n, $n = 0, \dots, M$ to the original problem (1.0.1), (1.4.1) by using an inverse transformation, i.e., from the variable Z back to the variable φ; for more details, see Chapter 3.

Given the above strategy for the development of a third order scheme, we will focus in Chapter 3 on the following main goals and questions:

- Construction of the scheme: We aim to design sufficiently accurate approximations for the (iterated) oscillatory integrals $\mathbf{M}^\varepsilon_p$ such that the resulting scheme is $\mathcal{O}(h^3)$ and asymptotically correct w.r.t. ε. The method should have at least the same asymptotic accuracy as the scheme from [AAN11], i.e. $\mathcal{O}(\varepsilon^3)$ as $\varepsilon \to 0$.

- The resulting third order scheme is much more involved than the second order scheme from [AAN11], as the established quadrature formulas for the matrices $\mathbf{M}_p^\varepsilon$ are more complex and involve more function evaluations. It is thus not clear a priori if the new derived method is more efficient. This shall be investigated through several numerical experiments with the help of work-precision diagrams. Can we find a significant efficiency gain for the novel third order method?

The main result of Chapter 3 is Theorem 3.4.1, which implies the following global error estimate:

$$\|\varphi(x_n) - \varphi_n\|_\infty \leq C\varepsilon^3 h^3 \max(\varepsilon, h)\,, \quad n = 0, \dots, M\,, \tag{1.4.7}$$

where $C > 0$ is a constant independent of ε and h. Given the $\max(\varepsilon, h)$-factor in the global error estimate (1.4.7), the novel scheme yields errors that locally even behave like $\mathcal{O}(\varepsilon^4)$ (for a fixed and sufficiently small h) or $\mathcal{O}(h^4)$ (for a fixed and sufficiently small ε), which may prove extra advantageous in applications. Indeed, although this $\max(\varepsilon, h)$-factor does not increase either the asymptotic (w.r.t. ε) or numerical (w.r.t. h) order of the method, it could still contribute to the accuracy of the method, recalling that both ε and h are small.

We note that estimate (1.4.7) implicitly assumes that the phase (1.2.9), which appears in the WKB-based transformation to obtain the IVP (1.4.2) as well as in the scheme update formula, is exactly computable. A generalization of the above estimate that also incorporates possible phase errors can be found in Theorem 3.4.3. We note that the strategy to obtain this more general result is the same as in [AKU22], where the authors proposed a refined error analysis for the second order method from [AAN11].

Several numerical experiments show that the new derived third order scheme is significantly more efficient than the second order scheme from [AAN11]. Indeed, when aiming for the same accuracy, the novel scheme is faster up to a factor of 40, according to the results obtained for the examples discussed in Chapter 3.

1.5 Optimally truncated WKB approximation: exponentially small errors

The WKB-based second order scheme from [AAN11] as well as the third order scheme we develop in Chapter 3 are efficient methods for solving (1.0.1), but they are both limited to an asymptotic accuracy of $\mathcal{O}(\varepsilon^3)$, as $\varepsilon \to 0$, see also (1.4.7). This is simply due to the fact that both methods rely on the same analytic pre-processing, which utilizes the phase (1.2.9) of the second order WKB approximation (1.2.8). Indeed, when constructing numerical methods based on the Picard approximation of the solution of the smoother problem (1.4.2) (as done in Chapter 3), it is the $\mathcal{O}(\varepsilon)$-factor on the r.h.s. of (1.4.2) that ultimately imposes a constraint on the maximum achievable error order w.r.t. ε. More precisely, using this approach, it is not feasible to achieve a higher asymptotic accuracy than $\mathcal{O}(\varepsilon^3)$.

One possible approach to obtain a higher asymptotic order is to implement directly an arbitrary order WKB approximation (1.2.7) as an approximate solution to (1.0.1). This approach strongly differs from the two methods employed in Chapter 2 and Chapter 3. In

particular, it does not align with classical numerical methods, as it does not involve the introduction of a grid size h, and consequently, there is no convergence as $h \to 0$. Instead, the resulting approximation error will be of order $\mathcal{O}(\varepsilon^N)$ as $\varepsilon \to 0$, where N refers to chosen number of terms in the truncated asymptotic WKB series, see (1.2.1)-(1.2.2). As this WKB series is typically divergent, it is evident that there is some best attainable accuracy for the WKB approximation, achieved with the optimal choice $N = N_{opt}$.

We will follow this approach in Chapter 4, where we implement (1.2.7) as an approximate solution of the IVP (1.0.1), (1.4.1) for a given interval $I = [\xi, \eta]$. To this end, the complex constants $\alpha_{N,\varepsilon}$ and $\beta_{N,\varepsilon}$ in (1.2.7) are determined such that (1.2.7) satisfies the initial conditions (1.4.1) exactly; for more details, see Chapter 4.

In particular, we mainly focus on the following goals and questions:

- Given a coefficient function $a(x)$, one objective is to analyze the growth behavior of the functions $S_n^{\pm}$ appearing in the WKB series (1.2.2) w.r.t. $n \in \mathbb{N}_0$. Under which assumptions on $a(x)$ can we derive explicit (w.r.t. n) upper bounds for the norms $\|S_n^{\pm}\|_{L^\infty(I)}$?

- By estimating the residual of the WKB approximation (1.2.7) w.r.t. the ODE (1.0.1), and by utilizing the established explicit upper bounds on the norms $\|S_n^{\pm}\|_{L^\infty(I)}$, our goal is to derive an error estimate for (1.2.7) which is explicit w.r.t. ε and N. Indeed, this explicitness is crucial for a further analysis of the approximation error w.r.t. the truncation order N.

- The computation of the functions $S_n^{\pm}$ in (1.2.7) involves several integrals which in general have to be approximated, incurring additional errors. Consequently, the question arises: What impact do these approximation errors have on the overall accuracy of (1.2.7)?

- The main goal is then to understand the ε-dependence of different truncation strategies for the asymptotic WKB series (1.2.2). Which choice of N is appropriate or even optimal, in the sense of minimizing the resulting approximation error for the WKB approximation (1.2.7)? Given the optimally truncated WKB approximation, what is the order of corresponding optimal error?

The key ingredient for our error analysis in Chapter 4 is the understanding of the growth of the functions $S_n^{\pm}$ w.r.t. $n \in \mathbb{N}_0$. The crucial assumption we introduce for this analysis is that the coefficient function a is analytic on a complex, bounded and simply connected neighbourhood $G \subset \mathbb{C}$ of the interval I. Indeed, this enables us to employ Cauchy's integral formula in order establish upper bounds for the norms $\|S_n^{\pm}\|_{L^\infty(I)}$, utilizing the recurrence relation (1.2.4)-(1.2.6). We find that these norms can be estimated as

$$\|S_n^{\pm}\|_{L^\infty(I)} \leq C K_2^n n^n \,, \quad n \in \mathbb{N}_0 \,, \tag{1.5.1}$$

where $C, K_2 > 0$ are positive constants independent of n; for more details, see Corollary 4.3.3. Further, we observe through several numerical experiments in Chapter 4 that the r.h.s. of (1.5.1) not only provides an upper bound, but also often represents the asymptotic behavior of $\|S_n^{\pm}\|_{L^\infty(I)}$, as $n \to \infty$.

Our first main result in Chapter 4 is stated in Theorem 4.3.7, which provides an explicit (w.r.t. ε and N) estimate for the error $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)}$ of the WKB approximation. As a consequence, the approximation error is of order $\mathcal{O}(\varepsilon^N)$, as $\varepsilon \to 0$. The explicitness of this estimate then allows to investigate the error w.r.t. the truncation order N. More precisely, the optimal truncation order N_{opt} can be predicted by minimizing the derived upper error bound w.r.t. N, and is found to be proportional to ε^{-1}.

The second main result in Chapter 4 is Corollary 4.4.1, which states that a truncation order $N = N(\varepsilon) \sim \varepsilon^{-1}$ leads to the following error estimate:

$$\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} \leq \frac{C}{\varepsilon^2} \exp\left(-\frac{r}{\varepsilon}\right) , \tag{1.5.2}$$

where $C, r > 0$ are constants independent of ε. Consequently, the error of the optimally truncated WKB approximation is of order $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, as $\varepsilon \to 0$.

We confirm the theoretical results established in Chapter 4 with several numerical experiments. Furthermore, we discuss in Appendix 4.A a class of coefficient functions $a(x)$, which yield convergent WKB series. In particular, in this case the function $n \mapsto \|S_n^{\pm}\|_{L^\infty(I)}$ does not grow as the r.h.s. of (1.5.1) may suggest.

1.6 State of the art

We note that the development of numerical methods for the oscillatory equation (1.0.1) is an active field of research. Hence, let us mention here recent non-standard approaches that are suitable for tackling problems corresponding to (1.0.1).

We start by citing approaches which also rely on the WKB approximation (1.2.7) or which are at least strongly related. In [JL03, Jah04, LJL05] a variety of *adiabatic integrators* were developed, which all rely on first transforming the given ODE into a smoother problem before the latter is solved numerically. Notably, their analytic transformation for eliminating the dominant oscillations is closely related to a zeroth order WKB approximation (i.e. $N = 0$ in (1.2.7)). A localized variant of this transformation also serves as the foundation of the *modified Magnus method* in [Ise02, §5].

A first order WKB approximation (i.e. $N = 1$ in (1.2.7)) was the basis in [Neg08, AMN07] for the construction of a WKB-based finite element method (FEM) for (1.0.1) (see also [Neg05, §2]). To this end, the authors introduced an appropriate finite dimensional space that is based on so-called "WKB hat-functions", which in the limit $h \to 0$ reduce to standard linear hat-functions. However, the construction of the WKB hat-functions requires a rather technical restriction on the mesh size h to ensure that the method is valid.

Relying on a second order WKB approximation (i.e. $N = 2$ in (1.2.7)) we mention again the second order scheme (w.r.t. the step size h) from [AAN11], which is the basis for our methods in Chapter 2 and Chapter 3. A variety of extensions or modifications of this work have also been recently developed. First, in [AN18] the authors analyzed the numerical coupling of oscillatory and evanescent regimes. To this end, they rely on a domain decomposition approach, using the WKB-based marching method from [AAN11] in the oscillatory regime and a WKB-based FEM similar to [Neg05, §2] for the evanescent regime. Notably, the authors assumed the regimes to be separated by jump discontinuities

in the coefficient function $a(x)$ to avoid turning points. An extension to this method was later proposed in [AD20], where the authors considered a turning point of first order in their domain. Assuming that the given coefficient function $a(x)$ is linear or quadratic in a small neighbourhood of the turning point, they analytically solved the problem in this neighbourhood in terms of Airy and parabolic cylinder functions, respectively. This analytical solution was then coupled to a numerical solution away from the turning point. Let us also mention again the work [AKU22], where the authors refined the error analysis from [AAN11] by incorporating also possible phase errors. As the phase (1.2.9) cannot be computed exactly in general, the authors employed a spectral integration routine which reduced the impact of phase errors within the WKB-based scheme to relative machine precision (compared to other terms). As already mentioned in Section 1.4, the refined error analysis from [AKU22] is also the basis for Theorem 3.4.3 in Chapter 3.

In [Gei11], efficient one-step methods were proposed, which are suitable for solving linear systems of equations, which contain the Schrödinger equation (1.0.1) as a special case. Inspired by the strategy in [AAN11], the basis for these methods was a WKB-type analytic transformation, which transforms the given ODE into a smoother one, with a system matrix of order $\mathcal{O}(\varepsilon^n)$, $n \in \mathbb{N}$. This transformation ansatz should be compared to the *super-adiabatic transformation* from [HLW06, §XIV].

We also cite again the article [HLH16], where the authors proposed the RKWKB method, a coupling of a WKB-based $\mathcal{O}(h)$ scheme with a Runge-Kutta method, combined with an adaptive step size control. This method was the motivation for our strategy in Chapter 2, where it is included in our numerical comparison. In their more recent paper [AHLH20], an extension of the RKWKB method was proposed in order to also be able to solve equations involving a first order term $\gamma(x)\varphi'(x)$. There, the authors also used a slightly different switching mechanism in comparison to [HLH16]. The method [AHLH20] was also implemented as an open-source numerical routine, called `oscode`, see [Ago20].

Similar to Chapter 4 in the present thesis, the question about the best attainable accuracy of the WKB approximation (1.2.7) was also addressed in [Win05], where the author compared the WKB series with the exact solution represented by a convergent Bremmer series. We note that the author also finds the optimal truncation order for the WKB series to be proportional to ε^{-1} (as in Chapter 4). However, the strategy to derive this result relies on several additional asymptotic approximations, and is therefore less rigorous than our strategy in Chapter 4.

Additionally, numerous numerical approaches for (1.0.1) exist that are not directly related to WKB theory. Instead of attempting to list all of them, we will mention only a few.

In [AB22], the authors proposed a numerical algorithm that switches adaptively between a defect correction iteration, based on an asymptotic expansion, for oscillatory regions of the solution and a conventional Chebyshev collocation solver for smoother regions. The basis for the defect correction is the residual of a nonlinear Riccati equation that is satisfied by the derivative of the phase function $z(x)$ in the substitution $\varphi(x) = \exp(z(x))$, where φ is the solution to the original ODE. While the method is demonstrated to be highly accurate and efficient, a comprehensive error analysis was deferred to future work. This numerical algorithm was also implemented as an open-source software, called `ricatti`, see [AB23].

Further, in [HBR15, BR16, Bre18], an algorithm was proposed which relies on finding a non-oscillatory phase function $\alpha(x)$ such that the functions $u(x) = \cos(\alpha(x))|\alpha'(x)|^{-1/2}$ and $v(x) = \sin(\alpha(x))|\alpha'(x)|^{-1/2}$ comprise a basis in the space of solutions of (1.0.1). While such a function $\alpha(x)$ may not always exist, they proved that if $a(x)$ is non-oscillatory, there exists a non-oscillatory function $\alpha(x)$ such that the above functions $u(x)$ and $v(x)$ approximate solutions of (1.0.1) with $\mathcal{O}(\mu^{-1}\varepsilon\exp(-\mu/\varepsilon))$ accuracy, where μ is a constant depending on a. Similar to our approach in Chapter 4 (with $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$ accuracy), this method involves no introduction of a grid size h and hence there is no convergence as $h \to 0$.

1.7 Structure and Authorship

In this section, we provide an overview of the main body of this thesis, which comprises three chapters. Additionally, we specify the authorship for each chapter.

In **Chapter** 2 we build upon the WKB-based one-step method from [AAN11] for solving the one-dimensional stationary Schrödinger equation in the highly oscillatory regime. We extend this WKB method by implementing an adaptive step size control combined with an automated methods switching, allowing the algorithm to switch to a standard Runge-Kutta method in smoother (i.e. less oscillatory) regions. We compare our approach to the similar strategy [HLH16, AHLH20] on two numerical examples and illustrate the advantages of our new approach w.r.t. accuracy and efficiency.

The content of this chapter is a joint work with Anton Arnold and Kirian Döpfner. The results were published in [KAD21]. The author of this thesis contributed by doing all numerical simulations and working out the draft of the paper. The coauthors contributed by providing the motivation for the paper and by reviewing the draft.

Chapter 3 is devoted to the development of a higher order extension (w.r.t. the step size h) of [AAN11] for solving the one-dimensional stationary Schrödinger equation in the highly oscillatory regime. By developing sufficiently accurate quadratures for several (iterated) oscillatory integrals occurring in the Picard approximation of the solution, we obtain a one-step method that is third order w.r.t. the step size. The accuracy and efficiency of the new method are illustrated through several numerical examples.

The content of this chapter is a joint work with Anton Arnold and was submitted for publication under [AK24]. The author of the present thesis contributed by working out the mathematical details, conducting all numerical simulations, and drafting of the paper. The coauthor contributed primarily by reviewing the draft of the paper.

In **Chapter** 4 we implement and analyze an arbitrary order WKB approximation for the one-dimensional stationary Schrödinger equation in the highly oscillatory regime. Given that the coefficient in the equation is analytic, we derive an explicit error estimate in terms of the small parameter ε and the truncation order N. For any fixed ε, we find the optimal truncation order N_{opt} to be proportional to ε^{-1}, yielding a corresponding optimal error of the order $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, with some parameter $r > 0$. The theoretical results presented in this chapter are confirmed by several numerical examples.

The content of this chapter is a joint work with Anton Arnold, Christian Klein, and Jens Markus Melenk. The results were submitted for publication under [AKKM24]. The author of this thesis contributed by working out the mathematical details, conducting all numerical simulations, and drafting of the paper. The coauthors were involved by working out mathematical details and reviewing the draft of the paper.

Bibliography

[AAN11] A. Arnold, N. B. Abdallah, and C. Negulescu. WKB-Based Schemes for the Oscillatory 1D Schrödinger Equation in the Semiclassical Limit. *SIAM J. Numer. Anal.*, 49:1436–1460, 2011.

[AB22] F. J. Agocs and A. H. Barnett. An adaptive spectral method for oscillatory second-order linear ODEs with frequency-independent cost. *arXiv*, abs/2212.06924, 2022.

[AB23] F. J. Agocs and A. H. Barnett. riccati: an adaptive, spectral solver for oscillatory ODEs. *Journal of Open Source Software*, 8(86):5430, 2023.

[AD20] A. Arnold and K. Döpfner. Stationary Schrödinger equation in the semiclassical limit: WKB-based scheme coupled to a turning point. *Calcolo*, 57:44, 2020.

[Ago20] F. J. Agocs. (py)oscode: fast solutions of oscillatory odes. *Journal of Open Source Software*, 5(56):2830, 2020.

[AHLH20] F. J. Agocs, W. J. Handley, A. N. Lasenby, and M. P. Hobson. Efficient method for solving highly oscillatory ordinary differential equations with applications to physical systems. *Phys. Rev. Res.*, 2:013030, Jan 2020.

[AK24] A. Arnold and J. Körner. WKB-based third order method for the highly oscillatory 1D stationary Schrödinger equation. *arXiv*, abs/2402.18406, 2024.

[AKKM24] A. Arnold, C. Klein, J. Körner, and J. M. Melenk. Optimally truncated WKB approximation for the 1D stationary Schrödinger equation in the highly oscillatory regime. *arXiv*, abs/2401.10141, 2024.

[AKU22] A. Arnold, C. Klein, and B. Ujvari. WKB-method for the 1D Schrödinger equation in the semi-classical limit: enhanced phase treatment. *BIT Numerical Mathematics*, 62:1–22, 2022.

[AMN07] N. B. Abdallah, M. Mouis, and C. Negulescu. An accelerated algorithm for 2d simulations of the quantum ballistic transport in nanoscale MOSFETs. *J. Comput. Phys.*, 225:74–99, 2007.

[AN18] A. Arnold and C. Negulescu. Stationary Schrödinger equation in the semiclassical limit: numerical coupling of oscillatory and evanescent regions. *Numerische Mathematik*, 138:501–536, 2018.

[BO99] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers I: asymptotic methods and perturbation theory*. Springer New York, NY, 1 edition, 1999.

[BR16] J. Bremer and V. Rokhlin. Improved estimates for nonoscillatory phase functions. *Discrete and Continuous Dynamical Systems*, 36(8):4101–4131, 2016.

[Bre18] J. Bremer. On the numerical solution of second order ordinary differential equations in the high-frequency regime. *Applied and Computational Harmonic Analysis*, 44(2):312–349, 2018.

[CS58] E. D. Courant and H. S. Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 281:360–408, 1958.

[DP80] J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980.

[Gea71] C. William Gear. *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice-Hall, 1st edition, 1971.

[Gei11] J. Geier. *Efficient integrators for linear highly oscillatory ODEs based on asymptotic expansions*. PhD thesis, Technische Universität Wien, Wien, 2011.

[HBR15] Z. Heitman, J. Bremer, and V. Rokhlin. On the existence of nonoscillatory phase functions for second order ordinary differential equations in the high-frequency regime. *Journal of Computational Physics*, 290:1–27, 2015.

[HLH16] W. J. Handley, A. N. Lasenby, and M. P. Hobson. The Runge-Kutta-Wentzel-Kramers-Brillouin method. *arXiv*, abs/1612.02288, 2016.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer-Verlag, 2nd edition, 2006.

[INO06] A. Iserles, S. P. Nørsett, and S. Olver. Highly oscillatory quadrature: the story so far. *Proceedings of ENUMATH 2005, Santiago de Compostela*, pages 97–118, 2006.

[Ise02] A. Iserles. On the Global Error of Discretization Methods for Highly-Oscillatory Ordinary Differential Equations. *BIT Numerical Mathematics*, 42:561–599, 2002.

[Jah04] T. Jahnke. Long-time-step integrators for almost-adiabatic quantum dynamics. *SIAM Journal on Scientific Computing*, 25(6):2145–2164, 2004.

[JL03] T. Jahnke and C. Lubich. Numerical integrators for quantum dynamics close to the adiabatic limit. *Numerische Mathematik*, 94:289–314, 2003.

[KAD21] J. Körner, A. Arnold, and K. Döpfner. WKB-based scheme with adaptive step size control for the Schrödinger equation in the highly oscillatory regime. *J. Comput. Appl. Math.*, 404:113905, 2021.

[Lew68] H. R. Lewis. Motion of a Time-Dependent Harmonic Oscillator, and of a Charged Particle in a Class of Time-Dependent, Axially Symmetric Electromagnetic Fields. *Phys. Rev.*, 172:1313–1315, 1968.

[LJL05] K. Lorenz, T. Jahnke, and C. Lubich. Adiabatic Integrators for Highly Oscillatory Second-Order Linear Differential Equations with Time-Varying Eigendecomposition. *BIT Numerical Mathematics*, 45:91–115, 2005.

[LL85] L. D. Landau and E. M. Lifschitz. *Quantenmechanik*. Akademie-Verlag, 1985.

[MJK13] J.-F. Mennemann, A. Jüngel, and H. Kosina. Transient Schrödinger-Poisson simulations of a high-frequency resonant tunneling diode oscillator. *J. Comput. Phys.*, 239:187–205, 2013.

[MS03] J. Martin and D. J. Schwarz. WKB approximation for inflationary cosmological perturbations. *Physical Review D*, 67:083512, 2003.

[Neg05] C. Negulescu. *Asymptotical models and numerical schemes for quantum systems*. PhD thesis, Université Paul Sabatier, Toulouse, 2005.

[Neg08] C. Negulescu. Numerical analysis of a multiscale finite element scheme for the resolution of the stationary schrödinger equation. *Numerische Mathematik*, 108:625–652, 2008.

[SHMS98] J. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman. Resonant tunneling diodes: models and properties. *Proc. IEEE*, 86:641–660, 1998.

[Win05] S. Winitzki. Cosmological particle production and the precision of the WKB approximation. *Physical Review D*, 72:104011, 2005.

2 WKB-based scheme with adaptive step size control for the Schrödinger equation in the highly oscillatory regime

The content of this chapter has been published in [KAD21].

2.1 Introduction

This chapter deals with the numerical solution of the highly oscillatory 1D Schrödinger equation

$$\varepsilon^2\varphi''(x) + a(x)\varphi(x) = 0\,, \quad x \in \mathbb{R}\,. \tag{2.1.1}$$

Here, $0 < \varepsilon \ll 1$ is the rescaled Planck constant ($\varepsilon := \frac{\hbar}{\sqrt{2m}}$) and φ the (possibly complex valued) Schrödinger wave function. The real valued coefficient function $a(x)$ is related to the potential. We shall assume here that it is bounded away from zero, i.e. $a(x) \geq \tau$ for some $\tau > 0$. The (local) wave length of a solution φ to (2.1.1) is given by $\lambda(x) = \frac{2\pi\varepsilon}{\sqrt{a(x)}}$. Hence, for small values of ε the solution is highly oscillatory, especially in the semi-classical limit $\varepsilon \to 0$.

Oscillatory problems like (2.1.1) appear in a wide range of applications, e.g., quantum mechanics, electron transport in semiconductor devices, and acoustic scattering. For instance, the state of an electron that is injected with the prescribed energy E from the right boundary into an electronic device (e.g., diode), modeled on the interval $[0, 1]$, is described by the equation (see [AAN11])

$$\begin{cases} -\varepsilon^2\psi_E''(x) + V(x)\psi_E(x) = E\psi_E(x)\,, & x \in (0,1)\,, \\ \psi_E'(0) + \mathrm{i}\,k(0)\psi_E(0) = 0\,, \\ \psi_E'(1) - \mathrm{i}\,k(1)\psi_E(1) = -2\,\mathrm{i}\,k(1)\,, \end{cases} \tag{2.1.2}$$

where $k(x) := \varepsilon^{-1}\sqrt{E - V(x)}$ is the wave vector and V denotes the electrostatic potential. Note that our assumption $a(x) \geq \tau > 0$ implies $E > V(x)$, so the solution ψ_E becomes oscillatory. Then, one is often interested in macroscopic quantities like the electron density n and the current density j, which are given by

$$n(x) = \int_0^\infty f(k)|\psi_{E(k)}(x)|^2\,\mathrm{d}k\,, \quad j(x) = \varepsilon\int_0^\infty f(k)\Im\left(\overline{\psi_{E(k)}(x)}\psi'_{E(k)}(x)\right)\,\mathrm{d}k\,, \tag{2.1.3}$$

where f represents the injection statistics of the electrons. Here, $\Im(\cdot)$ denotes the imaginary part and $E(k) = \varepsilon^2k^2 + V$ means the energy for a given wave vector k. In order to compute

the quantities (2.1.3), the Schrödinger equation (2.1.2) has to be solved many times, as a fine grid in $E(k)$ is needed. Hence, efficient methods for the solution of (2.1.2) are of great interest in such applications. Instead of solving the boundary value problem (2.1.2) directly, one can also solve the equivalent initial value problem, which results if equation (2.1.1) on the interval $(0,1)$ is augmented with the initial values $\varphi(0) = \varphi_0 = 1$ and $\varepsilon\varphi'(0) = \varphi_0' = -\mathrm{i}\sqrt{a(0)}$ with $a(x) = E - V(x)$. The solution φ of this initial value problem and that one of problem (2.1.2) are then related by

$$\psi_E(x) = -\frac{2\,\mathrm{i}\,k(1)}{\varphi'(1) - \mathrm{i}\,k(1)\varphi(1)}\varphi(x)\,,$$

according to [AAN11]. Indeed, the method proposed in this chapter will deal only with initial value problems for the Schrödinger equation (2.1.1), but through this equivalence it is equally suitable for solving problem (2.1.2).

In [AAN11, JL03, LJL05], efficient and accurate WKB-based (named after the physicists Wentzel, Kramers, Brillouin; cf. [LL85]) numerical schemes have been developed for (2.1.1) in the oscillatory regime. By transforming out the dominant oscillations, they allow to compute a solution using a coarse spatial grid with step size $h > \lambda$. In fact, the grid limitation can there be reduced to at least $h = \mathcal{O}(\sqrt{\varepsilon})$. Now, in this chapter we add on top of the algorithm from [AAN11] an adaptive step size control as well as a switching mechanism. This allows the algorithm to switch to a standard ODE method (e.g., Runge-Kutta) during the computation in order to avoid technical or efficiency problems in regions where the coefficient function $a(x)$ is very small or indeed equal to zero. We recall that the WKB-approximation is not valid close to turning points, i.e. where $a(x) = 0$. A switching mechanism was also used in [HLH16], where the authors presented another WKB-based numerical scheme for the initial value problem corresponding to (2.1.1). Therefore, one goal of this chapter is to compare numerical results from our method with the results given by the method from [HLH16], by considering two examples where the analytical solution is known.

Since numerical methods for oscillatory problems is an active field of research, let us mention some references that are intended for more general oscillatory problems, and hence also include adequate methods for (2.1.1): first the two monographs [WW18, HLW06]. The *adiabatic integrators* of [HLW06, §XIV] are in fact closely related to a zeroth order WKB-approximation, see (2.2.2)-(2.2.3), below. Concerning (highly oscillatory) *Hamiltonian boundary value methods* we cite [BIMR19, ABI20].

This chapter is organized as follows: In Section 2.2 we give a short review of the second order (w.r.t. ε) WKB-marching method from [AAN11]. Section 2.3 then describes the adaptive step size control algorithm as well as the switching mechanism. In Section 2.4 we recap the Runge-Kutta-WKB method from [HLH16] and point out the difference between their step size control and the one used in this chapter. In Section 2.5 we present numerical investigations on the error estimators of our algorithm as well as a comparison of the numerical results of our method and the method from [HLH16]. We then conclude in Section 2.6.

2.2 The WKB-marching method

We aim at solving the Schrödinger equation (2.1.1), augmented with the initial conditions

$$\varphi(x_0) = \varphi_0\,, \quad \varepsilon\varphi'(x_0) = \varphi_0' \tag{2.2.1}$$

with some $x_0 \in \mathbb{R}$. First we shall review the basics of the second order (w.r.t. ε) WKB-marching method from [AAN11] with focus on the algorithm. The motivation for this method was the construction of a numerical scheme that is uniformly correct in ε and sometimes even asymptotically correct, i.e. the numerical error goes to zero with $\varepsilon \to 0$ while the grid size h remains constant. For further details see [AAN11]. The method consists of two parts:

1. Analytic pre-processing of (2.1.1) by transforming the equation into a smoother (i.e. less oscillatory) problem that can be solved accurately and efficiently on a coarse grid.
2. Obtaining a numerical scheme by discretization of the smoother problem.

Analytic pre-processing. The well-known WKB-approximation (cf. [LL85]) for the oscillatory regime where $a(x) \geq \tau$ for some $\tau > 0$, is based on inserting the ansatz

$$\varphi(x) \sim \exp\left(\frac{1}{\varepsilon}\sum_{p=0}^{\infty}\varepsilon^p\phi_p(x)\right) \tag{2.2.2}$$

into equation (2.1.1). After a comparison of ε-powers one obtains the first three functions $\phi_p(x)$ as

$$\phi_0(x) = \pm\mathrm{i}\int_{x_0}^{x}\sqrt{a(y)}\,\mathrm{d}y\,, \tag{2.2.3}$$

$$\phi_1(x) = \ln(a(x)^{-\frac{1}{4}})\,, \tag{2.2.4}$$

$$\phi_2(x) = \mp\mathrm{i}\int_{x_0}^{x} b(y)\,\mathrm{d}y\,, \quad b(x) := -\frac{1}{2a(x)^{\frac{1}{4}}}\left(a(x)^{-\frac{1}{4}}\right)''\,. \tag{2.2.5}$$

Here the symbols $\pm$ and $\mp$ in (2.2.3) and (2.2.5) correspond to the fact that there is always a pair of approximate solutions to the Schrödinger equation (2.1.1), by analogy to the two fundamental solutions of (2.1.1). Hence the general solution is then a linear combination of the two. Therefore, a truncation of the sum (2.2.2) after $p = 2$ leads to the second order (w.r.t. ε) asymptotic WKB-approximation of $\varphi(x)$:

$$\varphi_2(x) = c_1\frac{\exp\left(\frac{\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)\right)}{a(x)^{\frac{1}{4}}} + c_2\frac{\exp\left(-\frac{\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)\right)}{a(x)^{\frac{1}{4}}}\,, \tag{2.2.6}$$

with constants $c_1, c_2 \in \mathbb{C}$ to be determined from initial or boundary conditions, and the phase is

$$\phi^\varepsilon(x) := \int_{x_0}^{x}\left(\sqrt{a(y)} - \varepsilon^2 b(y)\right)\,\mathrm{d}y\,. \tag{2.2.7}$$

In the WKB-marching method of [AAN11] this second order WKB-approximation is used to transform (2.1.1) into a smoother problem. To clarify our terminology, we point out that this method (as well as the Runge-Kutta-WKB method of Section 2.4) has both a *WKB-order* (w.r.t. ε; referring to the used cut-off in the asymptotic expansion (2.2.2)) and a *numerical order* (w.r.t. the step size h; referring to the convergence order). Firstly, using the notation

$$U(x) = \begin{pmatrix} u_1(x) \\ u_2(x) \end{pmatrix} := \begin{pmatrix} a(x)^{\frac{1}{4}} \varphi(x) \\ \dfrac{\varepsilon \left(a(x)^{\frac{1}{4}} \varphi(x) \right)'}{\sqrt{a(x)}} \end{pmatrix} , \tag{2.2.8}$$

the second order differential equation (2.1.1) with the initial conditions (2.2.1) can be reformulated as a system of first order differential equations:

$$\begin{cases} U'(x) = \left[\frac{1}{\varepsilon} \mathbf{A}_0(x) + \varepsilon \mathbf{A}_1(x) \right] U(x) , \quad x > x_0 , \\ U(x_0) = U_I . \end{cases} \tag{2.2.9}$$

Here, the two matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ are given by

$$\mathbf{A}_0(x) := \sqrt{a(x)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} ; \quad \mathbf{A}_1(x) := \begin{pmatrix} 0 & 0 \\ 2b(x) & 0 \end{pmatrix} .$$

Then, the first order system (2.2.9) for $U(x)$ is transformed by the change of variables

$$Z(x) = \begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix} := \exp\left(-\frac{\mathrm{i}}{\varepsilon} \mathbf{\Phi}^\varepsilon(x) \right) \mathbf{P} U(x) , \tag{2.2.10}$$

with the two matrices

$$\mathbf{P} := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i} & 1 \\ 1 & \mathrm{i} \end{pmatrix} ; \quad \mathbf{\Phi}^\varepsilon(x) := \begin{pmatrix} \phi^\varepsilon(x) & 0 \\ 0 & -\phi^\varepsilon(x) \end{pmatrix} ,$$

where ϕ^ε is the phase function defined in (2.2.7). This leads to the system

$$\begin{cases} Z'(x) = \varepsilon \mathbf{N}^\varepsilon(x) Z(x) , \quad x > x_0 , \\ Z(x_0) = Z_I = \mathbf{P} U_I , \end{cases} \tag{2.2.11}$$

where $\mathbf{N}^\varepsilon(x)$ is a (Hermitian) matrix with only off-diagonal non-zero entries:

$$N^\varepsilon_{1,2}(x) = b(x) \, \mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)} , \quad N^\varepsilon_{2,1}(x) = b(x) \, \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)} .$$

Since the transformation (2.2.10) eliminated the dominant oscillations, the system (2.2.11) can be solved numerically on a coarse grid $\{x_n, n \in \mathbb{N}_0\}$. Then the original solution can be recovered by the inverse transformation

$$U(x) = \mathbf{P}^{-1} \exp\left(\frac{\mathrm{i}}{\varepsilon} \mathbf{\Phi}^\varepsilon(x) \right) Z(x) .$$

It should be noted that, throughout the whole transformation from $U(x)$ to $Z(x)$, the phase integral $\phi^\varepsilon(x)$ is assumed to be known exactly. For a generalization using a spectral method to numerically compute the phase see [AKU22].

Numerical scheme. The derivation of the second order (in h) scheme for (2.2.11) is obtained via the second order Picard approximation

$$\widetilde{Z}_{n+1} := \widetilde{Z}_n + \varepsilon \int_{x_n}^{x_{n+1}} \mathbf{N}^\varepsilon(x)\,\mathrm{d}x\widetilde{Z}_n + \varepsilon^2 \int_{x_n}^{x_{n+1}} \mathbf{N}^\varepsilon(x) \int_{x_n}^{x} \mathbf{N}^\varepsilon(y)\,\mathrm{d}y\mathrm{d}x\widetilde{Z}_n\,. \tag{2.2.12}$$

Since the entries of $\mathbf{N}^\varepsilon(x)$ are highly oscillatory, (2.2.12) involves (iterated) oscillatory integrals. With ϕ^ε assumed to be known exactly, they are then approximated using similar techniques as the *asymptotic method* in [INO06]. The first order (in h) scheme for (2.2.11) is derived by only taking into account the first two terms from (2.2.12). For both of these schemes we introduce the following notations:

$$b_0(y) := \frac{b(y)}{2\left(\sqrt{a(y)} - \varepsilon^2 b(y)\right)}\,; \quad b_{k+1}(y) := \frac{1}{2\left(\phi^\varepsilon(y)\right)'}\frac{\mathrm{d}b_k}{\mathrm{d}y}(y)\,, \quad k = 0,1,2\,;$$
$$h_1(y) := \mathrm{e}^{\mathrm{i}y} - 1\,; \quad h_2(y) := \mathrm{e}^{\mathrm{i}y} - 1 - \mathrm{i}y\,.$$

Further, let $\{x_0, x_1, ..., x_N\}$ be a grid we want to compute the solution on, and $h := \max_{1\le n\le N} |x_n - x_{n-1}|$ be the step size. Then both schemes read as follows:

First order scheme: Let $Z_0 := Z_I$ be the initial condition and let $n = 0, ..., N-1$. Then the algorithm updates as

$$Z_{n+1} = \left(I + \mathbf{A}_n^1\right) Z_n\,, \tag{2.2.13}$$

with the (Hermitian) matrix

$$\mathbf{A}_n^1 := \varepsilon^3 b_1(x_{n+1}) \begin{pmatrix} 0 & \mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} h_1\left(-\frac{2}{\varepsilon}s_n\right) \\ \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} h_1\left(\frac{2}{\varepsilon}s_n\right) & 0 \end{pmatrix}$$
$$- \mathrm{i}\varepsilon^2 \begin{pmatrix} 0 & b_0(x_n)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} - b_0(x_{n+1})\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} \\ b_0(x_{n+1})\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} - b_0(x_n)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} & 0 \end{pmatrix},$$

and the phase increments

$$s_n := \phi^\varepsilon(x_{n+1}) - \phi^\varepsilon(x_n)\,.$$

Second order scheme: Let $Z_0 := Z_I$ be the initial condition and let $n = 0, ..., N-1$. Then the algorithm updates as

$$Z_{n+1} = \left(I + \mathbf{A}_{mod,n}^1 + \mathbf{A}_n^2\right) Z_n\,, \tag{2.2.14}$$

with the (Hermitian) matrix

$$
\begin{aligned}
&\mathbf{A}^1_{mod,n} := \\
&-\mathrm{i}\varepsilon^2 \begin{pmatrix} 0 & b_0(x_n)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} - b_0(x_{n+1})\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} \\ b_0(x_{n+1})\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} - b_0(x_n)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} & 0 \end{pmatrix} \\
&+\varepsilon^3 \begin{pmatrix} 0 & b_1(x_{n+1})\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} - b_1(x_n)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} \\ b_1(x_{n+1})\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_{n+1})} - b_1(x_n)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)} & 0 \end{pmatrix} \\
&+\mathrm{i}\varepsilon^4 b_2(x_{n+1}) \begin{pmatrix} 0 & -\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)}\, h_1\left(-\frac{2}{\varepsilon}s_n\right) \\ \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)}\, h_1\left(\frac{2}{\varepsilon}s_n\right) & 0 \end{pmatrix} \\
&-\varepsilon^5 b_3(x_{n+1}) \begin{pmatrix} 0 & \mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)}\, h_2\left(-\frac{2}{\varepsilon}s_n\right) \\ \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x_n)}\, h_2\left(\frac{2}{\varepsilon}s_n\right) & 0 \end{pmatrix},
\end{aligned}
$$

and the diagonal matrix

$$
\begin{aligned}
\mathbf{A}^2_n := &-\mathrm{i}\varepsilon^3 (x_{n+1}-x_n) \frac{b(x_{n+1})b_0(x_{n+1}) + b(x_n)b_0(x_n)}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\
&-\varepsilon^4 b_0(x_n)b_0(x_{n+1}) \begin{pmatrix} h_1\left(-\frac{2}{\varepsilon}s_n\right) & 0 \\ 0 & h_1\left(\frac{2}{\varepsilon}s_n\right) \end{pmatrix} \\
&+\varepsilon^5 b_1(x_{n+1}) \left[b_0(x_n) - b_0(x_{n+1})\right] \begin{pmatrix} h_2\left(-\frac{2}{\varepsilon}s_n\right) & 0 \\ 0 & -h_2\left(\frac{2}{\varepsilon}s_n\right) \end{pmatrix}.
\end{aligned}
$$

2.3 Step size control and switching mechanism

The WKB scheme is efficient in the highly oscillatory regime, but not applicable close to turning points, i.e. zeros of $a(x)$, see [AD20]. This is evident already from the transformation (2.2.8), which does not make sense when $a(x) \leq 0$. For mixed problems, e.g., the Airy equation on $\mathbb{R}_0^+$ (see Section 2.5.1), which has a turning point at $x = 0$, it is therefore convenient to couple two different methods: a method for highly oscillatory ODEs (e.g., WKB-based) away from the turning point, and a standard ODE method (e.g., Runge-Kutta) close to the turning point, where the solution is smooth anyhow. Here, we choose the well-known Runge-Kutta-Fehlberg 4(5) (RKF45) scheme (cf. [HNW00]) as the standard ODE method. The latter method will be applied directly on equation (2.1.1) and not on (2.2.11), since the WKB-transformation (2.2.8)-(2.2.11) is not permitted at turning points. The exact switching mechanism as well as the introduction of an adaptive step size control to the algorithm will be described in the two following subsections.

2.3.1 The adaptive step size controller

In order to compute the solution efficiently, an adaptive step size controller, based on an estimator for the local truncation error, will be added to the numerical methods. This control allows the step size to increase or decrease while aiming to keep the error estimator as close as possible within a given error tolerance. To illustrate the functionality of this

step size controller, we shall consider a numerical scheme of order k. We are then able to apply this step size control individually to the different methods mentioned above.
Let $Y_n^{(k)} = (\varphi_n^{(k)}, \varphi_n'^{(k)})$ and $Y_n^{(k+1)} = (\varphi_n^{(k+1)}, \varphi_n'^{(k+1)})$ be two numerical solutions of order k and $k+1$ to approximate the exact solution $Y(x_n) = (\varphi(x_n), \varphi'(x_n))$ of the initial value problem (2.1.1), (2.2.1). E.g., one could choose the WKB schemes (2.2.13) and (2.2.14) of h-order 1 and 2, respectively. Next we want to decide whether to accept the numerical solution at x_n (typically the more accurate solution $Y_n^{(k+1)}$) or rather to retry the computation with a modified step size. To this end we define the estimator for the local truncation error as

$$\mathrm{est}_n := \|Y_n^{(k)} - Y_n^{(k+1)}\|_\infty \,. \tag{2.3.1}$$

Let $h_{n,trial} := x_n - x_{n-1}$ be the (trial) step size which was used to compute the solutions at the current step n. We then use the common approach of varying the step size via the multiplicative control

$$h_{new} := \theta_n \cdot h_{n,trial} \,.$$

Here, we choose the factor θ_n to be based on the so-called *elementary controller* (e.g., see [Sö02]). Additionally, we introduce limitations in such way that the step size controller responds "smoothly" to abrupt changes in the solution behavior, that is, the ratio between two consecutive step sizes should not be exorbitantly large or small. That said, we choose the factor similar to [But08, p. 310] as

$$\theta_n := \max\left(0.5, \min\left(2, 0.9\left(\frac{\mathrm{ATol} + \mathrm{RTol}\cdot\|Y_n^{(k+1)}\|_\infty}{\mathrm{est}_n}\right)^{\frac{1}{k+1}}\right)\right) \,, \tag{2.3.2}$$

where $\mathrm{ATol} = \eta \cdot \mathrm{Tol}$ and $\mathrm{RTol} = \mathrm{Tol}$ are absolute and relative error tolerances for a given master tolerance Tol, the values 0.5 and 2 are design parameters that limit the ratio of two consecutive step sizes from below and above, and 0.9 is a common safety factor for increasing the probability of the next step to be accepted. Here, η is a scaling factor representing the gradual switch-over between absolute and relative errors, depending on the behavior of the solution. This is because for $\|Y_n^{(k+1)}\|_\infty \to 0$ the ATol term in the numerator in (2.3.2) is dominating, whereas for large values of $\|Y_n^{(k+1)}\|_\infty$ the RTol term is dominating. For our numerical simulations in Section 2.5 we choose $\eta = 10^{-2}$. If $\mathrm{ATol} + \mathrm{RTol}\cdot\|Y_n^{(k+1)}\|_\infty \geq \mathrm{est}_n$, we accept the n-th step with the step size defined as $h_n := h_{n,trial}$ and define the trial step size for the next step as

$$h_{n+1,trial} := h_{new} \,.$$

However, if $\mathrm{ATol} + \mathrm{RTol}\cdot\|Y_n^{(k+1)}\|_\infty < \mathrm{est}_n$, the n-th step gets rejected and a reattempt is done with the smaller (trial) step size $h_{n,trial}$ by updating its value as

$$h_{n,trial} \to h_{new} \,.$$

Since such an acceptance criterion is based on aiming to maintain the local error in each step as close as possible to the given error tolerances it is often referred to as *error per step (EPS)* control. In practice, using EPS control one can hope to achieve a global truncation error proportional to $\mathrm{Tol}^{k/(k+1)}$ (e.g., see [But08, p. 311]).

2.3.2 The switching mechanism

As already mentioned above, the algorithm shall automatically switch between two numerical methods, the WKB method in the oscillatory regime and another method, which is valid close to turning points. To realize this dynamical switching mechanism we follow a similar strategy as in [AHLH20]. To illustrate this procedure we now consider two numerical schemes of order $k^{(1)}$ and $k^{(2)}$, where the superscripts (1) and (2) correspond to the two schemes. In each step, the adaptive step size algorithm from the previous section is applied to both schemes individually up to the definition of the quantities $\theta_n^{(1)}$ and $\theta_n^{(2)}$, i.e., we just evaluate (2.3.1) and (2.3.2). Then, after checking the acceptance criterion $\mathrm{ATol} + \mathrm{RTol} \cdot \|Y_n^{(k^{(i)}+1)}\|_\infty \geq \mathrm{est}_n^{(i)}$ for each scheme $(i) \in \{(1),(2)\}$, the switching mechanism intervenes and it selects the acceptable numerical method that yields the larger value of $\theta_n^{(i)}$, for $i = 1, 2$, hence yielding the larger proposed new step size. We thus favor the method with the smaller error estimator, discounted by its respective order $k^{(i)}$. More precisely, we define

$$\Theta_n := \begin{cases} \theta_n^{(1)}\,, & \text{(1) accepted, (2) rejected} \\ \theta_n^{(2)}\,, & \text{(1) rejected, (2) accepted} \\ \max\left(\theta_n^{(1)}, \theta_n^{(2)}\right)\,, & \text{otherwise} \end{cases}$$

and store the information on the method of choice in that n-th step. Through this procedure the algorithm does not only use the error estimator to find the next step size, but also to decide between the two methods.

If at least one method was accepted, the algorithm sets

$$h_n := h_{n,trial}\,,$$
$$h_{n+1,trial} := \Theta_n \cdot h_{n,trial}\,.$$

Otherwise a reattempt is done with the smaller (trial) step size $h_{n,trial}$ by updating its value as

$$h_{n,trial} \to \Theta_n \cdot h_{n,trial}\,.$$

We remark that the coupling of two methods, as presented above, could incur additional computational costs, since in every step both methods have to be applied in order to compute Θ_n. However, in our case the WKB method (for highly oscillatory regions) and the standard ODE solver (for smoother regions) complement each other very well, yielding better results concerning accuracy as well as overall efficiency (see also Figures 2.5.7-2.5.8, 2.5.11 and 2.5.16).

2.4 The Runge-Kutta-WKB method

In this section we give a short review of the Runge-Kutta-WKB (RKWKB) method presented in [HLH16] for the initial value problem (2.1.1), (2.2.1). The method is based on a

dynamic switching mechanism between a standard Runge-Kutta scheme and a new stepping procedure that uses the WKB-ansatz (2.2.2) as an approximation of the true solution. This stepping procedure reads as follows:

$$\varphi_{n+1} := \gamma_+ f_+(x_n + h_n) + \gamma_- f_-(x_n + h_n)\,, \tag{2.4.1}$$
$$\varphi'_{n+1} := \delta_+ f'_+(x_n + h_n) + \delta_- f'_-(x_n + h_n)\,, \tag{2.4.2}$$
$$x_{n+1} := x_n + h_n\,, \tag{2.4.3}$$

where

$$\gamma_\pm := \frac{\varphi'_n f_\mp(x_n) - \varphi_n f'_\mp(x_n)}{f'_\pm(x_n) f_\mp(x_n) - f'_\mp(x_n) f_\pm(x_n)}\,, \tag{2.4.4}$$
$$\delta_\pm := \frac{\varphi''_n f'_\mp(x_n) - \varphi'_n f''_\mp(x_n)}{f''_\pm(x_n) f'_\mp(x_n) - f''_\mp(x_n) f'_\pm(x_n)}\,, \tag{2.4.5}$$

and φ''_n is computed from φ_n and equation (2.1.1). Here, $f_\pm$ are chosen as WKB-approximations (2.2.2) of some finite order. For instance, one gets the second (WKB-)order method by setting

$$f_\pm(x) := \frac{\exp\left(\pm \frac{\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)\right)}{a(x)^{\frac{1}{4}}}\,.$$

Note that this choice is equal to (2.2.6), but one can easily choose $f_\pm$ with higher (WKB-) orders. However, the stepping procedure (2.4.1)-(2.4.5) is always a first order (in h) numerical method. This holds because the coefficients $\gamma_\pm$ and $\delta_\pm$ are chosen in such way that one finds

$$\varphi_{n+1} = \varphi_n + \varphi'_n h + \mathcal{O}(h^2)\,,$$
$$\varphi'_{n+1} = \varphi'_n + \varphi''_n h + \mathcal{O}(h^2)\,,$$

from (2.4.1)-(2.4.3), as stated in [HLH16].

It is also worth noting that in [HLH16] the authors use a slightly different dynamic switching mechanism and a different step size control compared to the algorithm presented in Section 2.3. Firstly, their estimator of the relative error within the WKB-stepping procedure uses the difference between two numerical solutions $\varphi_n^{(k)}$ and $\varphi_n^{(k+1)}$ of different WKB-orders instead of different h-orders, simply since they do not have schemes of two different h-orders at their disposal. The algorithm then decides between a WKB step and a RK step by choosing the method with the smaller error estimator. In their more recent paper [AHLH20] the authors use a similar step size control and switching mechanism as presented in Section 2.3, but they do not limit the ratio of two consecutive step sizes and provide no option for controlling both absolute and relative errors as done in (2.3.2).

The goal of the following section is to compare numerical results from the WKB-marching method to results one gets using the RKWKB method. To both methods we will apply (exactly) the step size control and switching mechanism from Section 2.3, for the sake of comparability. Since the WKB-stepping procedure of the RKWKB method is always

WKB+RKF45	WKB-marching method (see Section 2.2) + step size and switching algorithms to RKF45 (see Section 2.3)
RKWKB	original method from [HLH16]: Runge-Kutta-WKB (see Section 2.4) + original step size and switching algorithms (see Section 2.4)
RKWKBmod	modified method from [HLH16]: Runge-Kutta-WKB (see Section 2.4) + modified step size and switching algorithms (see Section 2.3)
RKF45	Runge-Kutta-Fehlberg 4(5) scheme + step size algorithm (see Section 2.3)

Table 2.5.1: Terminology for the methods to be compared.

of first order w.r.t. h, the definition of the error estimator (2.3.1) does not make sense. Therefore we shall use two different WKB orders (instead of h-orders) to be able to compute the error estimator in this case, as also done in [HLH16]. Further, since our algorithm consists of a different step size update formula and switching criterion, we will call this modified RKWKB method simply RKWKBmod. Since this modification may produce slightly different numerical results compared to [HLH16, AHLH20] we will also include the original RKWKB method from [HLH16] into our comparison.

2.5 Numerical results: WKB-marching method vs. the Runge-Kutta-WKB method

In this section we will compare numerical results of the WKB-marching method and of the RKWKB method by applying both algorithms to two examples, where exact analytical solutions are available. The first example corresponds to a linear coefficient function $a(x)$ and is taken from [HLH16, AHLH20], whereas the second example involves a quadratic function $a(x)$ and appears in [AD20]. In both examples the phase integral (2.2.7) in the WKB basis functions (2.2.3)-(2.2.5) can easily be computed exactly, hence we do not need any numerical integration routine here. By contrast, in [AHLH20] they evaluate the WKB basis functions with a numerical integration routine and therefore get another source of error in their method. Moreover, we will always use the second order WKB-marching method, since no scheme of higher WKB-order has been developed yet, see [AAN11].

For clarity of the presentation we shall use in the sequel the terminology for the methods to be compared according to Table 2.5.1.

All simulations in this chapter are done with MATLAB version 9.10.0.1669831 (R2021a).

2.5.1 First example: Airy function

The first example investigated in [HLH16] is the Airy equation

$$\begin{cases} \varepsilon^2 \varphi''(x) + x\varphi(x) = 0\,, \quad x > 0\,, \\ \varphi(0) = \frac{3^{-2/3} + \mathrm{i}\, 3^{-1/6}}{\Gamma(\frac{2}{3})}\,, \\ \varepsilon\varphi'(0) = \frac{3^{-1/3} - \mathrm{i}\, 3^{1/6}}{\varepsilon^{-1/3}\Gamma(\frac{1}{3})}\,, \end{cases} \tag{2.5.1}$$

which results if one chooses the coefficient function $a(x) = x$. Here, Γ denotes the gamma function. The exact solution to the problem (2.5.1) is given by

$$\varphi_{exact}(x) = \mathrm{Ai}(-\frac{x}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}(-\frac{x}{\varepsilon^{2/3}})\,,$$

where Ai and Bi denote the Airy functions of first and second kind, respectively. This example demonstrates very well the advantages of a WKB method, since the solution becomes more and more oscillatory for large values of x. While standard adaptive ODE methods, e.g., Runge-Kutta methods, would need to decrease the step size more and more, a WKB-based method does not have to resolve the individual oscillations. Actually, it even allows to increase the step size for large x and is therefore highly efficient. This can be seen, e.g., in Figures 2.5.4-2.5.6.

Before starting to discuss the numerical solution of (2.5.1) we first remark that the *evaluation* of an oscillatory function (even of trigonometric functions such as sin or cos) is numerically ill posed for very large values of x. This is a generic problem and not related to the numerical solution of (2.5.1), or to the choice of the numerical method:

Remark 2.5.1. *We consider to evaluate $\varphi_{exact}(x)$ for an argument x that is only known with finite accuracy, specified by machine eps. Then, using the lowest order expansions for φ_{exact} and φ'_{exact} from 2.A, we have*

$$\varphi_{exact}(x) \sim \frac{\varepsilon^{1/6}}{\sqrt{\pi}x^{1/4}} \exp\left(\mathrm{i}\left(\frac{\pi}{4} - \frac{2}{3}\frac{x^{3/2}}{\varepsilon}\right)\right) \qquad as \quad x \to \infty\,,$$

and the relative error of $\varphi_{exact}(x(1+eps))$ is asymptotically $eps\, x^{3/2}/\varepsilon$.

Considering double precision, e.g., with MATLAB*'s $eps \approx 2.2 \cdot 10^{-16}$ and $x = 50$, this generic evaluation error is $7.8 \cdot 10^{-10}$ for $\varepsilon = 10^{-4}$ and $7.8 \cdot 10^{-11}$ for $\varepsilon = 10^{-3}$. These unavoidable errors limit the achievable accuracy, and it will play a role in Figures 2.5.10-2.5.11 below.*

As a first step we shall now test the reliability of our choice of error estimator (2.3.1) – but only for the WKB steps of WKB+RKF45 and RKWKBmod. Since we know the exact solution to this problem, we can compute the local truncation error in each step and are able to compare it to the error estimator. Moreover, we apply the adaptive step size control from Section 2.3 with the error tolerance $\mathrm{Tol} = 10^{-5}$ and set $\varepsilon = 1$.

According to the results of Figures 2.5.1 and 2.5.2, the error estimator is in excellent agreement with the local truncation error (for this example). Hence, it seems to be an adequate choice. We also find a very good agreement by plotting the respective relative errors for one single step as functions of the step size h for a fixed starting point x_0, see Figure 2.5.3. Again, the relative error is for both methods much smaller than one. Note that in the case of WKB+RKF45 the relative error goes to zero, if h tends to zero. Therefore, the estimator seems to be asymptotically correct.

Remark 2.5.2. *The computation of the local truncation error involves the evaluation of the Airy function, which seems to have an accuracy problem for large values of x when using the standard routine* `airy()` *in* MATLAB*. Hence we used a modified implementation*

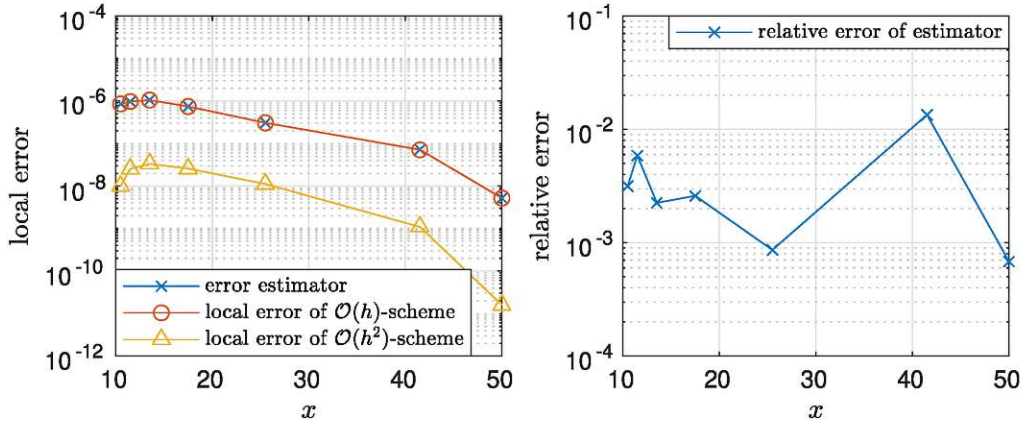

Figure 2.5.1: Left: The error estimator (2.3.1) in comparison to the actual local truncation error computed using only WKB steps for WKB+RKF45 of first order in h, (2.2.13), and second order in h, (2.2.14). Right: The relative error between the estimator and the local truncation error for the $\mathcal{O}(h)$-scheme. In both pictures we set $\mathrm{Tol} = 10^{-5}$ and $\varepsilon = 1$.

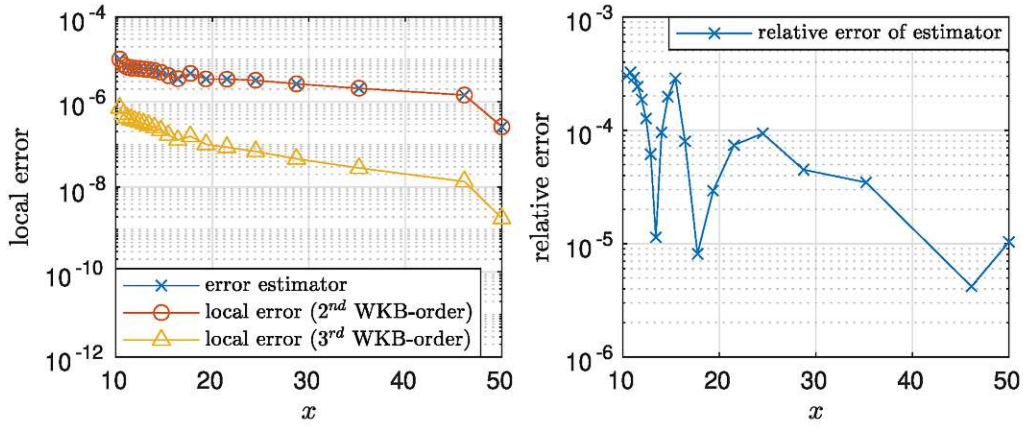

Figure 2.5.2: Left: The error estimator (2.3.1) in comparison to the actual local truncation error computed using only WKB steps in the RKWKBmod method. Since the method is of first order (in h), the estimator as well as the local errors were computed by using different WKB-orders instead (here order 2 and 3). Right: The relative error between the estimator and the local truncation error (2^{nd} WKB-order). In both pictures we set $\mathrm{Tol} = 10^{-5}$ and $\varepsilon = 1$.

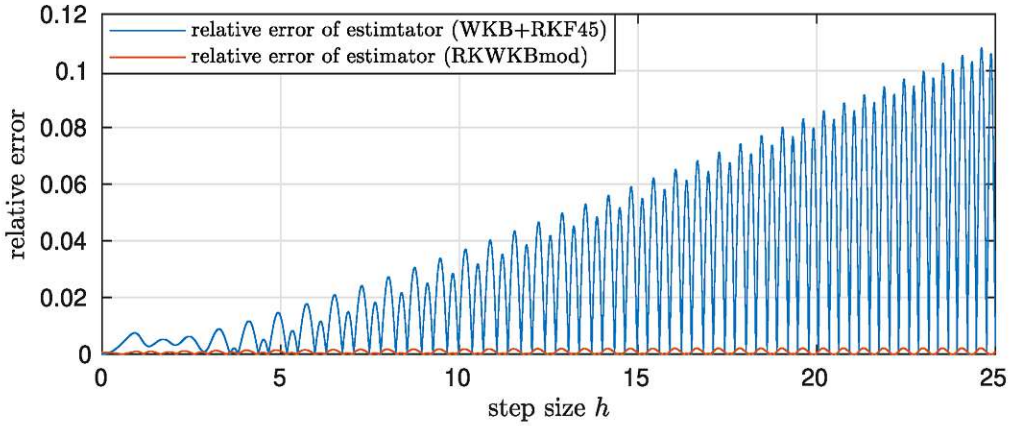

Figure 2.5.3: The relative error between the estimator (2.3.1) and the local truncation errors for WKB+RKF45 and RKWKBmod in one single WKB step as functions of the step size h for a fixed starting point $x_0 = 10$. Here, we set Tol $= 10^{-5}$ and $\varepsilon = 1$. For the computation of the estimator with the RKWKBmod method a 2^{nd} and 3^{rd} WKB-order was used.

	`airy()`	asymptotics
Ai	$x \in [0, 500]$	$x \in (500, \infty)$
Ai′	$x \in [0, 400]$	$x \in (400, \infty)$
Bi	$x \in [0, 500]$	$x \in (500, \infty)$
Bi′	$x \in [0, 400]$	$x \in (400, \infty)$

Table 2.5.2: Intervals for evaluating the Airy function of first and second kind as well as their derivatives. For small values of x the original function from MATLAB was used, and for large x the evaluation was performed using the asymptotic expressions (2.A.1)-(2.A.4) with truncation after $K = 3$.

of the Airy functions, which consists of the function `airy()` *from* MATLAB *for small values of x and asymptotic expressions for the Airy function for large values of x. More precisely, the evaluations were performed as given in Table 2.5.2. The asymptotic expressions are based on well-known expansions, which can be found in 2.A. The order for truncating the expansions was set to $K = 3$.*

Now, we will give numerical results for solving the initial value problem (2.5.1) for the Airy equation with WKB+RKF45 and RKWKBmod. We recall that the algorithm automatically chooses between RKF45 and the respective WKB steps. To illustrate this difference in the Figures 2.5.4-2.5.9 as well as Figures 2.5.13-2.5.14, we will mark RKF45 steps with red dots, WKB steps using WKB+RKF45 with blue squares, and WKB steps using the RKWKBmod method with green triangles. To exclude the turning point $x = 0$, we solve the Airy equation (2.5.1) in Figures 2.5.4 and 2.5.5 on $[0.1, 50]$. According to Figures 2.5.4 and 2.5.5, WKB+RKF45 seems to perform slightly better than the RKWKBmod method in this example, in matters of global error. But at the same time WKB+RKF45 needs significantly fewer steps than RKWKBmod. Also, within the algorithm using WKB+RKF45, the switch from RKF45 steps to WKB steps happens earlier as can be seen in Figure 2.5.4. In Figure 2.5.6, the relative global errors of both algorithms are compared on $[0.1, 10^8]$. We find that the algorithm using WKB+RKF45 made fewer steps (58 vs. 91) while producing

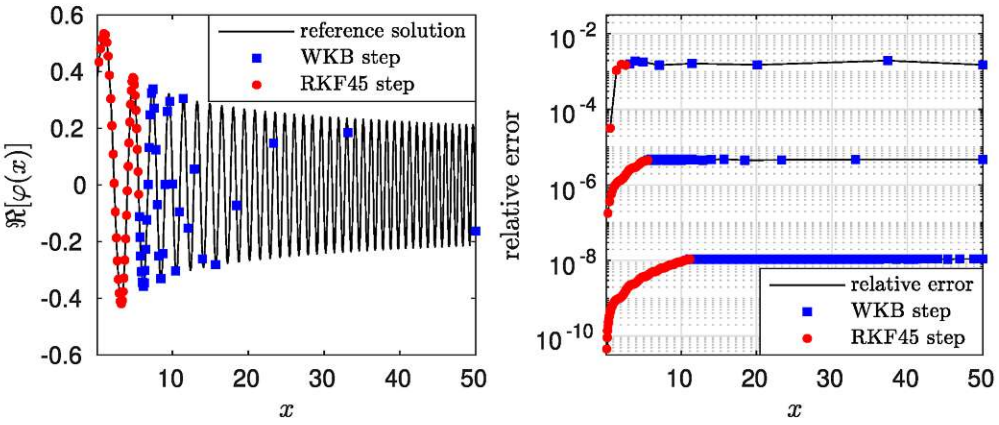

Figure 2.5.4: Left: Real part of the numerical solution obtained by using WKB+RKF45 compared to the (exact) reference solution (solid line) for Tol $= 10^{-6}$. Right: The global error for the choices Tol $= 10^{-3}, 10^{-6}, 10^{-9}$ (read from top to bottom). The respective overall number of steps made are 12, 77, and 856. For both pictures the initial step size was set to $h_{1,trial} = 0.5$ and the parameter ε was set to 1.

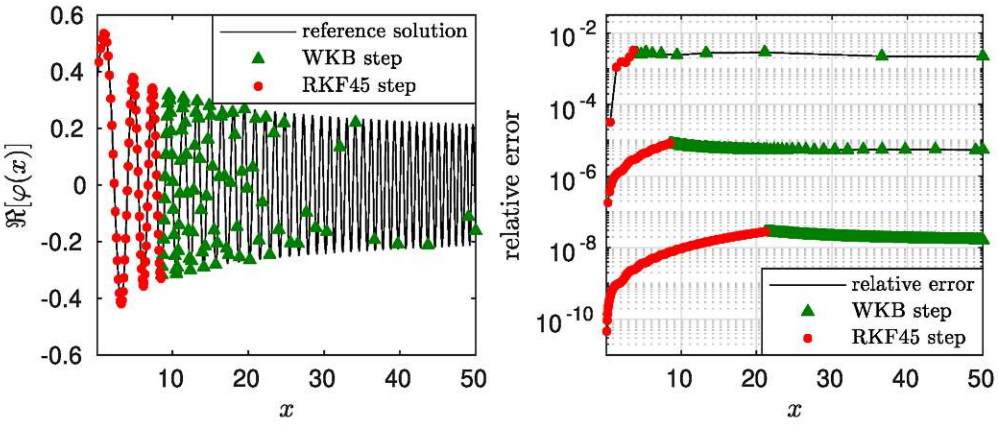

Figure 2.5.5: Left: Real part of the numerical solution obtained by using the RKWKBmod method compared to the (exact) reference solution (solid line) for Tol $= 10^{-6}$. Right: The global error for the choices Tol $= 10^{-3}, 10^{-6}, 10^{-9}$ (read from top to bottom). The respective overall number of steps made are 16, 171, and 2352. For both pictures the initial step size was set to $h_{1,trial} = 0.5$, the parameter ε was set to 1 and a third order WKB-ansatz was used.

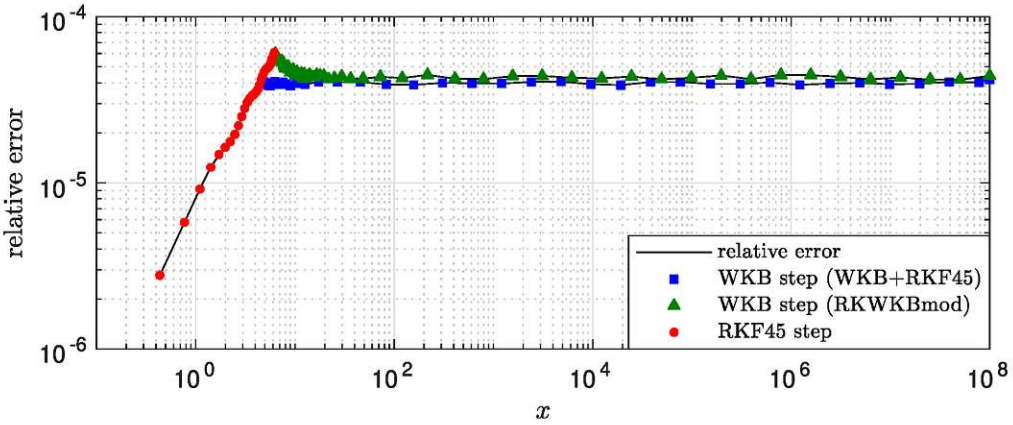

Figure 2.5.6: Global error comparison of WKB+RKF45 and RKWKBmod on $[0.1, 10^8]$ for $\varepsilon = 1$ using an initial step size $h_{1,trial} = 0.5$. The error tolerance was set to Tol $= 10^{-5}$. A third order WKB-ansatz was used for the RKWKBmod method. Overall the algorithm using WKB+RKF45 made 58 steps, whereas RKWKBmod made 91 steps.

a slightly lower global error at the same time. The almost identical grid spacing of both methods from $x \approx 300$ onwards is due to limiting the quotient of two consecutive step sizes, imposed in both methods. Furthermore, Figures 2.5.7 and 2.5.8 show the ratio between the WKB- and RK- error estimators as well as the ratio of the proposed step sizes (by the WKB and RKF schemes) for each step of the above simulations. For both WKB+RKF45 and RKWKBmod, these plots demonstrate well the superiority of the RK scheme in the less oscillatory regime, i.e. for x small, as the ratio of the error estimators is very large there. The switching points to the WKB schemes are well defined (again for both WKB methods) – due to the monotonous behavior of both the ratio of error estimators and the ratio of proposed step sizes. Note also that the step size ratios are bounded from above and below because of the limitations in (2.3.2). We also want to give a perception of how WKB+RKF45 can perform, if the phase integral (2.2.7) can not be evaluated exactly. For this purpose we use the *Clenshaw-Curtis* quadrature (cf. [CC60]) to approximate the phase (2.2.7). That spectral method was already used in combination with the WKB-marching method, see [AKU22, §5]. In [AKU22] they approximate the phase at first on Chebyshev collocation points throughout the *whole interval* and use barycentric interpolation for the ODE-grid points x_n. This was possible since they worked only on the "small" interval $[0, 1]$. In contrast, we will instead approximate the phase (2.2.7) in *each interval* $[x_n, x_{n+1}]$ individually, since we are dealing here with the "long" interval $[0.1, 10^8]$. Figure 2.5.9 gives a comparison of the global error for the Airy equation (2.5.1), when using the exact phase vs. a numerically computed phase. According to these results the relative φ-errors from computing the phase (2.2.7) numerically become visible only from $x \approx 10^7$ onwards.

We will next compare the numerical results of four methods, namely, WKB+RKF45, RKWKBmod, RKWKB, and RKF45 on the Airy equation for several values of the parameter ε. For RKF45, we do not use any smaller value of ε than 10^{-1}, as the CPU time gets exorbitantly large. In Figure 2.5.10 we compare the accuracy of each of the methods depending on Tol. We find that WKB+RKF45 outperforms the other methods, in matters

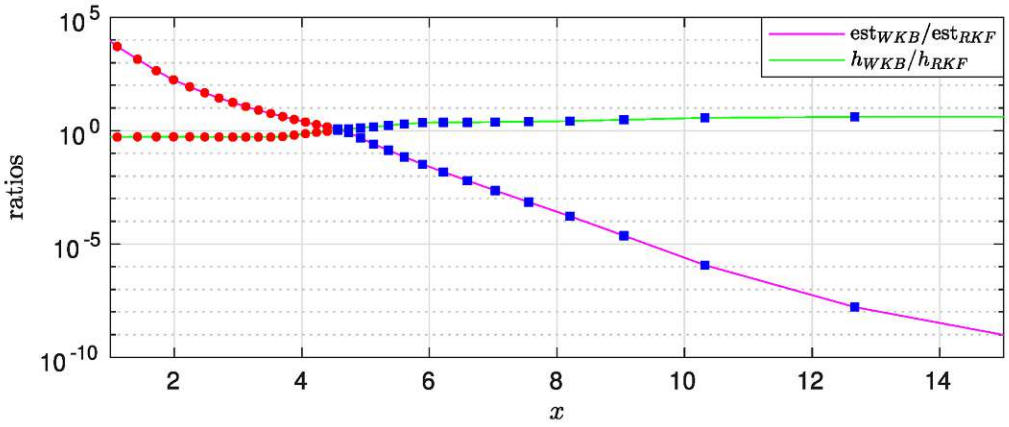

Figure 2.5.7: Numerical computation using WKB+RKF45: The magenta line indicates the ratio of the two error estimators (due to the WKB and RKF scheme). The green line gives the analogous ratio of the step sizes (locally) proposed by these two methods. For this computation we set $\mathrm{Tol} = 10^{-5}$ and $\varepsilon = 1$.

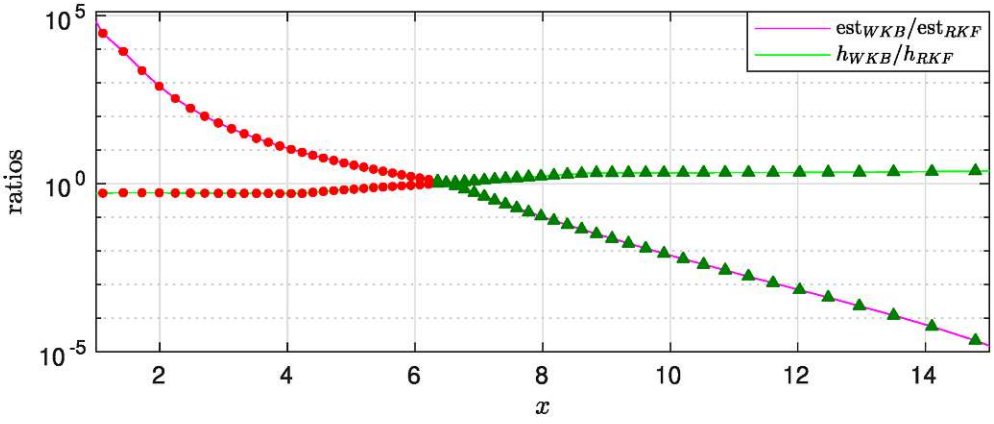

Figure 2.5.8: Numerical computation using the RKWKBmod method: The magenta line indicates the ratio of the two error estimators (due to the WKB and RKF scheme). The green line gives the analogous ratio of the step sizes (locally) proposed by these two methods. For this computation we set $\mathrm{Tol} = 10^{-5}$, $\varepsilon = 1$ and a third order WKB-ansatz was used.

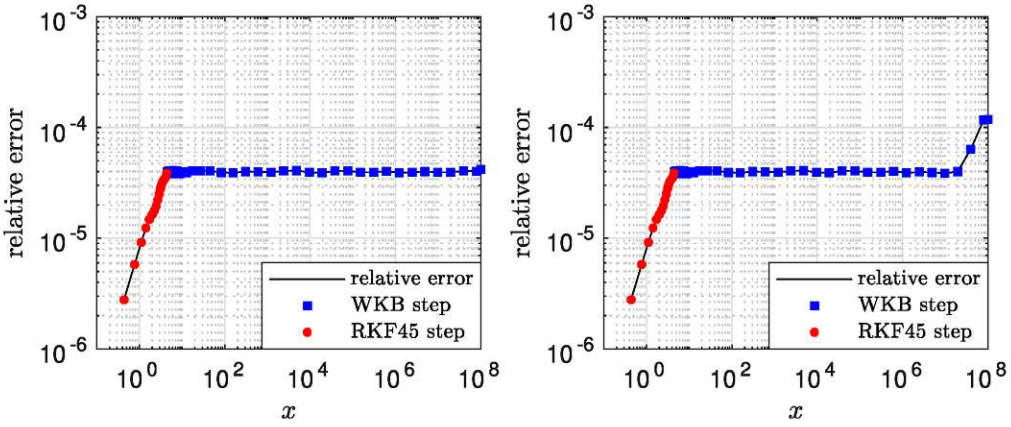

Figure 2.5.9: Left: Global (relative) error using WKB+RKF45 with exact phase. Right: Global (relative) error using WKB+RKF45 with numerically computed phase using the Clenshaw-Curtis quadrature with 15 integration nodes per step. For both computations we set $h_{1,trial} = 0.5$, Tol $= 10^{-5}$ and $\varepsilon = 1$.

of global errors, particularly for small values of ε and Tol. By Remark 2.5.1 the accuracy limit at $x = 50$ is $7.8 \cdot 10^{-10}$ for $\varepsilon = 10^{-4}$, which seems to explain the lower bound of the errors in Figure 2.5.10. Note also that for RKF45 the error increases like $\mathcal{O}(\varepsilon^{-1})$ for smaller ε-values, while it decreases for the WKB methods.

In Figure 2.5.11 we give a work-precision diagram; for a fair comparison between the methods, points showing the same error should be compared. There is a big difference between WKB+RKF45 and RKWKB(mod), regarding the CPU time: For $\varepsilon = 10^0, 10^{-1}, 10^{-2}$ this difference is particularly significant for small errors (stemming from small prescribed tolerances). For $\varepsilon = 10^{-3}$ WKB+RKF45 already beats RKWKB(mod) for all data points. For $\varepsilon = 10^{-4}$ the error intervals of WKB+RKF45 and RKWKB(mod) do not overlap. But for the same CPU time WKB+RKF45 yields much more accurate results than RKWKB or RKWKBmod. Overall we conclude from Figure 2.5.11 that WKB+RKF45 outperforms RKWKBmod and RKWKB significantly. Using RKF45, the CPU time increases drastically for smaller ε-values.

Figure 2.5.12 displays the respective number of steps needed in each method, as a function of the prescribed tolerance. Note that for WKB+RKF45 and RKWKBmod the number of steps is bounded from below. This is because of limiting the quotient of two consecutive step sizes and it is clearly visible for $\varepsilon = 10^{-3}, 10^{-4}$.

2.5.2 Second example: Parabolic cylinder function

The second example is taken from [AD20] and includes a quadratic coefficient function $a(x)$:

$$\begin{cases} \varepsilon^2 \varphi''(x) + \left(-\frac{1}{2}x^2 + x\right) \varphi(x) = 0\,, \quad x \in (0,2)\,, \\ \varphi(0) = \kappa U(\nu, z(0))\,, \\ \varphi'(0) = -\kappa 2^{-\frac{1}{4}} \varepsilon^{-\frac{1}{2}} U'(\nu, z(0))\,, \end{cases} \tag{2.5.2}$$

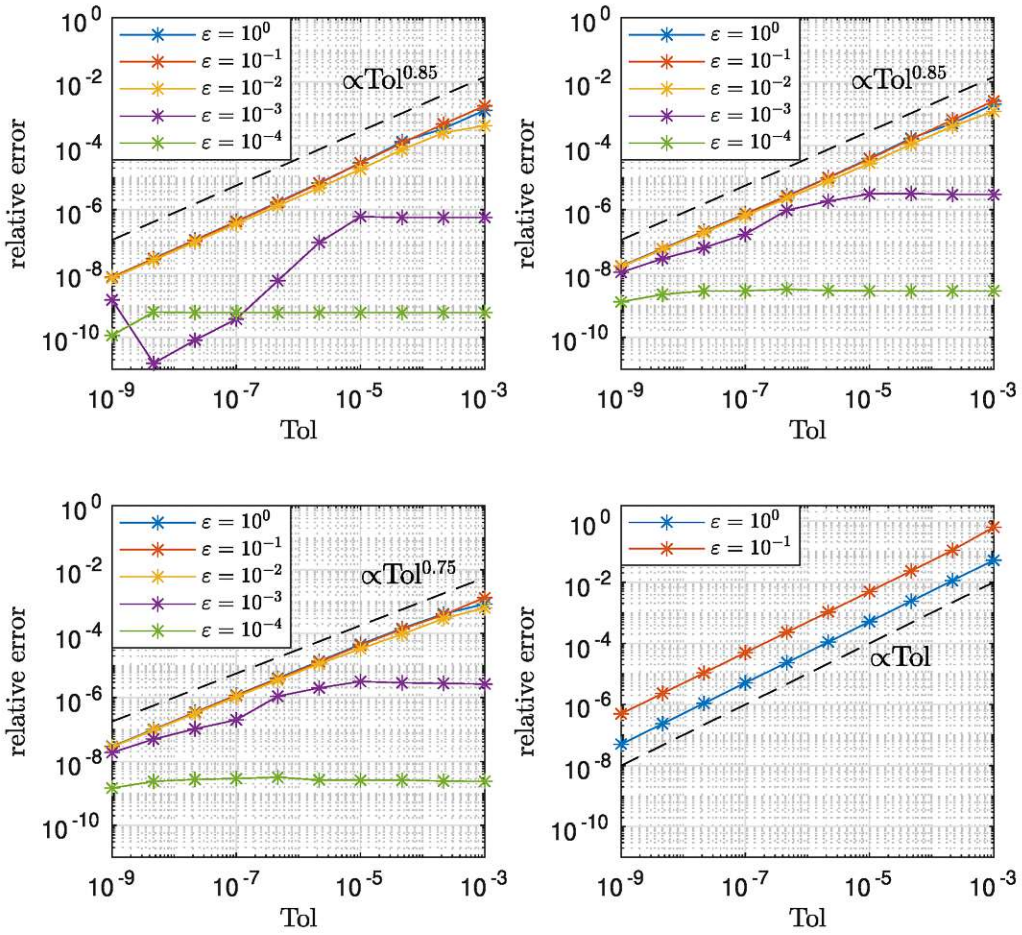

Figure 2.5.10: Global (relative) errors (in the l^2-norm) for the Airy equation (2.5.1) on $[0.1, 50]$ as a function of Tol, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45. For all methods we set $h_{1,trial} = 0.5$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

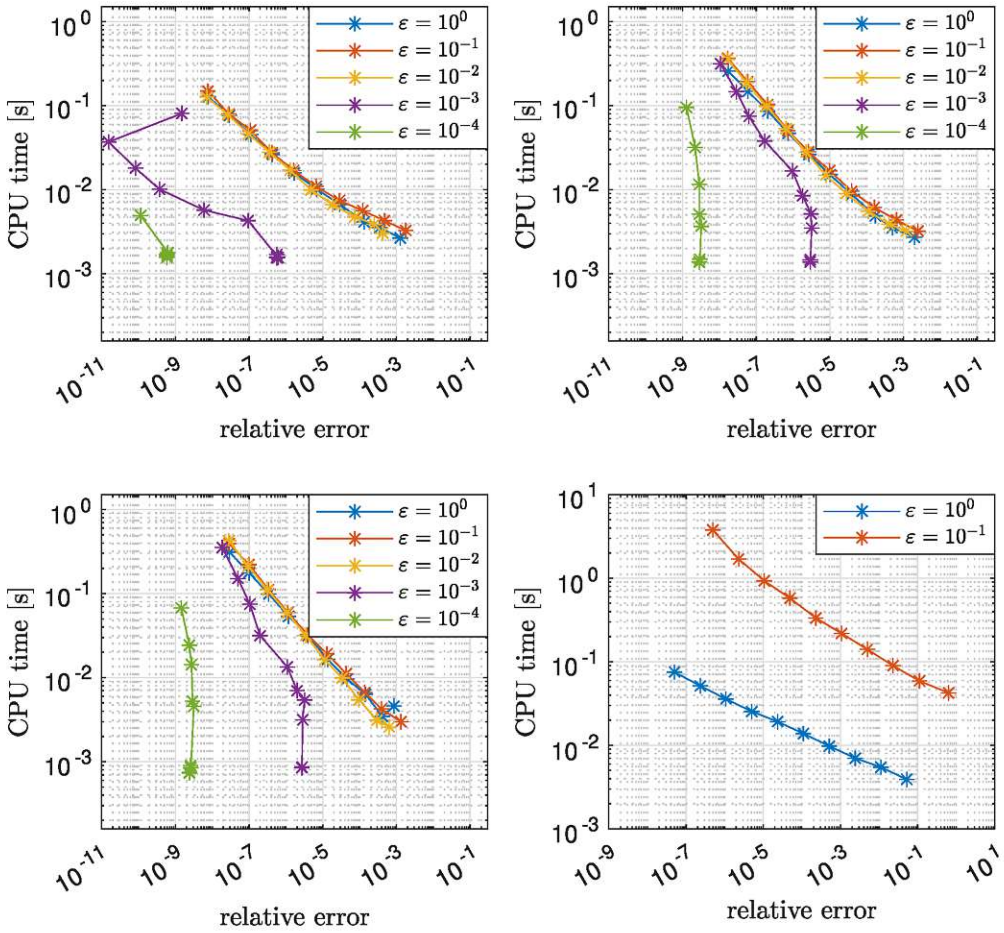

Figure 2.5.11: CPU times vs. global (relative) errors (in the l^2-norm) for the Airy equation (2.5.1) on $[0.1, 50]$, computed for 10 logarithmically evenly spaced values of Tol in the range $[10^{-9}, 10^{-3}]$, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45 (note the different scales). For all methods we set $h_{1,trial} = 0.5$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

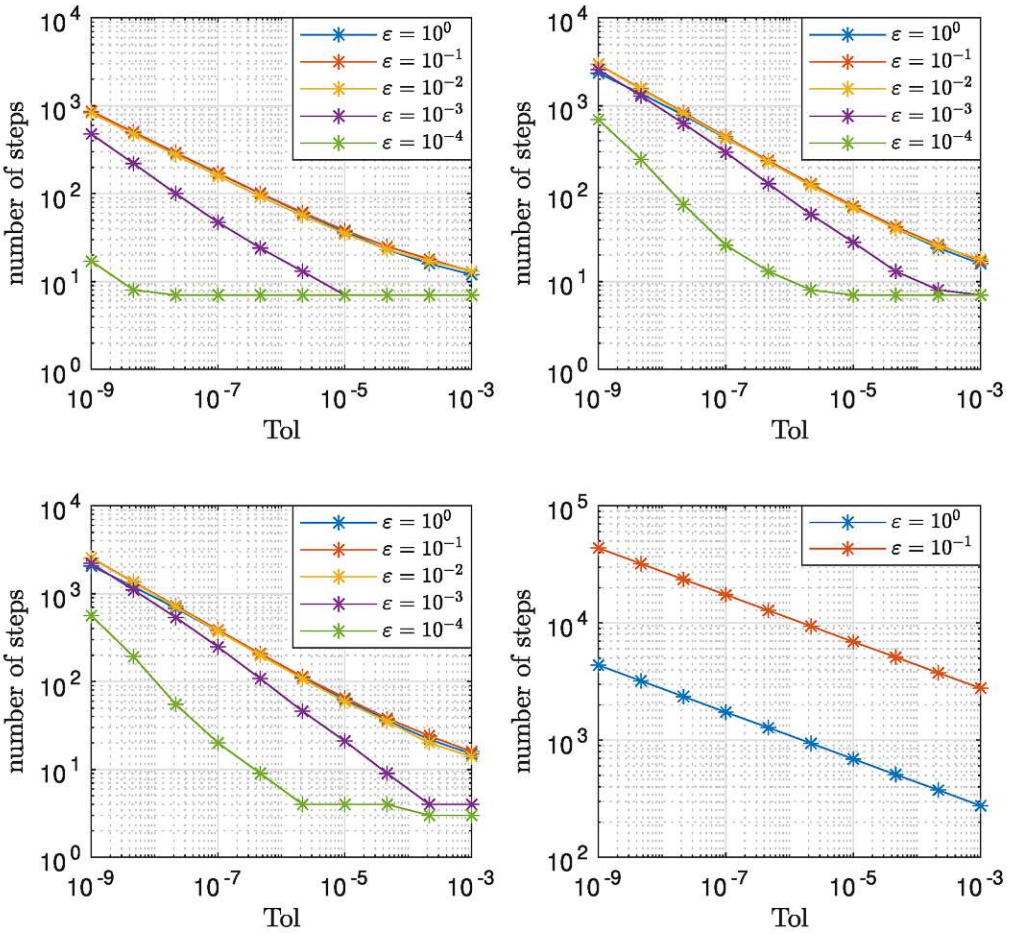

Figure 2.5.12: Number of steps needed for the Airy equation (2.5.1) on $[0.1, 50]$ as a function of Tol, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45. For all methods we set $h_{1,trial} = 0.5$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

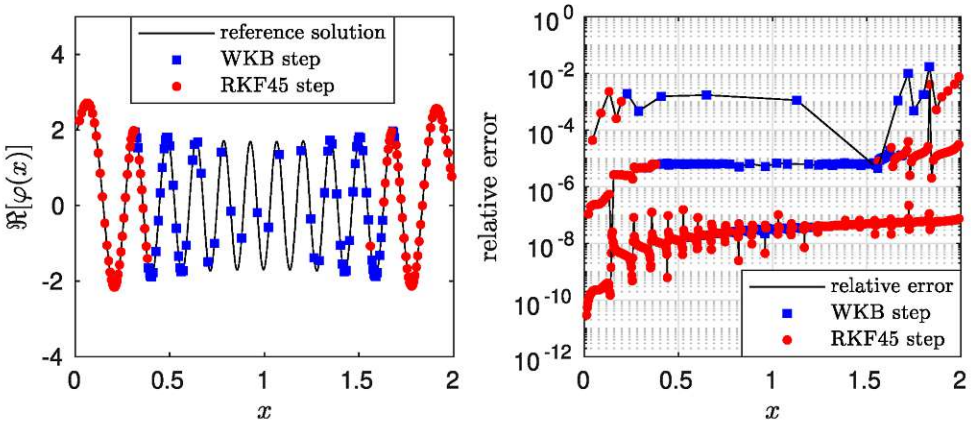

Figure 2.5.13: Left: Real part of the numerical solution obtained by using WKB+RKF45 compared to the (exact) reference solution (solid line) for Tol $= 10^{-6}$. Right: The global error for the choices Tol $= 10^{-3}, 10^{-6}, 10^{-9}$ (read from top to bottom). The respective overall number of steps made are 21, 166, and 1287. For both pictures the initial step size was set to $h_{1,trial} = 0.05$ and the parameter ε was set to 2^{-6}.

with

$$\nu := -\frac{1}{\sqrt{8\varepsilon}}\,, \quad z(x) := \frac{2^{\frac{1}{4}}}{\sqrt{\varepsilon}}(1-x)\,,$$

and

$$\kappa := \frac{2}{U(\nu,0) - \mathrm{i}\sqrt{\varepsilon}2^{\frac{3}{4}}U'(\nu,0)}\,.$$

The exact solution reads

$$\varphi_{exact}(x) = \kappa U(\nu, z(x))\,,$$

where $U(\nu, z)$ denotes the parabolic cylinder function (PCF) (cf. [OLBC10, §12]). As before, let us compare numerical results only for WKB+RKF45 and RKWKBmod at first. There are two turning points, namely at $x = 0$ and $x = 2$. Therefore, we expect the two methods to make RKF45 steps near the turning points and WKB steps between them. Numerical results for the specific choice $\varepsilon = 2^{-6}$ are presented in Figures 2.5.13-2.5.14.

According to Figures 2.5.13 and 2.5.14 no significant difference between WKB+RKF45 and RKWKBmod can be observed for $\varepsilon = 2^{-6}$, in matters of global error. But again, WKB+RKF45 needs significantly fewer steps than RKWKBmod. Within the algorithm using WKB+RKF45, the switch-over between RKF45 steps and WKB steps happens closer to the turning points.

We shall now present numerical results for the four methods WKB+RKF45, RKWKBmod, RKWKB, and RKF45 for several values of ε. In Figure 2.5.15, we compare the global errors of each method depending on Tol. Here, we observe only a small difference between WKB+RKF45 and RKWKBmod. Using RKWKB, less smooth error curves can bee seen, with large peaks for lower errors (i.e. lower tolerances). In contrast to every other method,

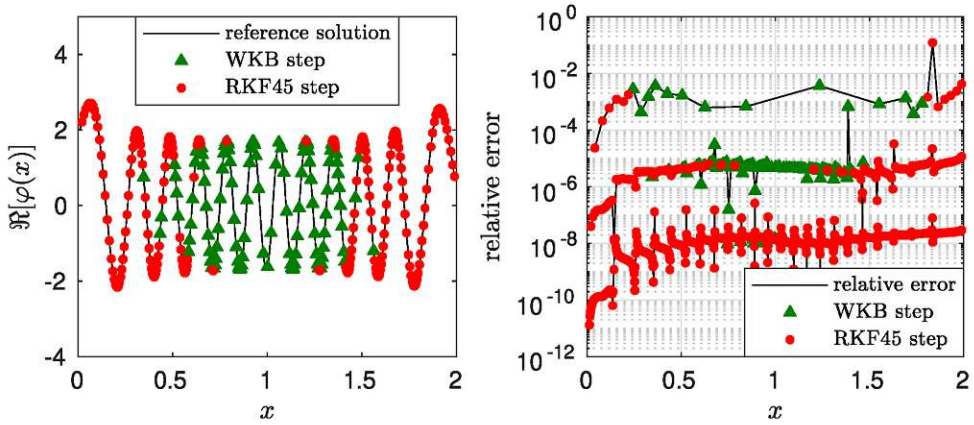

Figure 2.5.14: Left: Real part of the numerical solution obtained by using RKWKBmod compared to the (exact) reference solution (solid line) for Tol $= 10^{-6}$. Right: The global error for the choices Tol $= 10^{-3}, 10^{-6}, 10^{-9}$ (read from top to bottom). The respective overall number of steps made are 26, 326, and 1543. For both pictures the initial step size was set to $h_{1,trial} = 0.05$, the parameter ε was set to 2^{-6} and a third order WKB-ansatz was used.

WKB+RKF45 seems to produce quite ε-independent global error curves. For RKF45 one sees that the error again increases like $\mathcal{O}(\varepsilon^{-1})$ for smaller values of ε.

In the work-precision diagrams in Figure 2.5.16, we observe for WKB+RKF45 that the CPU times are quite independent of ε, whereas for RKWKB(mod) they grow with decreasing ε, particularly for small errors (stemming from small prescribed tolerances). Overall we conclude from Figure 2.5.16 that WKB+RKF45 outperforms RKWKBmod and RKWKB (particularly for small ε and small tolerances) while showing an ε-uniform behavior at the same time. Note also that, for RKF45, the CPU time increases drastically for smaller values of ε.

Figure 2.5.17 shows the respective number of steps needed in each method, as a function of the prescribed tolerance.

Remark 2.5.3. *The computation of the reference solution involves the evaluation of the PCF, which is not readily available in* MATLAB. *But the PCF can be related to the Kummer confluent hypergeometric function* ${}_1F_1$ *(see [OLBC10, §13]), which is available in* MATLAB *as* `hypergeom()`. *However, for small parameters* ε *this evaluation is very time consuming. For Figures 2.5.15-2.5.17 we thus computed the reference solutions by solving the initial value problem (2.5.2) with* MATLAB*'s routine* `ode45()` *(for all* ε*) using a very small error tolerance.*

2.6 Conclusion

We have introduced in this chapter an extension to the WKB-marching method from [AAN11] by including into the algorithm an adaptive step size controller as well as a switching mechanism. In numerical simulations based on two examples this method yielded smaller global errors (particularly for small tolerances and small ε-values) in comparison to the Runge-Kutta-WKB method from [HLH16, AHLH20], an alternative WKB-based

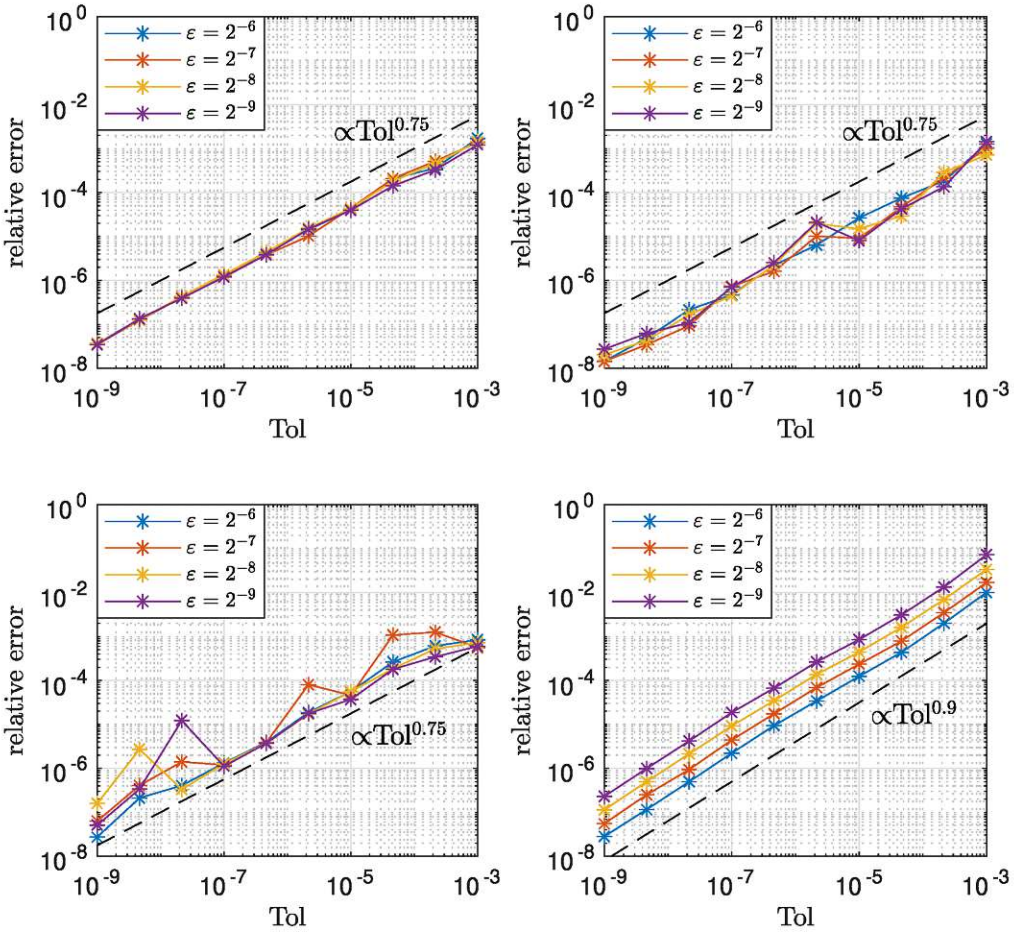

Figure 2.5.15: Global (relative) errors (in the l^2-norm) for equation (2.5.2) on $[0.01, 1.99]$ as a function of Tol, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45. For all methods we set $h_{1,trial} = 0.05$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

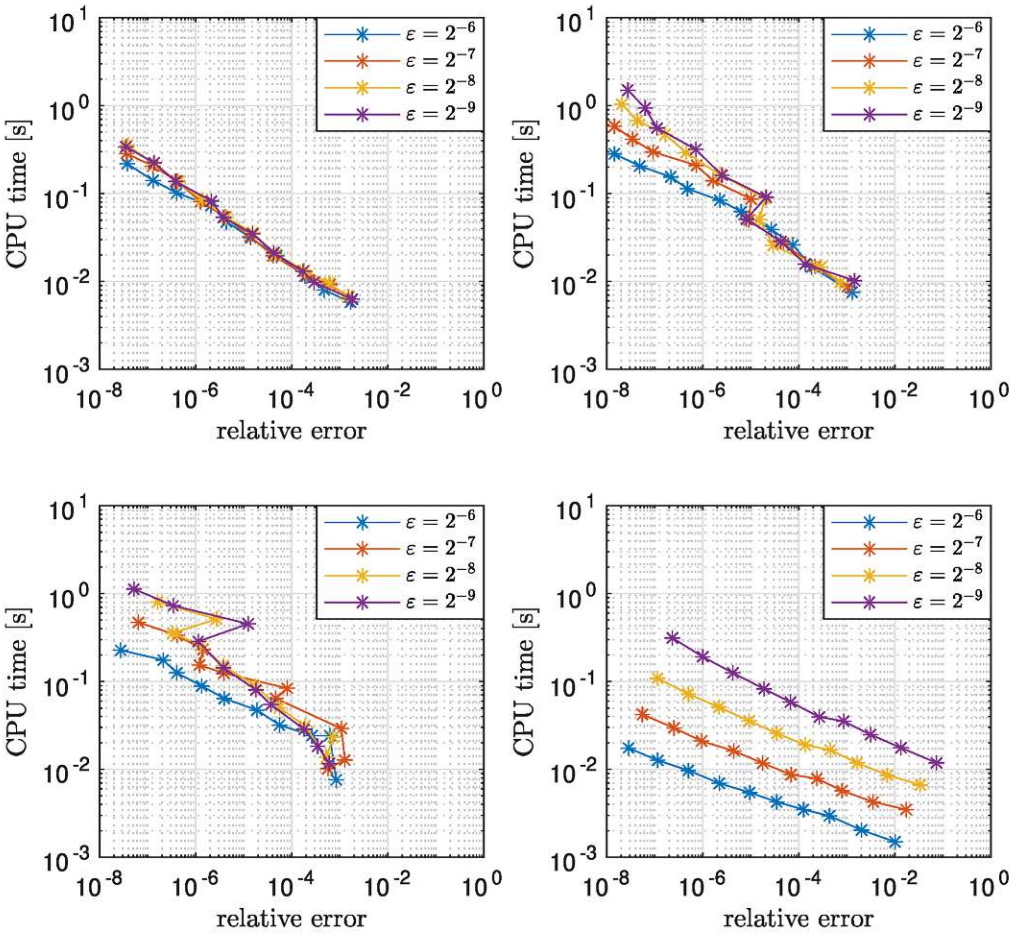

Figure 2.5.16: CPU times vs. global (relative) errors (in the l^2-norm) for equation (2.5.2) on $[0.01, 1.99]$, computed for 10 logarithmically evenly spaced values of Tol in the range $[10^{-9}, 10^{-3}]$, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45. For all methods we set $h_{1,trial} = 0.05$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

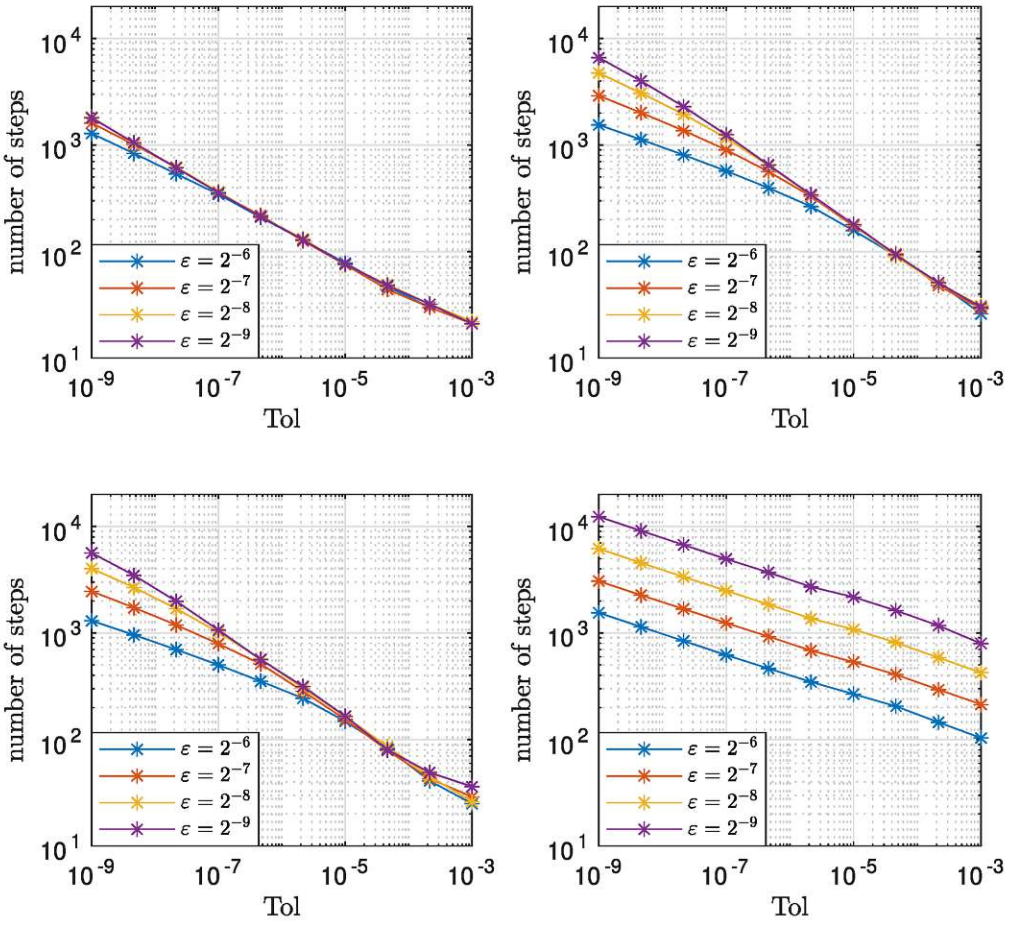

Figure 2.5.17: Number of steps needed for equation (2.5.2) on $[0.01, 1.99]$ as a function of Tol, for several ε-values. Top-left: WKB+RKF45. Top-right: RKWKBmod. Bottom-left: RKWKB. Bottom-right: RKF45. For all methods we set $h_{1,trial} = 0.05$. A third order WKB-ansatz for RKWKBmod and RKWKB was used.

scheme. Our tests revealed that the efficiency gain is mostly due to the different WKB method used here, while the different step size controls (here vs. [HLH16, AHLH20]) do not play a big role. Our switching mechanism ensures well defined switching points between WKB steps and Runge-Kutta-Fehlberg 4(5) steps for oscillatory and, respectively, smoother regions of the ODE-solution. Especially for the Airy equation on the large spatial interval $[0.1, 10^8]$ the efficiency of the method is demonstrated very well, as the scheme skips millions of oscillations within one step, while staying accurate at the same time. There is also a MATLAB program available in a GitHub repository[1], which also offers the possibility to compute the phase (2.2.7) numerically, as done for Figure 2.5.9.

[1] https://github.com/JannisKoerner/adaptive-WKB-marching-method

Appendix

2.A Asymptotic formulas for Airy functions

For real-valued x and $x \to \infty$, asymptotic expansions for the Airy functions and their first derivatives are given in [OLBC10, §9.7 (ii)]:

$$\mathrm{Ai}(-x) \sim \frac{1}{\sqrt{\pi} x^{\frac{1}{4}}} \left(\cos\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{u_{2k}}{\zeta^{2k}} + \sin\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{u_{2k+1}}{\zeta^{2k+1}} \right) , \tag{2.A.1}$$

$$\mathrm{Ai}'(-x) \sim \frac{x^{\frac{1}{4}}}{\sqrt{\pi}} \left(\sin\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{v_{2k}}{\zeta^{2k}} - \cos\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{v_{2k+1}}{\zeta^{2k+1}} \right) , \tag{2.A.2}$$

$$\mathrm{Bi}(-x) \sim \frac{1}{\sqrt{\pi} x^{\frac{1}{4}}} \left(-\sin\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{u_{2k}}{\zeta^{2k}} + \cos\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{u_{2k+1}}{\zeta^{2k+1}} \right) , \tag{2.A.3}$$

$$\mathrm{Bi}'(-x) \sim \frac{x^{\frac{1}{4}}}{\sqrt{\pi}} \left(\cos\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{v_{2k}}{\zeta^{2k}} + \sin\left(\zeta - \frac{\pi}{4}\right) \sum_{k=0}^{\infty} (-1)^k \frac{v_{2k+1}}{\zeta^{2k+1}} \right) . \tag{2.A.4}$$

Here, $\zeta := \frac{2}{3} x^{\frac{3}{2}}$ and the coefficients u_k and v_k are given by (see [OLBC10, §9.7 (i)]):

$$\begin{aligned} u_0 &= v_0 = 1 \,, \\ u_k &= \frac{(2k+1) \cdot (2k+3) \cdot (2k+5) \cdot \ldots \cdot (6k-1)}{216^k k!} \\ &= \frac{(6k-5) \cdot (6k-3) \cdot (6k-1)}{(2k-1) 216k} u_{k-1} \,, \quad k = 1, 2, \ldots \,, \\ v_k &= \frac{6k+1}{1-6k} u_k \,, \quad k = 1, 2, \ldots \,. \end{aligned}$$

Bibliography

[AAN11] A. Arnold, N. B. Abdallah, and C. Negulescu. WKB-Based Schemes for the Oscillatory 1D Schrödinger Equation in the Semiclassical Limit. *SIAM J. Numer. Anal.*, 49:1436–1460, 2011.

[ABI20] P. Amodio, L. Brugnano, and F. Iavernaro. Analysis of spectral Hamiltonian boundary value methods (SHBVMs) for the numerical solution of ODE problems. *Numerical Algorithms*, 83:1489–1508, 2020.

[AD20] A. Arnold and K. Döpfner. Stationary Schrödinger equation in the semi-classical limit: WKB-based scheme coupled to a turning point. *Calcolo*, 57:44, 2020.

[AHLH20] F. J. Agocs, W. J. Handley, A. N. Lasenby, and M. P. Hobson. Efficient method for solving highly oscillatory ordinary differential equations with applications to physical systems. *Phys. Rev. Res.*, 2:013030, Jan 2020.

[AKU22] A. Arnold, C. Klein, and B. Ujvari. WKB-method for the 1D Schrödinger equation in the semi-classical limit: enhanced phase treatment. *BIT Numerical Mathematics*, 62:1–22, 2022.

[BIMR19] L. Brugnano, F. Iavernaro, J. I. Montijano, and L. Rández. Spectrally accurate space-time solution of Hamiltonian PDEs. *Numerical Algorithms*, 81:1183–1202, 2019.

[But08] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley-Interscience, 2nd edition, 2008.

[CC60] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205, 1960.

[HLH16] W. J. Handley, A. N. Lasenby, and M. P. Hobson. The Runge-Kutta-Wentzel-Kramers-Brillouin method. *arXiv*, abs/1612.02288, 2016.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer-Verlag, 2nd edition, 2006.

[HNW00] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I.* Springer, 2000.

[INO06] A. Iserles, S. P. Nørsett, and S. Olver. Highly oscillatory quadrature: the story so far. *Proceedings of ENUMATH 2005, Santiago de Compostela*, pages 97–118, 2006.

[JL03] T. Jahnke and C. Lubich. Numerical integrators for quantum dynamics close to the adiabatic limit. *Numerische Mathematik*, 94:289–314, 2003.

[KAD21] J. Körner, A. Arnold, and K. Döpfner. WKB-based scheme with adaptive step size control for the Schrödinger equation in the highly oscillatory regime. *J. Comput. Appl. Math.*, 404:113905, 2021.

[LJL05] K. Lorenz, T. Jahnke, and C. Lubich. Adiabatic Integrators for Highly Oscillatory Second-Order Linear Differential Equations with Time-Varying Eigendecomposition. *BIT Numerical Mathematics*, 45:91–115, 2005.

[LL85] L. D. Landau and E. M. Lifschitz. *Quantenmechanik.* Akademie-Verlag, 1985.

[OLBC10] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. NIST Handbook of Mathematical Functions. Cambridge University Press, 2010.

[Sö02] G. Söderlind. Automatic Control and Adaptive Time-Stepping. *Numerical Algorithms*, 31:281–310, 2002.

[WW18] X. Wu and B. Wang. *Recent Developments in Structure-Preserving Algorithms for Oscillatory Differential Equations.* Springer, 2018.

3 WKB-based third order method for the highly oscillatory 1D stationary Schrödinger equation

The content of this chapter has been submitted for publication under [AK24].

3.1 Introduction

This chapter is concerned with the numerical treatment of the highly oscillatory 1D Schrödinger equation

$$\varepsilon^2 \varphi''(x) + a(x)\varphi(x) = 0\,, \quad x \in \mathbb{R}\,, \tag{3.1.1}$$

where the parameter $0 < \varepsilon \ll 1$ is assumed to be very small. Further, we assume the real-valued coefficient function a to be sufficiently smooth and bounded away from zero, i.e. $a(x) \geq a_0 > 0$. The Schrödinger equation (3.1.1) plays an important role within the context of quantum mechanical problems, e.g., for the simulation of electron transport in semiconductor devices [MJK13, Neg05, SHMS98]. In these applications, φ represents the (possibly complex-valued) Schrödinger wave function, and $a(x) := E - V(x)$ is associated with a prescribed electrostatic potential V, where $E \in \mathbb{R}$ denotes the injection energy of an electron of mass m. The small parameter $\varepsilon := \frac{\hbar}{\sqrt{2m}}$ is then proportional to the (reduced) Planck constant $\hbar$. We note that there are numerous additional applications of equation (3.1.1), including plasma physics [CS58, Lew68] and cosmology [MS03, Win05].

Since the (local) wave length of a solution φ to (3.1.1) is given by $\lambda(x) = (2\pi\varepsilon)/\sqrt{a(x)}$, the solution exhibits high oscillations for small values of ε, particularly in the semi-classical limit $\varepsilon \to 0$. Therefore, standard methods for solving (3.1.1) are typically constrained by choosing very small grid sizes $h = \mathcal{O}(\varepsilon)$ (e.g., see [IB95]) in order to obtain reasonably accurate numerical solutions. However, this leads to high inefficiencies for small values of ε. Hence, there is a keen interest in numerical methods that allow this grid size limitation to be reduced or even eliminated entirely.

In [JL03, LJL05] the authors proposed efficient and *uniformly accurate* (w.r.t. ε) schemes, which allow to reduce the grid size limitation to $h = \mathcal{O}(\sqrt{\varepsilon})$ while yielding global errors of order $\mathcal{O}(h^2)$. The same grid limitation is achieved with the WKB-based (named after the physicists Wentzel, Kramers, Brillouin; cf. [LL85]) second order (w.r.t. the step size h) one-step scheme from [AAN11]. This method relies on an analytical pre-processing of the ODE (3.1.1), utilizing a priori information of the solution by considering a second order (w.r.t. ε) WKB approximation. As a consequence, the dominant oscillations are eliminated, allowing the computation of a numerical solution on a coarse grid. In fact, the

method from [AAN11] is sometimes even *asymptotically correct*, meaning the numerical error goes to zero as ε approaches zero. This holds true under the condition that the integrals $\int^x \sqrt{a(\tau)}\,\mathrm{d}\tau$ and $\int^x a(\tau)^{-1/4}(a(\tau)^{-1/4})''\,\mathrm{d}\tau$ for the phase of the solution can be computed exactly. The method then yields numerical errors which are $\mathcal{O}(\varepsilon^3)$ as $\varepsilon \to 0$.

In [KAD21], an adaptive step size control and a switching mechanism were added on top of the second order scheme (w.r.t. the step size) from [AAN11]. The switching mechanism allowed the algorithm to switch to a standard ODE method (e.g., Runge-Kutta) during the computation and was implemented to avoid technical or efficiency issues in regions where the coefficient function $a(x)$ is very small or even equal to zero.

The aim of this chapter is to improve the method from [AAN11] by extending it to a third order scheme (w.r.t. the step size). This extension maintains the same analytical pre-processing of the given ODE (3.1.1) as in [AAN11]. The derivation of the third order scheme can then be realized mainly by two steps: Firstly, by keeping an additional term of the Picard approximation of the solution; and secondly, by a special treatment of the (iterated) oscillatory integrals which occur in each term of this Picard approximation. While the first step is straightforward, the second step is highly technical and consists of rather extensive computations in order to obtain sufficiently accurate quadratures for the oscillatory integrals. Here, the main strategies employed to achieve the desired accuracy are the same as in [AAN11].

This chapter is organized as follows: In Section 3.2 we provide a short review of the WKB-based transformation from [AAN11], which allows transforming (3.1.1) into a less oscillatory problem. Section 3.3 details the construction of sufficiently accurate approximations for the (iterated) oscillatory integrals that appear in the Picard approximation of the solution. Section 3.4 contains the definition as well as the error analysis of the numerical scheme, with the main result of this chapter stated in Theorem 3.4.1. In Section 3.5 we present numerical simulations and illustrate the theoretical results established in this chapter. We conclude in Section 3.6.

3.2 Review of the WKB-based transformation from [AAN11]

In this section we shall review the basics of the second order (w.r.t. ε) WKB-based transformation from [AAN11], which was used there to transform the Schrödinger equation (3.1.1) into a smoother (i.e. less oscillatory) equation. Based on this analytical pre-processing of (3.1.1), the authors developed a second order (w.r.t. the step size h) numerical scheme that is uniformly correct in ε and sometimes even asymptotically correct, i.e. the numerical error goes to zero with $\varepsilon \to 0$, while the grid size h remains constant.

Our aim here is to solve the following initial value problem (IVP) for the Schrödinger equation (3.1.1) on the spatial interval $I := [x_0, x_{end}]$:

$$\begin{cases} \varepsilon^2\varphi''(x) + a(x)\varphi(x) = 0\,, \quad x \in I = [x_0, x_{end}]\,, \\ \varphi(x_0) = \varphi_0\,, \\ \varepsilon\varphi'(x_0) = \varphi_1\,. \end{cases} \tag{3.2.1}$$

Before we explain the reformulation of this IVP, we introduce the following assumption.

Hypothesis A. *Let $a \in C^7(I)$ be a fixed (real-valued) function, which satisfies $a(x) \geq a_0 > 0$ for all $x \in I$. Further, let $0 < \varepsilon \leq \varepsilon_0$ for some sufficiently small ε_0.*

The basis of the transformation from [AAN11] relies on the well-known WKB-ansatz (cf. [LL85]) for the ODE (3.1.1):

$$\varphi(x) \sim \exp\left(\frac{1}{\varepsilon}\sum_{p=0}^{\infty}\varepsilon^p \phi_p(x)\right), \quad \varepsilon \to 0\,; \quad \phi_p(x) \in \mathbb{C}\,. \tag{3.2.2}$$

Here, the series in the exponential function is asymptotic w.r.t. ε and typically divergent[1]. To derive approximate solutions, it is therefore necessary to truncate the series after some finite order. By substituting the WKB-ansatz into equation (3.1.1), a comparison of ε-powers leads to the first three functions $\phi_p(x)$ as

$$\phi_0(x) = \pm \mathrm{i}\int_{x_0}^{x} \sqrt{a(y)}\,\mathrm{d}y\,, \tag{3.2.3}$$

$$\phi_1(x) = \ln(a(x)^{-\frac{1}{4}}) - \ln(a(x_0)^{-\frac{1}{4}})\,, \tag{3.2.4}$$

$$\phi_2(x) = \mp \mathrm{i}\int_{x_0}^{x} b(y)\,\mathrm{d}y\,, \quad b(x) := -\frac{1}{2a(x)^{\frac{1}{4}}}\left(a(x)^{-\frac{1}{4}}\right)''\,. \tag{3.2.5}$$

Here, the two different signs $\pm$ and $\mp$ in (3.2.3) and (3.2.5) imply that there is always a pair of approximate solutions to the Schrödinger equation (3.1.1). This corresponds to the fact that the second order ODE (3.1.1) has two fundamental solutions. By truncating the series in (3.2.2) after $p = 2$, one obtains the second order (w.r.t. ε) asymptotic WKB approximation of $\varphi(x)$ as

$$\varphi_2^{WKB}(x) = \alpha\frac{\exp\left(\frac{\mathrm{i}}{\varepsilon}\phi^{\varepsilon}(x)\right)}{a(x)^{\frac{1}{4}}} + \beta\frac{\exp\left(-\frac{\mathrm{i}}{\varepsilon}\phi^{\varepsilon}(x)\right)}{a(x)^{\frac{1}{4}}}\,, \tag{3.2.6}$$

with constants $\alpha, \beta \in \mathbb{C}$. The phase is

$$\phi^{\varepsilon}(x) := \int_{x_0}^{x}\left(\sqrt{a(y)} - \varepsilon^2 b(y)\right)\mathrm{d}y\,. \tag{3.2.7}$$

Note that Hypothesis A implies that ϕ^{ε} is bounded on I. The WKB-based transformation now consists of two steps: First, using the notation

$$U(x) = \begin{pmatrix} u_1(x) \\ u_2(x) \end{pmatrix} := \begin{pmatrix} a(x)^{\frac{1}{4}}\varphi(x) \\ \frac{\varepsilon\left(a(x)^{\frac{1}{4}}\varphi(x)\right)'}{\sqrt{a(x)}} \end{pmatrix}\,, \tag{3.2.8}$$

the second order IVP (3.2.1) can be reformulated as a system of first order differential equations:

$$\begin{cases} U'(x) = \left[\frac{1}{\varepsilon}\mathbf{A}_0(x) + \varepsilon\mathbf{A}_1(x)\right]U(x)\,, \quad x \in I = [x_0, x_{end}]\,, \\ U(x_0) = U_0\,, \end{cases} \tag{3.2.9}$$

[1] See [AKKM24] for a class of coefficient functions a, which lead to convergent WKB series.

where the two matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ are given by

$$\mathbf{A}_0(x) := \sqrt{a(x)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} ; \quad \mathbf{A}_1(x) := \begin{pmatrix} 0 & 0 \\ 2b(x) & 0 \end{pmatrix} .$$

Second, the first order system (3.2.9) is transformed by an additional change of variables

$$Z(x) = \begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix} := \exp\left(-\frac{\mathrm{i}}{\varepsilon} \mathbf{\Phi}^\varepsilon(x) \right) \mathbf{P} U(x) , \tag{3.2.10}$$

with the two matrices

$$\mathbf{P} := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i} & 1 \\ 1 & \mathrm{i} \end{pmatrix} ; \quad \mathbf{\Phi}^\varepsilon(x) := \begin{pmatrix} \phi^\varepsilon(x) & 0 \\ 0 & -\phi^\varepsilon(x) \end{pmatrix} .$$

Here, the function ϕ^ε is the phase of the second order WKB approximation (3.2.6) as defined in (3.2.7). The resulting system for Z then reads

$$\begin{cases} Z'(x) = \varepsilon \mathbf{N}^\varepsilon(x) Z(x) , & x \in I = [x_0, x_{end}] , \\ Z(x_0) = Z_0 = \mathbf{P} U_0 , \end{cases} \tag{3.2.11}$$

where $\mathbf{N}^\varepsilon(x)$ is a (Hermitian) matrix with only off-diagonal non-zero entries:

$$N^\varepsilon_{1,2}(x) = b(x) \, \mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)} , \quad N^\varepsilon_{2,1}(x) = b(x) \, \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi^\varepsilon(x)} . \tag{3.2.12}$$

As stated in [AAN11], the IVP (3.2.11) admits a unique solution with the following estimates:

Proposition 3.2.1. *[AAN11, Theorem 3.1] Let Hypothesis A be satisfied. Then the problem (3.2.11) has a unique solution $Z \in C^6(I)$ with the explicit form*

$$Z(x) = Z_0 + \sum_{p=1}^{\infty} \varepsilon^p \mathbf{M}^\varepsilon_p(x; x_0) Z_0 , \tag{3.2.13}$$

where the matrices $\mathbf{M}^\varepsilon_p$, $p \geq 1$ are given by

$$\begin{aligned} \mathbf{M}^\varepsilon_p(\eta; \xi) &= \int_\xi^\eta \int_\xi^{y_1} \cdots \int_\xi^{y_{p-1}} \mathbf{N}^\varepsilon(y_1) \cdots \mathbf{N}^\varepsilon(y_p) \, \mathrm{d}y_p \cdots \mathrm{d}y_1 , \\ \mathbf{M}^\varepsilon_p(\eta; \xi) &= \int_\xi^\eta \mathbf{N}^\varepsilon(y) \mathbf{M}^\varepsilon_{p-1}(y; \xi) \, \mathrm{d}y , \quad \mathbf{M}^\varepsilon_0 = \mathbf{I} . \end{aligned} \tag{3.2.14}$$

Here, $\mathbf{I}$ denotes the 2×2 identity matrix. Moreover we have the estimates

$$\| Z - Z_0 \|_{L^\infty(I)} \leq C\varepsilon^2 , \quad \| Z' \|_{L^\infty(I)} \leq C\varepsilon , \quad \| Z'' \|_{L^\infty(I)} \leq C , \tag{3.2.15}$$

with a constant $C > 0$ independent of ε.

According to Proposition 3.2.1, the solution Z of IVP (3.2.11) oscillates around the initial value Z_0 with an amplitude of at most $\mathcal{O}(\varepsilon^2)$. In [AAN11], this fact was the motivation for the construction of a uniformly correct scheme in ε, and it shall also be the motivation for the scheme we develop in Section 3.3. The idea is that, instead of solving IVP (3.2.1), we aim to solve the transformed problem (3.2.11). Indeed, since the transformation (3.2.10) eliminated the dominant oscillations, the IVP (3.2.11) can be solved numerically on a coarse grid $\{x_n\}$. Then, the originally desired solution φ can be obtained by the inverse transformation

$$U(x) = \mathbf{P}^{-1} \exp\left(\frac{\mathrm{i}}{\varepsilon}\mathbf{\Phi}^{\varepsilon}(x)\right) Z(x) \tag{3.2.16}$$

and $\varphi(x) = a(x)^{-\frac{1}{4}} u_1(x)$.

3.3 Construction of the numerical method

The aim of this section is to construct a third order (w.r.t. the step size) one-step scheme for solving the Schrödinger equation-IVP (3.2.1). To this end, we build upon the WKB-based transformation from Section 3.2, in the same way as it was done in [AAN11]. That is, instead of solving (3.2.1) directly, we will solve the transformed problem (3.2.11).

We now consider a discretization $\{x_0, x_1, ..., x_N\}$ of the interval $I = [x_0, x_{end}]$ and set $h := \max_{1\leq n\leq N} |x_n - x_{n-1}|$ as the maximum step size. Further, throughout this whole section we will use the abbreviations $\xi := x_n$ and $\eta := x_{n+1}$.

The development of the one-step method relies on deriving specific quadrature rules for the matrix-valued integrals (3.2.14), with error estimates depending on the two small parameters ε and h. This leads us to introduce the following notation, which is analogous to the usual 'big-O' notation for a single parameter:

Definition 3.3.1. *Let V be a vector space (over $\mathbb{C}$) with norm $\|\cdot\|$ and let $0 < \varepsilon_0, h_0 < 1$. Consider two functions $f : (0,\varepsilon_0) \times (0,h_0) \to V$ and $g : (0,\varepsilon_0) \times (0,h_0) \to \mathbb{R}$. We say that $f(\varepsilon,h) = \mathcal{O}_{\varepsilon,h}(g(\varepsilon,h))$, if and only if there exists a constant $C > 0$, such that $\|f(\varepsilon,h)\| \leq Cg(\varepsilon,h)$, for any $(\varepsilon,h) \in (0,\varepsilon_0) \times (0,h_0)$.*

Now, in order to derive a numerical scheme of order P (w.r.t. the step size h) for solving the IVP (3.2.11), we shall start with the P-th order Picard approximation

$$Z(\eta) \approx Z(\xi) + \sum_{p=1}^{P} \varepsilon^p \mathbf{M}_p^{\varepsilon}(\eta;\xi) Z(\xi) \tag{3.3.1}$$

of (3.2.13), which corresponds to making a single step within the grid, namely, from ξ to η. According to [AAN11, Eq. (2.17)], the remainder of this truncated series is of order $\mathcal{O}_{\varepsilon,h}(\varepsilon^{P+1}h^P \min(\varepsilon,h))$, i.e., we have

$$\left\| \sum_{p=P+1}^{\infty} \varepsilon^p \mathbf{M}_p^{\varepsilon}(\eta;\xi) \right\|_{\infty} \leq C\varepsilon^{P+1}h^P \min(\varepsilon,h)\,, \tag{3.3.2}$$

where the constant $C \geq 0$ is independent of ε and h. Here, $\|\cdot\|_\infty$ denotes the ∞-matrix norm in $\mathbb{C}^{2\times 2}$. Since the entries (3.2.12) of the matrix $\mathbf{N}^\varepsilon(y)$ are highly oscillatory, the matrices $\mathbf{M}_p^\varepsilon(\eta;\xi)$ in (3.3.1) and (3.3.3) involve (iterated) oscillatory integrals and can therefore not be computed exactly in general. To design suitable quadrature rules for the oscillatory integrals $\mathbf{M}_p^\varepsilon(\eta;\xi)$ we shall use techniques such as the *asymptotic method* from [INO06], which mainly relies on integration by parts. If we want to keep the error order (3.3.2) also when approximating the oscillatory integrals, the desired error order (w.r.t. ε and h) for the approximation of the integral $\mathbf{M}_p^\varepsilon$ is $\mathcal{O}_{\varepsilon,h}(\varepsilon^{P+1-p}h^P \min(\varepsilon,h))$.

To derive a third order (w.r.t. h) scheme for (3.2.11), we only take into account the first three terms of the sum in (3.3.1). Indeed, using estimate (3.3.2) we then have

$$\begin{aligned} Z(\eta) = &\left[\mathbf{I} + \varepsilon\mathbf{M}_1^\varepsilon(\eta;\xi) + \varepsilon^2\mathbf{M}_2^\varepsilon(\eta;\xi) + \varepsilon^3\mathbf{M}_3^\varepsilon(\eta;\xi)\right] Z(\xi) \\ &+ \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h^3 \min(\varepsilon,h))\,. \end{aligned} \tag{3.3.3}$$

The goal of the following three subsections is to construct approximations for the matrices $\mathbf{M}_1^\varepsilon$, $\mathbf{M}_2^\varepsilon$, and $\mathbf{M}_3^\varepsilon$, respectively. Unfortunately, achieving the above mentioned desired approximation error order for each matrix is not always possible, as some terms involved in the computations constrain the maximum achievable order w.r.t. ε (this was also the case in the construction of the second order scheme in [AAN11]). However, we will be able to construct approximations for $\mathbf{M}_p^\varepsilon$, $p = 1,2,3$, which lead to a third order scheme with a local discretization error of the order $\mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^4 \max(\varepsilon,h))$.

3.3.1 Approximation of the matrix $\mathbf{M}_1^\varepsilon$

We start with the approximation of the integral

$$\mathbf{M}_1^\varepsilon(\eta;\xi) = \int_\xi^\eta \mathbf{N}^\varepsilon(y)\,\mathrm{d}y\,.$$

As we already mentioned above, the desired error order is $\mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^3 \min(\varepsilon,h))$ for this approximation.

For the phase ϕ^ε let us suppress the superscript ε and just write ϕ from now on. Further, we denote with

$$s_n := \phi(\eta) - \phi(\xi) = \mathcal{O}(h)\,, \quad h \to 0\,,$$

the phase increments and introduce the abbreviations

$$b_0(x) := \frac{b(x)}{2\phi'(x)}\,, \quad b_{p+1}(x) := \frac{b_p'(x)}{2\phi'(x)}\,, \quad p \in \mathbb{N}_0\,.$$

Note that Hypothesis A implies that for $0 < \varepsilon \leq \varepsilon_0$, where ε_0 is chosen sufficiently small, it holds $\phi'(x) \geq C_\phi > 0$ for some constant C_ϕ and for all $x \in I$, see (3.2.7). Hence both the function b and the functions b_p are uniformly bounded (w.r.t. ε) on the interval I. Moreover, we define the auxiliary function

$$h_p(x) := \mathrm{e}^{\mathrm{i}x} - \sum_{k=0}^{p-1} \frac{(\mathrm{i}\,x)^k}{k!}\,, \quad p \in \mathbb{N}_0$$

(for $p = 0$ the sum drops). In the following we will frequently use that

$$h_{p-1}(f(x)) = -\mathrm{i}(f'(x))^{-1}\frac{\mathrm{d}}{\mathrm{d}x}h_p(f(x)))\,, \quad p \in \mathbb{N}\,,$$

for any differentiable function f on I and any $x \in I$ with $f'(x) \neq 0$, and

$$h_p(x) = \mathcal{O}(\min(x^p, x^{p-1}))\,, \quad x \geq 0\,, \quad p \in \mathbb{N}\,. \tag{3.3.4}$$

The following lemma provides an approximation to $\mathbf{M}_1^\varepsilon$, with an error of the order $\mathcal{O}_{\varepsilon,h}(\varepsilon^P h^{\widetilde{P}}\min(\varepsilon, h))$, for any $P \in \mathbb{N}_0$, $\widetilde{P} \in \mathbb{N}$.

Lemma 3.3.2. *Let Hypothesis A be satisfied. For any $P \in \mathbb{N}_0$ and $\widetilde{P} \in \mathbb{N}$ define*

$$\begin{aligned} Q_1^{P,\widetilde{P}}(\eta,\xi) := &-\sum_{p=1}^{P}(\mathrm{i}\,\varepsilon)^p\left(b_{p-1}(\eta)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\eta)} - b_{p-1}(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\right) \\ &- \mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\sum_{p=1}^{\widetilde{P}}(\mathrm{i}\,\varepsilon)^{p+P}b_{p+P-1}(\eta)h_p\left(\frac{2}{\varepsilon}s_n\right) \end{aligned} \tag{3.3.5}$$

(for $P = 0$ the first sum drops) and

$$\mathbf{Q}_1^{P,\widetilde{P}}(\eta,\xi) := \begin{pmatrix} 0 & \overline{Q_1^{P,\widetilde{P}}(\eta,\xi)} \\ Q_1^{P,\widetilde{P}}(\eta,\xi) & 0 \end{pmatrix}. \tag{3.3.6}$$

Then there exists $C \geq 0$ independent of $\varepsilon \in (0, \varepsilon_0]$, h, and n, such that

$$\|\mathbf{M}_1^\varepsilon(\eta,\xi) - \mathbf{Q}_1^{P,\widetilde{P}}(\eta,\xi)\|_\infty \leq C\varepsilon^P h^{\widetilde{P}}\min(\varepsilon, h)\,. \tag{3.3.7}$$

Proof. Recall from (3.2.12) that $\mathbf{N}^\varepsilon(y)$ is a Hermitian matrix with only off-diagonal non-zero entries, namely $N_{2,1}^\varepsilon(y) = b(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)} = \overline{N_{1,2}^\varepsilon(y)}$. Hence, it is sufficient to prove that $\|m_1^\varepsilon(\eta,\xi) - Q_1^{P,\widetilde{P}}(\eta,\xi)\|_\infty \leq C\varepsilon^P h^{\widetilde{P}}\min(\varepsilon, h)$, where $m_1^\varepsilon(\eta,\xi) := (\mathbf{M}_1^\varepsilon(\eta;\xi))_{2,1}$. To this end, we start by expanding the integral $m_1^\varepsilon(\eta,\xi)$ by making P steps of the so-called *asymptotic method (AM)* for oscillatory integrals (cf. [INO06]), which relies on repeated integration by parts:

$$\begin{aligned} m_1^\varepsilon(\eta,\xi) &= \int_\xi^\eta b(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y \\ &= -(\mathrm{i}\,\varepsilon)\int_\xi^\eta b_0(y)\frac{\mathrm{d}}{\mathrm{d}y}\left[\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\right]\mathrm{d}y \\ &= -(\mathrm{i}\,\varepsilon)\left[b_0(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\right]_\xi^\eta + (\mathrm{i}\,\varepsilon)\int_\xi^\eta b_0'(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y \\ &= -\sum_{p=1}^{P}(\mathrm{i}\,\varepsilon)^p\left[b_{p-1}(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\right]_\xi^\eta + T_P^\varepsilon(\eta,\xi)\,. \end{aligned} \tag{3.3.8}$$

Here, $T_P^\varepsilon(\eta,\xi)$ is the remaining integral after the last integration by parts. It is evident from (3.3.8) that each step of integration by parts increases the ε-order of the remaining integral by one, since the functions b_p are uniformly bounded w.r.t. ε. Next we approximate $T_P^\varepsilon(\eta,\xi)$ by making $\widetilde{P}$ steps of the so-called *shifted asymptotic method (SAM)* (see [AAN11]), which also relies on repeated integration by parts. However, in contrast to the AM, the idea of the SAM is to shift the oscillatory factor such that an additional zero within the integration interval is created (here we create it at $y=\xi$). This increases the h-order of the remainder by one in each step:

$$\begin{aligned}
T_P^\varepsilon(\eta,\xi) &= (\mathrm{i}\varepsilon)^P \int_\xi^\eta b'_{P-1}(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y \\
&= -(\mathrm{i}\varepsilon)^{P+1}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \int_\xi^\eta b_P(y)\frac{\mathrm{d}}{\mathrm{d}y}\left[h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\right]\mathrm{d}y \\
&= -(\mathrm{i}\varepsilon)^{P+1}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left\{ b_P(\eta)h_1\left(\frac{2}{\varepsilon}s_n\right) - \int_\xi^\eta b'_P(y)h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y \right\} \\
&= -(\mathrm{i}\varepsilon)^{P+1}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left\{ b_P(\eta)h_1\left(\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\varepsilon)b_{P+1}(\eta)h_2\left(\frac{2}{\varepsilon}s_n\right) \right. \\
&\qquad \left. - (\mathrm{i}\varepsilon)\int_\xi^\eta b'_{P+1}(y)h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y \right\} \\
&= -(\mathrm{i}\varepsilon)^{P+1}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \sum_{p=1}^{\widetilde{P}} (\mathrm{i}\varepsilon)^{p-1} b_{p+P-1}(\eta)h_p\left(\frac{2}{\varepsilon}s_n\right) \\
&\quad + \mathcal{O}_{\varepsilon,h}(\varepsilon^P h^{\widetilde{P}}\min(\varepsilon,h))
\end{aligned} \tag{3.3.9}$$

(for $P=0$ we set $b'_{-1} := b$). Here, we used for the remainder integral in the last equation $h_{\widetilde{P}}\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) = \mathcal{O}_{\varepsilon,h}(\min(h^{\widetilde{P}}\varepsilon^{-\widetilde{P}}, h^{\widetilde{P}-1}\varepsilon^{-(\widetilde{P}-1)}))$, which is a consequence of (3.3.4). This concludes the proof. □

Remark 3.3.3. *A drawback of using quadrature (3.3.5) is that one has to provide explicit formulas for the functions b_p. Indeed, since these functions are defined in a recursive manner by differentiating a quotient of two functions, the number of terms involved in the formulas grows fast with p. For an efficient implementation we recommend to express the functions b_p only through the functions $b^{(k)}$, $k=0,\dots,p$, and $\phi^{(k)}$, $k=1,\dots,p+1$, as this keeps the formulas much shorter compared to when expressing b_p through a and its derivatives $a^{(k)}$, $k=1,\dots,p+2$.*

In the following, almost all approximations of occurring integrals will be derived by using the AM or the SAM. Recall that each step of the AM in (3.3.8) increased the ε-order of the remainder by one, whereas each step of the SAM in (3.3.9) increased the h-order by one. Thus, by combining both methods appropriately, one might hope to be able to achieve the desired mixed (w.r.t. ε and h) error order also for the approximations of the matrices $\mathbf{M}_2^\varepsilon$ and $\mathbf{M}_3^\varepsilon$. Unfortunately, this will not be the case as we will see in the following two subsections. Finally, we note that this strategy of combining the AM with the SAM was already used in [AAN11] for the construction of their second order (w.r.t. h) scheme.

3.3.2 Approximation of the matrix $\mathbf{M}_2^\varepsilon$

Next, according to (3.3.3), we would like to approximate the integral

$$\mathbf{M}_2^\varepsilon(\eta;\xi) = \int_\xi^\eta \mathbf{N}^\varepsilon(y)\mathbf{M}_1^\varepsilon(y;\xi)\,\mathrm{d}y \tag{3.3.10}$$

with the error order $\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3 \min(\varepsilon,h))$. But this will not be possible since the matrix product $\mathbf{N}^\varepsilon(y)\mathbf{M}_1^\varepsilon(y;\xi)$ contains terms which involve non-oscillatory integrals. Indeed, the basis for the AM to increase the ε-order in each step of (3.3.8) was the oscillatory factor $\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}$ of the integrand. However, for non-oscillatory integrals, it is only feasible to derive approximations with arbitrary h-order, but not with arbitrary ε-order. Nonetheless, in this subsection, we will derive an approximation for $\mathbf{M}_2^\varepsilon$, which ultimately leads to a third order scheme (w.r.t. h) that is uniformly correct w.r.t. ε – in fact with a local discretization error of the order $\mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^4 \max(\varepsilon,h))$. This will be achieved by employing a sufficiently accurate quadrature formula for the non-oscillatory integrals. To this end, let us introduce the notation

$$Q_S[f](\eta,\xi) := \frac{\eta-\xi}{6}\left(f(\xi) + 4f\left(\frac{\xi+\eta}{2}\right) + f(\eta)\right),$$

which denotes Simpson's rule applied to the integral $\int_\xi^\eta f(y)\,\mathrm{d}y$, admitting an error of the order $\mathcal{O}(h^5)$.

Since $\mathbf{N}^\varepsilon$ and $\mathbf{M}_1^\varepsilon$ are off-diagonal Hermitian matrices, the matrix $\mathbf{M}_2^\varepsilon$ is diagonal and the entries are conjugate of one another. Hence, we shall in the following only study one entry, namely,

$$m_2^\varepsilon(\eta,\xi) := (\mathbf{M}_2^\varepsilon(\eta;\xi))_{1,1} = \int_\xi^\eta (\mathbf{N}^\varepsilon(y))_{1,2}\,(\mathbf{M}_1^\varepsilon(y;\xi))_{2,1}\,\mathrm{d}y\,. \tag{3.3.11}$$

Here, it seems reasonable to insert for $(\mathbf{M}_1^\varepsilon(y;\xi))_{2,1} = m_1^\varepsilon(y,\xi)$ the approximation from Lemma 3.3.2 with $P = \widetilde{P} = 2$, namely $Q_1^{2,2}(y,\xi)$, as the order of its approximation error increases exactly to $\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3 \min(\varepsilon,h))$, when integrating over the subinterval $[\xi,\eta]$. From (3.3.11), (3.2.12), and (3.3.7) we have:

$$\begin{aligned}
m_2^\varepsilon(\eta,\xi) &= \int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\left\{Q_1^{2,2}(y,\xi) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^2 \min(\varepsilon,h))\right\}\mathrm{d}y \\
&= \int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\Bigg\{ -(\mathrm{i}\varepsilon)\left(b_0(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)} - b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\right) \\
&\quad -(\mathrm{i}\varepsilon)^2\left(b_1(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)} - b_1(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\right) - (\mathrm{i}\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_2(y)h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) \\
&\quad -(\mathrm{i}\varepsilon)^4\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_3(y)h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^2 \min(\varepsilon,h))\Bigg\}\mathrm{d}y \\
&= J_1 + J_2 + J_3 + J_4 + J_5 + J_6 + \mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3 \min(\varepsilon,h))\,,
\end{aligned} \tag{3.3.12}$$

where

$$J_1 := -(\mathrm{i}\,\varepsilon)\int_\xi^\eta b(y)b_0(y)\,\mathrm{d}y = -(\mathrm{i}\,\varepsilon)Q_S[bb_0](\eta,\xi) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^5)\,, \tag{3.3.13}$$

$$J_2 := (\mathrm{i}\,\varepsilon)b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y = (\mathrm{i}\,\varepsilon)b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\overline{m_1^\varepsilon(\eta,\xi)}\,,$$

$$J_3 := -(\mathrm{i}\,\varepsilon)^2\int_\xi^\eta b(y)b_1(y)\,\mathrm{d}y = -(\mathrm{i}\,\varepsilon)^2 Q_S[bb_1](\eta,\xi) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^5)\,, \tag{3.3.14}$$

$$J_4 := (\mathrm{i}\,\varepsilon)^2 b_1(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y = (\mathrm{i}\,\varepsilon)^2 b_1(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\overline{m_1^\varepsilon(\eta,\xi)}\,,$$

$$J_5 := -(\mathrm{i}\,\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_2(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y\,,$$

$$J_6 := -(\mathrm{i}\,\varepsilon)^4\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_3(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y\,.$$

Our next goal is to find suitable approximations for the integrals J_2, J_4, J_5, and J_6. Using Lemma 3.3.2 with $P=1$ and $\widetilde{P}=3$ we obtain by using $\overline{h_p(x)} = h_p(-x)$:

$$J_2 = (\mathrm{i}\,\varepsilon)^2 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left[b_0(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\right]_\xi^\eta - (\mathrm{i}\,\varepsilon)^3 b_0(\xi)\Bigg\{b_1(\eta)h_1\left(-\frac{2}{\varepsilon}s_n\right)$$
$$-(\mathrm{i}\,\varepsilon)b_2(\eta)h_2\left(-\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\,\varepsilon)^2 b_3(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\Bigg\} + \mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3\min(\varepsilon,h)). \tag{3.3.15}$$

Similarly, applying Lemma 3.3.2 with $P=0$ and $\widetilde{P}=3$ we find that

$$J_4 = (\mathrm{i}\,\varepsilon)^3 b_1(\xi)\Bigg\{b_0(\eta)h_1\left(-\frac{2}{\varepsilon}s_n\right) - (\mathrm{i}\,\varepsilon)b_1(\eta)h_2\left(-\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\,\varepsilon)^2 b_2(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\Bigg\}$$
$$+\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3\min(\varepsilon,h))\,. \tag{3.3.16}$$

For the approximation of the integral J_5, we begin by using the simple identity

$$\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) = -\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,. \tag{3.3.17}$$

Then, by making two SAM-steps, we derive

$$J_5 = (\mathrm{i}\,\varepsilon)^4\Bigg\{b_0(\eta)b_2(\eta)h_2\left(-\frac{2}{\varepsilon}s_n\right) - (\mathrm{i}\,\varepsilon)\left(b_1(\eta)b_2(\eta)+b_0(\eta)b_3(\eta)\right)h_3\left(-\frac{2}{\varepsilon}s_n\right)\Bigg\}$$
$$+\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3\min(\varepsilon,h)). \tag{3.3.18}$$

In order to approximate J_6 we use the identity

$$\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)$$
$$= -h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right) + \frac{2\,\mathrm{i}}{\varepsilon}(\phi(\xi)-\phi(y))h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,, \tag{3.3.19}$$

which allows us to split the integral into $J_6 = \widetilde{J}_6 + \widehat{J}_6$, with

$$\widetilde{J}_6 := (\mathrm{i}\,\varepsilon)^4 \int_\xi^\eta b(y)b_3(y)h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y\,,$$
$$\widehat{J}_6 := 2(\mathrm{i}\,\varepsilon)^3 \int_\xi^\eta b(y)b_3(y)\left(\phi(\xi)-\phi(y)\right)h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y\,.$$

One SAM-step for $\widetilde{J}_6$ leads to

$$\widetilde{J}_6 = (\mathrm{i}\,\varepsilon)^5 b_0(\eta)b_3(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2h^3\min(\varepsilon,h))\,. \tag{3.3.20}$$

For the approximation of $\widehat{J}_6$ we start with one SAM-step to see that

$$\widehat{J}_6 = 2(\mathrm{i}\,\varepsilon)^4\Bigg\{b_0(\eta)b_3(\eta)(-s_n)h_2\left(-\frac{2}{\varepsilon}s_n\right) - \underbrace{\int_\xi^\eta \left[b_0(y)b_3(y)\left(\phi(\xi)-\phi(y)\right)\right]' h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y}_{=:I}\Bigg\}\,. \tag{3.3.21}$$

Now, using $b_3(y)\phi'(y) = \frac{1}{2}b_2'(y)$, we split up the integral I as

$$\begin{aligned}
I &= \int_\xi^\eta \left(b_0'(y)b_3(y)+b_0(y)b_3'(y)\right)\left(\phi(\xi)-\phi(y)\right)h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y\\
&\quad -\frac{1}{2}\int_\xi^\eta b_0(y)b_2'(y)h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y\\
&= (\mathrm{i}\,\varepsilon)\left(b_1(\eta)b_3(\eta)+b_0(\eta)b_4(\eta)\right)(-s_n)h_3\left(-\frac{2}{\varepsilon}s_n\right)\\
&\quad -(\mathrm{i}\,\varepsilon)\int_\xi^\eta \left[\left(b_1(y)b_3(y)+b_0(y)b_4(y)\right)\left(\phi(\xi)-\phi(y)\right)\right]' h_3\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\,\mathrm{d}y\\
&\quad -\frac{1}{2}(\mathrm{i}\,\varepsilon)b_0(\eta)b_3(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2\min(h^4\varepsilon^{-4},h^3\varepsilon^{-3}))\,,
\end{aligned} \tag{3.3.22}$$

where we used one SAM-step for each integral in the last equation. Note that, combined with the $\mathcal{O}(\varepsilon^4)$-factor from (3.3.21), the first term in (3.3.22) is already of the desired order $\mathcal{O}_{\varepsilon,h}(\varepsilon^2h^3\min(\varepsilon,h))$ and can thus be omitted. Using (3.3.4) we see that the second term in (3.3.22) has the same order, and hence can also be neglected. Thus, we have

$$\begin{aligned}
\widehat{J}_6 &= 2(\mathrm{i}\,\varepsilon)^4\left\{b_0(\eta)b_3(\eta)(-s_n)h_2\left(-\frac{2}{\varepsilon}s_n\right)+\frac{1}{2}(\mathrm{i}\,\varepsilon)b_0(\eta)b_3(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\right\}\\
&\quad + \mathcal{O}_{\varepsilon,h}(\varepsilon^2h^3\min(\varepsilon,h))\,,
\end{aligned} \tag{3.3.23}$$

and hence, by combining (3.3.20) and (3.3.23),

$$\begin{aligned}
J_6 &= 2(\mathrm{i}\,\varepsilon)^4 b_0(\eta)b_3(\eta)(-s_n)h_2\left(-\frac{2}{\varepsilon}s_n\right) + 2(\mathrm{i}\,\varepsilon)^5 b_0(\eta)b_3(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\\
&\quad + \mathcal{O}_{\varepsilon,h}(\varepsilon^2h^3\min(\varepsilon,h))\,.
\end{aligned} \tag{3.3.24}$$

Finally, we summarize the approximations (3.3.12)-(3.3.16), (3.3.18), and (3.3.24) in the following lemma.

Lemma 3.3.4. *Let Hypothesis A be satisfied and define*

$$\begin{aligned} Q_2(\eta,\xi) := &-\mathrm{i}\varepsilon Q_S[bb_0](\eta,\xi) \\ &-\varepsilon^2\left[b_0(\xi)b_0(\eta)h_0\left(-\frac{2}{\varepsilon}s_n\right) - b_0(\xi)^2 - Q_S[bb_1](\eta,\xi)\right] \\ &+\mathrm{i}\varepsilon^3\left[b_0(\xi)b_1(\eta) - b_1(\xi)b_0(\eta)\right]h_1\left(-\frac{2}{\varepsilon}s_n\right) \\ &+\varepsilon^4\left[\left(b_0(\xi)+b_0(\eta)\right)b_2(\eta) - b_1(\xi)b_1(\eta) - 2b_0(\eta)b_3(\eta)s_n\right]h_2\left(-\frac{2}{\varepsilon}s_n\right) \\ &+\mathrm{i}\varepsilon^5\left[\left(b_0(\eta)-b_0(\xi)\right)b_3(\eta) - \left(b_1(\eta)-b_1(\xi)\right)b_2(\eta)\right]h_3\left(-\frac{2}{\varepsilon}s_n\right) \end{aligned} \tag{3.3.25}$$

and

$$\mathbf{Q}_2(\eta,\xi) := \begin{pmatrix} Q_2(\eta,\xi) & 0 \\ 0 & \overline{Q_2(\eta,\xi)} \end{pmatrix}. \tag{3.3.26}$$

Then there exists a constant $C \geq 0$ independent of $\varepsilon \in (0,\varepsilon_0]$, h, and n such that

$$\|\mathbf{M}_2^\varepsilon(\eta,\xi) - \mathbf{Q}_2(\eta,\xi)\|_\infty \leq C\varepsilon h^4 \max(\varepsilon,h)\,. \tag{3.3.27}$$

Proof. Since Q_2 is defined precisely through the approximations (3.3.12)-(3.3.16), (3.3.18), and (3.3.24) by neglecting the $\mathcal{O}_{\varepsilon,h}(\cdot)$-terms, estimate (3.3.27) follows by noticing that $\mathcal{O}_{\varepsilon,h}(\varepsilon^2h^3\min(\varepsilon,h)) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^5) + \mathcal{O}_{\varepsilon,h}(\varepsilon^2h^5) = \mathcal{O}_{\varepsilon,h}(\varepsilon h^4\max(\varepsilon,h))$. □

3.3.3 Approximation of the matrix $\mathbf{M}_3^\varepsilon$

Finally, in this subsection, we construct an approximation for the integral

$$\mathbf{M}_3^\varepsilon(\eta;\xi) = \int_\xi^\eta \mathbf{N}^\varepsilon(y)\mathbf{M}_2^\varepsilon(y;\xi)\,\mathrm{d}y\,. \tag{3.3.28}$$

While the desired error order for this approximation would be $\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))$ (see the discussion after (3.3.2)), this will again not be possible here – like in Subsection 3.3.2. Since the matrix $\mathbf{M}_2^\varepsilon$ is diagonal, $\mathbf{M}_3^\varepsilon$ is off-diagonal. We will now study $m_3^\varepsilon(\eta,\xi) := (\mathbf{M}_3^\varepsilon(\eta,\xi))_{2,1}$ as the entry $(\mathbf{M}_3^\varepsilon(\eta,\xi))_{1,2}$ is just its complex conjugate. To approximate m_3^ε, we have to insert an approximation for $\mathbf{M}_2^\varepsilon$ in the expression for m_3^ε and could of course simply use (3.3.26). However, we will instead insert a weaker approximation as this will not result in a reduced error order for the resulting scheme and, at the same time, it leads to a shorter quadrature formula for m_3^ε. The weaker approximation for $\mathbf{M}_2^\varepsilon$ can be derived by using $\mathbf{Q}_1^{1,1}$

(see (3.3.6)) to approximate $\mathbf{M}_1^\varepsilon$ and this was already used in [AAN11, Eq. (2.29)-(2.30)]:

$$\begin{aligned}(\mathbf{M}_2^\varepsilon(\eta;\xi))_{1,1} = m_2^\varepsilon(\eta,\xi) = &-(\mathrm{i}\,\varepsilon)\frac{\eta-\xi}{2}\left[b(\eta)b_0(\eta)+b(\xi)b_0(\xi)\right]\\ &+(\mathrm{i}\,\varepsilon)^2 b_0(\xi)b_0(\eta)h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(\eta))\right)\\ &+(\mathrm{i}\,\varepsilon)^3 b_1(\eta)\left[b_0(\eta)-b_0(\xi)\right]h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(\eta))\right)\\ &+\mathcal{O}_{\varepsilon,h}(\varepsilon h^3)\end{aligned} \tag{3.3.29}$$

Inserting the approximation (3.3.29) of $m_2^\varepsilon(y,\xi)$ for $(\mathbf{M}_2^\varepsilon(y;\xi))_{1,1}$ thus yields

$$\begin{aligned}m_3^\varepsilon(\eta,\xi) &= \int_\xi^\eta (\mathbf{N}^\varepsilon(y))_{2,1}\,(\mathbf{M}_2^\varepsilon(y;\xi))_{1,1}\,\mathrm{d}y\\ &= K_1+K_2+K_3+K_4+K_5+\mathcal{O}_{\varepsilon,h}(\varepsilon h^4)\,,\end{aligned} \tag{3.3.30}$$

where

$$\begin{aligned}K_1 &:= -\frac{(\mathrm{i}\,\varepsilon)}{2}\int_\xi^\eta b(y)^2 b_0(y)(y-\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y\,,\\ K_2 &:= -\frac{(\mathrm{i}\,\varepsilon)}{2}b(\xi)b_0(\xi)\int_\xi^\eta b(y)(y-\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y\,,\\ K_3 &:= (\mathrm{i}\,\varepsilon)^2 b_0(\xi)\int_\xi^\eta b(y)b_0(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y\,,\\ K_4 &:= -(\mathrm{i}\,\varepsilon)^3 b_0(\xi)\int_\xi^\eta b(y)b_1(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y\,,\\ K_5 &:= (\mathrm{i}\,\varepsilon)^3\int_\xi^\eta b(y)b_0(y)b_1(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y\,.\end{aligned}$$

For the approximation of the integrals K_i, $i=1,\dots,5$ it is convenient to introduce the abbreviations

$$c_0(y) := \frac{b(y)^2 b_0(y)}{2\phi'(y)}\,,\quad c_1(y) := \frac{c_0'(y)}{2\phi'(y)}\,,\quad d_0(y) := \frac{c_0(y)}{2\phi'(y)}\,, \tag{3.3.31}$$

$$d_1(y) := \frac{d_0'(y)}{2\phi'(y)}\,,\quad e_0(y) := \frac{c_1(y)}{2\phi'(y)}\,,\quad f_0(y) := \frac{b_0(y)}{2\phi'(y)}\,, \tag{3.3.32}$$

$$f_1(y) := \frac{f_0'(y)}{2\phi'(y)}\,,\quad g_0(y) := \frac{b_1(y)}{2\phi'(y)}\,,\quad \kappa(y) := b(y)b_1(y)\,, \tag{3.3.33}$$

$$\kappa_0(y) := \frac{\kappa(y)}{2\phi'(y)}\,,\quad l(y) := b(y)b_0(y)b_1(y)\,,\quad l_0(y) := \frac{l(y)}{2\phi'(y)}\,. \tag{3.3.34}$$

Note that all of these functions are uniformly bounded w.r.t. ε (with $0<\varepsilon\le\varepsilon_0$ for some sufficiently small ε_0) on the interval I. Indeed, this is again guaranteed by Hypothesis A.

Now, let us start with the approximation of K_1. By making three SAM-steps, we find

$$\begin{aligned}
K_1 &= \frac{(\mathrm{i}\,\varepsilon)^2}{2}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{c_0(\eta)(\eta-\xi)h_1\left(\frac{2}{\varepsilon}s_n\right)\right.\\
&\quad\left.+(\mathrm{i}\,\varepsilon)\int_\xi^\eta (c_1(y)(y-\xi)+d_0(y))\,\frac{\mathrm{d}}{\mathrm{d}y}\left[h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\right]\,\mathrm{d}y\right\}\\
&= \frac{(\mathrm{i}\,\varepsilon)^2}{2}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{c_0(\eta)(\eta-\xi)h_1\left(\frac{2}{\varepsilon}s_n\right)+(\mathrm{i}\,\varepsilon)\left(c_1(\eta)(\eta-\xi)+d_0(\eta)\right)h_2\left(\frac{2}{\varepsilon}s_n\right)\right.\\
&\quad\left.+(\mathrm{i}\,\varepsilon)^2\int_\xi^\eta\left(\frac{c_1'(y)}{2\phi'(y)}(y-\xi)+e_0(y)+d_1(y)\right)\frac{\mathrm{d}}{\mathrm{d}y}\left[h_3\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\right]\,\mathrm{d}y\right\}\\
&= \frac{(\mathrm{i}\,\varepsilon)^2}{2}\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{c_0(\eta)(\eta-\xi)h_1\left(\frac{2}{\varepsilon}s_n\right)+(\mathrm{i}\,\varepsilon)\left(c_1(\eta)(\eta-\xi)+d_0(\eta)\right)h_2\left(\frac{2}{\varepsilon}s_n\right)\right.\\
&\quad\left.+(\mathrm{i}\,\varepsilon)^2\left(e_0(\eta)+d_1(\eta)\right)h_3\left(\frac{2}{\varepsilon}s_n\right)\right\}+\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,. \qquad (3.3.35)
\end{aligned}$$

Next, the integral K_2 can be treated in the same way as K_1, i.e., by making three SAM-steps:

$$\begin{aligned}
K_2 &= \frac{(\mathrm{i}\,\varepsilon)^2}{2}b(\xi)b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{b_0(\eta)(\eta-\xi)h_1\left(\frac{2}{\varepsilon}s_n\right)+(\mathrm{i}\,\varepsilon)\left(b_1(\eta)(\eta-\xi)+f_0(\eta)\right)h_2\left(\frac{2}{\varepsilon}s_n\right)\right.\\
&\quad\left.+(\mathrm{i}\,\varepsilon)^2\left(g_0(\eta)+f_1(\eta)\right)h_3\left(\frac{2}{\varepsilon}s_n\right)\right\}+\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,. \qquad (3.3.36)
\end{aligned}$$

For the approximation of K_3 we first make use of (3.3.17). Then, we make two SAM-steps, yielding

$$\begin{aligned}
K_3 &= -(\mathrm{i}\,\varepsilon)^2 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_0(y)h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\,\mathrm{d}y\\
&= (\mathrm{i}\,\varepsilon)^3 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{b_0(\eta)^2h_2\left(\frac{2}{\varepsilon}s_n\right)+2(\mathrm{i}\,\varepsilon)b_0(\eta)b_1(\eta)h_3\left(\frac{2}{\varepsilon}s_n\right)\right\}\\
&\quad+\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,. \qquad (3.3.37)
\end{aligned}$$

In order to construct an approximation of K_4, we use a similar splitting as for J_6 in Subsection 3.3.2. Indeed, applying the identity

$$\begin{aligned}
\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right) &= -\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\\
&\quad+\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\frac{2\,\mathrm{i}}{\varepsilon}\left(\phi(y)-\phi(\xi)\right)h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\,, \qquad (3.3.38)
\end{aligned}$$

we can write $K_4 = \widetilde{K}_4 + \widehat{K}_4$, where

$$\begin{aligned}\widetilde{K}_4 &:= (\mathrm{i}\varepsilon)^3 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \int_\xi^\eta \kappa(y) h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) \mathrm{d}y \\ &= -(\mathrm{i}\varepsilon)^4 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\kappa_0(\eta) h_3\left(\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}\left(\varepsilon h^3 \min(\varepsilon,h)\right), \end{aligned} \tag{3.3.39}$$

and

$$\begin{aligned}\widehat{K}_4 &:= 2(\mathrm{i}\varepsilon)^2 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \int_\xi^\eta \kappa(y)\left(\phi(y)-\phi(\xi)\right) h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) \mathrm{d}y \\ &= -2(\mathrm{i}\varepsilon)^3 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left\{ \kappa_0(\eta) s_n h_2\left(\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\varepsilon)\frac{1}{2}\kappa_0(\eta) h_3\left(\frac{2}{\varepsilon}s_n\right)\right\} \\ &\quad + \mathcal{O}_{\varepsilon,h}\left(\varepsilon h^3 \min(\varepsilon,h)\right). \end{aligned} \tag{3.3.40}$$

Here, we made one SAM-step for $\widetilde{K}_4$ and treated $\widehat{K}_4$ in the same way as $\widehat{J}_6$ in the previous subsection. Thus, by combining (3.3.39) and (3.3.40), we obtain

$$\begin{aligned}K_4 &= -2(\mathrm{i}\varepsilon)^3 b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left\{ \kappa_0(\eta) s_n h_2\left(\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\varepsilon)\kappa_0(\eta) h_3\left(\frac{2}{\varepsilon}s_n\right)\right\} \\ &\quad + \mathcal{O}_{\varepsilon,h}\left(\varepsilon h^3 \min(\varepsilon,h)\right). \end{aligned} \tag{3.3.41}$$

Finally, the approximation of K_5 can be derived in the same way as the one of K_4, by noticing that l and l_0 now play the roles of κ and κ_0, respectively. Therefore, it holds that

$$\begin{aligned}K_5 &= 2(\mathrm{i}\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left\{ l_0(\eta) s_n h_2\left(\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\varepsilon) l_0(\eta) h_3\left(\frac{2}{\varepsilon}s_n\right)\right\} \\ &\quad + \mathcal{O}_{\varepsilon,h}\left(\varepsilon h^3 \min(\varepsilon,h)\right). \end{aligned} \tag{3.3.42}$$

We close this subsection with the following lemma, which summarizes the approximations (3.3.35)-(3.3.37), (3.3.41), and (3.3.42).

Lemma 3.3.5. *Let Hypothesis A be satisfied and define*

$$\begin{aligned}Q_3(\eta,\xi) &:= -\varepsilon^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\frac{\eta-\xi}{2}\left(c_0(\eta)+b(\xi)b_0(\xi)b_0(\eta)\right) h_1\left(\frac{2}{\varepsilon}s_n\right) \\ &\quad - \mathrm{i}\varepsilon^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left[\frac{1}{2}\left[c_1(\eta)(\eta-\xi)+d_0(\eta)+b(\xi)b_0(\xi)\left(b_1(\eta)(\eta-\xi)+f_0(\eta)\right)\right]\right. \\ &\qquad \left. + b_0(\xi)b_0(\eta)^2 + 2s_n\left(l_0(\eta)-b_0(\xi)\kappa_0(\eta)\right)\right] h_2\left(\frac{2}{\varepsilon}s_n\right) \\ &\quad + \varepsilon^4\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \left[\frac{1}{2}\left[e_0(\eta)+d_1(\eta)+b(\xi)b_0(\xi)\left(g_0(\eta)+f_1(\eta)\right)\right]\right. \\ &\qquad \left. + 2\left[b_0(\xi)b_0(\eta)b_1(\eta)+\left(l_0(\eta)-b_0(\xi)\kappa_0(\eta)\right)\right]\right] h_3\left(\frac{2}{\varepsilon}s_n\right) \end{aligned} \tag{3.3.43}$$

and

$$\mathbf{Q}_3(\eta,\xi) := \begin{pmatrix} 0 & \overline{Q_3(\eta,\xi)} \\ Q_3(\eta,\xi) & 0 \end{pmatrix} . \tag{3.3.44}$$

Then there exists a constant $C \geq 0$ independent of $\varepsilon \in (0,\varepsilon_0]$, h, and n such that

$$\|\mathbf{M}_3^\varepsilon(\eta,\xi) - \mathbf{Q}_3(\eta,\xi)\|_\infty \leq C\varepsilon h^4 . \tag{3.3.45}$$

3.4 Numerical scheme and error analysis

Based on the Picard approximation (3.3.3) as well as the quadratures (3.3.6), (3.3.26), and (3.3.44) we can now define the numerical scheme: Let

$$\mathbf{A}_n^1 := \varepsilon \mathbf{Q}_1^{3,3}(x_{n+1},x_n) , \quad \mathbf{A}_n^2 := \varepsilon^2 \mathbf{Q}_2(x_{n+1},x_n) , \quad \mathbf{A}_n^3 := \varepsilon^3 \mathbf{Q}_3(x_{n+1},x_n) . \tag{3.4.1}$$

Given the initial value Z_0 we define

$$Z_{n+1} := \left(\mathbf{I} + \mathbf{A}_n^1 + \mathbf{A}_n^2 + \mathbf{A}_n^3\right) Z_n , \quad n = 0,\dots,N-1 . \tag{3.4.2}$$

The numerical solution of (3.2.9) can then be obtained through the inverse transform (3.2.16):

$$U_n := \mathbf{P}^{-1} \exp\left(\frac{\mathrm{i}}{\varepsilon}\mathbf{\Phi}^\varepsilon(x_n)\right) Z_n , \quad n = 0,\dots,N . \tag{3.4.3}$$

The method (3.4.2)-(3.4.3) satisfies the following global error estimates:

Theorem 3.4.1. *Let Hypothesis A be satisfied. Let Z and U be the exact solutions of the IVPs (3.2.11) and (3.2.9), respectively. Then, for Z_n and U_n being computed through the scheme (3.4.2)-(3.4.3), there exists a generic constant $C \geq 0$ independent of $\varepsilon \in (0,\varepsilon_0]$, h, and n such that*

$$\|Z(x_n) - Z_n\|_\infty \leq C\varepsilon^3 h^3 \max(\varepsilon,h) , \quad n = 0,\dots,N , \tag{3.4.4}$$

$$\|U(x_n) - U_n\|_\infty \leq C\varepsilon^3 h^3 \max(\varepsilon,h) , \quad n = 0,\dots,N . \tag{3.4.5}$$

Proof. From the definitions (3.3.5), (3.3.25), and (3.3.43) it is evident that

$$\begin{aligned} \|\mathbf{A}_n^1\|_\infty &= \|\varepsilon \mathbf{Q}_1^{3,3}(x_{n+1},x_n)\|_\infty = \mathcal{O}_{\varepsilon,h}(\varepsilon \min(\varepsilon,h)) , \\ \|\mathbf{A}_n^2\|_\infty &= \|\varepsilon^2 \mathbf{Q}_2(x_{n+1},x_n)\|_\infty = \mathcal{O}_{\varepsilon,h}(\varepsilon^3 h) , \\ \|\mathbf{A}_n^3\|_\infty &= \|\varepsilon^3 \mathbf{Q}_3(x_{n+1},x_n)\|_\infty = \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h \min(\varepsilon,h)) , \end{aligned} \tag{3.4.6}$$

which implies that the one-step method (3.4.2) is stable, with an ε-independent stability constant. Further, due to (3.3.3), the consistency error for $n = 0,\dots,N-1$ reads

$$\begin{aligned} e_n &:= Z(x_{n+1}) - \left(\mathbf{I} + \mathbf{A}_n^1 + \mathbf{A}_n^2 + \mathbf{A}_n^3\right) Z(x_n) \\ &= Z(x_n) + \left[\varepsilon \mathbf{M}_1^\varepsilon(x_{n+1};x_n) + \varepsilon^2 \mathbf{M}_2^\varepsilon(x_{n+1};x_n) + \varepsilon^3 \mathbf{M}_3^\varepsilon(x_{n+1};x_n)\right] Z(x_n) \\ &\quad - \left(\mathbf{I} + \mathbf{A}_n^1 + \mathbf{A}_n^2 + \mathbf{A}_n^3\right) Z(x_n) + \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h^3 \min(\varepsilon,h)) \\ &= \left[\varepsilon \mathbf{M}_1^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^1 + \varepsilon^2 \mathbf{M}_2^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^2 + \varepsilon^3 \mathbf{M}_3^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^3\right] Z(x_n) \\ &\quad + \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h^3 \min(\varepsilon,h)) , \end{aligned} \tag{3.4.7}$$

Now, Lemma 3.3.2, Lemma 3.3.4, and Lemma 3.3.5 imply

$$\begin{aligned}\varepsilon \mathbf{M}_1^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^1 &= \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h^3 \min(\varepsilon,h))\,,\\ \varepsilon^2 \mathbf{M}_2^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^2 &= \mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^4 \max(\varepsilon,h))\,,\\ \varepsilon^3 \mathbf{M}_3^\varepsilon(x_{n+1};x_n) - \mathbf{A}_n^3 &= \mathcal{O}_{\varepsilon,h}(\varepsilon^4 h^4)\,,\end{aligned} \tag{3.4.8}$$

which, together with (3.4.7), yields

$$e_n = \mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^4 \max(\varepsilon,h))\,. \tag{3.4.9}$$

Hence, the one-step method (3.4.2) is consistent and therefore convergent with the global error estimate (3.4.4). Estimate (3.4.5) now follows from (3.4.4) and the inverse transformation (3.2.16), using the unitarity of the matrices $\mathbf{P}^{-1}$ and $\exp\left(\frac{\mathrm{i}}{\varepsilon}\mathbf{\Phi}^\varepsilon\right)$. This proves the claim. □

Remark 3.4.2. *We emphasize that, in practical applications, when solving IVP (3.2.1) using a WKB-based method like the one presented above, one may want to use a grid size $h \gg \varepsilon$. In such scenarios the global error of the third order scheme (3.4.2) is even proportional to h^4, according to the estimates (3.4.4) and (3.4.5). This property is observable in Figure 3.5.3 of the next section.*

3.4.1 Refined error estimate incorporating phase errors

In the framework of Theorem 3.4.1 we implicitly assumed that the phase (3.2.7) is exactly available. Indeed, this is the case in several relevant applications, e.g., for so-called RTD-models (e.g., see [MJK13, SHMS98]), where the coefficient function a is typically piecewise linear. Thus, in such scenarios the method (3.4.2)-(3.4.3) is asymptotically correct w.r.t. ε.

In general, however, the integral within the definition of the phase (3.2.7) cannot be expected to be exactly computable. In [AKU22], the authors extended the error analysis of the numerical methods from [AAN11] to the case of a numerically computed phase. We emphasize that, with the exact same strategy, the estimates (3.4.4) and (3.4.5) can be generalized. To this end, let us first collect the basic assumptions from [AKU22].

Firstly, for an approximate phase $\tilde{\phi} \approx \phi$ we write $\tilde{\phi}(x) = \tilde{\phi}_1(x) - \varepsilon^2\tilde{\phi}_2(x)$, with $\tilde{\phi}_1(x)$ and $\tilde{\phi}_2(x)$ being numerical approximations to $\int_{x_0}^x \sqrt{a(y)}\,\mathrm{d}y$ and $\int_{x_0}^x b(y)\,\mathrm{d}y$, respectively. Then we impose the following assumption:

Hypothesis B. *Let $\tilde{\phi}_1, \tilde{\phi}_2 \in C^6(I)$ and $\tilde{\phi}_1' \geq \widetilde{C} > 0$.*

In particular, Hypothesis B implies that there are positive constants E, E' and E'' such that the approximate phase $\tilde{\phi}$ satisfies the following $L^\infty(I)$-error bounds uniformly in $\varepsilon \in (0, \tilde{\varepsilon}_0]$:

$$\|\phi - \tilde{\phi}\|_{L^\infty(I)} \leq E\,, \quad \|\phi' - \tilde{\phi}'\|_{L^\infty(I)} \leq E'\,, \quad \|\phi'' - \tilde{\phi}''\|_{L^\infty(I)} \leq E''\,, \tag{3.4.10}$$

where $\tilde{\varepsilon}_0 \leq \varepsilon_0$ is sufficiently small. Note that the functions a, b, and b_p, $p = 0,\ldots,5$, as well as all functions defined in (3.3.31)-(3.3.34) can be computed exactly, since ϕ' is explicitly known. However, since we assume here an approximate phase $\tilde{\phi}$, it appears more consistent

to use instead the numerical approximations $\tilde{a}$, $\tilde{b}$, $\tilde{b}_p$ (for $p = 0, \dots, 5$), $\tilde{c}_0$, $\tilde{c}_1$, $\tilde{d}_0$, $\tilde{d}_1$, $\tilde{e}_0$, $\tilde{f}_0$, $\tilde{f}_1$, $\tilde{g}_0$, $\tilde{\kappa}_0$, and $\tilde{l}_0$, which can be obtained using exact derivatives of $\tilde{\phi}$ (e.g., when using the spectral integration from [AKU22, §4] to obtain $\tilde{\phi}$, the exact derivatives are readily available). Together with the approximate phase increments $\tilde{s}_n := \tilde{\phi}(x_{n+1}) - \tilde{\phi}(x_n)$ this leads to the approximate matrices $\tilde{\mathbf{A}}_n^1 \approx \mathbf{A}_n^1$, $\tilde{\mathbf{A}}_n^2 \approx \mathbf{A}_n^2$ and $\tilde{\mathbf{A}}_n^3 \approx \mathbf{A}_n^3$ which are defined in analogy to (3.4.1). The third order method involving the approximate phase then reads

$$\tilde{Z}_{n+1} := \left(\mathbf{I} + \tilde{\mathbf{A}}_n^1 + \tilde{\mathbf{A}}_n^2 + \tilde{\mathbf{A}}_n^3\right) \tilde{Z}_n \,, \quad n = 0, \dots, N-1 \,, \tag{3.4.11}$$

with $\tilde{Z}_0 := Z_0$ being the initial value. The corresponding (perturbed) inverse transform to the U-variable is then given by (see (3.2.16))

$$\tilde{U}_n = \mathbf{P}^{-1} \exp\left(\frac{\mathrm{i}}{\varepsilon} \tilde{\mathbf{\Phi}}^{\varepsilon}(x_n)\right) \tilde{Z}_n \,, \quad n = 0, \dots, N \,, \tag{3.4.12}$$

with $\tilde{\mathbf{\Phi}}^{\varepsilon} := \operatorname{diag}(\tilde{\phi}, -\tilde{\phi})$ being the perturbed phase matrix.

Now, given the above considerations, we stress that, by using the exact same arguments as in the proof of [AKU22, Theorem 3.2], we can extend the estimates (3.4.4)-(3.4.5) to incorporate also the phase errors:

Theorem 3.4.3 (Adaption of Thm. 3.2 from [AKU22]). *Let Hypothesis A be satisfied and let the approximate phase $\tilde{\phi}$ satisfy Hypothesis B. Let Z and U be the exact solutions of the IVPs (3.2.11) and (3.2.9), respectively. Then, for $\tilde{Z}_n$ and $\tilde{U}_n$ being computed through the scheme (3.4.11)-(3.4.12), there exists a generic constant $C \geq 0$ independent of $\varepsilon \in (0, \tilde{\varepsilon}_0]$, h, and n such that*

$$\begin{aligned} \|Z(x_n) - \tilde{Z}_n\| \leq\ & C\varepsilon^3 h^3 \max(\varepsilon, h) \\ & + C\varepsilon \left[\min(\varepsilon, E) + \varepsilon(E' + E'')\right] \,, \quad n = 0, \dots, N \,, \end{aligned} \tag{3.4.13}$$

$$\begin{aligned} \|U(x_n) - \tilde{U}_n\| \leq\ & C\frac{E}{\varepsilon} + C\varepsilon^3 h^3 \max(\varepsilon, h) \\ & + C\varepsilon \left[\min(\varepsilon, E) + \varepsilon(E' + E'')\right] \,, \quad n = 0, \dots, N \,. \end{aligned} \tag{3.4.14}$$

Here, the constants E, E', and E'' are from (3.4.10).

Let us compare this result with the estimates (3.4.4) and (3.4.5): The new (additional) second term in (3.4.13) is caused by the impact of the approximate phase during the computation of the step update (3.4.11), whereas the new first term in (3.4.14) is due to the inverse transform (3.4.12) involving the perturbed phase matrix $\tilde{\mathbf{\Phi}}^{\varepsilon}(x_n)$. Obviously, the $\mathcal{O}(E/\varepsilon)$-term is rather unfavorable. In order to reduce its impact as much as possible, one should hence aim for a highly accurate approximation $\tilde{\phi}$, e.g., by employing a spectral method. Indeed, such an approach has already proven useful in [AKU22], where the $\mathcal{O}(E/\varepsilon)$-term in their estimate was numerically invisible, since it was reduced below relative machine precision (compared to the other terms).

3.4.2 Simplified third order scheme

In this subsection, we present a simplified third order method for solving the IVP (3.2.11). The basis for this method is the observation, that the inability to achieve the originally desired error order for the approximations of $\mathbf{M}_2^\varepsilon$ and $\mathbf{M}_3^\varepsilon$ in Section 3.3, was due to the presence of non-oscillatory integrals in the computations. Indeed, it is evident from Subsection 3.3.2 that the approximation of the non-oscillatory integral J_1 in (3.3.13) constrains the maximum achievable error order (w.r.t. ε) for the approximation of $\mathbf{M}_2^\varepsilon$ to $\mathcal{O}(\varepsilon)$. Consequently, this implies that the maximum achievable error order (w.r.t. ε) for any numerical scheme resulting from this approximation is $\mathcal{O}(\varepsilon^3)$. Given this unavoidable constraint, it seems unnecessary and superfluous to approximate the other terms with higher accuracy. Thus, we may want to weaken several approximations from Section 3.3 in order to derive a simplified scheme, which remains third order w.r.t. h, while still yielding the same asymptotic accuracy as the third order scheme (3.4.2)-(3.4.3), i.e. $\mathcal{O}(\varepsilon^3)$ as $\varepsilon \to 0$. This can be achieved as follows:

Step 1 (approximation of $\mathbf{M}_1^\varepsilon$ in (3.3.3)): Instead of $\mathbf{Q}_1^{3,3}$, we shall use $\mathbf{Q}_1^{2,3}$ to approximate $\mathbf{M}_1^\varepsilon$. This incurs an approximation error of order $\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3 \min(\varepsilon, h))$ (instead of $\mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^3 \min(\varepsilon, h))$), see Lemma 3.3.2.

Step 2 (approximation of $\mathbf{M}_2^\varepsilon$ in (3.3.3)): To approximate $\mathbf{M}_2^\varepsilon$, we shall use a simplified quadrature $\mathbf{Q}_{2,simp}$ (instead of $\mathbf{Q}_2$), which can be derived by inserting $\mathbf{Q}_1^{1,2}$ (instead of $\mathbf{Q}_1^{2,2}$) as an approximation for $\mathbf{M}_1^\varepsilon$ in (3.3.10). The non-oscillatory integrals that then occur are again approximated applying Simpson's rule, and the oscillatory integrals are all approximated with an error order $\mathcal{O}_{\varepsilon,h}(\varepsilon h^3 \min(\varepsilon, h))$ (instead of $\mathcal{O}_{\varepsilon,h}(\varepsilon^2 h^3 \min(\varepsilon, h))$), using only SAM-steps.

Step 3 (approximation of $\mathbf{M}_3^\varepsilon$ in (3.3.3)): To approximate $\mathbf{M}_3^\varepsilon$, we shall use a simplified quadrature $\mathbf{Q}_{3,simp}$ (instead of $\mathbf{Q}_3$), which can be derived as follows: First, by inserting $\mathbf{Q}_1^{0,1}$ (instead of $\mathbf{Q}_1^{1,1}$) as an approximation for $\mathbf{M}_1^\varepsilon$ in (3.3.10), we derive with only one SAM-step an approximation $\widetilde{\mathbf{Q}}_{2,simp}$ of $\mathbf{M}_2^\varepsilon$, with an error order $\mathcal{O}_{\varepsilon,h}(h^2 \min(\varepsilon, h))$. Then, by inserting $\widetilde{\mathbf{Q}}_{2,simp}$ in (3.3.28) one obtains an oscillatory integral that can be approximated using only SAM-steps with an error order $\mathcal{O}_{\varepsilon,h}(h^3 \min(\varepsilon, h))$. This yields the quadrature $\mathbf{Q}_{3,simp}$.

We stress that the implementation and analysis of Step 2 and Step 3 is fully straightforward and very similar to the computations in Subsection 3.3.2 and Subsection 3.3.3, while Step 1 is trivial. The resulting quadratures $\mathbf{Q}_{2,simp}$ and $\mathbf{Q}_{3,simp}$, along with their corresponding error estimates, are provided by the following lemma.[2]

[2]The proof of Lemma 3.4.4 is deferred to Appendix 3.A.

Lemma 3.4.4. *Let Hypothesis A be satisfied and define*

$$\begin{aligned} Q_{2,simp}(\eta,\xi) := &-\mathrm{i}\,\varepsilon Q_S[bb_0](\eta,\xi) - \varepsilon^2 b_0(\xi)b_0(\eta)h_1\left(-\frac{2}{\varepsilon}s_n\right) \\ &-\mathrm{i}\,\varepsilon^3\left[b_0(\eta)\left(b_1(\eta)-2b_2(\eta)s_n\right)-b_0(\xi)b_1(\eta)\right]h_2\left(-\frac{2}{\varepsilon}s_n\right) \\ &+\varepsilon^4\left[b_2(\eta)\left(b_0(\xi)+b_0(\eta)\right)-b_1(\eta)^2\right]h_3\left(-\frac{2}{\varepsilon}s_n\right), \end{aligned} \tag{3.4.15}$$

$$Q_{3,simp}(\eta,\xi) := -2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_0(\eta)^3\left(\varepsilon^2 s_n h_2\left(\frac{2}{\varepsilon}s_n\right)+\mathrm{i}\,\varepsilon^3 h_3\left(\frac{2}{\varepsilon}s_n\right)\right) \tag{3.4.16}$$

and

$$\mathbf{Q}_{2,simp}(\eta,\xi) := \begin{pmatrix} Q_{2,simp}(\eta,\xi) & 0 \\ 0 & \overline{Q_{2,simp}(\eta,\xi)} \end{pmatrix}, \tag{3.4.17}$$

$$\mathbf{Q}_{3,simp}(\eta,\xi) := \begin{pmatrix} 0 & \overline{Q_{3,simp}(\eta,\xi)} \\ Q_{3,simp}(\eta,\xi) & 0 \end{pmatrix}. \tag{3.4.18}$$

Then there exists a generic constant $C \geq 0$ independent of $\varepsilon \in (0,\varepsilon_0]$, h, and n such that

$$\|\mathbf{M}_2^\varepsilon(\eta,\xi) - \mathbf{Q}_{2,simp}(\eta,\xi)\|_\infty \leq C\varepsilon h^4, \tag{3.4.19}$$

$$\|\mathbf{M}_3^\varepsilon(\eta,\xi) - \mathbf{Q}_{3,simp}(\eta,\xi)\|_\infty \leq Ch^3\min(\varepsilon,h). \tag{3.4.20}$$

The resulting simplified third order scheme then reads: Let $\mathbf{A}_n^{1,simp} := \varepsilon\mathbf{Q}_1^{2,3}(x_{n+1},x_n)$, $\mathbf{A}_n^{2,simp} := \varepsilon^2\mathbf{Q}_{2,simp}(x_{n+1},x_n)$, and $\mathbf{A}_n^{3,simpl} := \varepsilon^3\mathbf{Q}_{3,simp}(x_{n+1},x_n)$. Given the initial value Z_0 we define

$$Z_{n+1} := \left(\mathbf{I}+\mathbf{A}_n^{1,simp}+\mathbf{A}_n^{2,simp}+\mathbf{A}_n^{3,simp}\right)Z_n, \quad n=0,\dots,N-1. \tag{3.4.21}$$

The simplified method (3.4.21), (3.4.3) satisfies the following global error estimate, which can be proven analogously to Theorem 3.4.1:

Theorem 3.4.5. *Let Hypothesis A be satisfied. Let Z and U be the exact solutions of the IVPs (3.2.11) and (3.2.9), respectively. Then, for Z_n and U_n being computed through the scheme (3.4.21), (3.4.3), there exists a generic constant $C \geq 0$ independent of $\varepsilon \in (0,\varepsilon_0]$, h, and n such that*

$$\|Z(x_n) - Z_n\|_\infty \leq C\varepsilon^3 h^3, \quad n=0,\dots,N, \tag{3.4.22}$$

$$\|U(x_n) - U_n\|_\infty \leq C\varepsilon^3 h^3, \quad n=0,\dots,N. \tag{3.4.23}$$

When comparing estimates (3.4.22)-(3.4.23) with estimates (3.4.4)-(3.4.5) for the scheme (3.4.2)-(3.4.3), we observe that the downside of using the simplified scheme (3.4.21), (3.4.3) is the loss of the $\mathcal{O}(\max(\varepsilon,h))$-factor. Indeed, while both methods have the same asymptotical (w.r.t. ε) as well as numerical (w.r.t. h) order, this factor might still be beneficial,

recalling that both parameters ε and h are small. Further, we note that, in case of a numerically computed phase, Theorem 3.4.5 can be generalized analogously to Theorem 3.4.3. That is, by using the same strategy as in Subsection 3.4.1, we have the following error estimates for the simplified scheme including an approximated phase:

$$\begin{aligned}\|Z(x_n) - \tilde{Z}_n\|_\infty \leq\ & C\varepsilon^3 h^3 \\ & + C\varepsilon \left[\min(\varepsilon, E) + \varepsilon(E' + E'')\right] , \quad n = 0, \ldots, N , \end{aligned} \tag{3.4.24}$$

$$\begin{aligned}\|U(x_n) - \tilde{U}_n\|_\infty \leq\ & C\frac{E}{\varepsilon} + C\varepsilon^3 h^3 \\ & + C\varepsilon \left[\min(\varepsilon, E) + \varepsilon(E' + E'')\right] , \quad n = 0, \ldots, N . \end{aligned} \tag{3.4.25}$$

3.5 Numerical results

In this section we present and compare numerical results by applying the novel schemes from Section 3.4 to exemplary IVPs. Since a large part of the development of the two presented third order schemes (3.4.2)-(3.4.3) and (3.4.21), (3.4.3) was based on the strategies from [AAN11], we shall also include their second order method into our comparison. This will also illustrate the efficiency gain of the new third order methods over the second order method from [AAN11].

For clarity of the presentation we shall in the following denote with *WKB3* the third order scheme (3.4.2)-(3.4.3), and with *WKB3s* the simplified third order scheme (3.4.21), (3.4.3). Further, we denote with *WKB2* the second order method from [AAN11], which yields global errors (w.r.t. the U-variable) of order $\mathcal{O}_{\varepsilon,h}(\varepsilon^3 h^2)$, provided that the phase (3.2.7) is explicitly available. We note that a single step of each of these three methods involves a different number of computational operations as well as function evaluations. Hence it is not clear a priori which of the three methods is the most efficient. We shall thus assess the overall efficiency of each method by comparing respective work-precision diagrams.

All computations here are carried out using MATLAB version 23.2.0.2459199 (R2023b). Moreover, for the results of Figures 3.5.3-3.5.4, which involve very small approximation errors, we used the Advanpix Multiprecision Computing Toolbox for MATLAB [Adv23] with quadruple-precision to avoid rounding errors. We note that this increases the computational times when compared to using the standard double-precision arithmetic of MATLAB.

3.5.1 First example: Airy equation

For the Airy equation, i.e., with $a(x) = x$ we consider the following IVP:

$$\begin{cases} \varepsilon^2 \varphi''(x) + x\varphi(x) = 0 , \quad x \in [1, 2] , \\ \varphi(1) = \mathrm{Ai}(-\frac{1}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}(-\frac{1}{\varepsilon^{2/3}}) , \\ \varepsilon\varphi'(1) = \varepsilon^{1/3}\left(\mathrm{Ai}'(-\frac{1}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}'(-\frac{1}{\varepsilon^{2/3}})\right) . \end{cases} \tag{3.5.1}$$

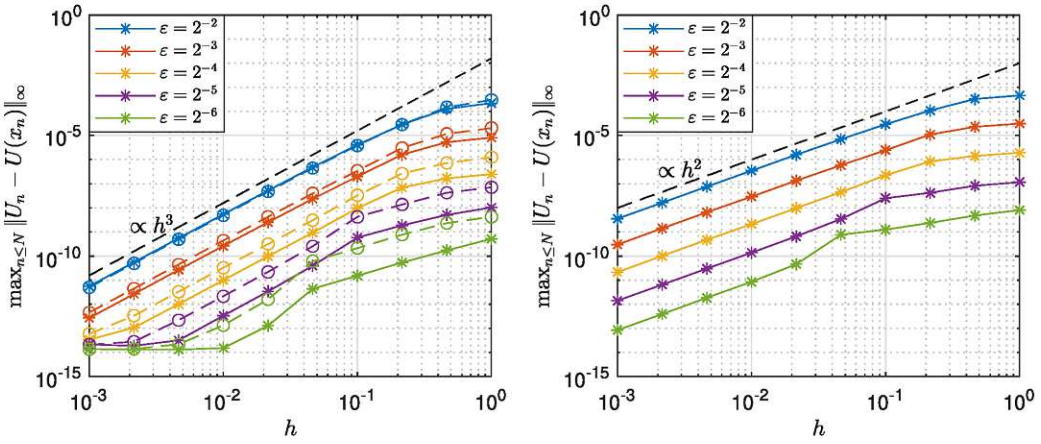

Figure 3.5.1: Global errors $\max_{n\leq N}\|U(x_n) - U_n\|_\infty$ for the Airy equation (3.5.1) as functions of the step size h, for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

Here, the exact solution is given by

$$\varphi_{exact}(x) = \mathrm{Ai}(-\frac{x}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}(-\frac{x}{\varepsilon^{2/3}})\,,$$

where Ai and Bi denote the Airy functions of first and second kind, respectively (e.g., see [OLBC10, Chap. 9], [KAD21, Sec. 5.1]). Here the phase (3.2.7) is exactly computable.

Let us first investigate the convergence results from Theorems 3.4.1 and 3.4.5. In Figure 3.5.1 we plot the global error $\max_{n\leq N}\|U(x_n) - U_n\|_\infty$ as a function of the step size h, for several values of ε. The left plot shows the error for WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles), and the right plot shows the error for WKB2. As indicated by the dashed black line in the left plot, the error for WKB3 and WKB3s clearly decays like $\mathcal{O}(h^3)$, in accordance with the third order estimates (3.4.5) and (3.4.23). For small step sizes h and $\varepsilon = 2^{-5}, 2^{-6}$ the error saturates at approximately 10^{-14}, due to rounding errors. Further, for a fixed step size h, the error for all three methods decreases with ε, demonstrating the ε-asymptotical correctness of each method. According to estimate (3.4.5), we expect the error for WKB3 to behave here like $\mathcal{O}(\varepsilon^4)$, since no shown error curve decays like $\mathcal{O}(h^4)$, suggesting that the $\mathcal{O}_{\varepsilon,h}(\max(\varepsilon, h))$-factor in (3.4.5) is equal to $\mathcal{O}(\varepsilon)$ for the shown ε-values. Further, estimate (3.4.23) suggests that the error for WKB3s decreases like $\mathcal{O}(\varepsilon^3)$, which is also the expected error behavior of WKB2, according to [AAN11]. However, Figure 3.5.1 reveals that the error for each method shows an ε-order that is 0.3 - 0.9 higher than expected theoretically. This is due to the oscillatory behavior of the consistency error of each method, which leads to cancellation effects in successive integration steps. Indeed, this phenomenon was already observed and analyzed in [AAN11, §3.3], and appears in all figures of this chapter.

Figure 3.5.2 contains work-precision diagrams, which correspond to the computations from Figure 3.5.1. That is, CPU times (measured in seconds) are plotted against the global errors, resulting from using the different step sizes h. Clearly, in the same amount of time, WKB3 and WKB3s both reach much smaller global errors, when compared to WKB2. For instance, with a CPU time of approximately 10^{-3} seconds, the global errors using WKB3

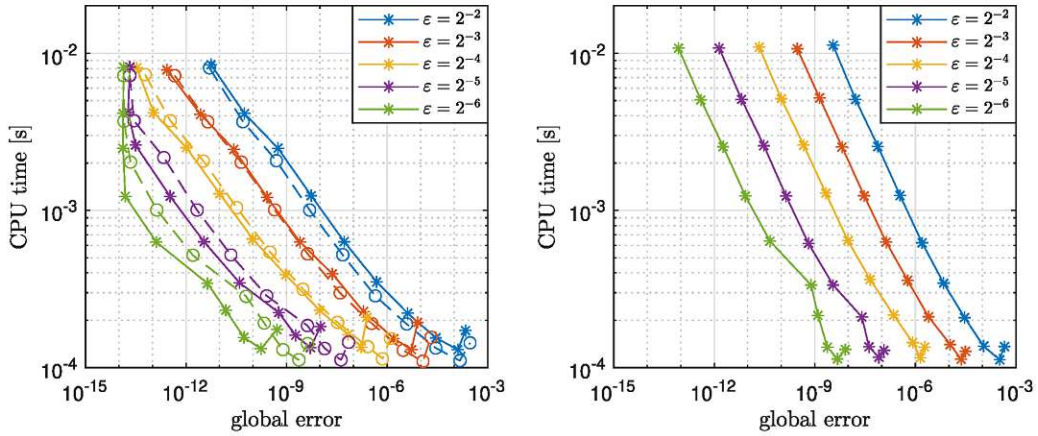

Figure 3.5.2: CPU times vs. global errors $\max_{n\leq N}\|U(x_n) - U_n\|_\infty$ for the Airy equation (3.5.1) on the spatial interval $[1, 2]$, for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

or WKB3s lie between 10^{-14} - 10^{-13} and 10^{-8} for the different ε-values, whereas the errors lie between 10^{-11} and 10^{-7} - 10^{-6} when using WKB2. Vice versa, to reach a fixed accuracy, WKB3 and WKB3s need significantly less CPU time than WKB2. For example, to produce a global error of approximately 10^{-8} for $\varepsilon = 2^{-2}$, WKB3 and WKB3s need approximately 10^{-3} seconds, whereas WKB2 needs around $5 \cdot 10^{-3}$ seconds. The time difference is even larger when considering smaller values of ε: E.g., to reach an accuracy of approximately 10^{-13} for $\varepsilon = 2^{-6}$, WKB3 and WKB3s need around $6 \cdot 10^{-4}$ and 10^{-3} seconds, respectively. By contrast, the computational time to reach the same accuracy with WKB2 is around 10^{-2} seconds. Hence, we conclude from Figure 3.5.2 that WKB3 and WKB3s are both more efficient than WKB2 (for this example), with CPU times being smaller up to a factor of ten. The difference between WKB3 and WKB3s is not so obvious: Indeed, for $\varepsilon = 2^{-2}$ and 2^{-3} we conclude from the left plot of Figure 3.5.2 that WKB3s needs less time to reach the same accuracy as WKB3. However, for the smaller values $\varepsilon = 2^{-6}$ and $\varepsilon = 2^{-5}$, WKB3 becomes more efficient.

In Figures 3.5.3-3.5.4 analogous plots are shown for much smaller values of ε. According to the convergence plot on the left of Figure 3.5.3, when $\varepsilon \ll h$, the error for WKB3 even decreases like $\mathcal{O}(h^4)$, e.g., for $\varepsilon = 10^{-4}, 10^{-5}, 10^{-6}$, as indicated by the bottom dashed black line. This behavior agrees well with estimate (3.4.5), see also Remark 3.4.2. Indeed, the $\mathcal{O}_{\varepsilon,h}(\max(\varepsilon, h))$-factor from estimate (3.4.5) is equal to $\mathcal{O}(h)$ in these cases. Hence, for large step sizes h, we cannot expect an $\mathcal{O}(\varepsilon^4)$ error behavior for WKB3 anymore (as in Figure 3.5.1). Instead, we observe that (e.g., for $h = 1$), the error for WKB3, WKB3s, and WKB2 (see the right plot) roughly decreases like $\mathcal{O}(\varepsilon^3)$. This is in good agreement with estimates (3.4.5) and (3.4.23). In the work-precision diagrams of Figure 3.5.4 one can clearly observe that WKB3 and WKB3s are again more efficient than WKB2. E.g., with a CPU time of approximately 3 seconds, the global errors for the different ε-values lie between 10^{-29} and 10^{-13} - 10^{-12} when using WKB3 and WKB3s, whereas they lie between 10^{-25} and 10^{-11} when using WKB2. Moreover, we observe also a big difference between WKB3 and WKB3s for $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$: Indeed, for each of these ε-values, and

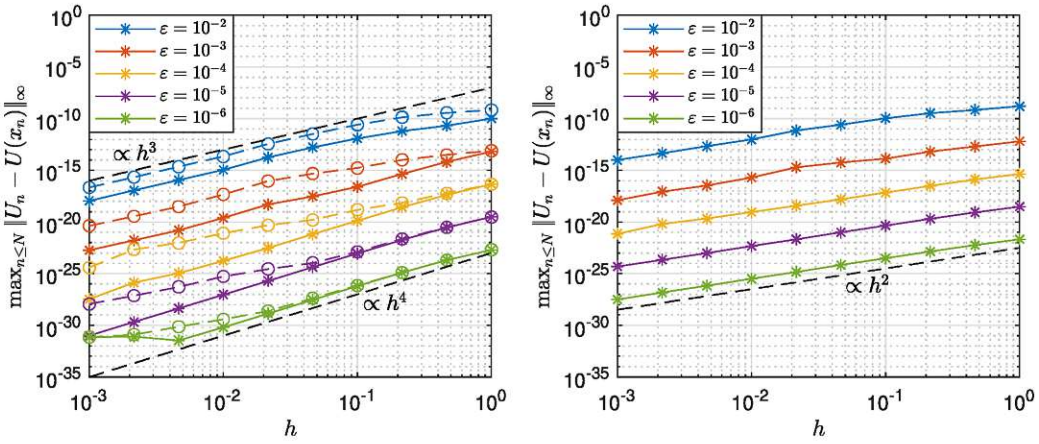

Figure 3.5.3: Global errors $\max_{n\le N}\|U(x_n) - U_n\|_\infty$ for the Airy equation (3.5.1) on the spatial interval $[1, 2]$ as a function of the step size h, for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

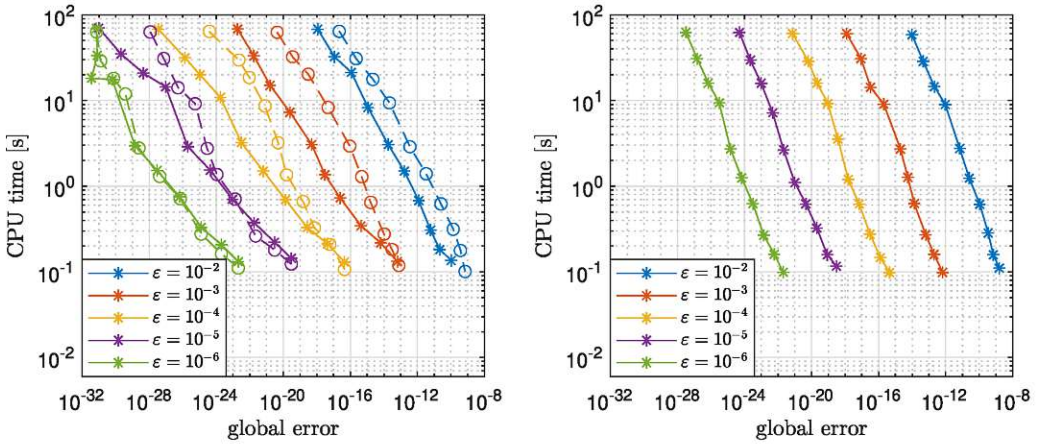

Figure 3.5.4: CPU times vs. global errors $\max_{n\le N}\|U(x_n) - U_n\|_\infty$ for the Airy equation (3.5.1) on the spatial interval $[1, 2]$ as a function of the step size h, for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

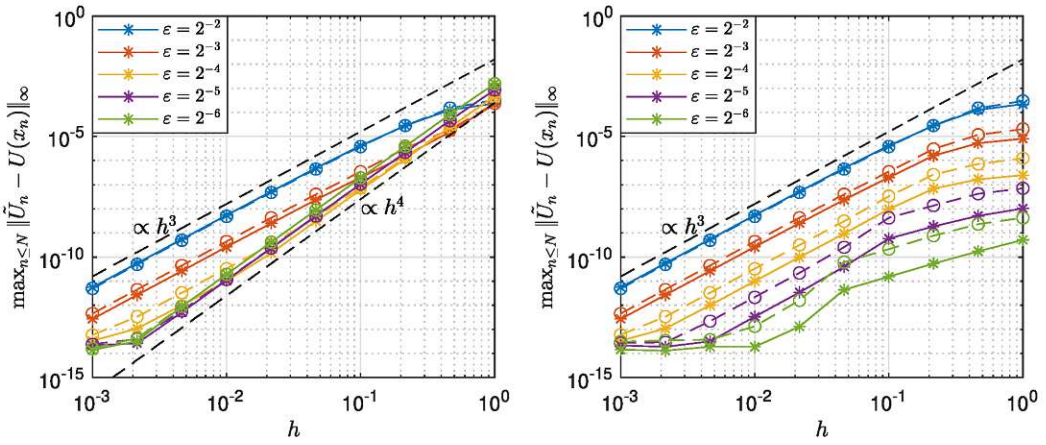

Figure 3.5.5: Global errors $\max_{n\le N}\|U(x_n) - \tilde{U}_n\|_\infty$ for the Airy equation (3.5.1) on the spatial interval $[1,2]$ as a function of the step size h, for several ε-values. The solid lines with asterisks correspond to WKB3, and the dashed lines with circles correspond to WKB3s. Here, we used an approximate phase $\tilde{\phi}$ computed with two different methods: Left: Composite Simpson's rule. Right: Clenshaw-Curtis algorithm based on $N_{cheb} = 17$ Chebyshev grid points, with barycentric interpolation.

rather small global errors (incurred by small step sizes h), WKB3 needs significantly less CPU time than WKB3s. This is due to the $\mathcal{O}(h^4)$ behavior of the error for WKB3 in these cases. Vice versa, for a fixed accuracy, the difference between the needed CPU times for WKB3, WKB3s, and WKB2 is very large. For instance, for $\varepsilon = 10^{-4}$ and a global error of approximately 10^{-21}, the CPU times for WKB3, WKB3s, and WKB2 are approximately 1.5 seconds, 8 seconds, and 60 seconds, respectively. Hence, for the same accuracy, WKB3 is faster up to a factor of 40, when compared to WKB2.

Next, let us investigate numerically the estimates (3.4.14), (3.4.25). To this end, we compute in the following the phase (3.2.7) numerically (even though it is exactly computable for this example) and investigate results obtained with WKB3 and WKB3s. In Figure 3.5.5 we show on the left the global error of WKB3 and WKB3s when using the composite Simpson rule to compute the approximate phase $\tilde{\phi}$. Note that this implies $\|\phi - \tilde{\phi}\|_{L^\infty(I)} = \mathcal{O}(h^4)$. Indeed, for $\varepsilon = 2^{-4}, 2^{-5}, 2^{-6}$ and $h \ge 10^{-2}$ the error curves for WKB3 and WKB3s behave like the first error term in (3.4.14) and (3.4.25), respectively, i.e. like $\mathcal{O}_{\varepsilon,h}(h^4/\varepsilon)$ due to Simpson's rule, as indicated by the lower dashed black line. For $h = 1$ and $\varepsilon \le 2^{-3}$ we even observe an inversion of all error curves for WKB3 and WKB3s, i.e., the error increases with ε. However, for small step sizes and large ε (e.g., $h \le 10^{-1}$ and $\varepsilon = 2^{-3}$) the second error term in the estimates (3.4.14), (3.4.25) becomes dominant. Indeed, as indicated by the upper dashed black line, the error curves for both methods are clearly third order. Further, in this case the inversion of the error curves w.r.t. ε disappears. Moreover, for small values of ε and h (e.g., $\varepsilon \le 2^{-4}$ and $h = 10^{-3}$) the error curves reach a saturation level at approximately 10^{-14}-10^{-13}, due to rounding errors.

For the right plot of Figure 3.5.5 we computed the approximate phase $\tilde{\phi}$ with a spectral method, as already used in [AKU22]. That is, we use the well-known Clenshaw-Curtis algorithm [CC60] to approximate the phase (3.2.7) on a Chebyshev grid for the whole interval $[1,2]$ with $N_{cheb} = 17$ points, and then use barycentric interpolation (e.g., see [BT04]) to obtain the approximate phase $\tilde{\phi}$ at the uniform grid which is used for the WKB

schemes. We note that, using this method, the phase is approximated to machine precision. Indeed, we observe in the right plot of Figure 3.5.5 that the first error term in the estimates (3.4.14), (3.4.25) is numerically invisible, as the entire plot is almost indistinguishable from the left plot of Figure 3.5.1, where the exact phase was used.

3.5.2 Second example

As a second example let us consider the following IVP:

$$\begin{cases} \varepsilon^2 \varphi''(x) + \mathrm{e}^x \varphi(x) = 0\,, \quad x \in [0,1]\,, \\ \varphi(0) = 1\,, \\ \varepsilon \varphi'(0) = 0\,. \end{cases} \tag{3.5.2}$$

Here, the exact solution is given by[3]

$$\varphi_{exact}(x) = \frac{J_0\left(\frac{2}{\varepsilon}\,\mathrm{e}^{x/2}\right) Y_1\left(\frac{2\sqrt{\mathrm{e}}}{\varepsilon}\right) - Y_0\left(\frac{2}{\varepsilon}\,\mathrm{e}^{x/2}\right) J_1\left(\frac{2\sqrt{\mathrm{e}}}{\varepsilon}\right)}{Y_1\left(\frac{2\sqrt{\mathrm{e}}}{\varepsilon}\right) J_0\left(\frac{2}{\varepsilon}\right) - J_1\left(\frac{2\sqrt{\mathrm{e}}}{\varepsilon}\right) Y_0\left(\frac{2}{\varepsilon}\right)}\,, \tag{3.5.3}$$

where J_ν and Y_ν denote the Bessel functions of first and second kind of order ν, respectively (e.g., see [OLBC10, Chap. 10]). Again the phase (3.2.7) is exactly computable here.

In Figure 3.5.6 we plot again the global error $\max_{n \le N} \|U(x_n) - U_n\|_\infty$ as a function of the step size h, for several ε-values. The left plot shows the results for WKB3 and WKB3s, and the right plot shows the results for WKB2. The dashed black line in the left plot confirms the third order estimates (3.4.5), (3.4.23), as the error for WKB3 and WKB3s clearly decays like $\mathcal{O}(h^3)$. For small values of h and small values of ε (e.g. $h \le 10^{-2}$ and $\varepsilon = 2^{-5}$) the error curves include rounding errors, due to the used double precision arithmetic. Moreover, when using the step size $h = 1$, the error for WKB3 is $\mathcal{O}(\varepsilon^4)$, whereas the error for WKB3s is $\mathcal{O}(\varepsilon^3)$. This is again due to the $\mathcal{O}_{\varepsilon,h}(\max(\varepsilon, h))$-factor in estimate (3.4.5), which for the shown ε-values yields an $\mathcal{O}(\varepsilon)$-factor.

Figure 3.5.7 shows work-precision diagrams, which correspond to the computations from Figure 3.5.6. We observe that, in the same amount of time, WKB3 and WKB3s both yield much smaller global errors than WKB2. E.g., with a CPU time of approximately 10^{-3}-$1.5 \cdot 10^{-3}$ seconds, the global errors for WKB3 and WKB3s are between 10^{-14} and $5 \cdot 10^{-11}$ for the different ε-values. In contrast, the errors for WKB2 lie between 10^{-13} and $5 \cdot 10^{-9}$. Vice versa, in order to attain a fixed accuracy, WKB3 and WKB3s need significantly less CPU time, when compared to WKB2. This can be observed, e.g., for $\varepsilon = 2^{-2}$ and an accuracy of approximately 10^{-10}: Then, WKB3 and WKB3s both need less than 10^{-3} seconds, whereas WKB2 needs approximately $5 \cdot 10^{-2}$ seconds. This time difference is even larger for smaller values of ε. We conclude from Figure 3.5.7 that WKB3 and WKB3s are more efficient than WKB2 (in the present example). Similar to the previous example, the difference between WKB3 and WKB3s is not so obvious here. Indeed, we observe on the left of Figure 3.5.7 that the blue curves for $\varepsilon = 2^{-2}$ are almost identical. But for $\varepsilon = 2^{-4}$ and a global error of approximately $3 \cdot 10^{-12}$ the CPU time for WKB3 is $4 \cdot 10^{-4}$, whereas

[3] We computed the exact solution by using the Symbolic Math Toolbox of MATLAB.

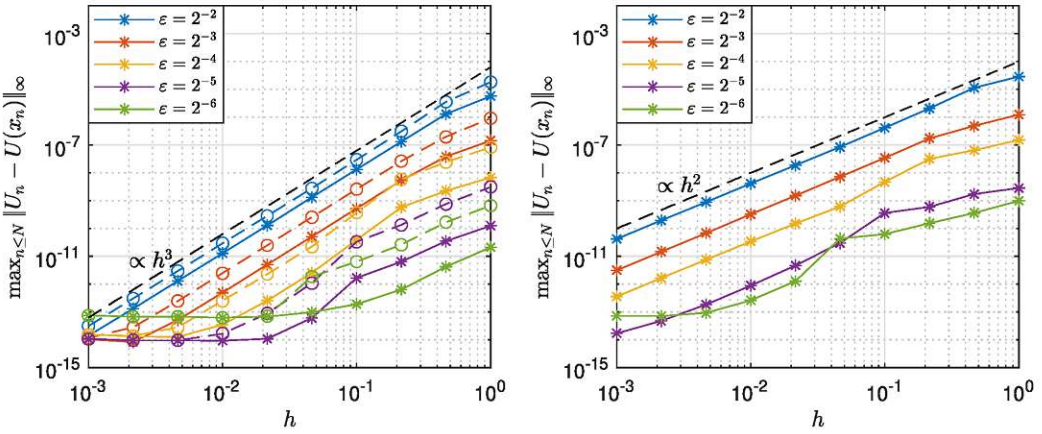

Figure 3.5.6: Global errors $\max_{n\leq N}\|U(x_n) - U_n\|_\infty$ for the IVP (3.5.2) as functions of the step size h, for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

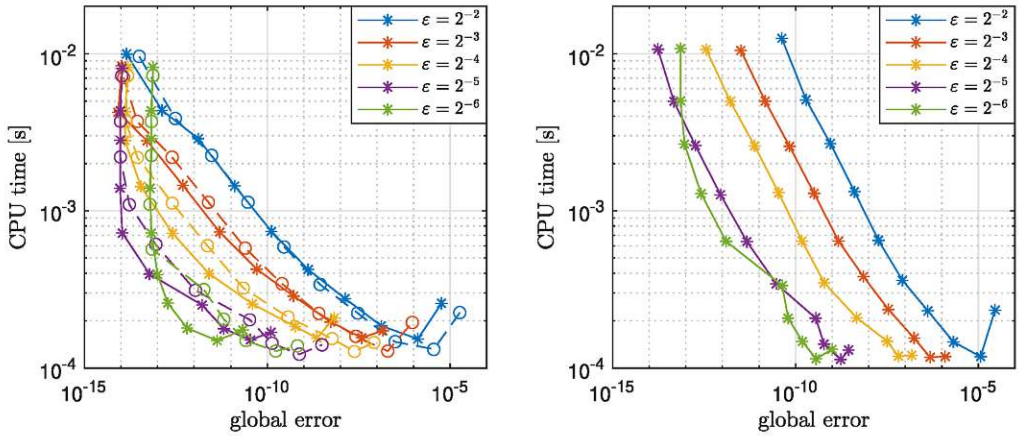

Figure 3.5.7: CPU times vs. global errors $\max_{n\leq N}\|U(x_n) - U_n\|_\infty$ for the IVP (3.5.2), for several ε-values. Left: WKB3 (solid lines with asterisks) and WKB3s (dashed lines with circles). Right: WKB2.

the CPU time for WKB3s is $6 \cdot 10^{-4}$. Overall, we conclude that the efficiency difference between WKB3 and WKB3s is small for this example. However, for a given accuracy and small ε-values, WKB3 tends to be faster than WKB3s.

Next, we compute for the schemes WKB3 and WKB3s the phase (3.2.7) numerically in two ways and investigate the estimates (3.4.14), (3.4.25). For this, we shall again apply the composite Simpson rule as well as the Clenshaw-Curtis algorithm along with barycentric interpolation. In Figure 3.5.8 we show on the left the global error $\max_{n\leq N}\|U(x_n) - \tilde{U}_n\|_\infty$ for WKB3 and WKB3s as a function of the step size when using Simpson's rule. We observe that for both methods the first error term in (3.4.14), (3.4.25), i.e. the $\mathcal{O}_{\varepsilon,h}(h^4/\varepsilon)$-term, dominates for $\varepsilon \leq 2^{-3}$ and all used step sizes h. Indeed, this is indicated by the dashed black line which decays like $\mathcal{O}(h^4)$. For $\varepsilon = 2^{-2}$, however, this error term merely dominates for step sizes $h \geq 10^{-1}$. Indeed, when using smaller step sizes, i.e. $h \leq 10^{-1}$, the error for $\varepsilon = 2^{-2}$ behaves like $\mathcal{O}(h^3)$ for both methods. Note also that for $h = 1$ we can observe again an inversion of the shown error curves w.r.t. ε, when compared to the left plot of

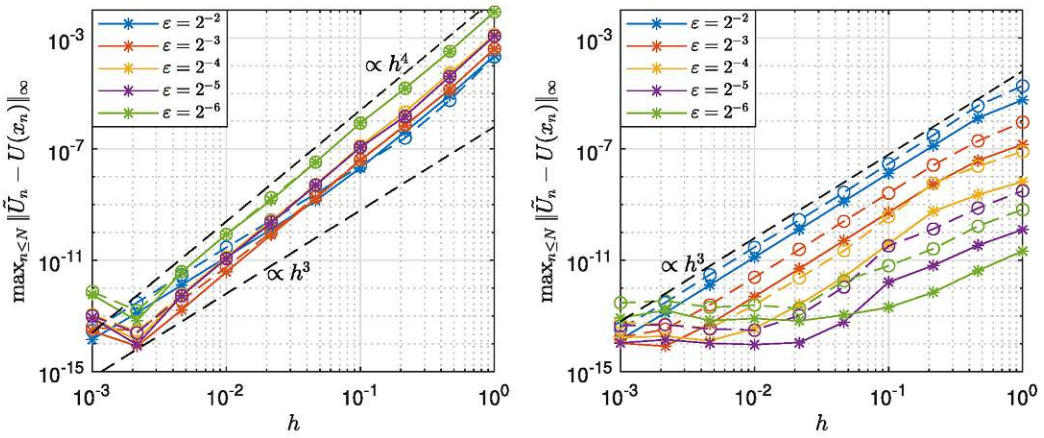

Figure 3.5.8: Global errors $\max_{n\leq N}\|U(x_n) - \tilde{U}_n\|_\infty$ for the IVP (3.5.2) as a function of the step size h, for several ε-values. The solid lines with asterisks correspond to WKB3, and the dashed lines with circles correspond to WKB3s. Here, we used an approximate phase $\tilde{\phi}$ computed with two different methods: Left: Composite Simpson's rule. Right: Clenshaw-Curtis algorithm based on $N_{cheb} = 17$ Chebyshev grid points, with barycentric interpolation.

Figure 3.5.6.

On the right of Figure 3.5.8 we show the errors for WKB3 and WKB3s when the phase is obtained with the spectral method. As indicated by the dashed black line, the first error term in (3.4.14), (3.4.25) is essentially eliminated. The error curves for both methods are clearly third order and the entire plot is very similar to the left plot of Figure 3.5.6.

3.6 Conclusion

In this chapter we developed a third order one-step method to efficiently compute solutions to the highly oscillatory 1D Schrödinger equation (3.1.1). The method is based on the WKB-transformation from [AAN11], which was already used there for the development of a first and second order scheme. Like those two methods, the presented method has the property of asymptotical correctness w.r.t. the small parameter ε in case of an explicitly computable phase. Additionally, in scenarios where $\varepsilon \ll h$ (with h being the grid size), the error even decays with fourth order w.r.t. h.

Appendix

3.A Proof of Lemma 3.4.4

Proof. We will follow Step 2 and Step 3 from Subsection 3.4.2 in order to derive appropriate approximations for the matrices $\mathbf{M}_2^\varepsilon$ and $\mathbf{M}_3^\varepsilon$, respectively. Similar to Subsection 3.3.2 and Subsection 3.3.3, we shall study here only the entries $m_2^\varepsilon(\eta,\xi) = (\mathbf{M}_2^\varepsilon(\eta;\xi))_{1,1}$ and $m_3^\varepsilon(\eta,\xi) = (\mathbf{M}_3^\varepsilon(\eta;\xi))_{2,1}$ in order to obtain the quadratures (3.4.15)-(3.4.16), respectively.

Step 2 (approximation of $\mathbf{M}_2^\varepsilon$ in (3.3.3)): To approximate $\mathbf{M}_2^\varepsilon(\eta;\xi)$, we insert $\mathbf{Q}_1^{1,2}(y,\xi)$ (see Lemma 3.3.2) as an approximation of $\mathbf{M}_1^\varepsilon(y,\xi)$ in (3.3.10):

$$
\begin{aligned}
m_2^\varepsilon(\eta,\xi) &= \int_\xi^\eta (\mathbf{N}^\varepsilon(y))_{1,2}\,(\mathbf{M}_1^\varepsilon(y;\xi))_{2,1}\,\mathrm{d}y \\
&= \int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\left\{Q_1^{1,2}(y,\xi) + \mathcal{O}(\varepsilon h^2\min(\varepsilon,h))\right\}\mathrm{d}y \\
&= \int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\Bigg\{ -(\mathrm{i}\varepsilon)\left(b_0(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)} - b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\right) \\
&\qquad\qquad -(\mathrm{i}\varepsilon)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_1(y)h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) \\
&\qquad\qquad -(\mathrm{i}\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_2(y)h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right) \\
&\qquad\qquad + \mathcal{O}_{\varepsilon,h}(\varepsilon h^2\min(\varepsilon,h))\Bigg\}\mathrm{d}y \\
&= D_1 + D_2 + D_3 + D_4 + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,, \qquad (3.A.1)
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 &:= -(\mathrm{i}\varepsilon)\int_\xi^\eta b(y)b_0(y)\,\mathrm{d}y = -(\mathrm{i}\varepsilon)Q_S[bb_0](\eta,\xi) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^5)\,, \qquad (3.A.2)\\
D_2 &:= (\mathrm{i}\varepsilon)b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,\mathrm{d}y = (\mathrm{i}\varepsilon)b_0(\xi)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,\overline{m_1^\varepsilon(\eta,\xi)}\,,\\
D_3 &:= -(\mathrm{i}\varepsilon)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_1(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y\,,\\
D_4 &:= -(\mathrm{i}\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_2(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y\,.
\end{aligned}
$$

Our next goal is to find approximations for the integrals D_2, D_3, and D_4, which incur errors of the order $\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))$. Using Lemma 3.3.2 with $P=0$ and $\widetilde{P}=3$ we obtain by

using $\overline{h_p(x)} = h_p(-x)$:

$$D_2 = (\mathrm{i}\,\varepsilon)^2 b_0(\xi)\left\{ b_0(\eta)h_1\left(-\frac{2}{\varepsilon}s_n\right) - (\mathrm{i}\,\varepsilon)b_1(\eta)h_2\left(-\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\,\varepsilon)^2 b_2(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\right\} + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3 \min(\varepsilon,h)). \tag{3.A.3}$$

For the approximation of the integral D_3, we first use the identity (3.3.17) and make then two SAM-steps to obtain

$$D_3 = (\mathrm{i}\,\varepsilon)^3\left\{ b_0(\eta)b_1(\eta)h_2\left(-\frac{2}{\varepsilon}s_n\right) - (\mathrm{i}\,\varepsilon)\left(b_1(\eta)^2 + b_0(\eta)b_2(\eta)\right)h_3\left(-\frac{2}{\varepsilon}s_n\right)\right\} + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3 \min(\varepsilon,h))\,. \tag{3.A.4}$$

Next, we use (3.3.19) to split the integral D_4 into $D_4 = \widetilde{D}_4 + \widehat{D}_4$, with

$$\begin{aligned}\widetilde{D}_4 &:= (\mathrm{i}\,\varepsilon)^3 \int_\xi^\eta b(y)b_2(y)h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y \\ &= (\mathrm{i}\,\varepsilon)^4 b_0(\eta)b_2(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,,\end{aligned} \tag{3.A.5}$$

and

$$\begin{aligned}\widehat{D}_4 &:= 2(\mathrm{i}\,\varepsilon)^2 \int_\xi^\eta b(y)b_2(y)\left(\phi(\xi)-\phi(y)\right)h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y \\ &= 2(\mathrm{i}\,\varepsilon)^3\left\{ b_0(\eta)b_2(\eta)(-s_n)h_2\left(-\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\,\varepsilon)\frac{1}{2}b_0(\eta)b_2(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right)\right\} \\ &\quad + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,.\end{aligned} \tag{3.A.6}$$

Here, we made one SAM-step for $\widetilde{D}_4$ and treated $\widehat{D}_4$ similar to $\widehat{K}_4$ in Subsection 3.3.3. Combining (3.A.5)-(3.A.6) thus yields

$$D_4 = 2(\mathrm{i}\,\varepsilon)^3 b_0(\eta)b_2(\eta)(-s_n)h_2\left(-\frac{2}{\varepsilon}s_n\right) + 2(\mathrm{i}\,\varepsilon)^4 b_0(\eta)b_2(\eta)h_3\left(-\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h))\,. \tag{3.A.7}$$

Since $Q_{2,simp}$ in (3.4.15) is defined precisely through the approximations (3.A.1)-(3.A.4), and (3.A.7) by neglecting the $\mathcal{O}_{\varepsilon,h}(\cdot)$-terms, estimate (3.4.19) follows by noticing that $\mathcal{O}_{\varepsilon,h}(\varepsilon h^3\min(\varepsilon,h)) + \mathcal{O}_{\varepsilon,h}(\varepsilon h^5) = \mathcal{O}_{\varepsilon,h}(\varepsilon h^4)$.

Step 3 (approximation of $\mathbf{M}_3^\varepsilon$ in (3.3.3)): To approximate $\mathbf{M}_3^\varepsilon(\eta;\xi)$, we use an approximation $\widetilde{\mathbf{Q}}_{2,simp}$ for $\mathbf{M}_2^\varepsilon$ in (3.3.28), which can be derived by inserting $\mathbf{Q}_1^{0,1}(y,\xi)$ (see

Lemma 3.3.2) as an approximation of $\mathbf{M}_1^\varepsilon(y,\xi)$ in (3.3.10):

$$\begin{aligned}
m_2^\varepsilon(\eta,\xi) &= \int_\xi^\eta (\mathbf{N}^\varepsilon(y))_{1,2}\,(\mathbf{M}_1^\varepsilon(y;\xi))_{2,1}\,\mathrm{d}y \\
&= \int_\xi^\eta b(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\left\{Q_1^{0,1}(y,\xi)+\mathcal{O}_{\varepsilon,h}(h\min(\varepsilon,h))\right\}\mathrm{d}y \\
&= -(\mathrm{i}\,\varepsilon)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_0(y)\,\mathrm{e}^{-\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y \\
&\quad + \mathcal{O}_{\varepsilon,h}(h^2\min(\varepsilon,h)) \\
&= (\mathrm{i}\,\varepsilon)\int_\xi^\eta b(y)b_0(y)h_1\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y + \mathcal{O}_{\varepsilon,h}(h^2\min(\varepsilon,h)) \\
&= (\mathrm{i}\,\varepsilon)^2 b_0(\eta)^2 h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(\eta))\right) + \mathcal{O}_{\varepsilon,h}(h^2\min(\varepsilon,h))\,. \qquad (3.A.8)
\end{aligned}$$

Here, we used the identity (3.3.17) in the fourth equality, and made one SAM-step in the last equality. Thus, by inserting

$$\widetilde{Q}_{2,simp}(\eta,\xi) := (\mathrm{i}\,\varepsilon)^2 b_0(\eta)^2 h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(\eta))\right), \qquad (3.A.9)$$

$$\widetilde{\mathbf{Q}}_{2,simp}(\eta,\xi) := \begin{pmatrix} \widetilde{Q}_{2,simp}(\eta,\xi) & 0 \\ 0 & \overline{\widetilde{Q}_{2,simp}(\eta,\xi)} \end{pmatrix}, \qquad (3.A.10)$$

as an approximation for $\mathbf{M}_2^\varepsilon$ in (3.3.28), incurring an error of $\mathcal{O}_{\varepsilon,h}(h^2\min(\varepsilon,h))$ (see (3.A.8)), we obtain

$$\begin{aligned}
m_3^\varepsilon(\eta,\xi) &= \int_\xi^\eta (\mathbf{N}^\varepsilon(y))_{2,1}\,(\mathbf{M}_2^\varepsilon(y;\xi))_{1,1}\,\mathrm{d}y \\
&= \int_\xi^\eta b(y)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\left\{\widetilde{Q}_{2,simp}(y,\xi)+\mathcal{O}_{\varepsilon,h}(h^2\min(\varepsilon,h))\right\}\mathrm{d}y \\
&= (\mathrm{i}\,\varepsilon)^2\int_\xi^\eta b(y)b_0(y)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(y)}\,h_2\left(\frac{2}{\varepsilon}(\phi(\xi)-\phi(y))\right)\mathrm{d}y \\
&\quad + \mathcal{O}_{\varepsilon,h}(h^3\min(\varepsilon,h))\,. \qquad (3.A.11)
\end{aligned}$$

Let us abbreviate the integral on the r.h.s. of (3.A.11) with F_1. By employing the identity (3.3.38), we split this integral into $F_1 = \widetilde{F}_1 + \widehat{F}_1$, with

$$\begin{aligned}
\widetilde{F}_1 &:= -(\mathrm{i}\,\varepsilon)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\int_\xi^\eta b(y)b_0(y)^2h_2\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\mathrm{d}y \\
&= (\mathrm{i}\,\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_0(\eta)^3h_3\left(\frac{2}{\varepsilon}s_n\right) + \mathcal{O}_{\varepsilon,h}(h^3\min(\varepsilon,h))\,, \qquad (3.A.12)
\end{aligned}$$

and

$$\begin{aligned}\widehat{F}_1 &:= -2(\mathrm{i}\,\varepsilon)\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)} \int_\xi^\eta b(y)b_0(y)^2(\phi(y)-\phi(\xi))h_1\left(\frac{2}{\varepsilon}(\phi(y)-\phi(\xi))\right)\,\mathrm{d}y \\ &= 2(\mathrm{i}\,\varepsilon)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\left\{b_0(\eta)^3 s_n h_2\left(\frac{2}{\varepsilon}s_n\right) + (\mathrm{i}\,\varepsilon)\frac{1}{2}b_0(\eta)^3 h_3\left(\frac{2}{\varepsilon}s_n\right)\right\} \\ &\quad + \mathcal{O}_{\varepsilon,h}(h^3\min(\varepsilon,h))\,. \end{aligned} \tag{3.A.13}$$

Here, we made one SAM-step for $\widetilde{F}_1$ and treated $\widehat{F}_1$ similar to $\widehat{K}_4$ in Subsection 3.3.3. Combining (3.A.11)-(3.A.13) thus yields

$$\begin{aligned} m_3^\varepsilon(\eta,\xi) &= 2(\mathrm{i}\,\varepsilon)^2\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_0(\eta)^3 s_n h_2\left(\frac{2}{\varepsilon}s_n\right) + 2(\mathrm{i}\,\varepsilon)^3\,\mathrm{e}^{\frac{2\mathrm{i}}{\varepsilon}\phi(\xi)}\,b_0(\eta)^3 h_3\left(\frac{2}{\varepsilon}s_n\right) \\ &\quad + \mathcal{O}_{\varepsilon,h}(h^3\min(\varepsilon,h)) \\ &= Q_{3,simp}(\eta,\xi) + \mathcal{O}_{\varepsilon,h}(h^3\min(\varepsilon,h))\,, \end{aligned} \tag{3.A.14}$$

with $Q_{3,simp}(\eta,\xi)$ defined in (3.4.16). This proves estimate (3.4.20) and concludes the proof. □

Bibliography

[AAN11] A. Arnold, N. B. Abdallah, and C. Negulescu. WKB-Based Schemes for the Oscillatory 1D Schrödinger Equation in the Semiclassical Limit. *SIAM J. Numer. Anal.*, 49:1436–1460, 2011.

[Adv23] Advanpix LLC. *Multiprecision Computing Toolbox for* MATLAB *version 5.1.0.15432.* Advanpix LLC, 2023.

[AK24] A. Arnold and J. Körner. WKB-based third order method for the highly oscillatory 1D stationary Schrödinger equation. *arXiv*, abs/2402.18406, 2024.

[AKKM24] A. Arnold, C. Klein, J. Körner, and J. M. Melenk. Optimally truncated WKB approximation for the 1D stationary Schrödinger equation in the highly oscillatory regime. *arXiv*, abs/2401.10141, 2024.

[AKU22] A. Arnold, C. Klein, and B. Ujvari. WKB-method for the 1D Schrödinger equation in the semi-classical limit: enhanced phase treatment. *BIT Numerical Mathematics*, 62:1–22, 2022.

[BT04] J.-P. Berrut and L. N. Trefethen. Barycentric lagrange interpolation. *SIAM Rev.*, 46:501–517, 2004.

[CC60] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205, 1960.

[CS58] E. D. Courant and H. S. Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 281:360–408, 1958.

[IB95] F. Ihlenburg and I. Babuška. Finite element solution of the Helmholtz equation with high wave number Part I: The h-version of the FEM. *Computers and Mathematics with Applications*, 30(9):9–37, 1995.

[INO06] A. Iserles, S. P. Nørsett, and S. Olver. Highly oscillatory quadrature: the story so far. *Proceedings of ENUMATH 2005, Santiago de Compostela*, pages 97–118, 2006.

[JL03] T. Jahnke and C. Lubich. Numerical integrators for quantum dynamics close to the adiabatic limit. *Numerische Mathematik*, 94:289–314, 2003.

[KAD21] J. Körner, A. Arnold, and K. Döpfner. WKB-based scheme with adaptive step size control for the Schrödinger equation in the highly oscillatory regime. *J. Comput. Appl. Math.*, 404:113905, 2021.

[Lew68] H. R. Lewis. Motion of a Time-Dependent Harmonic Oscillator, and of a Charged Particle in a Class of Time-Dependent, Axially Symmetric Electromagnetic Fields. *Phys. Rev.*, 172:1313–1315, 1968.

[LJL05] K. Lorenz, T. Jahnke, and C. Lubich. Adiabatic Integrators for Highly Oscillatory Second-Order Linear Differential Equations with Time-Varying Eigendecomposition. *BIT Numerical Mathematics*, 45:91–115, 2005.

[LL85] L. D. Landau and E. M. Lifschitz. *Quantenmechanik.* Akademie-Verlag, 1985.

[MJK13] J.-F. Mennemann, A. Jüngel, and H. Kosina. Transient Schrödinger-Poisson simulations of a high-frequency resonant tunneling diode oscillator. *J. Comput. Phys.*, 239:187–205, 2013.

[MS03] J. Martin and D. J. Schwarz. WKB approximation for inflationary cosmological perturbations. *Physical Review D*, 67:083512, 2003.

[Neg05] C. Negulescu. *Asymptotical models and numerical schemes for quantum systems.* PhD thesis, Université Paul Sabatier, Toulouse, 2005.

[OLBC10] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. NIST Handbook of Mathematical Functions. Cambridge University Press, 2010.

[SHMS98] J. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman. Resonant tunneling diodes: models and properties. *Proc. IEEE*, 86:641–660, 1998.

[Win05] S. Winitzki. Cosmological particle production and the precision of the WKB approximation. *Physical Review D*, 72:104011, 2005.

4 Optimally truncated WKB approximation for the 1D stationary Schrödinger equation in the highly oscillatory regime

The content of this chapter has been submitted for publication under [AKKM24].

4.1 Introduction

In this chapter we are concerned with the numerical solution of the highly oscillatory and stationary 1D Schrödinger equation

$$\begin{cases} \varepsilon^2 \varphi''(x) + a(x)\varphi(x) = 0\,, \quad x \in I := [\xi, \eta]\,, \\ \varphi(\xi) = \varphi_0\,, \\ \varepsilon \varphi'(\xi) = \varphi_1\,. \end{cases} \tag{4.1.1}$$

Here, $0 < \varepsilon \ll 1$ is a very small parameter and a is a real-valued function satisfying $a(x) \geq a_0 > 0$ and, for a quantum mechanical problem, it is related to the potential. The constants $\varphi_0, \varphi_1 \in \mathbb{C}$ may depend on ε but are assumed to be ε-uniformly bounded. It is known that the (local) wave length λ of the solution φ to (4.1.1) is proportional to ε. More precisely, it can be expressed as $\lambda(x) = (2\pi\varepsilon)/\sqrt{a(x)}$. Consequently, for a small parameter ε the solution becomes highly oscillatory, particularly in the semi-classical limit $\varepsilon \to 0$.

Highly oscillatory problems such as (4.1.1) occur across a broad range of applications, e.g., plasma physics [CS58, Lew68], inflationary cosmology [MS03, Win05] and electron transport in semiconductor devices such as resonant tunneling diodes [MJK13, SHMS98, Neg05]. More specifically, the state of an electron of mass m that is injected with the prescribed energy E from the right boundary into an electronic device (e.g., diode), modeled on the interval $[\xi, \eta]$, can be described by the following boundary value problem (BVP) (e.g., see [AAN11] or [Neg05, Chap. 2]):

$$\begin{cases} -\varepsilon^2 \psi_E''(x) + V(x)\psi_E(x) = E\psi_E(x)\,, \quad x \in (\xi, \eta)\,, \\ \psi_E'(\xi) + \mathrm{i}\, k(\xi)\psi(\xi) = 0\,, \\ \psi_E'(\eta) - \mathrm{i}\, k(\eta)\psi(\eta) = -2\,\mathrm{i}\, k(\eta)\,. \end{cases} \tag{4.1.2}$$

Here, $\varepsilon := \hbar/\sqrt{2m}$ is proportional to the (reduced) Planck constant $\hbar$, the quantity $k(x) := \varepsilon^{-1}\sqrt{E - V(x)}$ is the wave vector and the real-valued function V denotes the electrostatic potential. In the context of (4.1.2), our assumption $a(x) \geq a_0 > 0$ simply reads $E > V(x)$, which means that we are in the oscillatory regime. One is then often interested in

macroscopic quantities such as the charge density n and the current density j, which are given by

$$n(x) = \int_0^\infty |\psi_E(x)|^2 f(E)\,\mathrm{d}E\,, \quad j(x) = \varepsilon \int_0^\infty \mathrm{Im}(\overline{\psi_E(x)}\psi_E'(x)) f(E)\,\mathrm{d}E\,. \tag{4.1.3}$$

Here, f is the distribution function which represents the injection statistics of the electron and $\mathrm{Im}(\cdot)$ denotes the imaginary part. Thus, in order to compute the quantities (4.1.3), one has to use a very fine grid in E which means that the BVP (4.1.2) has to be solved many times. Consequently, there exists a substantial demand for efficient numerical methods that are suitable for solving problems like (4.1.2). Further, we note that the BVP (4.1.2) is strongly connected to IVP (4.1.1). Indeed, for suitable initial values, namely, $\varphi_0 = 1$ and $\varphi_1 = -\mathrm{i}\sqrt{a(\xi)}$, the solution φ of IVP (4.1.1) and the solution ψ_E of BVP (4.1.2) are related by

$$\psi_E(x) = -\frac{2\mathrm{i}\,k(\eta)}{\varphi'(\eta) - \mathrm{i}\,k(\eta)\varphi(\eta)}\varphi(x)\,. \tag{4.1.4}$$

Thus, any numerical method for solving IVP (4.1.1) is also suitable for the numerical treatment of BVP (4.1.2).

4.1.1 Background and approach

Since the solution φ to (4.1.1) exhibits rapid oscillations when ε is small, standard numerical methods for ODEs become inefficient as they are typically constrained by grid limitations $h = \mathcal{O}(\varepsilon)$ (h denoting the step size), in order to resolve the oscillations accurately. By contrast, in [LJL05, JL03] *uniformly accurate* (w.r.t. ε) marching methods were proposed which yield global errors of order $\mathcal{O}(h^2)$ and allow to reduce the grid limitation to at least $h = \mathcal{O}(\sqrt{\varepsilon})$. The WKB-based (named after the physicists Wentzel, Kramers, Brillouin; cf. [LL85]) one-step method from [AAN11] is even *asymptotically correct*, i.e. the numerical error goes to zero with $\varepsilon \to 0$, provided that the integrals $\int^x \sqrt{a(\tau)}\,\mathrm{d}\tau$ and $\int^x a(\tau)^{-1/4}(a(\tau)^{-1/4})''\,\mathrm{d}\tau$ for the phase of the solution can be computed exactly. More precisely, the method then yields an error which is of order $\mathcal{O}(\varepsilon^3)$ as $\varepsilon \to 0$ and $\mathcal{O}(h^2)$ as $h \to 0$. If these integrals cannot be evaluated exactly, the asymptotically correct error behavior can be (almost) recovered by employing spectral methods for the integrals, as shown in [AKU22]. Further, in [AB22] the authors propose a numerical algorithm, which switches adaptively between a defect correction iteration (which builds on an asymptotic expansion) for oscillatory regions of the solution, and a conventional Chebyshev collocation solver for smoother regions. Although the method is demonstrated to be highly accurate and efficient, a full error analysis was left for future work.

Our approach here is to implement directly a WKB approximation for the solution of (4.1.1), which is asymptotically correct and of arbitrary order w.r.t. ε. The essence of the method is rather an analytic approximation via an asymptotic WKB series with optimal truncation. As such, the main goal is to understand the asymptotic ε-dependence of this truncation strategy and of the resulting error. Thus our strategy is not a classical numerical method with some chosen grid size h and convergence as $h \to 0$. Instead, the resulting approximation error will be of order $\mathcal{O}(\varepsilon^N)$ as $\varepsilon \to 0$, where N refers to the used truncation

order in the underlying asymptotic WKB series, see (4.2.1)-(4.2.2) below. As N can be chosen freely, this approach may prove very practical for applications, especially when the model parameter ε is very small. Since the computation of the terms of the asymptotic series involves several integrals, we will employ highly accurate spectral methods, as already proven useful in [AKU22].

The key question when implementing this WKB approximation is which choice of N is adequate or even optimal, in the sense of minimizing the resulting approximation error. Indeed, since the asymptotic WKB series is typically divergent, the error cannot simply be reduced further by increasing the value of N. This question about the best attainable accuracy of the WKB approximation was already addressed in [Win05], where the author compared the WKB series with the exact solution represented by a convergent Bremmer series, or more precisely, by an asymptotic expansion of that Bremmer series. The author finds that in cases where the coefficient function a is analytic, the optimal truncation order is proportional to ε^{-1}, yielding a corresponding optimal accuracy which is exponentially small w.r.t. ε. However, to derive these results, the author makes several additional asymptotic approximations. In this chapter, on the other hand, we shall follow a more rigorous strategy by providing error estimates for the WKB approximation which are explicit w.r.t. ε *and* N. We note, however, that the key assumption from [Win05], i.e., a being analytic, will also be crucial for the strategy in this chapter.

4.1.2 Main results

Our first main result is Theorem 4.3.7, which provides an explicit (w.r.t. ε and N) error estimate for the WKB approximation, and implies that the approximation error is of order $\mathcal{O}(\varepsilon^N)$. The explicitness of this estimate then allows the investigation of the error w.r.t. the truncation order N. Indeed, the optimal truncation order N_{opt} can be predicted by minimizing the established upper error bound w.r.t. N or by determining the smallest term of the asymptotic series, and is found to be proportional to ε^{-1}. This leads to our second main result, namely, Corollary 4.4.1. It states that, for an adequate choice of $N = N(\varepsilon) \sim \varepsilon^{-1}$, the error of the WKB approximation is of order $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, $r > 0$ being some constant. As a consequence, also the error of the optimally truncated WKB approximation is of order $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, see also Remark 4.4.2.

This chapter is organized as follows: In Section 4.2 we introduce the N-th order (w.r.t. ε) WKB approximation as an approximate solution of IVP (4.1.1). Section 4.3 then contains a detailed error analysis for the WKB approximation and includes explicit (w.r.t. ε and the truncation order N) error estimates. In Section 4.4 we specify the computation of the WKB approximation. This includes the description of the chosen methods for the computation of the terms of the underlying asymptotic series as well as a reasonable truncation strategy. In Section 4.5 we illustrate the theoretical results established in this chapter by several numerical examples. We conclude in Section 4.6.

4.2 WKB approximation

In this section we introduce the *WKB approximation* as an approximate solution of IVP (4.1.1). The basis for its construction is the well-known WKB-ansatz (cf. [BO99, LL85]),

which for the ODE (4.1.1) reads[1]

$$\varphi(x) \sim \exp\left(\frac{1}{\varepsilon}S(x)\right), \quad \varepsilon \to 0, \tag{4.2.1}$$

where S is a complex-valued function containing information of the phase as well as the amplitude of the solution φ. To derive WKB approximations it is then convenient to express S as an asymptotic expansion[2] w.r.t. the small parameter ε:

$$S(x) \sim \sum_{n=0}^{\infty} \varepsilon^n S_n(x), \quad \varepsilon \to 0; \quad S_n(x) \in \mathbb{C}. \tag{4.2.2}$$

It should be noted that this asymptotic series is typically divergent (as usual for asymptotic series) and must therefore be truncated in order to obtain an approximate solution.

By substituting the ansatz (4.2.1)-(4.2.2) into (4.1.1), one obtains (formally)

$$\left(\sum_{n=0}^{\infty} \varepsilon^n S_n'(x)\right)^2 + \sum_{n=0}^{\infty} \varepsilon^{n+1} S_n''(x) + a(x) = 0. \tag{4.2.3}$$

A comparison of ε-powers then yields the following well-known recurrence relation for the functions S_n':

$$S_0' = \pm \mathrm{i}\sqrt{a}, \tag{4.2.4}$$

$$S_1' = -\frac{S_0''}{2S_0'} = -\frac{a'}{4a} = -\frac{1}{4}(\ln(a))', \tag{4.2.5}$$

$$S_n' = -\frac{1}{2S_0'}\left(\sum_{j=1}^{n-1} S_j' S_{n-j}' + S_{n-1}''\right), \quad n \geq 2. \tag{4.2.6}$$

The computation of each S_n, $n \geq 0$, thus involves one integration constant. Further, the repeated differentiation in (4.2.6) indicates that a WKB approximation relying on $N+1$ terms in the truncated series (4.2.2) requires $a \in C^N(I)$. Moreover, the fact that the r.h.s. of (4.2.4) has two different signs implies that there are two sequences of functions, which solve (4.2.4)-(4.2.6). This corresponds to the fact that there are two fundamental solutions of the ODE in (4.1.1). Let us denote by $(S_n^-)_{n\in\mathbb{N}_0}$ the sequence induced by the choice $S_0' = -\mathrm{i}\sqrt{a}$. The one following from $S_0' = \mathrm{i}\sqrt{a}$ will be denoted by $(S_n^+)_{n\in\mathbb{N}_0}$. Then, a simple observation is the following proposition.

[1] We say that two functions $f, g : I \times (0, \varepsilon_0) \to \mathbb{C}$ are *asymptotically equivalent* as $\varepsilon \to 0$, if and only if for any $x \in I$ it holds $f(x,\varepsilon) - g(x,\varepsilon) = o(g(x,\varepsilon))$ as $\varepsilon \to 0$. In this case we write $f(x,\varepsilon) \sim g(x,\varepsilon)$, $\varepsilon \to 0$.

[2] We say that a function $f : I \times (0, \varepsilon_0) \to \mathbb{C}$ has an *asymptotic expansion* as $\varepsilon \to 0$, if and only if there exist sequences of functions $(f_n : I \to \mathbb{C})_{n\in\mathbb{N}_0}$ and $(\phi_n : (0,\varepsilon_0) \to \mathbb{C})_{n\in\mathbb{N}_0}$ satisfying for all $n \in \mathbb{N}_0$ and $x \in I$ that $\phi_{n+1}(\varepsilon) f_{n+1}(x) = o(\phi_n(\varepsilon) f_n(x))$ as $\varepsilon \to 0$, such that for all $N \geq 0$ it holds $f(x,\varepsilon) - \sum_{n=0}^{N} \phi_n(\varepsilon) f_n(x) = o(\phi_N(\varepsilon) f_N(x))$ as $\varepsilon \to 0$. In this case we write $f(x,\varepsilon) \sim \sum_{n=0}^{\infty} \phi_n(\varepsilon) f_n(x)$, $\varepsilon \to 0$. We call an asymptotic expansion *uniform* w.r.t. $x \in I$, if all the order symbols hold uniformly in $x \in I$.

Proposition 4.2.1.

$$(S_{2n}^-)'(x) = -(S_{2n}^+)'(x) \in \mathrm{i}\,\mathbb{R}\,, \tag{4.2.7}$$
$$(S_{2n+1}^-)'(x) = (S_{2n+1}^+)'(x) \in \mathbb{R}\,, \tag{4.2.8}$$

for all $x \in I$ and $n \geq 0$.

Proof. The statement can easily be verified by induction on $n \in \mathbb{N}_0$. □

Since both sequences $(S_n^\pm)_{n\in\mathbb{N}_0}$ lead to an approximate solution of the ODE in (4.1.1), the general approximate solution can be written as the linear combination

$$\varphi \approx \varphi_N^{WKB} := \alpha_{N,\varepsilon} \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^-\right) + \beta_{N,\varepsilon} \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^+\right)\,, \tag{4.2.9}$$

with arbitrary $\alpha_{N,\varepsilon}, \beta_{N,\varepsilon} \in \mathbb{C}$. Note that all integration constants in the computation of S_n^- and S_n^+ can be "absorbed" into $\alpha_{N,\varepsilon}$ and $\beta_{N,\varepsilon}$, respectively. Hence, these integration constants can be set to zero without loss of generality. More precisely, we define

$$S_n^\pm(x) := \int_\xi^x \left(S_n^\pm\right)'(\tau)\,\mathrm{d}\tau\,. \tag{4.2.10}$$

With this, Proposition 4.2.1 implies

$$S_{2n}^-(x) = -S_{2n}^+(x) \in \mathrm{i}\,\mathbb{R}\,, \tag{4.2.11}$$
$$S_{2n+1}^-(x) = S_{2n+1}^+(x) \in \mathbb{R}\,, \tag{4.2.12}$$

for all $x \in I$ and $n \geq 0$. Hence, functions with even indices only contribute to the phase of the WKB approximation φ_N^{WKB}, whereas functions with odd indices only provide corrections to the amplitude.

Note that in general the constants $\alpha_{N,\varepsilon}$ and $\beta_{N,\varepsilon}$ can be uniquely determined by initial or boundary conditions. Here, for the WKB approximation (4.2.9) to satisfy the initial conditions in (4.1.1), we set

$$\alpha_{N,\varepsilon} = \frac{\varphi_0 \left(\sum_{n=0}^N \varepsilon^n (S_n^+)'(\xi)\right) - \varphi_1}{\sum_{n=0}^N \varepsilon^n \left((S_n^+)'(\xi) - (S_n^-)'(\xi)\right)}\,, \tag{4.2.13}$$

$$\beta_{N,\varepsilon} = \frac{\varphi_1 - \varphi_0 \left(\sum_{n=0}^N \varepsilon^n (S_n^-)'(\xi)\right)}{\sum_{n=0}^N \varepsilon^n \left((S_n^+)'(\xi) - (S_n^-)'(\xi)\right)}\,. \tag{4.2.14}$$

In the following we will often simply write S_n whenever one could insert either S_n^- or S_n^+.

According to [BO99, Sec. 10.2], for the WKB-ansatz (4.2.1)-(4.2.2) to be valid on the whole interval I, it is necessary that the series $\sum_{n=0}^\infty \varepsilon^{n-1} S_n(x)$ is a uniform asymptotic expansion of $\varepsilon^{-1} S(x)$ as $\varepsilon \to 0$. This implies that for any $n \in \mathbb{N}_0$ the relation

$$\varepsilon^n S_{n+1}(x) = o(\varepsilon^{n-1} S_n(x))\,, \quad \varepsilon \to 0\,, \tag{4.2.15}$$

must hold uniformly in $x \in I$. Note that this condition is violated if the interval I includes so-called *turning points*, i.e., points $x_0 \in I$ with $a(x_0) = 0$. Indeed, this is already evident from (4.2.5), which implies that S_1 blows up at such turning points.

4.3 Error analysis

In this section we aim to find an explicit (w.r.t. ε and the truncation order N) error estimate for the WKB approximation (4.2.9). One key ingredient will be the following a priori estimate for the solution φ of the inhomogeneous analog of the Schrödinger equation-IVP (4.1.1).

Proposition 4.3.1. *Let $a \in W^{1,\infty}(I)$ with $a(x) \geq a_0 > 0$ and $f \in C(I)$. Further, let $\varphi \in C^2(I)$ be the solution of the inhomogeneous IVP*

$$\begin{cases} \varepsilon^2 \varphi'' + a(x)\varphi = f(x)\,, & x \in I\,, \\ \varphi(\xi) = \hat{\varphi}_0\,, \\ \varepsilon\varphi'(\xi) = \hat{\varphi}_1\,, \end{cases}$$

with constants $\hat{\varphi}_0, \hat{\varphi}_1 \in \mathbb{C}$. Then there exists $C > 0$ independent of ε such that

$$\|\varphi\|_{L^\infty(I)} \leq \frac{C}{\varepsilon}\|f\|_{L^2(I)} + C\left(|\hat{\varphi}_1| + |\hat{\varphi}_0|\right)\,, \tag{4.3.1}$$

$$\|\varepsilon\varphi'\|_{L^\infty(I)} \leq \frac{C}{\varepsilon}\|f\|_{L^2(I)} + C\left(|\hat{\varphi}_1| + |\hat{\varphi}_0|\right)\,. \tag{4.3.2}$$

Proof. Estimates (4.3.1)-(4.3.2) can be derived by finding an upper bound for the real-valued function $E(x) := \varepsilon^2|\varphi'|^2 + a|\varphi|^2$. At first, it holds that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}E(x) &= \varepsilon^2\frac{\mathrm{d}}{\mathrm{d}x}|\varphi'|^2 + a\frac{\mathrm{d}}{\mathrm{d}x}|\varphi|^2 + a'|\varphi|^2 \\ &= 2\,\mathrm{Re}((\varepsilon^2\varphi'' + a\varphi)\overline{\varphi'}) + a'|\varphi|^2 \\ &= 2\,\mathrm{Re}(f\overline{\varphi'}) + a'|\varphi|^2 \\ &\leq 2|f||\varphi'| + \|a'\|_{L^\infty(I)}|\varphi|^2\,. \end{aligned} \tag{4.3.3}$$

Using Young's inequality, we obtain

$$2|f||\varphi'| \leq \frac{1}{\varepsilon^2}|f|^2 + \varepsilon^2|\varphi'|^2\,. \tag{4.3.4}$$

Moreover, $a(x) \geq a_0 > 0$ implies that

$$\|a'\|_{L^\infty(I)}|\varphi|^2 \leq \frac{\|a'\|_{L^\infty(I)}}{a_0}a|\varphi|^2\,. \tag{4.3.5}$$

Thus, from (4.3.3)-(4.3.5) we obtain with $c := \max(1, \frac{\|a'\|_{L^\infty(I)}}{a_0}) \geq 1$

$$\frac{\mathrm{d}}{\mathrm{d}x}E(x) \leq \frac{1}{\varepsilon^2}|f|^2 + cE(x)\,. \tag{4.3.6}$$

Applying Gronwall's inequality, we therefore get

$$\begin{aligned} E(x) &\leq \left(\frac{1}{\varepsilon^2}\|f\|^2_{L^2(I)} + E(\xi)\right)\mathrm{e}^{c(x-\xi)} \\ &\leq \mathrm{e}^{c(\eta-\xi)}\left(\frac{1}{\varepsilon^2}\|f\|^2_{L^2(I)} + |\hat{\varphi}_1|^2 + a(\xi)|\hat{\varphi}_0|^2\right)\,, \end{aligned} \tag{4.3.7}$$

which implies the estimates (4.3.1)-(4.3.2). □

In order to derive an error estimate for the WKB approximation (4.2.9), which is explicit not only w.r.t. ε but also w.r.t. the truncation order N, it is essential to control the growth of the functions S_n w.r.t. $n \in \mathbb{N}_0$. As a first step, we aim to establish upper bounds for the derivatives S_n' which are given by recurrence relation (4.2.4)-(4.2.6). To this end, we employ a strategy similar to [Mel97, Lemma 2], which relies heavily on Cauchy's integral formula. To enable us to apply this tool, we shall assume that S_0' is not only defined on the real interval I, but also on a complex neighbourhood $G \subset \mathbb{C}$ of I. This leads us to introduce the following assumption.

Hypothesis C. *A Let S_0' be analytic on a complex, bounded, simply connected neighbourhood $G \subset \mathbb{C}$ of I, satisfying $S_0'(z) \neq 0$ for any $z \in G$.*

As a consequence of Hypothesis C, a and all S_n, $n \in \mathbb{N}$ are analytic on G. In particular, each S_n is bounded on I.

For the next lemma, we introduce the open set

$$G_\delta := \{z \in G \mid \operatorname{dist}(z, \partial G) > \delta\} \tag{4.3.8}$$

for some $\delta > 0$.

Lemma 4.3.2. *Let Hypothesis C be satisfied and let $0 < \delta \leq 1$ be such that $G_\delta \neq \emptyset$. Then there exist constants $0 < \widehat{K}_\delta \leq \widehat{K}$ with $\widehat{K}$ depending only on G and S_0' such that*

$$\|S_n'\|_{L^\infty(G_\delta)} \leq \|S_0'\|_{L^\infty(G)} \widehat{K}_\delta^n n^n \delta^{-n}\,, \quad n \in \mathbb{N}_0\,. \tag{4.3.9}$$

Here we define 0^0 as 1.

Proof. Define the auxiliary functions $\widehat{S}_n' := -(2S_0')^{-1} S_n'$. By using (4.2.4)-(4.2.6) we then find that the functions $\widehat{S}_n'$ satisfy the following recurrence relation

$$\widehat{S}_0' = -\frac{1}{2}\,, \tag{4.3.10}$$

$$\widehat{S}_n' = \left(\sum_{j=1}^{n-1} \widehat{S}_j' \widehat{S}_{n-j}'\right) + \left(2S_0'\right)^{-2} \left(-2S_0' \widehat{S}_{n-1}'\right)'\,, \quad n \geq 1\,. \tag{4.3.11}$$

Note that since S_0' is analytic on G, it follows from recurrence relation (4.2.4)-(4.2.6) that S_n' and hence also $\widehat{S}_n'$ are analytic on G, for every $n \in \mathbb{N}_0$. We will now prove by induction on n that

$$\|\widehat{S}_n'\|_{L^\infty(G_\delta)} \leq \frac{1}{2} \widehat{K}_\delta^n n^n \delta^{-n}\,, \quad n \in \mathbb{N}_0\,. \tag{4.3.12}$$

Obviously, this estimate does hold for $n = 0$, according to (4.3.10). Assume now that the estimate in (4.3.12) holds for $0 \leq j \leq n-1$ with some fixed $n \geq 1$. We will now prove it for n. Let $0 < \kappa < 1$ and $z \in G_\delta$. We denote with $\partial B_{\kappa\delta}(z)$ a circle of radius $\kappa\delta$ around z, see the left of Figure 4.3.1. Then Cauchy's integral formula implies

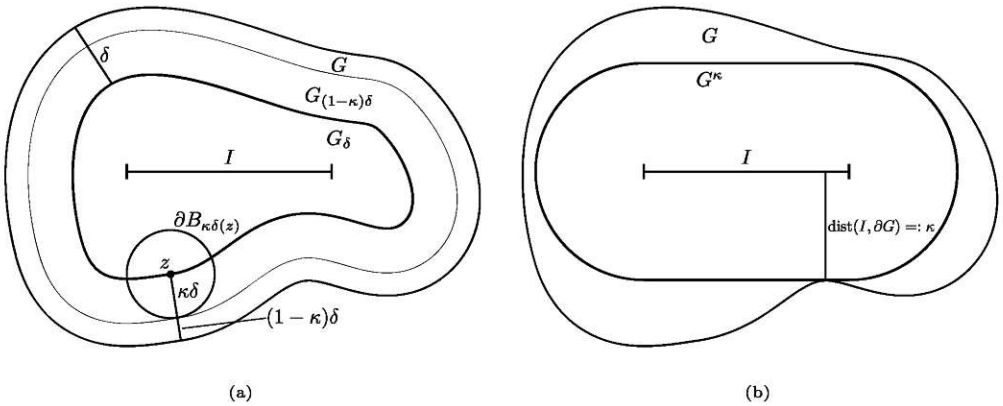

Figure 4.3.1: (a) Exemplary sketch of the situation from the proof of Lemma 4.3.2: $G_\delta \subset G_{(1-\kappa)\delta} \subset G$, where G is a complex neighbourhood of the interval I. Here, the point $z \in G_\delta$ is very close to the boundary ∂G_δ, which makes it clear why one has to consider $G_{(1-\kappa)\delta}$ in the r.h.s. of (4.3.14). (b) Every possible candidate G for the minimum on the l.h.s. of (4.3.22) can be reduced to a set $G^\kappa := \{z \in \mathbb{C} \mid \mathrm{dist}(z, I) < \kappa\}$, where $\kappa := \mathrm{dist}(I, \partial G) > 0$.

$$|(-2S_0'\widehat{S}_{n-1}')'(z)| = \frac{1}{2\pi}\left|\int_{\partial B_{\kappa\delta}(z)} \frac{-2S_0'(\zeta)\widehat{S}_{n-1}'(\zeta)}{(\zeta - z)^2}\,\mathrm{d}\zeta\right| \leq \frac{2\pi\kappa\delta}{2\pi} 2\|S_0'\|_{L^\infty(G)}\|\widehat{S}_{n-1}'\|_{L^\infty(\partial B_{\kappa\delta}(z))}(\kappa\delta)^{-2}\,. \tag{4.3.13}$$

This, together with the fact that $\partial B_{\kappa\delta}(z) \subseteq G_{(1-\kappa)\delta}$, now yields

$$\|(-2S_0'\widehat{S}_{n-1}')'\|_{L^\infty(G_\delta)} \leq 2\|S_0'\|_{L^\infty(G)}(\kappa\delta)^{-1}\|\widehat{S}_{n-1}'\|_{L^\infty(G_{(1-\kappa)\delta})}\,. \tag{4.3.14}$$

By applying estimate (4.3.14) and the induction hypothesis to (4.3.11), we find

$$\begin{aligned}\|\widehat{S}_n'\|_{L^\infty(G_\delta)} &\leq \sum_{j=1}^{n-1}\|\widehat{S}_j'\|_{L^\infty(G_\delta)}\|\widehat{S}_{n-j}'\|_{L^\infty(G_\delta)} \\ &\quad + \frac{1}{4}\|(S_0')^{-2}\|_{L^\infty(G_\delta)} 2\|S_0'\|_{L^\infty(G)}(\kappa\delta)^{-1}\|\widehat{S}_{n-1}'\|_{L^\infty(G_{(1-\kappa)\delta})} \\ &\leq \frac{1}{4}\widehat{K}_\delta^n\delta^{-n}\sum_{j=1}^{n-1} j^j(n-j)^{n-j} \\ &\quad + \frac{1}{4}\|(S_0')^{-2}\|_{L^\infty(G_\delta)}\|S_0'\|_{L^\infty(G)}\delta^{-n}\widehat{K}_\delta^{n-1}\frac{(n-1)^{n-1}}{\kappa(1-\kappa)^{n-1}}\,.\end{aligned} \tag{4.3.15}$$

Since $j^j(n-j)^{n-j} \leq (n-1)^{n-1}$ for all $1 \leq j \leq n-1$, we can bound the sum in the first term of (4.3.15) by $(n-1)^n$. Thus, we obtain

$$\|\widehat{S}_n'\|_{L^\infty(G_\delta)} \leq \frac{1}{2}\widehat{K}_\delta^n n^n\delta^{-n}\left[\frac{1}{2}\left(\frac{n-1}{n}\right)^n + \frac{\|(S_0')^{-2}\|_{L^\infty(G_\delta)}\|S_0'\|_{L^\infty(G)}}{2\widehat{K}_\delta\kappa(1-\kappa)^{n-1}n}\left(\frac{n-1}{n}\right)^{n-1}\right]. \tag{4.3.16}$$

It now suffices to show that the expression in the square brackets is less than or equal to 1. By further estimating $\left(\frac{n-1}{n}\right)^n \leq \frac{1}{\mathrm{e}}$, and choosing $\kappa = \frac{1}{n}$, we get

$$\|\widehat{S}_n'\|_{L^\infty(G_\delta)} \leq \frac{1}{2}\widehat{K}_\delta^n n^n \delta^{-n}\left[\frac{1}{2\,\mathrm{e}} + \frac{\|(S_0')^{-2}\|_{L^\infty(G_\delta)}\|S_0'\|_{L^\infty(G)}}{2\widehat{K}_\delta}\right] \tag{4.3.17}$$

Thus it is sufficient to choose

$$\widehat{K}_\delta := \frac{\mathrm{e}}{2\,\mathrm{e}-1}\|(S_0')^{-2}\|_{L^\infty(G_\delta)}\|S_0'\|_{L^\infty(G)}\,. \tag{4.3.18}$$

Finally, we note that $\widehat{K}_\delta \leq \widehat{K} := \frac{\mathrm{e}}{2\,\mathrm{e}-1}\|(S_0')^{-2}\|_{L^\infty(G)}\|S_0'\|_{L^\infty(G)}$ for any $0 < \delta \leq 1$. This concludes the proof. □

A simple but important implication of Lemma 4.3.2 is the fact that we are now able to provide estimates not only for all the derivatives of S_n but also for S_n itself:

Corollary 4.3.3. *Let Hypothesis C be satisfied. Then there exist constants $K_1, K_2 > 0$ depending only on G and S_0' such that*

$$\|S_n\|_{L^\infty(I)} \leq (\eta - \xi)\|S_0'\|_{L^\infty(G)}K_2^n n^n\,, \quad n \in \mathbb{N}_0\,, \tag{4.3.19}$$

$$\|S_n^{(k)}\|_{L^\infty(I)} \leq \|S_0'\|_{L^\infty(G)}(k-1)!\,K_1^{k-1}K_2^n n^n\,, \quad n \in \mathbb{N}_0\,, \quad k \in \mathbb{N}\,. \tag{4.3.20}$$

Here we define 0^0 as 1.

Proof. Since G is a complex neighbourhood of I, there is some $0 < \delta \leq 1$ such that $I \subset G_{2\delta}$. To prove estimate (4.3.20), we start with the trivial estimate $\|S_n^{(k)}\|_{L^\infty(I)} \leq \|S_n^{(k)}\|_{L^\infty(G_{2\delta})}$. Then, for any $k \in \mathbb{N}$ and $z \in G_{2\delta}$, Cauchy's integral formula implies

$$|S_n^{(k)}(z)| = \frac{(k-1)!}{2\pi}\left|\int_{\partial B_\delta(z)}\frac{S_n'(z)}{(\zeta - z)^k}\,\mathrm{d}\zeta\right| \leq (k-1)!\,\delta^{-k+1}\|S_n'\|_{L^\infty(G_\delta)}\,. \tag{4.3.21}$$

By applying Lemma 4.3.2 on the r.h.s. of (4.3.21) we conclude that (4.3.20) holds with $K_1 := 1/\delta$ and $K_2 := \widehat{K}_\delta/\delta$. Estimate (4.3.19) then follows from (4.3.20) for $k = 1$ and by the definition of S_n, see (4.2.10). □

Remark 4.3.4. *Of course, it is of great interest to find a constant K_2 from Corollary 4.3.3 which is as small as possible. To this end one would have to minimize the constant $\widehat{K}_\delta/\delta$ in estimate (4.3.9). In particular, one has to fix some complex neighbourhood G of I as well as a constant $0 < \delta \leq 1$ such that it holds $I \subset G_\delta$. Further, the proof of Lemma 4.3.2 indicates that $\widehat{K}_\delta$ can be reduced by choosing G small, see (4.3.18). However, this means that one is forced to reduce also the value of δ. Hence, this procedure usually results in a trade-off between the magnitudes of $\widehat{K}_\delta$ and δ. More precisely, one would have to solve the following minimization problem:*

$$\min_{\substack{0<\delta\leq 1\\ G\subset\mathbb{C}\\ I\subset G_\delta}}\frac{\|(S_0')^{-2}\|_{L^\infty(G_\delta)}\|S_0'\|_{L^\infty(G)}}{\delta} = \|(S_0')^{-2}\|_{L^\infty(I)}\min_{\substack{0<\delta\leq 1\\ G^\delta\subset\mathbb{C}}}\frac{\|S_0'\|_{L^\infty(G^\delta)}}{\delta}\,, \tag{4.3.22}$$

where $G^\delta := \{z \in \mathbb{C} \mid \operatorname{dist}(z, I) < \delta\}$. *Equality in (4.3.22) holds for the following reasons: First, on the l.h.s. of (4.3.22) one only needs to consider sets* $G \subset \mathbb{C}$ *of the form* $G = G^\kappa := \{z \in \mathbb{C} \mid \operatorname{dist}(z, I) < \kappa\}$, *with* $\kappa > 0$. *For any* $0 < \delta \leq 1$ *such that* $I \subset G_\delta$, *this follows since the numerator on the l.h.s. of (4.3.22) is not increased when replacing* G *and* G_δ *by* $G^{\operatorname{dist}(I,\partial G)}$ *and* $(G^{\operatorname{dist}(I,\partial G)})_\delta$, *respectively, see the right of Figure 4.3.1. The condition* $I \subset G_\delta$ *then simply reads* $\kappa > \delta$. *Second, since for a fixed* $0 < \delta \leq 1$ *and* $\delta < \kappa_1 < \kappa_2$ *it holds that* $G^{\kappa_1} \subset G^{\kappa_2}$ *and* $(G^{\kappa_1})_\delta \subset (G^{\kappa_2})_\delta$, *it is sufficient to consider simply the sets* $G^{\delta+\epsilon}$, *with* $\epsilon > 0$ *being an arbitrarily small number. The equality in (4.3.22) then follows from the fact that* $\lim_{\epsilon\to 0} G^{\delta+\epsilon} = G^\delta$ *and* $\lim_{\epsilon\to 0}(G^{\delta+\epsilon})_\delta = \lim_{\epsilon\to 0} G^\epsilon = \bigcap_{\epsilon>0} G^\epsilon = I$.

We will later make use of the residual of the WKB approximation (4.2.9) w.r.t. the ODE in (4.1.1). For this, the following lemma will be helpful.

Lemma 4.3.5. *Denote with* $L_\varepsilon := \varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} + a(x)$ *the linear operator appearing in the Schrödinger equation (4.1.1) and let* $\widetilde{\varphi}_N := \exp\left(\sum_{n=0}^N \varepsilon^{n-1} S_n\right)$, $N \in \mathbb{N}_0$. *Then it holds*

$$L_\varepsilon \widetilde{\varphi}_N = \widetilde{\varphi}_N f_{N,\varepsilon}\,, \tag{4.3.23}$$

where

$$f_{N,\varepsilon} = \varepsilon^{N+1}(-2S_0' S_{N+1}') + \sum_{n=2}^{N} \sum_{k=2+N-n}^{N} \varepsilon^{n+k} S_n' S_k'\,; \tag{4.3.24}$$

for $N < 2$ *the double sum drops.*

Proof. First we observe that

$$\begin{aligned} L_\varepsilon \widetilde{\varphi}_N &= \varepsilon^2 \widetilde{\varphi}_N'' + a(x)\widetilde{\varphi}_N \\ &= \varepsilon^2 \widetilde{\varphi}_N \left(\left(\frac{1}{\varepsilon} \sum_{n=0}^N \varepsilon^n S_n' \right)^2 + \frac{1}{\varepsilon} \sum_{n=0}^N \varepsilon^n S_n'' \right) + a(x)\widetilde{\varphi}_N \\ &= \widetilde{\varphi}_N \left(\sum_{0\leq n,k\leq N} \varepsilon^{n+k} S_n' S_k' + \sum_{n=0}^N \varepsilon^{n+1} S_n'' + a(x) \right)\,. \end{aligned} \tag{4.3.25}$$

Let us denote the second factor in (4.3.25) by $f_{N,\varepsilon}$. We will now show that $f_{N,\varepsilon}$ reduces to (4.3.24). To this end, let us first rewrite $f_{N,\varepsilon}$ as

$$\begin{aligned} f_{N,\varepsilon} = (S_0'^2 + a) &+ \left(\sum_{\substack{1\leq n+k\leq N \\ 0\leq n,k\leq N}} \varepsilon^{n+k} S_n' S_k' + \sum_{n=0}^{N-1} \varepsilon^{n+1} S_n'' \right) \\ &+ \left(\sum_{\substack{n+k=N+1 \\ 0\leq n,k\leq N}} \varepsilon^{N+1} S_n' S_k' + \varepsilon^{N+1} S_N'' \right) + \sum_{\substack{n+k>N+1 \\ 0\leq n,k\leq N}} \varepsilon^{n+k} S_n' S_k'\,. \end{aligned} \tag{4.3.26}$$

Now, the first term in (4.3.26) vanishes due to (4.2.4). The second term also vanishes since

$$\begin{aligned}\sum_{\substack{1\le n+k\le N\\ 0\le n,k\le N}} \varepsilon^{n+k} S_n' S_k' &= \sum_{n=0}^{N-1} \varepsilon^{n+1} \sum_{j=0}^{n+1} S_j' S_{n+1-j}' \\ &= \sum_{n=0}^{N-1} \varepsilon^{n+1} \left(2S_0' S_{n+1}' + \sum_{j=1}^{n} S_j' S_{n+1-j}' \right) \\ &= -\sum_{n=0}^{N-1} \varepsilon^{n+1} S_n'' \,, \end{aligned} \tag{4.3.27}$$

where we used in the last equation recurrence relation (4.2.6) for the function S_{n+1}'. Finally, by using (4.2.6) for the function S_{N+1}', the third term in (4.3.26) simplifies to $\varepsilon^{N+1}(-2S_0'S_{N+1}')$. The claim follows. □

Recalling that $S_0(x) \in \mathrm{i}\,\mathbb{R}$ we note that $\widetilde{\varphi}_N(x)$ is, for fixed $x \in I$, uniformly bounded w.r.t. $\varepsilon \in (0,1]$. Thus the r.h.s. of (4.3.23) is of the order $\mathcal{O}(\varepsilon^{N+1})$, and we conclude from Lemma 4.3.5 that the function $\widetilde{\varphi}_N$ satisfies the ODE $L_\varepsilon \varphi = 0$ asymptotically, as $\varepsilon \to 0$. This is one of the main properties we can utilize to show that also the numerical error of the WKB approximation (4.2.9) will approach 0 as $\varepsilon \to 0$, at least for $N \ge 1$. To this end we need the following lemma.

Lemma 4.3.6. *Let Hypothesis C be satisfied and define* $\varphi_N^{\pm} := \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^{\pm}\right)$, $N \in \mathbb{N}_0$. *Then there exist constants* $\varepsilon_0 \in (0,1)$ *and* $C > 0$ *such that it holds for* $\varepsilon \in (0, \varepsilon_0]$*:*

$$\begin{aligned}\|\alpha_{N,\varepsilon}\varphi_N^-\|_{L^\infty(I)} \le C &\left(|\varphi_0|\, \|S_0'\|_{L^\infty(G)} \sum_{n=0}^{N} \varepsilon^n K_2^n n^n + |\varphi_1| \right) \\ &\times \exp\left((\eta - \xi) \|S_0'\|_{L^\infty(G)} \sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} K_2^{2n+1} (2n+1)^{2n+1} \right) , \end{aligned} \tag{4.3.28}$$

with $\alpha_{N,\varepsilon}$ *from (4.2.13). For* $N = 0$ *the last sum drops. The same estimate holds for* $\|\beta_{N,\varepsilon}\varphi_N^+\|_{L^\infty(I)}$. *In particular, since the initial values* φ_0 *and* φ_1 *are assumed to be uniformly bounded w.r.t.* ε, *so is* φ_N^{WKB} *in* $L^\infty(I)$.

Proof. We will prove only the estimate for $\alpha_{N,\varepsilon}\varphi_N^-$. For $\beta_{N,\varepsilon}\varphi_N^+$ it is fully analogous. First notice that Proposition 4.2.1 implies that

$$|\alpha_{N,\varepsilon}| = \frac{\left| \varphi_0 \sum_{n=0}^{N} \varepsilon^n (S_n^+)'(\xi) - \varphi_1 \right|}{2 \left| \sum_{n=0}^{\lfloor \frac{N}{2} \rfloor} \varepsilon^{2n} (S_{2n}^+)'(\xi) \right|} . \tag{4.3.29}$$

Due to $a(x) \geq a_0 > 0$, we have $|(S_0^+)'(\xi)| \geq \sqrt{a_0} > 0$. Thus, there exists $\varepsilon_0 \in (0,1)$ sufficiently small such that

$$\begin{aligned}\left|\sum_{n=0}^{\lfloor \frac{N}{2} \rfloor} \varepsilon^{2n} (S_{2n}^+)'(\xi)\right| &\geq \left|(S_0^+)'(\xi)\right| - \sum_{n=1}^{\lfloor \frac{N}{2} \rfloor} \varepsilon^{2n} \left|(S_{2n}^+)'(\xi)\right| \\ &\geq \sqrt{a_0} - \sum_{n=1}^{\lfloor \frac{N}{2} \rfloor} \varepsilon^{2n} \left|(S_{2n}^+)'(\xi)\right| \\ &\geq \frac{1}{2C} \end{aligned} \tag{4.3.30}$$

for all $\varepsilon \in (0, \varepsilon_0]$ and some $C > 0$ (since $(S_{2n}^+)'$ is bounded on I). Hence, we obtain

$$|\alpha_{N,\varepsilon}| \leq C \left| \varphi_0 \sum_{n=0}^{N} \varepsilon^n (S_n^+)'(\xi) - \varphi_1 \right|. \tag{4.3.31}$$

Next, (4.2.11)-(4.2.12) imply $|\varphi_N^-(x)| \leq \exp\left(\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} \left|S_{2n+1}^-(x)\right|\right)$ for all $x \in I$. Together with (4.3.31) this yields

$$|\alpha_{N,\varepsilon} \varphi_N^-(x)| \leq C \left(|\varphi_0| \sum_{n=0}^{N} \varepsilon^n \left|(S_n^+)'(\xi)\right| + |\varphi_1| \right) \exp\left(\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} \left|S_{2n+1}^-(x)\right| \right) \tag{4.3.32}$$

for all $x \in I$. Applying Corollary 4.3.3 then yields the claim. □

Finally, we provide an error estimate for the WKB approximation (4.2.9).

Theorem 4.3.7. *Let Hypothesis C be satisfied and let $\varphi \in C^2(I)$ be the solution of IVP (4.1.1). There exist constants $\varepsilon_0 \in (0,1)$ and $C > 0$ independent of N and ε such that it holds for $\varepsilon \in (0, \varepsilon_0]$:*

$$\begin{aligned}\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} \leq C \|S_0'\|_{L^\infty(G)}^2 &\left(|\varphi_0| \, \|S_0'\|_{L^\infty(G)} \sum_{n=0}^{N} \varepsilon^n K_2^n n^n + |\varphi_1| \right) \\ &\times \exp\left((\eta - \xi) \|S_0'\|_{L^\infty(G)} \sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} K_2^{2n+1} (2n+1)^{2n+1} \right) \\ &\times \left(\varepsilon^N K_2^{N+1} (N+1)^{N+1} + \sum_{n=2}^{N} \sum_{k=2+N-n}^{N} \varepsilon^{n+k-1} K_2^{n+k} n^n k^k \right). \end{aligned} \tag{4.3.33}$$

For $N = 0$ the sum in the exponential function drops, and for $N < 2$ the double sum drops.

Proof. To compute the residual of the WKB approximation (4.2.9), we notice that $\varphi_N^{WKB} = \alpha_{N,\varepsilon}\varphi_N^- + \beta_{N,\varepsilon}\varphi_N^+$, where $\varphi_N^\pm = \exp\left(\sum_{n=0}^N \varepsilon^{n-1} S_n^\pm\right)$. By applying Lemma 4.3.5, we obtain

$$\begin{aligned} L_\varepsilon(\varphi - \varphi_N^{WKB}) &= -\alpha_{N,\varepsilon} L_\varepsilon \varphi_N^- - \beta_{N,\varepsilon} L_\varepsilon \varphi_N^+ \\ &= -\alpha_{N,\varepsilon}\varphi_N^- f_{N,\varepsilon}^- - \beta_{N,\varepsilon}\varphi_N^+ f_{N,\varepsilon}^+ , \end{aligned} \tag{4.3.34}$$

where the functions $f_{N,\varepsilon}^\pm$ are given by (4.3.24) when inserting $S_n^\pm$ for S_n. Further, since φ_N^{WKB} satisfies the initial conditions in (4.1.1), we have $(\varphi - \varphi_N^{WKB})(\xi) = 0$ and $\varepsilon(\varphi - \varphi_N^{WKB})'(\xi) = 0$. Thus, Proposition 4.3.1 for $\hat\varphi_0, \hat\varphi_1 = 0$ implies (note that $f_{N,\varepsilon}^\pm \in C(I)$ and $a \in W^{1,\infty}(I)$) the existence of some $C > 0$ independent of N and ε such that

$$\begin{aligned} \|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} &\le \frac{C}{\varepsilon}\|\alpha_{N,\varepsilon}\varphi_N^- f_{N,\varepsilon}^- + \beta_{N,\varepsilon}\varphi_N^+ f_{N,\varepsilon}^+\|_{L^2(I)} \\ &\le \frac{\widetilde{C}}{\varepsilon}\left(\|\alpha_{N,\varepsilon}\varphi_N^-\|_{L^\infty(I)}\|f_{N,\varepsilon}^-\|_{L^\infty(I)} + \|\beta_{N,\varepsilon}\varphi_N^+\|_{L^\infty(I)}\|f_{N,\varepsilon}^+\|_{L^\infty(I)}\right) , \end{aligned} \tag{4.3.35}$$

where $\widetilde{C} := \sqrt{\eta - \xi}C$. Further, according to Corollary 4.3.3,

$$\begin{aligned} \|f_{N,\varepsilon}^\pm\|_{L^\infty(I)} &\le 2\varepsilon^{N+1}\|S_0'\|_{L^\infty(I)}\|S_{N+1}'\|_{L^\infty(I)} + \sum_{n=2}^N \sum_{k=2+N-n}^N \varepsilon^{n+k}\|S_n'\|_{L^\infty(I)}\|S_k'\|_{L^\infty(I)} \\ &\le \|S_0'\|_{L^\infty(G)}^2 \left(2\varepsilon^{N+1}K_2^{N+1}(N+1)^{N+1} + \sum_{n=2}^N \sum_{k=2+N-n}^N \varepsilon^{n+k}K_2^{n+k}n^n k^k\right) . \end{aligned} \tag{4.3.36}$$

Estimate (4.3.33) now follows from (4.3.35)-(4.3.36) by applying Lemma 4.3.6. This concludes the proof. □

Remark 4.3.8. *As a consequence of Theorem 4.3.7, we have that*

$$\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} = \mathcal{O}(\varepsilon^N) , \quad \varepsilon \to 0 . \tag{4.3.37}$$

4.3.1 Refined error estimate incorporating quadrature errors

Theorem 4.3.7 yields an explicit (w.r.t. ε and N) error estimate for the WKB approximation (4.2.9). However, in practice one cannot expect to be able to compute (4.2.9) exactly. Indeed, even though for a given function a one can compute the derivatives $(S_n^\pm)'$ exactly through (4.2.4)-(4.2.6), one still has to deal with the integrals $\int_\xi^x (S_n^\pm)'\,\mathrm{d}\tau$ in (4.2.10) in order to compute the functions $S_n^\pm$. For a detailed description of the method we use to approximate these integrals, we refer to Section 4.4.1.

For now, let us assume we are given numerical approximations $\widetilde{S}_n^\pm$, $n \in \mathbb{N}_0$, of the functions $S_n^\pm$, which satisfy $\|\widetilde{S}_n^\pm - S_n^\pm\|_{L^\infty(I)} \le e_n$ with positive constants e_n. We then define the corresponding "perturbed" WKB approximation as

$$\widetilde{\varphi}_N^{WKB} := \alpha_{N,\varepsilon}\exp\left(\sum_{n=0}^N \varepsilon^{n-1}\widetilde{S}_n^-\right) + \beta_{N,\varepsilon}\exp\left(\sum_{n=0}^N \varepsilon^{n-1}\widetilde{S}_n^+\right) . \tag{4.3.38}$$

Notice that we use here the exact constants $\alpha_{N,\varepsilon}$ and $\beta_{N,\varepsilon}$ as given by formulas (4.2.13)-(4.2.14) (since the values $(S_n^{\pm})'(\xi)$ are exactly known from (4.2.4)-(4.2.6)).

We are now interested in an error estimate for the perturbed WKB approximation (4.3.38). Such an estimate is provided by the following theorem:

Theorem 4.3.9. *Let Hypothesis C be satisfied and let $\varphi \in C^2(I)$ be the solution of IVP (4.1.1). Further assume $\widetilde{S}_{2n}^{\pm}(x) \in \mathrm{i}\mathbb{R}$, $n \in \mathbb{N}_0$, for any $x \in I$. Then, there exist constants $\varepsilon_0 \in (0,1)$ and $C > 0$ independent of N and ε such that it holds for $\varepsilon \in (0, \varepsilon_0]$:*

$$\begin{aligned}\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)} \leq \exp\left((\eta-\xi)\|S_0'\|_{L^\infty(G)} \sum_{n=0}^{\lfloor\frac{N-1}{2}\rfloor} \varepsilon^{2n} K_2^{2n+1}(2n+1)^{2n+1}\right) \\ \times \left[C\|S_0'\|_{L^\infty(G)}^2 \left(|\varphi_0|\,\|S_0'\|_{L^\infty(G)} \sum_{n=0}^{N} \varepsilon^n K_2^n n^n + |\varphi_1|\right)\right. \\ \times \left(\varepsilon^N K_2^{N+1}(N+1)^{N+1} + \sum_{n=2}^{N}\sum_{k=2+N-n}^{N} \varepsilon^{n+k-1} K_2^{n+k} n^n k^k\right) \\ \left. + (|\alpha_{N,\varepsilon}| + |\beta_{N,\varepsilon}|)\left(\sum_{n=0}^{N} \varepsilon^{n-1} e_n\right) \exp\left(\sum_{n=0}^{\lfloor\frac{N-1}{2}\rfloor} \varepsilon^{2n} e_{2n+1}\right)\right].\end{aligned} \tag{4.3.39}$$

For $N = 0$ the sums in the exponential functions drop, and for $N < 2$ the double sum drops.

Proof. Obviously, we have that

$$\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)} \leq \|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} + \|\varphi_N^{WKB} - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}\,. \tag{4.3.40}$$

Now, the first term in (4.3.40) can be estimated using Theorem 4.3.7 and enforces the restriction $\varepsilon \in (0, \varepsilon_0]$. For the second term we estimate

$$\begin{aligned}\|\varphi_N^{WKB} - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)} \leq |\alpha_{N,\varepsilon}| \left\|\exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^-\right) - \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} \widetilde{S}_n^-\right)\right\|_{L^\infty(I)} \\ + |\beta_{N,\varepsilon}| \left\|\exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^+\right) - \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} \widetilde{S}_n^+\right)\right\|_{L^\infty(I)}.\end{aligned} \tag{4.3.41}$$

Let us introduce the abbreviation $\Delta S_n^{\pm} := \widetilde{S}_n^{\pm} - S_n^{\pm}$ and estimate

$$
\begin{aligned}
&\left\| \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} S_n^{\pm}\right) - \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} \widetilde{S}_n^{\pm}\right) \right\|_{L^\infty(I)} \\
&= \left\| \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} \left(S_n^{\pm} + t\Delta S_n^{\pm}\right)\right) \mathrm{d}t \right\|_{L^\infty(I)} \\
&\leq \left(\int_0^1 \left\| \exp\left(\sum_{n=0}^{N} \varepsilon^{n-1} \left(S_n^{\pm} + t\Delta S_n^{\pm}\right)\right) \right\|_{L^\infty(I)} \mathrm{d}t \right) \left(\sum_{n=0}^{N} \varepsilon^{n-1} \|\Delta S_n^{\pm}\|_{L^\infty(I)} \right) \\
&\leq \left(\int_0^1 \exp\left(\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} \left(\|S_{2n+1}^{\pm}\|_{L^\infty(I)} + t\|\Delta S_{2n+1}^{\pm}\|_{L^\infty(I)} \right) \right) \mathrm{d}t \right) \left(\sum_{n=0}^{N} \varepsilon^{n-1} \|\Delta S_n^{\pm}\|_{L^\infty(I)} \right) \\
&\leq \exp\left(\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} \left(\|S_{2n+1}^{\pm}\|_{L^\infty(I)} + \|\Delta S_{2n+1}^{\pm}\|_{L^\infty(I)} \right) \right) \left(\sum_{n=0}^{N} \varepsilon^{n-1} \|\Delta S_n^{\pm}\|_{L^\infty(I)} \right) \\
&\leq \exp\left((\eta - \xi) \|S_0'\|_{L^\infty(G)} \sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} K_2^{2n+1} (2n+1)^{2n+1} \right) \exp\left(\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} e_{2n+1} \right) \\
&\quad \times \left(\sum_{n=0}^{N} \varepsilon^{n-1} e_n \right) ,
\end{aligned}
\tag{4.3.42}
$$

where we used in the third step that $S_{2n}^{\pm}(x) + t\Delta S_{2n}^{\pm}(x) \in \mathrm{i}\,\mathbb{R}$ for every $t \in [0,1]$ and $x \in I$, which is a direct consequence of (4.2.11) and the assumption $\widetilde{S}_{2n}^{\pm}(x) \in \mathrm{i}\,\mathbb{R}$. Moreover, in the last step we used Corollary 4.3.3. The claim now follows by combining (4.3.40)-(4.3.42). □

Let us compare the extended error estimate (4.3.39) with (4.3.33). The new (additional) second term inside the square brackets in (4.3.39) is due to the perturbed functions $\widetilde{S}_n^{\pm}$ and includes the approximation error bounds e_n. In particular, the factor $\sum_{n=0}^{N} \varepsilon^{n-1} e_n$ is rather unfavorable, as it is of order $\mathcal{O}(\varepsilon^{-1})$, as $\varepsilon \to 0$. We note that the appearance of this $\mathcal{O}(\varepsilon^{-1})$-term in estimate (4.3.39) is strongly related to the appearance of the $\mathcal{O}(\varepsilon^{-1})$-terms in [AAN11, Theorem 3.1], [AKU22, Theorem 3.2] and [JL03, Eq. (35)]. There it implied an upper step size limit $h \leq \bar{h}(\varepsilon) = \varepsilon^{\gamma}$ with some $\gamma \in (0,1)$. Similarly, it would require here some ε-dependent upper bound on the quadrature error e_0 of $\widetilde{S}_0^{\pm}$. We specify this observation in the following remark.

Remark 4.3.10. *It is evident from (4.3.31) that $\alpha_{N,\varepsilon} = \mathcal{O}(1)$, $\varepsilon \to 0$. The same holds for $\beta_{N,\varepsilon}$. Hence, we see from (4.3.39) that*

$$
\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)} = \mathcal{O}(\varepsilon^N) + \sum_{n=0}^{N} \mathcal{O}(\varepsilon^{n-1}) e_n \,, \quad \varepsilon \to 0 \,. \tag{4.3.43}
$$

Thus, asymptotically, as $\varepsilon \to 0$, the approximation error of $\widetilde{S}_0^{\pm}$ has the biggest impact on the overall error since it is multiplied by a factor $\mathcal{O}(\varepsilon^{-1})$. In order to recover an

overall $\mathcal{O}(\varepsilon^N)$ error behavior, as in Theorem 4.3.7, one should hence aim for highly accurate approximations of the functions $S_n^{\pm}$, with an ε-dependent error order of at most $e_n = \mathcal{O}(\varepsilon^{N-n+1})$.

4.4 Computation of the WKB approximation

In this section we present the methods we use to compute the (perturbed) WKB approximation (4.2.9), (4.3.38). This process can be divided into two steps. First, the computation of the functions S_n. Second, an adequate truncation of the asymptotic series (4.2.2).

4.4.1 Computation of the functions S_n

The computation of the functions S_n relies on recurrence relation (4.2.4)-(4.2.6) as well as on definition (4.2.10). Since the latter involves the evaluation of an integral, one cannot expect to be able to compute S_n exactly, in general. Consequently, we will instead compute approximations $\widetilde{S}_n \approx S_n = \int_\xi^x S_n' \,\mathrm{d}\tau$ which satisfy the assumption $\widetilde{S}_{2n}(x) \in \mathrm{i}\,\mathbb{R}$, $n \in \mathbb{N}_0$, such that the resulting error for the corresponding perturbed WKB approximation can be controlled by Theorem 4.3.9.

As the first step, we compute the derivatives S_n' through (4.2.4)-(4.2.6) exactly, employing symbolical computation[3]. Secondly, we employ a highly accurate quadrature for approximating the integral in (4.2.10). For this, we use the well-known Clenshaw-Curtis algorithm [CC60], which we shall briefly explain in the following.

The basic idea of this method is an expansion of the integrand f in terms of Chebyshev polynomials the integrals of which are known. More precisely, one considers a truncated Chebyshev series for the integrand, i.e. $f(l) \approx \sum_{r=0}^{M} a_r T_r(l)$, $l \in [-1,1]$, where $T_r(l) = \cos(r \arccos(l))$, $r \in \mathbb{N}_0$, are the Chebyshev polynomials. Here, the spectral coefficients a_r are determined with a collocation method on the Chebyshev collocation points $l_k = \cos(k\pi/M)$, $k = 0, \dots, M$, by solving the $M+1$ equations $f(l_k) = \sum_{r=0}^{M} a_r \cos\left(\frac{rk\pi}{M}\right)$ for the a_r, $r = 0, \dots, M$. Therefore, the spectral coefficients can be computed by the discrete cosine transformation (DCT) of the function f sampled at the collocations points. We note that the DCT is related to the discrete Fourier transform and can be computed efficiently using the fast Fourier transform algorithm after some preprocessing (e.g., see [Tre00, Chap. 8]).

Then, the antiderivative of f can be approximated again by a Chebyshev sum,

$$\int_{-1}^{l} f(\tau) \,\mathrm{d}\tau \approx \sum_{r=0}^{M} b_r T_r(l)\,, \tag{4.4.1}$$

where the coefficients b_r are related to the a_r, see [CC60] for the detailed formulas. In [Cha68] it was shown that the Clenshaw-Curtis method approximates integrals of analytic functions with spectral accuracy, i.e., the numerical error is decreasing exponentially with the number of modes M.

[3] As an alternative to (4.2.4)-(4.2.6), in [RR00] the authors established an almost explicit formula for the derivatives S_n', depending on a and its derivatives $a', \dots, a^{(n)}$. Although not used here, this approach may prove advantageous with regard to the computational time.

An integration over the interval $x \in [\xi, \eta]$ is realized by mapping $x = \eta(1+l)/2+\xi(1-l)/2$, $l \in [-1,1]$ to the interval $[-1,1]$. Thus, by sampling the derivatives S_n' at the transformed Chebyshev points x_k, $k = 0, \dots, M$ in the interval $I = [\xi, \eta]$, we obtain the approximations $\widetilde{S}_n(x_k) \approx S_n(x_k)$. Notably, the coefficients b_r are such that the r.h.s. of (4.4.1) vanishes at $l = -1$, implying $\widetilde{S}_n(\xi) = 0$. Hence, the perturbed WKB approximation (4.3.38) satisfies the first initial condition in (4.1.1), namely, $\widetilde{\varphi}_N^{WKB}(\xi) = \alpha_{N,\varepsilon} + \beta_{N,\varepsilon} = \varphi_0$. Finally, it is worth mentioning that when employing the Clenshaw-Curtis algorithm for the integrals in (4.2.10), it follows that $\widetilde{S}_{2n}(x_k) \in \mathrm{i}\,\mathbb{R}$. As a consequence, the error of the corresponding perturbed WKB approximation (4.3.38) can be controlled with the aid of Theorem 4.3.9.

We note that an alternative and efficient way of approximating the functions S_n can be realized without the need for symbolical computation of the derivatives S_n'. Indeed, one can instead employ a spectral method to perform the differentiation of the predecessor S_{n-1}' in the recursion (4.2.6). For instance, by using the $(M+1)\times(M+1)$ Chebyshev differentiation matrices $\mathbf{D}_M$ as described in [Tre00, Chap. 6], one can efficiently approximate the derivative of a function at Chebyshev grid points $l_k \in [-1,1]$, $k = 0, 1, \dots, M$. Thus, to approximate the derivative of a function sampled at transformed Chebyshev points $x_k \in [\xi, \eta]$, it is necessary to use the scaled matrix $\widetilde{\mathbf{D}}_M := \frac{2}{\eta-\xi}\mathbf{D}_M$. Following recurrence relation (4.2.4)-(4.2.6), we can therefore approximate the derivatives S_n' sampled at Chebyshev points x_k through the following pointwise definition on the grid:

$$\bar{S}_1'(x_k) := -\frac{\sum_{l=0}^{M}(\widetilde{\mathbf{D}}_M)_{k+1,l+1}S_0'(x_l)}{2S_0'(x_k)}\,, \tag{4.4.2}$$

$$\bar{S}_n'(x_k) := -\frac{\sum_{j=1}^{n-1}\bar{S}_j'(x_k)\bar{S}_{n-j}'(x_k) + \sum_{l=0}^{M}(\widetilde{\mathbf{D}}_M)_{k+1,l+1}\bar{S}_{n-1}'(x_l)}{2S_0'(x_k)}\,, \quad n \geq 2\,, \tag{4.4.3}$$

for $k = 0, \dots, M$. One then obtains approximations $\widetilde{S}_n(x_k) \approx S_n(x_k)$ by employing the Clenshaw-Curtis algorithm using the approximations $\bar{S}_n'(x_k) \approx S_n'(x_k)$, $k = 0, \dots, M$.

However, note that approximating S_n' using (4.4.2)-(4.4.3) can lead to a rapid accumulation of errors, as repeated numerical differentiation is intrinsically unstable. The reason for this behavior lies in the ill-conditioned Chebyshev differentiation matrices $\mathbf{D}_M$. It is known that the condition number of these matrices is of order $\mathcal{O}(M^2)$ (e.g., see [BE92, Fun87]). In a finite precision approach this leads to a big loss, which means that roughly two orders of magnitude are lost in accuracy in each application of the recurrence relation to compute the S_n'. Consequently, it is recommendable to employ this approach primarily for small values of N.

4.4.2 Truncation of the WKB series

When truncating the asymptotic series

$$f \sim \sum_{n=0}^{\infty} \varepsilon^n f_n\,, \quad \varepsilon \to 0 \tag{4.4.4}$$

after some finite order N, one would like to analyze the difference $f - \sum_{n=0}^{N}\varepsilon^n f_n$. But since the function S in (4.2.1)-(4.2.2) remains unknown, we shall investigate the numerical error of the WKB approximation, as started in Section 4.3.

Recall that for a fixed $N \geq 0$, Theorem 4.3.7 guarantees that $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} = \mathcal{O}(\varepsilon^N)$ as $\varepsilon \to 0$, see also Remark 4.3.8. In practical applications, however, the situation is exactly the opposite, namely, the small parameter ε is fixed and N can be chosen freely. Note also that just including more terms into the series (4.2.2) does not necessarily reduce the error of the WKB approximation, simply since the asymptotic series is typically divergent. Hence the question arises which choice of N will minimize $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)}$, often referred to as *optimal truncation*. In this context, we denote with $N_{opt} = N_{opt}(\varepsilon) := \operatorname{argmin}_{N\in\mathbb{N}_0}\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)}$ the *optimal truncation order*. In general, an *optimally truncated* asymptotic series is sometimes referred to as *superasymptotics* (e.g., see [Boy99]). The corresponding error of an optimally truncated series is then typically of the form $\sim \exp(-c/\varepsilon)$, as $\varepsilon \to 0$, with some constant $c > 0$.

In practice, a useful heuristic for finding the optimal truncation order for a fixed ε is given in [Boy99]. It suggests that it can be obtained by truncating the asymptotic series before its smallest term. In our case, we would hence have to find the minimizer $N_{heu} = N_{heu}(\varepsilon)$ of $n \mapsto \varepsilon^n \|S_{n+1}\|_{L^\infty(I)}$. This can either be found by "brute force", comparing the size of each term up to some prescribed maximal order N_{max}, or by utilizing Corollary 4.3.3 to (roughly) predict N_{heu}. Indeed, for any $N \in \mathbb{N}_0$, estimate (4.3.19) implies

$$\varepsilon^N \|S_{N+1}\|_{L^\infty(I)} \leq (\eta - \xi)\|S_0'\|_{L^\infty(G)} \varepsilon^N K_2^{N+1} (N+1)^{N+1} . \tag{4.4.5}$$

Treating N as a continuous variable for a moment, we find the minimum of

$$g(N) := \ln\left(\varepsilon^N K_2^{N+1}(N+1)^{N+1}\right) \tag{4.4.6}$$

at

$$\widehat{N}_{heu} = \widehat{N}_{heu}(\varepsilon) = \frac{1}{\mathrm{e} K_2 \varepsilon} - 1 . \tag{4.4.7}$$

Hence, the minimum of the right-hand side of (4.4.5) is

$$(\eta - \xi)\|S_0'\|_{L^\infty(G)} \exp\left(g(\widehat{N}_{heu})\right) = \frac{(\eta - \xi)\|S_0'\|_{L^\infty(G)}}{\varepsilon} \exp\left(-\frac{1}{\mathrm{e} K_2 \varepsilon}\right) . \tag{4.4.8}$$

So, the first term of the remainder of the asymptotic series appearing in the WKB-ansatz (4.2.1)-(4.2.2), truncated at the nearest integer value to $\widehat{N}_{heu}$, is exponentially small w.r.t. ε. Recalling that the term $\exp(g(N)) = \varepsilon^N K_2^{N+1}(N+1)^{N+1}$ also appears in estimate (4.3.33), we therefore might also expect the error $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)}$ to be exponentially small w.r.t. ε, if N is chosen adequately. Indeed, this is guaranteed by the following corollary of Theorem 4.3.7.

Corollary 4.4.1. *Let Hypothesis C be satisfied and let $\varphi \in C^2(I)$ be the solution of IVP (4.1.1). Then there exist $\tilde{\varepsilon}_0 \in (0,1)$ and $N = N(\varepsilon) \in \mathbb{N}$ such that it holds for $\varepsilon \in (0, \tilde{\varepsilon}_0]$:*

$$\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} \leq \frac{C}{\varepsilon^2} \exp\left(-\frac{r}{\varepsilon}\right) , \tag{4.4.9}$$

with constants $C, r > 0$ independent of ε.

Proof. We prove estimate (4.4.9) by applying Theorem 4.3.7 for a specific choice of $N = N(\varepsilon)$. First, choose $0 < \tilde{\varepsilon}_0 < \min(\varepsilon_0, \frac{1}{K_2})$ with $\varepsilon_0 \in (0,1)$ being the constant from Theorem 4.3.7 and K_2 from Corollary 4.3.3. Then there exists some constant $c \in [\mathrm{e}\, K_2\tilde{\varepsilon}_0, \mathrm{e})$ implying that $N := \lfloor \frac{c}{\mathrm{e}\, K_2\varepsilon} \rfloor - 1 \geq 0$ for any $\varepsilon \in (0, \tilde{\varepsilon}_0]$. The idea is now to majorize, for this choice of N, several sums in (4.3.33) by convergent geometric series. First, we have

$$\sum_{n=0}^{N} (\varepsilon K_2 n)^n \leq \sum_{n=0}^{N} (\varepsilon K_2 (N+1))^n \leq \sum_{n=0}^{\infty} \left(\frac{c}{\mathrm{e}}\right)^n = \frac{1}{1-\frac{c}{\mathrm{e}}}\,, \tag{4.4.10}$$

where we used $\varepsilon K_2(N+1) \leq \frac{c}{\mathrm{e}}$. Similarly, we get

$$\begin{aligned}
\sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} \varepsilon^{2n} (K_2(2n+1))^{2n+1} &\leq K_2 \sum_{n=0}^{\lfloor \frac{N-1}{2} \rfloor} (\varepsilon K_2 (N+1))^{2n} (2n+1) \\
&\leq K_2 \sum_{n=0}^{N} \left(\frac{c}{\mathrm{e}}\right)^n (n+1) \\
&\leq K_2 \left(\sum_{n=0}^{\infty} \left(\frac{c}{\mathrm{e}}\right)^n n + \sum_{n=0}^{\infty} \left(\frac{c}{\mathrm{e}}\right)^n \right) \\
&= K_2 \frac{\mathrm{e}^2}{(\mathrm{e}-c)^2}\,,
\end{aligned} \tag{4.4.11}$$

where we used the geometric series variant $\sum_{n=0}^{\infty} q^n n = \frac{q}{(1-q)^2}$ for any $q \in \mathbb{R}$ with $|q| < 1$. At this point, Theorem 4.3.7 and (4.4.10)-(4.4.11) imply for $\varepsilon \in (0, \tilde{\varepsilon}_0]$

$$\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} \leq C\varepsilon^N K_2^{N+1} (N+1)^{N+1} + C \sum_{n=2}^{N} \sum_{k=2+N-n}^{N} \varepsilon^{n+k-1} K_2^{n+k} n^n k^k\,, \tag{4.4.12}$$

where $C > 0$ is some constant independent of ε. Now, for the first term in (4.4.12) we have that

$$\varepsilon^N K_2^{N+1} (N+1)^{N+1} \leq \frac{1}{\varepsilon} \left(\frac{c}{\mathrm{e}}\right)^{\lfloor \frac{c}{\mathrm{e}\, K_2 \varepsilon} \rfloor} \leq \frac{\mathrm{e}}{c\varepsilon} \left(\frac{c}{\mathrm{e}}\right)^{\frac{c}{\mathrm{e}\, K_2 \varepsilon}} = \frac{\mathrm{e}}{c\varepsilon} \exp\left(-\frac{r}{\varepsilon}\right)\,, \tag{4.4.13}$$

with $r := \frac{c\ln(\mathrm{e}/c)}{\mathrm{e}\,K_2} > 0$. Finally, the second term in (4.4.12) can be estimated as follows:

$$\begin{aligned}
\sum_{n=2}^{N}\sum_{k=2+N-n}^{N} \varepsilon^{n+k-1}K_2^{n+k}n^n k^k &\leq \frac{1}{\varepsilon}\sum_{n=2}^{N}\sum_{k=2+N-n}^{N}(\varepsilon K_2 N)^n(\varepsilon K_2 N)^k \\
&= \frac{1}{\varepsilon}(\varepsilon K_2 N)^{N+2}\sum_{n=0}^{N-2}\sum_{k=0}^{n}(\varepsilon K_2 N)^k \\
&\leq \frac{1}{\varepsilon}(\varepsilon K_2 N)^{N+2}\sum_{n=0}^{N-2}\sum_{k=0}^{\infty}\left(\frac{c}{\mathrm{e}}\right)^k \\
&= \frac{1}{\varepsilon}(\varepsilon K_2 N)^{N+2}\frac{N-1}{1-\frac{c}{\mathrm{e}}} \\
&\leq \frac{K_2}{1-\frac{c}{\mathrm{e}}}\left(\frac{c}{\mathrm{e}\,K_2\varepsilon}\right)^2\left(\frac{c}{\mathrm{e}}\right)^{\left\lfloor\frac{c}{\mathrm{e}\,K_2\varepsilon}\right\rfloor} \\
&\leq \frac{c}{(\mathrm{e}-c)K_2\varepsilon^2}\left(\frac{c}{\mathrm{e}}\right)^{\frac{c}{\mathrm{e}\,K_2\varepsilon}} \\
&= \frac{c}{(\mathrm{e}-c)K_2\varepsilon^2}\exp\left(-\frac{r}{\varepsilon}\right). \qquad (4.4.14)
\end{aligned}$$

We observe that the r.h.s. of (4.4.13) can be bounded by the r.h.s. of (4.4.14) (up to a multiplicative constant) for $\varepsilon \in (0, \tilde{\varepsilon}_0]$. Thus, the claim follows. □

Remark 4.4.2. *We note that the specific value N from the proof of Corollary 4.4.1 is not necessarily equal to the optimal truncation order N_{opt}. However, as a consequence of Corollary 4.4.1, and by the definition of N_{opt}, we conclude that*

$$\|\varphi - \varphi_{N_{opt}}^{WKB}\|_{L^\infty(I)} = \mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon)), \quad \varepsilon \to 0, \quad r > 0. \qquad (4.4.15)$$

Finally, we note that, apart from N_{heu} and $\widehat{N}_{heu}$, another option for predicting the optimal truncation order N_{opt} is to find the minimizer of error estimate (4.3.33) (for ε fixed), say $\widehat{N}_{opt} = \widehat{N}_{opt}(\varepsilon)$, although this rather complicated expression can only be minimized numerically by brute force.

At this point, it seems convenient to summarize the notations and meanings of the different mentioned truncation orders which aim to estimate N_{opt} – see Table 4.4.1. In the next section we will compare results for each truncation order from Table 4.4.1, since it is not clear a priori which of these orders provides the most accurate prediction of N_{opt}. Nonetheless, let us note that in our experiments N_{opt}, $\widehat{N}_{opt}$, and N_{heu} can only be determined by brute force, while $\widehat{N}_{heu}$ is given explicitly by formula (4.4.7).

4.5 Numerical simulations

In this section we present several numerical simulations to illustrate some of the theoretical results we derived in Section 4.3. To this end, we will compute the (perturbed) WKB approximation as described in Section 4.4.1. That is, the functions S_n' are pre-computed

N_{opt}	minimizer of $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)}$ (optimal truncation order)
$\widehat{N}_{opt}$	minimizer of error estimate (4.3.33) (prediction of N_{opt})
N_{heu}	minimizer of $\varepsilon^N \|S_{N+1}\|_{L^\infty(I)}$ (heuristic prediction of N_{opt})
$\widehat{N}_{heu}$	minimizer of the r.h.s. of (4.4.5) (prediction of N_{heu})

Table 4.4.1: Terminology for the different truncation orders mentioned in Section 4.4. The numbers $\widehat{N}_{opt}$ and $\widehat{N}_{heu}$ are predictions for N_{opt} and N_{heu} by means of (4.3.33) and (4.4.5), respectively.

symbolically and are then integrated numerically using the Clenshaw-Curtis algorithm based on a Chebyshev grid with $M+1$ grid points, where M will be specified later. All computations are carried out using MATLAB version 9.13.0.2049777 (R2022b). Further, since we are dealing with very small errors for the WKB approximation, especially when investigating the optimal truncation order, we use the Advanpix Multiprecision Computing Toolbox for MATLAB [Adv23] with quadruple-precision to avoid rounding errors.

4.5.1 Example 1: Airy equation

Consider the initial value problem

$$\begin{cases} \varepsilon^2 \varphi''(x) + x\varphi(x) = 0\,, \quad x \in [1,2]\,, \\ \varphi(1) = \mathrm{Ai}(-\frac{1}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}(-\frac{1}{\varepsilon^{2/3}})\,, \\ \varepsilon\varphi'(1) = -\varepsilon^{1/3}\left(\mathrm{Ai}'(-\frac{1}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}'(-\frac{1}{\varepsilon^{2/3}})\right)\,, \end{cases} \tag{4.5.1}$$

where the exact solution is given by

$$\varphi_{exact}(x) = \mathrm{Ai}(-\frac{x}{\varepsilon^{2/3}}) + \mathrm{i}\,\mathrm{Bi}(-\frac{x}{\varepsilon^{2/3}})\,. \tag{4.5.2}$$

Here, Ai and Bi denote the Airy functions of first and second kind, respectively (e.g., see [OLBC10, Chap. 9]). Note that for this example, where $a(x) = x$, the derivatives S_n' are given by powers of x (up to a constant factor). Hence, the functions $S_n^\pm$ can be computed exactly from (4.2.10); however, we shall use them here only as reference solutions for the approximations $\widetilde{S}_n^\pm$. Indeed, for a fixed number $M+1$ of Chebyshev grid points, we are then able to compute explicitly the approximation error $\|S_n^\pm - \widetilde{S}_n^\pm\|_{L^\infty(I)} =: e_n$. Since $\widetilde{S}_n^\pm$ is only available at the grid points, we actually compute the discrete analog of this norm.

On the left of Figure 4.5.1 the real part of φ_{exact} is plotted for the choice $\varepsilon = 2^{-8}$, which illustrates well the highly oscillatory behavior of the solution. Let us first investigate numerically the result from Corollary 4.3.3. For this, let us compute a constant K_2, as indicated by the proof of Corollary 4.3.3 and Remark 4.3.4. Indeed, by using the minimization strategy from (4.3.22), we find that (note that here $S_0'(z) = \pm\mathrm{i}\sqrt{z}$; $\delta = 1$)

$$K_2 = \frac{\mathrm{e}}{2\,\mathrm{e}-1} \min_{0<\delta\le 1} \frac{\sqrt{2+\delta}}{\delta} = \frac{\sqrt{3}\,\mathrm{e}}{2\,\mathrm{e}-1} \approx 1.0612 \tag{4.5.3}$$

is a suitable constant within the context of Corollary 4.3.3. On the right of Figure 4.5.1, we present the $L^\infty(I)$-norms of the functions S_n' and the approximations $\widetilde{S}_n$ when using

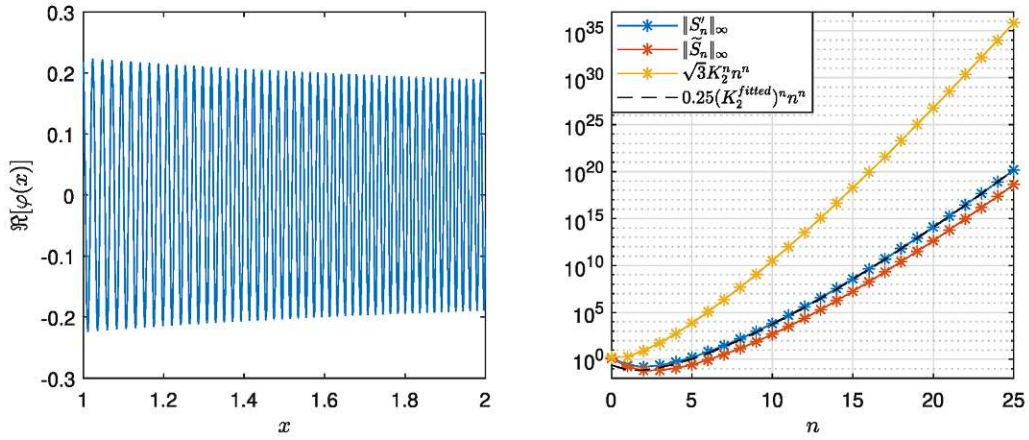

Figure 4.5.1: Left: Real part of the exact solution (4.5.2) of IVP (4.5.1) for the choice $\varepsilon = 2^{-8}$. Right: $L^\infty(I)$-norm of S_n' and $\widetilde{S}_n$ as functions of n, for the example $a(x) = x$ on the interval $[1, 2]$.

$M = 25$, along with the theoretical bound (4.3.20) on $\|S_n'\|_{L^\infty(I)}$. We observe that the true norms consistently remain below the theoretical bound. Additionally, we include as a dashed line the theoretical bound (4.3.20) when replacing K_2 and $\|S_0'\|_{L^\infty(G)}$ by the fitted values $K_2^{fitted} = 10/37 \approx 0.27$ and 0.25, respectively. We observe very good agreement between the norms $\|S_n'\|_{L^\infty(I)}$ and the dashed line. This demonstrates well that, in the present example, the norms $\|S_n'\|_{L^\infty(I)}$ grow as Corollary 4.3.3 suggests, i.e., $\|S_n'\|_{L^\infty(I)} \sim CK_2^n n^n$ as $n \to \infty$, for some $C, K_2 > 0$. In general, however, this is not the case. We refer to Appendix 4.A and Section 4.5.3 for an example, where the functions S_n' and S_n even decay, as $n \to \infty$.

Next, we investigate numerically the behavior of the WKB approximation error $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ as a function of ε. We may compare the results with the error "estimate" (4.3.43). As a first test, we set $M = 8$ to compute $\widetilde{\varphi}_N^{WKB}$. This results in an approximation error for S_n of $e_n \approx 10^{-8}$, $n = 0, \dots, 4$. On the left of Figure 4.5.2 we plot for $N = 0, \dots, 4$ the error as a function of ε: For $N = 2, 3, 4$ and small values of ε, the $\mathcal{O}(\varepsilon^{-1})e_0$-term is dominant. In contrast, for $N = 0$ and $N = 1$ this error term is not visible for the given range of ε-values such that the $\mathcal{O}(\varepsilon^N)$-term is dominant. As a second test, we set again $M = 8$, but now use in $\widetilde{\varphi}_N^{WKB}$ the exactly computed function S_0. The $\mathcal{O}(\varepsilon^{-1})e_0$-term from (4.3.43) is thus eliminated. On the right of Figure 4.5.2 we show again the error $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ as a function of ε: For $N = 2, 3, 4$ and small ε-values, the $\mathcal{O}(\varepsilon^0)e_1$-term, which is the next term in the sum in (4.3.43), now dominates. Indeed, the error curves show an almost constant value of approximately $2 \cdot 10^{-9}$ for small values of ε. For larger ε, the error curves behave like $\mathcal{O}(\varepsilon^N)$. As a third test, we set $M = 25$ and approximate again all functions S_n, $n = 0, \dots, 4$ (as in the first test). The corresponding approximation errors of S_n are $e_n \approx 10^{-23}$, $n = 0, \dots, 4$. On the left of Figure 4.5.3 we present the resulting WKB approximation errors. We observe that, on this scale, all $\mathcal{O}(\varepsilon^{n-1})e_n$-terms in the sum of (4.3.43) are essentially eliminated, since all the shown error curves behave like $\mathcal{O}(\varepsilon^N)$. Overall, we observe very good agreement between the numerical results of each of the three tests and the statements from Theorem 4.3.9 and Remark 4.3.10.

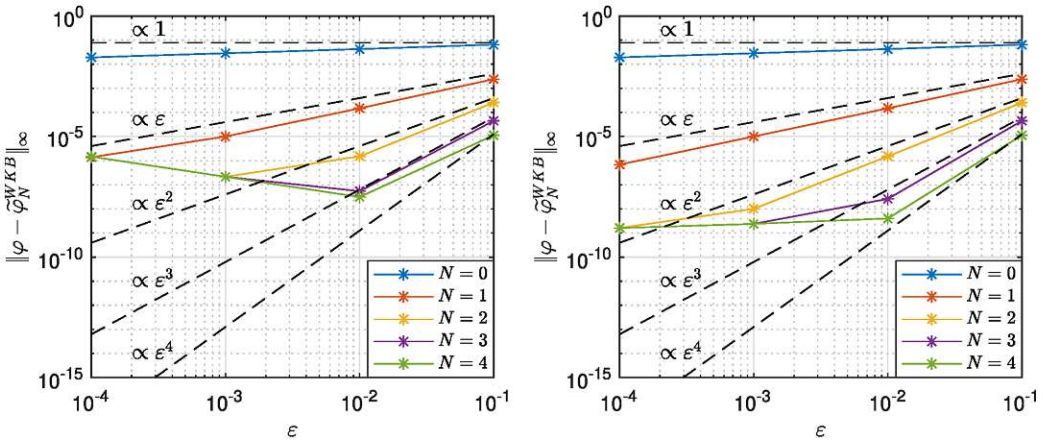

Figure 4.5.2: $L^\infty(I)$-norm of the error of the WKB approximation as a function of ε, for the IVP (4.5.1) and several choices of N. Left: $M = 8$. Right: $M = 8$; using the exactly computed function S_0.

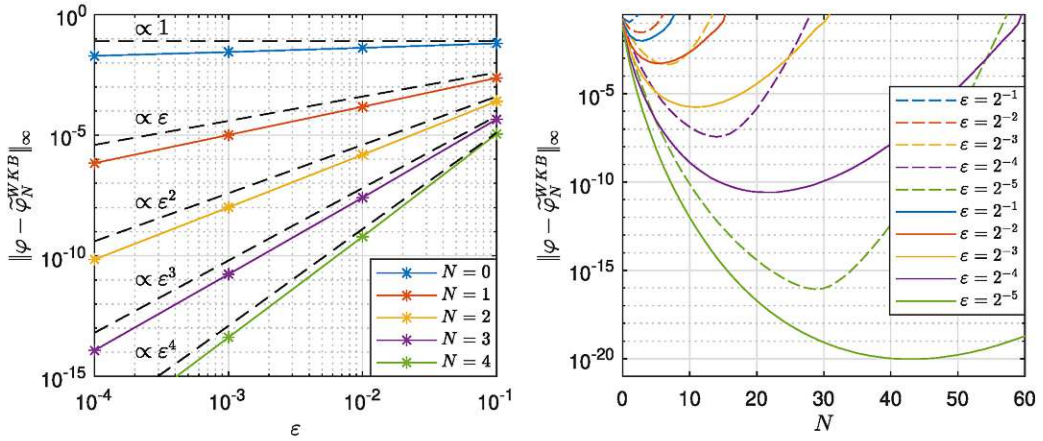

Figure 4.5.3: Left: $L^\infty(I)$-norm of the error of the WKB approximation as a function of ε, for the IVP (4.5.1) and several choices of N. Here, we set $M = 25$. Right: $L^\infty(I)$-norm of the error of the WKB approximation as a function of N, for the IVP (4.5.1) and several choices of ε. The dash-dotted lines correspond to the error estimate according to Theorem 4.3.7 and the solid lines correspond to the actual error of the WKB approximation.

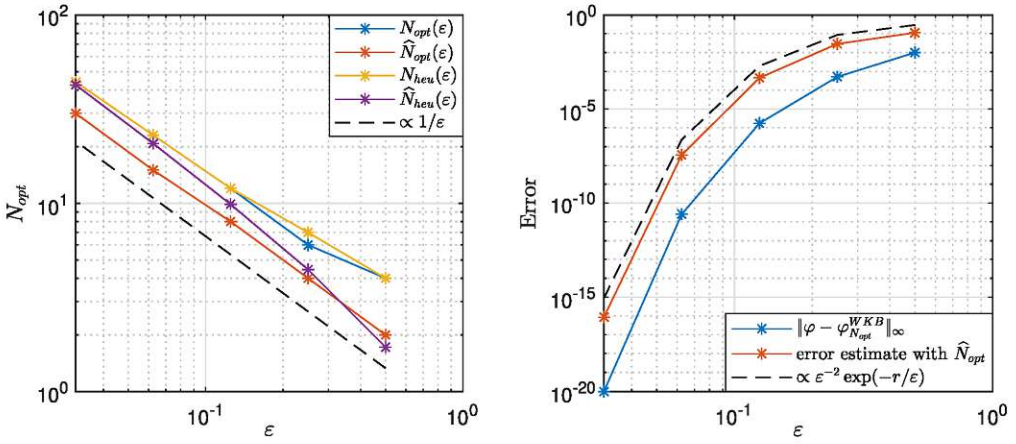

Figure 4.5.4: Left: The optimal truncation order N_{opt} as well as the predicted "optimal" orders $\widehat{N}_{opt}$, N_{heu}, and $\widehat{N}_{heu}$ as functions of ε. The dashed line is proportional to $1/\varepsilon$. Right: The optimal error achieved by using N_{opt} as well as error estimate (4.3.33) when using $N = \widehat{N}_{opt}$, both as functions of ε. The dashed line is proportional to $\varepsilon^{-2}\exp(-\frac{13}{10\varepsilon})$.

Next we investigate the error of the (perturbed) WKB approximation as a function of the truncation order N. For this, we set again $M = 25$, yielding approximation errors of S_n as $e_n \approx 10^{-23}$ for $n = 0, 1, \ldots$. We may therefore neglect the errors caused by approximating the functions S_n. On the right of Figure 4.5.3 we plot the actual error $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ and its error estimate (4.3.33) while again using $K_2^{fitted} = 10/37$, both as functions of N, for several ε-values. We observe that, even when using the fitted constant K_2^{fitted}, the "optimal" truncation order $\widehat{N}_{opt}$, as predicted by the estimate (4.3.33), is smaller than N_{opt} (determined as the argmin of the actual error curve). For instance, we have $\widehat{N}_{opt}(2^{-4}) \approx 14 < 22 \approx N_{opt}(2^{-4})$ and $\widehat{N}_{opt}(2^{-5}) \approx 29 < 44 \approx N_{opt}(2^{-5})$, respectively. In Figure 4.5.4 we plot on the left the optimal truncation order $N_{opt}(\varepsilon)$ as a function of ε as well as its predictions $\widehat{N}_{opt}(\varepsilon)$, $N_{heu}(\varepsilon)$, and $\widehat{N}_{heu}(\varepsilon)$. The plot suggests that N_{opt}, $\widehat{N}_{opt}$, and N_{heu} are proportional to ε^{-1}, as $\varepsilon \to 0$ (for $\widehat{N}_{heu}$ this is already evident from (4.4.7)). Further, for $\varepsilon = 2^{-1}, 2^{-3}, 2^{-4}, 2^{-5}$ we observe that $N_{opt} = N_{heu}$. On the right of Figure 4.5.4 we plot the corresponding optimal error which is achieved by using N_{opt} as well as error estimate (4.3.33) when using $N = \widehat{N}_{opt}$, both as a function of ε. As indicated by the dashed line, the optimal error decays like $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, with $r \approx 13/10$ being a fitted value, in good agreement with Remark 4.4.2.

4.5.2 Example 2

As our second example let us consider the initial value problem

$$\begin{cases} \varepsilon^2\varphi''(x) + \mathrm{e}^{5x}\,\varphi(x) = 0\,, \quad x \in [0,1]\,, \\ \varphi(0) = 1\,, \\ \varepsilon\varphi'(0) = 0\,, \end{cases} \tag{4.5.4}$$

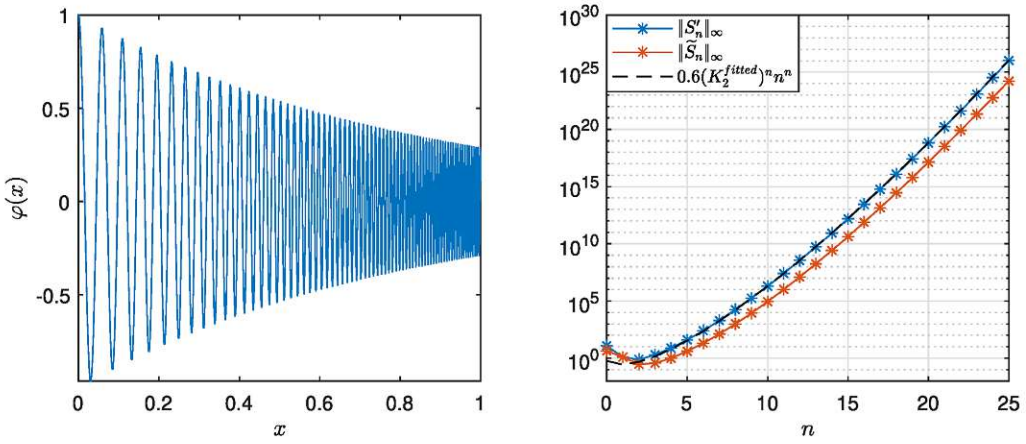

Figure 4.5.5: Left: Exact solution (4.5.5) of IVP (4.5.4) for the choice $\varepsilon = 10^{-2}$. Right: $L^\infty(I)$-norm of S_n' and $\widetilde{S}_n$ as functions of n, for the example $a(x) = \exp(5x)$ on the interval $I = [0, 1]$.

where the exact solution is given by[4]

$$\varphi_{exact}(x) = \frac{J_0\left(\frac{2}{5\varepsilon}\,\mathrm{e}^{5x/2}\right) Y_1\left(\frac{2}{5\varepsilon}\right) - Y_0\left(\frac{2}{5\varepsilon}\,\mathrm{e}^{5x/2}\right) J_1\left(\frac{2}{5\varepsilon}\right)}{J_0\left(\frac{2}{5\varepsilon}\right) Y_1\left(\frac{2}{5\varepsilon}\right) - J_1\left(\frac{2}{5\varepsilon}\right) Y_0\left(\frac{2}{5\varepsilon}\right)}\,. \tag{4.5.5}$$

Here, J_ν and Y_ν denote the Bessel functions of first and second kind of order ν, respectively (e.g., see [OLBC10, Chap. 10]).

On the left of Figure 4.5.5 the exact solution φ_{exact} is plotted for the choice $\varepsilon = 10^{-2}$. Throughout the whole interval, due to the fast growth of the function $a(x) = \exp(5x)$, the solution exhibits a rapid increase of its oscillatory behavior. Further, we plot on the right of Figure 4.5.5 the $L^\infty(I)$-norms of the derivatives S_n' and the approximations $\widetilde{S}_n$ when using $M = 30$. As indicated by the dashed line, the smallest (fitted) constant K_2 such that estimate (4.3.20) holds is $K_2^{fitted} \approx 9/20$ (here we also replaced $\|S_0'\|_{L^\infty(G)}$ in (4.3.20) by the fitted value 0.6). In Figure 4.5.6 we present the WKB approximation error $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ as a function of ε and may again compare the results with the error "estimate" (4.3.43). We observe that, on this scale, all $\mathcal{O}(\varepsilon^{n-1})e_n$-terms are essentially eliminated, since all the shown error curves behave like $\mathcal{O}(\varepsilon^N)$. Overall, we observe very good agreement with the statements from Theorem 4.3.9 and Remark 4.3.10. Finally, we plot in Figure 4.5.7 on the left the optimal truncation order N_{opt} as well as its predictions $\widehat{N}_{opt}$, N_{heu}, and $\widehat{N}_{heu}$, as functions of ε. We find that N_{opt} is proportional to ε^{-1}, as $\varepsilon \to 0$. On the right of Figure 4.5.7 we present the corresponding optimal error as well as error estimate (4.3.33) when using $N = \widehat{N}_{opt}$, both as a function of ε. As the dashed line indicates, the error decays like $\mathcal{O}(\varepsilon^{-2}\exp(-r/\varepsilon))$, with $r \approx 3/4$ being a fitted value. This is in good agreement with Remark 4.4.2.

[4] We found the exact solution by using the Symbolic Math Toolbox of MATLAB.

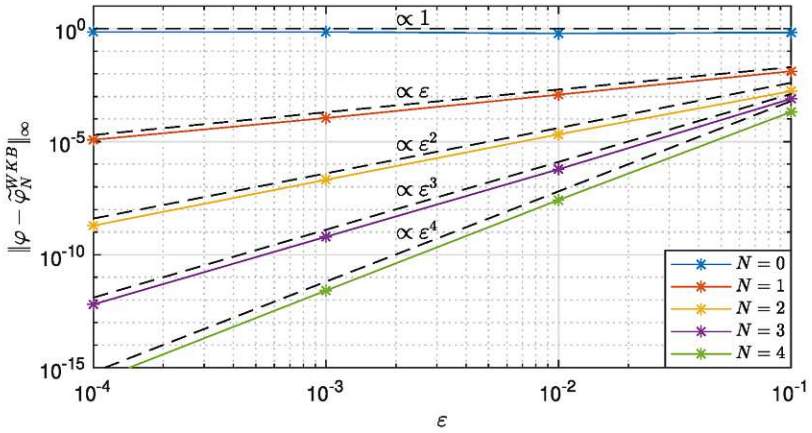

Figure 4.5.6: $L^\infty(I)$-norm of the error of the WKB approximation as a function of ε, for the IVP (4.5.4) and several choices of N. Here, we set $M = 30$.

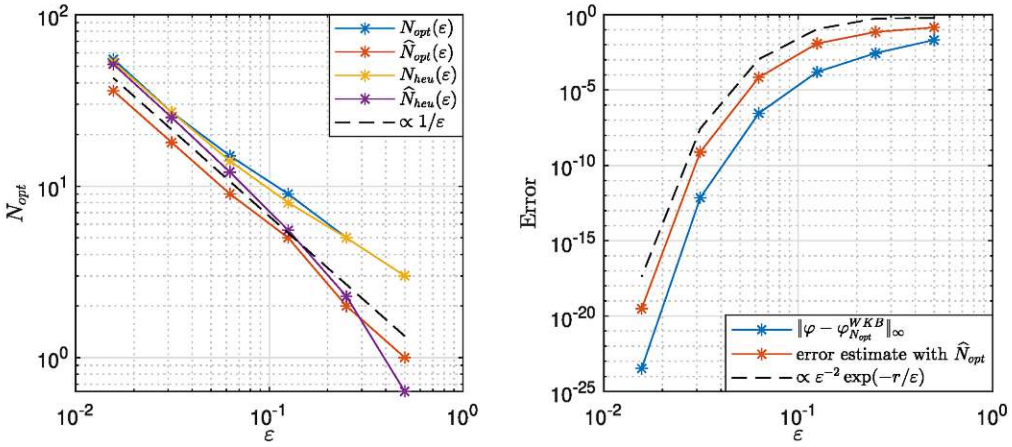

Figure 4.5.7: Left: The optimal truncation order N_{opt} as well as the predicted "optimal" orders $\widehat{N}_{opt}$, N_{heu}, and $\widehat{N}_{heu}$ as functions of ε. The dashed line is proportional to $1/\varepsilon$. Right: The optimal error achieved by using N_{opt} as well as error estimate (4.3.33) when using $N = \widehat{N}_{opt}$, both as functions of ε. The dashed line is proportional to $\varepsilon^{-2}\exp(-\frac{3}{4\varepsilon})$.

4.5.3 Example 3: Convergent WKB approximation

As a final example, let us consider $a(x) = (1+x+x^2)^{-2}$. We are interested in investigating the initial value problem

$$\begin{cases} \varepsilon^2\varphi''(x) + (1+x+x^2)^{-2}\varphi(x) = 0\,, \quad x \in [0,1]\,, \\ \varphi(0) = 1\,, \\ \varepsilon\varphi'(0) = 1\,, \end{cases} \tag{4.5.6}$$

where the exact solution φ_{exact} is given by[5]

$$\begin{aligned} \varphi_{exact}(x) = \frac{a(x)^{-1/4}}{\sqrt{3}\gamma(\varepsilon)} \sin\left(\gamma(\varepsilon)\left(\arctan\left(\frac{2x+1}{\sqrt{3}}\right) - \frac{\pi}{6}\right)\right) \\ - a(x)^{-1/4}\cos\left(\gamma(\varepsilon)\left(\arctan\left(\frac{2x+1}{\sqrt{3}}\right) - \frac{\pi}{6}\right)\right)\,, \end{aligned} \tag{4.5.7}$$

where $\gamma(\varepsilon) := \sqrt{3\varepsilon^2+4}/(\sqrt{3}\varepsilon)$.

This example is special in the sense that $a(x) = (1+x+x^2)^{-2}$ belongs to the class of functions represented as $(C_1 + C_2x + C_3x^2)^{-2}$, with constants C_i, $i = 1,2,3$, satisfying $|C_2| + |C_3| > 0$ and $C_2^2 \neq 4C_1C_3$. Further details regarding this class of functions are discussed in Appendix 4.A, particularly with regard to the corresponding WKB series. Notably, for such functions it holds that $S_1' \not\equiv 0$, $S_2' \not\equiv 0$ and $S_3' \equiv 0$, see Remark 4.A.2. Moreover, according to Proposition 4.A.1 and Remark 4.A.3, it follows that

$$S_{2n}' = \mathcal{O}\left((n-1)^{-3/2}|C_1C_3 - C_2^2/4|^{n-1}\right)\,, \quad n \to \infty\,, \tag{4.5.8}$$

$$S_{2n+1}' \equiv 0\,, \quad n \geq 1\,. \tag{4.5.9}$$

Consequently, this implies that the underlying asymptotic series (4.2.2) is (geometrically) convergent for any $\varepsilon \leq |C_1C_3 - C_2^2/4|^{-1/2}$, see again Remark 4.A.3. In this case, given that $|C_1C_3 - C_2^2/4| = 3/4$, the functions S_{2n}' (and hence S_{2n}) exhibit exponential decay as $n \to \infty$, uniformly in $x \in I$. The corresponding WKB series is convergent for any $\varepsilon \in (0, 2/\sqrt{3}]$.

In Figure 4.5.8 on the left we plot φ_{exact} for the choice $\varepsilon = 2^{-9}$. Moreover, on the right of Figure 4.5.8 we plot the $L^\infty(I)$-norm of S_n' and $\widetilde{S}_n$, both as a function of n. Here, we set $M = 30$ for the numerical integration of the functions S_n'. We observe that the norms indeed decay exponentially, in agreement with Remark 4.A.3. Here, the dashed line is precisely given by the r.h.s. of (4.A.7) with $C_1 = C_2 = C_3 = 1$. In Figure 4.5.9 on the left we plot for $N = 0,\dots,4$ the error of the WKB approximation $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ as a function of ε. By comparing the results with (4.3.43), we observe that all $\mathcal{O}(\varepsilon^{n-1})e_n$-terms are essentially eliminated. Further, the error curves for $N = 0,1,3$ behave like $\mathcal{O}(\varepsilon^N)$ whereas the curves corresponding to the choices $N = 2,4$ behave like $\mathcal{O}(\varepsilon^{N+1})$. This is because the given function a implies $S_{2n+1}' \equiv 0$, for any $n \geq 1$, which means $\varphi_N^{WKB} = \varphi_{N+1}^{WKB}$ for any even $N \geq 2$.

[5] We found the exact solution by using the Symbolic Math Toolbox of MATLAB.

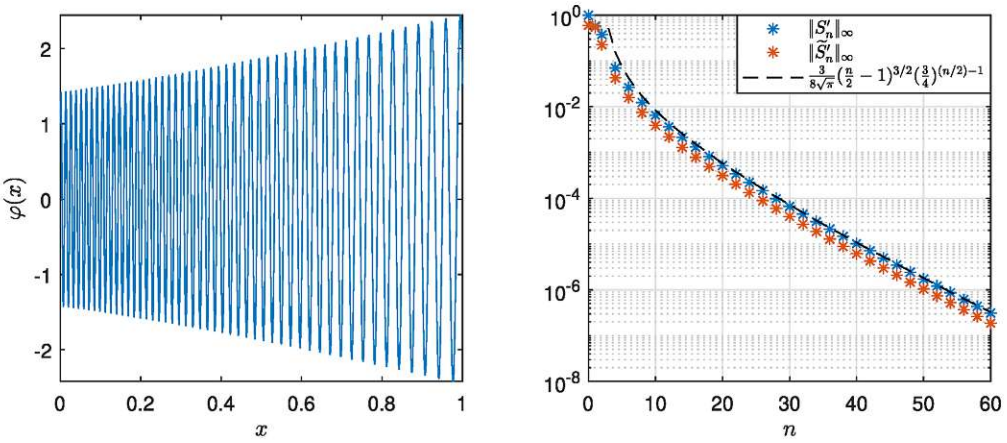

Figure 4.5.8: Left: Exact solution (4.5.7) of IVP (4.5.6) for the choice $\varepsilon = 2^{-9}$. Right: $L^\infty(I)$-norm of S_n' and $\widetilde{S}_n$ as functions of even n, for the example $a(x) = (1+x+x^2)^{-2}$ on the interval $I = [0,1]$. The dashed line is proportional to the r.h.s. of (4.A.7) with $C_1 = C_2 = C_3 = 1$.

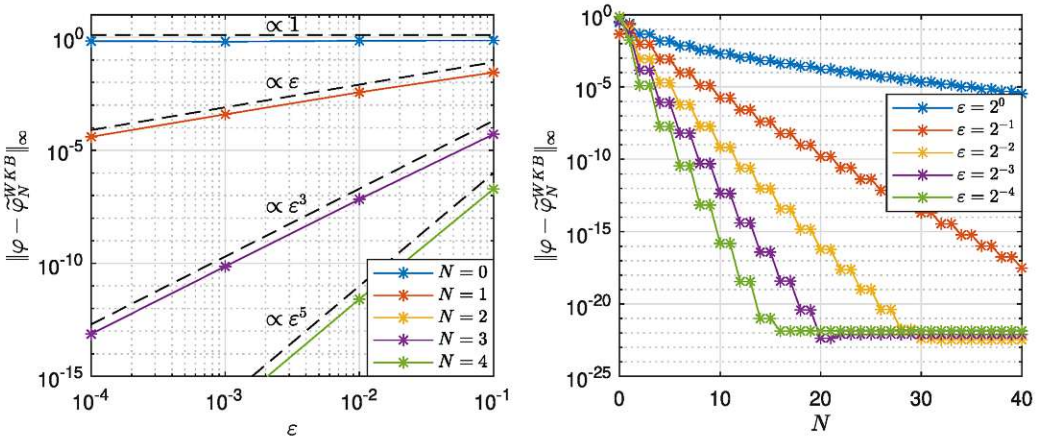

Figure 4.5.9: Left: $L^\infty(I)$-norm of the error of the WKB approximation as a function of ε, for the IVP (4.5.6) and several choices of N. Here, we set $M = 30$. The yellow curve for $N = 2$ is the same as for $N = 3$ and hence not visible in the shown plot. Right: $L^\infty(I)$-norm of the error of the WKB approximation as a function of N, for the IVP (4.5.6) and several choices of ε. Here, we set $M = 30$.

Finally, on the right of Figure 4.5.9 we plot the error $\|\varphi - \widetilde{\varphi}_N^{WKB}\|_{L^\infty(I)}$ as a function of the truncation order N, for several ε-values. We observe that all shown error curves are decreasing functions in N, up to the point where they reach values of approximately 10^{-22}. This is due to the approximation of the functions S_n. More precisely, the first term of the sum in (4.3.43), namely, the $\mathcal{O}(\varepsilon^{-1})e_0$-term corresponding to the approximation of S_0, becomes dominant at this point. For this reason, the minimum achievable error level is growing with decreasing ε. Besides from this saturation effect, the plot aligns well with Remark 4.A.3, suggesting that the WKB approximation converges to the exact solution of IVP (4.5.6) as $N \to \infty$, for all displayed ε-values. Furthermore, one can observe again the fact that $\varphi_N^{WKB} = \varphi_{N+1}^{WKB}$ for even $N \geq 2$, as indicated by the step-like behavior of all shown error curves.

4.6 Conclusion

In this chapter we analyzed the WKB approximation of the solution to a highly oscillatory initial value problem. Assuming that the potential in the equation is analytic, we found explicit upper bounds for the terms occurring in the asymptotic WKB series of the approximate solution. Building on that, we proved error estimates which are explicit not only w.r.t. the small parameter ε but also w.r.t. N, the chosen number of terms in the truncated asymptotic series. We showed that the optimal truncation order N_{opt} is proportional to ε^{-1}, and this results in an approximation error that is exponentially small w.r.t. ε. We confirmed our theoretical results by several numerical experiments.

Appendix

4.A Convergent WKB series

In this appendix we provide examples where the asymptotic series (4.2.2) is convergent in $L^\infty(I)$. In practice, the norms $\|S_n'\|_{L^\infty(I)}$ (and $\|S_n\|_{L^\infty(I)}$) often decrease up to a certain number of n before they start to increase rapidly, e.g., see the right plot of Figure 4.5.1. However, there are examples where one can easily verify that this is not the case. For instance, consider the simplest case in which $a \equiv a_0$ is constant. By (4.2.5) this is equivalent to $S_1' \equiv 0$, which by (4.2.6) then implies $S_n' \equiv 0$ for every $n \geq 1$. Similarly, one easily verifies that $S_2' \equiv 0$ is equivalent to a having the form $a(x) = (C_1 + C_2 x)^{-4}$ for some constants C_1 and C_2, see also [BO99, Problem 10.2]. It then holds $S_n' \equiv 0$ for every $n \geq 2$. Thus, in both of the just mentioned cases, the asymptotic series (4.2.2) terminates automatically and is therefore convergent. The corresponding WKB approximation (4.2.9) with $N \geq 0$ (respectively $N \geq 1$) is then the exact solution to IVP (4.1.1). Indeed, revisiting (4.3.24), it is clear that the r.h.s. in (4.3.35) then vanishes, i.e. $\|\varphi - \varphi_N^{WKB}\|_{L^\infty(I)} = 0$.

In the subsequent discussion, we will give examples of convergent WKB series which do not terminate automatically.

Proposition 4.A.1. *Let $S_3' \equiv 0$. Then it holds*

$$S_{2n}' = S_2' \left(-\frac{S_2'}{2S_0'} \right)^{n-1} a_n \,, \tag{4.A.1}$$

$$S_{2n+1}' \equiv 0 \,, \tag{4.A.2}$$

for $n \geq 2$. Here, the sequence a_n is recursively defined by $a_1 := 1$ and

$$a_{n+1} := \sum_{j=1}^{n} a_j a_{n+1-j} \,, \quad n \geq 1 \,. \tag{4.A.3}$$

Proof. It is easy to check, that (4.A.1) and (4.A.2) hold for $n = 2$. We proceed now by induction on n. To this end, assume that formulas (4.A.1) and (4.A.2) hold for all $2 \leq k \leq n$ for some fixed $n \geq 2$. We shall now prove them for $n+1$. The induction hypothesis implies

that $S''_{2n+1} \equiv 0$ as well as $S'_j \equiv 0$ for all odd indices j such that $1 \leq j \leq 2n+1$. Hence,

$$\begin{aligned} S'_{2n+2} &= -\frac{1}{2S'_0}\left(\sum_{j=1}^{2n+1} S'_j S'_{2n+2-j} + S''_{2n+1}\right) \\ &= -\frac{1}{2S'_0}\sum_{j=1}^{n} S'_{2j} S'_{2(n+1-j)} \\ &= S'_2\left(-\frac{S'_2}{2S'_0}\right)^n \sum_{j=1}^{n} a_j a_{n+1-j} \\ &= S'_2\left(-\frac{S'_2}{2S'_0}\right)^n a_{n+1}\,, \end{aligned} \tag{4.A.4}$$

where we have again used the induction hypothesis in the third equation. Differentiating (4.A.4) and using $\frac{S''_0}{S'_0} = -2S'_1$ we further obtain

$$S''_{2n+2} = \left(-\frac{S'_2}{2S'_0}\right)^n \left(2nS'_1S'_2 + (n+1)S''_2\right)a_{n+1}\,. \tag{4.A.5}$$

Moreover, the induction hypothesis implies $S'_j S'_{2n+3-j} \equiv 0$ for $2 \leq j \leq 2n+1$ since either j or $2n+3-j$ is odd. Therefore, we get

$$\begin{aligned} S'_{2n+3} &= -\frac{1}{2S'_0}\left(\sum_{j=1}^{2n+2} S'_j S'_{2n+3-j} + S''_{2n+2}\right) \\ &= -\frac{1}{2S'_0}\left(2S'_1 S'_{2n+2} + S''_{2n+2}\right) \\ &= \left(-\frac{S'_2}{2S'_0}\right)^n \left(-\frac{1}{2S'_0}\right)\left(2(n+1)S'_1S'_2 + (n+1)S''_2\right)a_{n+1} \\ &= \left(-\frac{S'_2}{2S'_0}\right)^n (n+1)S'_3 a_{n+1} \\ &\equiv 0\,, \end{aligned} \tag{4.A.6}$$

by assumption on S'_3. This concludes the proof. □

Remark 4.A.2. *Proposition 4.A.1 assumes $S'_3 \equiv 0$, which is equivalent to $a(x)$ satisfying the third order nonlinear ODE $15a'^3 + 4a^2a''' - 18aa'a'' = 0$. With the aid of* MATLAB*'s Symbolic Math Toolbox we find that the general solution to this ODE is given by $a(x) = (C_1 + C_2x + C_3x^2)^{-2}$, where C_1, C_2 and C_3 are constants. A simple computation then shows that if $|C_2| + |C_3| > 0$ and $C_2^2 \neq 4C_1C_3$, the coefficient function a does not have one of the two forms mentioned before Proposition 4.A.1, i.e., $S'_1 \not\equiv 0$ and $S'_2 \not\equiv 0$. Thus, due to Proposition 4.A.1, the corresponding WKB series does not terminate in this case.*

Remark 4.A.3. *The numbers $a_n =: c_{n-1}$ in Proposition 4.A.1 are the so-called Catalan numbers (e.g., see [GKP91]), which are known to grow asymptotically as $c_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$,*

for $n \to \infty$. *Let us assume that* $a(x) = (C_1 + C_2 x + C_3 x^2)^{-2}$ *such that* $S_3' \equiv 0$, *see Remark 4.A.2. We then have* $S_2'(x)/S_0'(x) = C_1C_3/2 - C_2^2/8$. *According to (4.A.1), we thus have for* $n \geq 2$

$$\begin{aligned}\|S_{2n}'\|_{L^\infty(I)} &\leq \|S_2'\|_{L^\infty(I)} \left\| \frac{S_2'}{2S_0'} \right\|_{L^\infty(I)}^{n-1} c_{n-1} \\ &\sim \frac{\|S_2'\|_{L^\infty(I)}}{(n-1)^{3/2}\sqrt{\pi}} \left\| \frac{2S_2'}{S_0'} \right\|_{L^\infty(I)}^{n-1}, \quad n \to \infty \\ &= \frac{\|S_2'\|_{L^\infty(I)}}{(n-1)^{3/2}\sqrt{\pi}} \left| C_1C_3 - \frac{C_2^2}{4} \right|^{n-1}. \end{aligned} \tag{4.A.7}$$

By definition (4.2.10), we conclude that

$$\|S_{2n}\|_{L^\infty(I)} = \mathcal{O}\left((n-1)^{-3/2} |C_1C_3 - C_2^2/4|^{n-1} \right), \quad n \to \infty. \tag{4.A.8}$$

Thus, the constants C_i, $i = 1, 2, 3$, *determine whether the function* $\|S_{2n}\|_{L^\infty(I)}$ *is exponentially growing or decaying, as* $n \to \infty$. *Note that Proposition 4.A.1 also implies that* $S_{2n+1} \equiv 0$ *for* $n \geq 1$. *A short calculation then shows that (4.A.8) implies that the corresponding WKB series* $\exp(\sum_{n=0}^\infty \varepsilon^{n-1} S_n(x))$ *is (geometrically) convergent for any* $x \in I$, *if* $\varepsilon \leq |C_1C_3 - C_2^2/4|^{-1/2}$.

Bibliography

[AAN11] A. Arnold, N. B. Abdallah, and C. Negulescu. WKB-Based Schemes for the Oscillatory 1D Schrödinger Equation in the Semiclassical Limit. *SIAM J. Numer. Anal.*, 49:1436–1460, 2011.

[AB22] F. J. Agocs and A. H. Barnett. An adaptive spectral method for oscillatory second-order linear ODEs with frequency-independent cost. *arXiv*, abs/2212.06924, 2022.

[Adv23] Advanpix LLC. *Multiprecision Computing Toolbox for* MATLAB *version 5.1.0.15432*. Advanpix LLC, 2023.

[AKKM24] A. Arnold, C. Klein, J. Körner, and J. M. Melenk. Optimally truncated WKB approximation for the 1D stationary Schrödinger equation in the highly oscillatory regime. *arXiv*, abs/2401.10141, 2024.

[AKU22] A. Arnold, C. Klein, and B. Ujvari. WKB-method for the 1D Schrödinger equation in the semi-classical limit: enhanced phase treatment. *BIT Numerical Mathematics*, 62:1–22, 2022.

[BE92] K. S. Breuer and R. M. Everson. On the errors incurred calculating derivatives using Chebyshev polynomials. *Journal of Computational Physics*, 99:56–67, 1992.

[BO99] C. M. Bender and Steven A. Orszag. *Advanced mathematical methods for scientists and engineers I: asymptotic methods and perturbation theory*. Springer New York, NY, 1 edition, 1999.

[Boy99] J. P. Boyd. The Devil's Invention: Asymptotic, Superasymptotic and Hyperasymptotic Series. *Acta Applicandae Mathematica*, 56:1–98, 1999.

[CC60] C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2:197–205, 1960.

[Cha68] M. M. Chawla. Error estimates for the Clenshaw-Curtis quadrature. *Mathematics of Computation*, 22:651–656, 1968.

[CS58] E. D. Courant and H. S. Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 281:360–408, 1958.

[Fun87] D. Funaro. A Preconditioning Matrix for the Chebyshev Differencing Operator. *SIAM Journal on Numerical Analysis*, 24:1024–1031, 1987.

[GKP91] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics - a foundation for computer science.* Addison-Wesley Publishing Company, 1991.

[JL03] T. Jahnke and C. Lubich. Numerical integrators for quantum dynamics close to the adiabatic limit. *Numerische Mathematik*, 94:289–314, 2003.

[Lew68] H. R. Lewis. Motion of a Time-Dependent Harmonic Oscillator, and of a Charged Particle in a Class of Time-Dependent, Axially Symmetric Electromagnetic Fields. *Phys. Rev.*, 172:1313–1315, 1968.

[LJL05] K. Lorenz, T. Jahnke, and C. Lubich. Adiabatic Integrators for Highly Oscillatory Second-Order Linear Differential Equations with Time-Varying Eigendecomposition. *BIT Numerical Mathematics*, 45:91–115, 2005.

[LL85] L. D. Landau and E. M. Lifschitz. *Quantenmechanik.* Akademie-Verlag, 1985.

[Mel97] J. M. Melenk. On the robust exponential convergence of *hp* finite element methods for problems with boundary layers. *IMA Journal of Numerical Analysis*, 17:577–601, 1997.

[MJK13] J.-F. Mennemann, A. Jüngel, and H. Kosina. Transient Schrödinger-Poisson simulations of a high-frequency resonant tunneling diode oscillator. *J. Comput. Phys.*, 239:187–205, 2013.

[MS03] J. Martin and D. J. Schwarz. WKB approximation for inflationary cosmological perturbations. *Physical Review D*, 67:083512, 2003.

[Neg05] C. Negulescu. *Asymptotical models and numerical schemes for quantum systems.* PhD thesis, Université Paul Sabatier, Toulouse, 2005.

[OLBC10] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. NIST Handbook of Mathematical Functions. Cambridge University Press, 2010.

[RR00] M. Robnik and V. G. Romanovski. Some properties of WKB series. *Journal of Physics A*, 33:5093–5104, 2000.

[SHMS98] J. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman. Resonant tunneling diodes: models and properties. *Proc. IEEE*, 86:641–660, 1998.

[Tre00] L. N. Trefethen. Spectral Methods in MATLAB. vol. 10 of Software, Environments, and Tools, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[Win05] S. Winitzki. Cosmological particle production and the precision of the WKB approximation. *Physical Review D*, 72:104011, 2005.