

# WKB-METHOD FOR THE 1D SCHRÖDINGER EQUATION IN THE SEMI-CLASSICAL LIMIT: ENHANCED PHASE TREATMENT

ANTON ARNOLD\*, CHRISTIAN KLEIN†, AND BERNHARD UJVARI‡

**Abstract.** This paper is concerned with the efficient numerical computation of solutions to the 1D stationary Schrödinger equation in the semiclassical limit in the highly oscillatory regime. A previous approach to this problem based on explicitly incorporating the leading terms of the WKB approximation is enhanced in two ways: first a refined error analysis for the method is presented for a not explicitly known WKB phase, and secondly the phase and its derivatives will be computed with spectral methods. The efficiency of the approach is illustrated for several examples.

**Key words.** Uniformly accurate scheme, Schrödinger equation, highly oscillating wave functions, higher order WKB-approximation, asymptotically correct finite difference scheme, spectral methods

**AMS subject classifications.** 35Q40, 81Q20, 65M70, 65L11

**1. Introduction.** This paper is concerned with the numerical solution of highly oscillating differential equations of the type

$$(1.1) \quad \varepsilon^2 \varphi''(x) + a(x)\varphi(x) = 0, \quad x \in \mathbb{R},$$

where  $0 < \varepsilon \ll 1$  is a very small parameter and  $a(x) \geq a_0 > 0$  a sufficiently smooth function. Such models play an important role in electromagnetic and acoustic scattering (1D Maxwell and Helmholtz equations in the high frequency regime), as well as wave propagation problems in quantum and plasma physics. More concretely, (1.1) has been used for the simulation of electron transport in nano-scale semiconductor devices (like 1D quantum models of resonant tunneling diodes [27, 23] or for the longitudinal mode(s) in 2D and 3D quantum waveguides [21]). In such applications  $\varepsilon := \hbar/\sqrt{2m}$  is proportional to the (reduced) Planck constant  $\hbar$ ,  $a(x) := E - V(x)$ , with some prescribed, real valued, electrostatic potential  $V$ , and  $E \in \mathbb{R}$  is the injection energy of the electrons (with effective mass  $m$ ) into the device from the metallic leads on both sides. Since  $E$  may take arbitrarily large values, (1.1) appears as a Schrödinger equation in the semi-classical regime, i.e. for  $\varepsilon \rightarrow 0$ .

For  $\varepsilon \ll 1$  the wave length  $\lambda = \frac{2\pi\varepsilon}{\sqrt{a(x)}}$  is very small, such that the solution  $\varphi$  becomes highly oscillating. In classical ODE-schemes (like in [13, 14]) such a situation requires a very fine mesh in order to accurately resolve the oscillations, see Fig. 1.1. Hence, standard numerical methods would be very costly and inefficient here.

A possible remedy is to use analytic a-priori information on the solution to simplify its numerical treatment. Within this spirit, various numerical strategies for highly oscillatory problems from quantum mechanics or of Hamiltonian structure have been developed in recent years. One group of approaches is based on *adiabatic integrators* (see [19, 18, 22]; §XIV of [11]), which is closely related to using the zeroth order WKB-approximation (cf. [20])  $\varphi(x) \approx C \exp\left(\pm \frac{i}{\varepsilon} \int_0^x \sqrt{a(\tau)} d\tau\right)$  to eliminate the dominant oscillations; the resulting smoother problem is then solved numerically. A

---

\*Institut für Analysis und Scientific Computing, Technische Universität Wien, Wiedner Hauptstr. 8-10, A-1040 Wien, Austria ([anton.arnold@tuwien.ac.at](mailto:anton.arnold@tuwien.ac.at)).

†Institut de Mathématiques de Bourgogne, Université de Bourgogne-Franche-Comté, 9 avenue Alain Savary, France. ([Christian.Klein@u-bourgogne.fr](mailto:Christian.Klein@u-bourgogne.fr)).

‡Institut für Analysis und Scientific Computing, Technische Universität Wien, Wiedner Hauptstr. 8-10, A-1040 Wien, Austria ([e0326211@student.tuwien.ac.at](mailto:e0326211@student.tuwien.ac.at)).

localized variant of this transformation is also the background for the *modified Magnus method* developed in §5 of [15]. But, according to the detailed numerical comparison in §5.1 of [22], their *adiabatic Magnus method* is actually even more accurate than both the standard [16, 24] and modified Magnus methods [15], particularly for very small  $\varepsilon$ . *Modulated Fourier expansions* are closely related techniques that apply to the case  $a = \text{const}$ , but allowing for nonlinearities in the ODE (see e.g. §XIII of [11]; [7]).

Other numerical approaches for (1.1) include a macroscopic reformulation [8] and, more recently, WKB-approximations (see [25, 1, 12]). The last three papers are concerned only with the oscillatory case (i.e.  $a(x) > 0$ ). The evanescent case (with  $a(x) < 0$ ) is equally important in applications (e.g. for quantum tunneling), but needs a different numerical approach, see [3] for a FEM with WKB-type basis functions. For the inclusion of a turning point we refer to [2].

WKB-based methods rest upon using asymptotically correct solution formulas for (1.1) in the semi-classical limit  $\varepsilon \rightarrow 0$ , in order to simplify the numerical problem. With the WKB-ansatz

$$(1.2) \quad \varphi(x) \sim \exp\left(\frac{1}{\varepsilon} \sum_{p=0}^{\infty} \varepsilon^p \phi_p(x)\right)$$

and a comparison of coefficients one obtains the well-known WKB-approximation in arbitrary order of accuracy w.r.t. powers of  $\varepsilon$  (cf. [20]). This a-priori knowledge on the (complex valued) solution  $\varphi(x)$  allows to separate its scales: The microscale behavior, involving the high oscillations of  $\varphi$  (or its argument,  $\arg \varphi$ ) are analytically rather accurately known. Hence, it can be eliminated from the problem to a large extent. But the macroscale behavior, corresponding to the smooth variations of  $|\varphi(x)|$  has to be obtained numerically. W.r.t. this scale separation, the numerical WKB-method has a conceptual similarity to the *heterogeneous multiscale method* (see [9] for a review). In the WKB-method, however, it is not necessary to solve the microscale problem numerically – its contribution can be obtained analytically.

In this paper we shall extend and refine the asymptotically correct (as  $\varepsilon \rightarrow 0$ ) WKB-scheme from [1]. It is based on the second order WKB-approximation for (1.1) which reads

$$(1.3) \quad \varphi(x) \approx C \frac{\exp\left(\pm \frac{i}{\varepsilon} \int_0^x [\sqrt{a(\tau)} - \varepsilon^2 \beta(\tau)] d\tau\right)}{\sqrt[4]{a(x)}},$$

with

$$(1.4) \quad \beta := -\frac{1}{2a^{1/4}}(a^{-1/4})''.$$

Since [22] is also based on a WKB-approximation (of zero order), its strategy is closely related to the procedure in [1]. But the latter paper yields a refinement to higher  $\varepsilon$ -order. This improvement is illustrated by the numerical comparison in §7.2 in [10].

The hybrid method from [1] consists of two steps: first an analytic preprocessing of (1.1) to transform it into a smoother, i.e. less oscillatory problem. Then, the numerical solution of the transformed ODE problem can be carried out on a coarse grid with high accuracy. It is based on a truncated Picard iteration. This explicit representation of an approximate solution involves oscillatory integrals with two small parameters,  $\varepsilon$  and the step size  $h$ . Hence, the key issue in the numerical step is to

construct approximations of these integrals that are  $\varepsilon$ -uniform as well as accurate enough w.r.t.  $h$ , since the latter determines the local discretization error for the ODE scheme. In [1] both a first and second order method (w.r.t.  $h$ ) were derived.

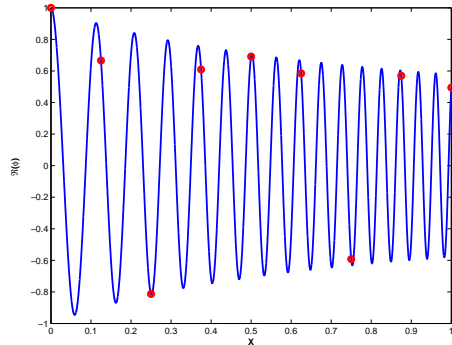


FIG. 1.1. In standard numerical methods highly oscillating solutions require a very fine mesh to resolve the oscillations. However, with the analytic pre-processing of the WKB-marching method from [1] an accurate solution can be obtained on a coarse grid (dots). Plotted is the solution  $\Re[\varphi(x)]$  of (1.1) with  $\varepsilon = 0.01$ ,  $h = 0.125$ , and  $a(x) = (x + \frac{1}{2})^2$ .

In the analytic step, the second order differential equation (1.1) is first transformed to a system of first order differential equations using the still oscillatory variable

$$(1.5) \quad U(x) = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} := \begin{pmatrix} a^{1/4}\varphi(x) \\ \frac{\varepsilon(a^{1/4}\varphi)'(x)}{\sqrt{a(x)}} \end{pmatrix}.$$

In the essential second step, the dominant oscillations are transformed out of  $U(x)$ , yielding a first order ODE for a new, smooth variable  $Z(x)$ . This transformation is made precise in §2.1, and the numerical solution of the ODE for  $Z(x)$  is presented in §2.2.

The numerical analysis of the WKB-method in [1, Th. 3.1] led to the following error estimates for the first and, resp., second order methods:

$$(1.6) \quad \|U(x_n) - U_n\| \leq C \frac{h^\gamma}{\varepsilon} + C\varepsilon^2 \min(\varepsilon, h), \quad 1 \leq n \leq N,$$

and

$$(1.7) \quad \|U(x_n) - U_n\| \leq C \frac{h^\gamma}{\varepsilon} + C\varepsilon^3 h^2, \quad 1 \leq n \leq N,$$

where  $U_n$  is a numerical approximation for the solution  $U(x)$  at the grid point  $x_n$ . Here and in the sequel,  $C$  denotes generic, but not necessarily equal constants that are independent of grid index  $n$ ,  $h$ , and  $\varepsilon$ . Moreover,  $\gamma > 0$  is the order of the chosen numerical integration method for computing the phase integral

$$(1.8) \quad \phi(x) := \int_0^x \left( \sqrt{a(\tau)} - \varepsilon^2 \beta(\tau) \right) d\tau,$$

which is a smooth function of both  $x$  and  $\varepsilon$ . Here,  $\|\cdot\|$  denotes any vector norm in  $\mathbb{C}^2$ .

The second terms in (1.6) and (1.7) are the errors due to the WKB method, and they decay like  $\mathcal{O}(\varepsilon^3)$ , even when the step size  $h$  is kept constant. By contrast, the first term is critical in the semi-classical limit as it grows for  $\varepsilon \rightarrow 0$ . This error is due to using an approximate phase for the analytic transformation step (from  $U$  to  $Z$  and back). To limit the size of the first term one has two options: One can either choose an  $\varepsilon$ -dependent step size (like  $h = \mathcal{O}(\sqrt{\varepsilon})$  when using e.g. the Simpson rule with  $\gamma = 4$  for the phase computation, cf. [22, 1]). Note that this restriction is much weaker than having to resolve each oscillation (by using  $h = \mathcal{O}(\varepsilon)$ ). Moreover it would then render the numerical scheme still second order in  $h$ , uniformly for  $\varepsilon$  satisfying  $h = \mathcal{O}(\sqrt{\varepsilon})$ , see (1.7). Alternatively one can use a highly accurate method to compute the phase (1.8) – like a spectral method, and this will be our approach here.

The goal of this paper is twofold: In [1], numerical errors in the phase were only considered for the backward transformation from  $Z_n$ , the numerical approximation of  $Z$ , to  $U_n$ . Hence we shall first complete that error analysis by taking into account that also the first (analytic) transformation (from (1.1) to the ODE for  $Z$ ) is typically affected by an inaccurate phase. Secondly, we shall combine the WKB-method with a spectral method for computing the phase function  $\phi(x)$ , yielding spectral accuracy. With little effort, this will reduce the first term in the error estimates (1.6) and (1.7) to the order:  $eps * \mathcal{O}(\varepsilon)$ , where  $eps$  denotes the machine precision. This makes these error components irrelevant for most practical computations.

This paper is organized as follows: In §2 we briefly review the WKB-method from [1], and in §3 we extend the error analysis to include the phase error in  $\phi$ . In §4 we review the Chebychev collocation method along with the Clenshaw-Curtis algorithm to compute the phase integral. In §5 we illustrate the efficiency of the combined WKB-spectral method on some numerical examples, and we conclude in §6.

**2. Review of the WKB-method for the stationary Schrödinger equation.** In this section we briefly review the WKB-based numerical method from [1] for solving the following scalar, highly oscillating initial value problem (IVP):

$$(2.1) \quad \begin{cases} \varepsilon^2 \varphi''(x) + a(x)\varphi(x) = 0, & x \in (0, 1), \\ \varphi(0) = \varphi_0, \\ \varepsilon \varphi'(0) = \varphi_1, \end{cases}$$

with possibly complex valued initial conditions. For the rest of the paper we make the following assumptions on the coefficient function  $a(x)$ :

**Hypothesis A** *Let  $a \in C^\infty[0, 1]$  be a fixed smooth (real valued) function, satisfying  $a(x) \geq a_0 > 0$  in  $[0, 1]$ , which means that we are in the oscillatory regime. Besides, let  $0 < \varepsilon < \varepsilon_0$  be an arbitrary real number with*

$$0 < \varepsilon_0 < \varepsilon_1 := \min \left\{ 1, \min_{x \in [0, 1]} [a(x)^{1/4} \beta_+(x)^{-1/2}] \right\},$$

where  $\beta_+(x) := \max(0, \beta(x))$ .

This hypothesis implies that there are positive constants  $C_0, C_1$  such that

$$0 < C_0 \leq \phi'(x) = \sqrt{a(x)} - \varepsilon^2 \beta(x) \leq C_1, \quad x \in [0, 1],$$

and that  $\phi'' \in L^\infty(0, 1)$ .

Clearly,  $a$  could only be piecewise  $C^\infty$ . Then the ODE problem could just be restarted at an interface point of reduced regularity.

The WKB-method consists of two steps: first an analytic transformation of (2.1) into a smoother problem, and then the numerical discretization of the latter.

**2.1. Reformulation of the continuous problem.** This transformation involves three steps: Using the substitution (1.5), the ODE (2.1) is first transformed to a system of first order differential equations:

$$(2.2) \quad \begin{cases} U'(x) = \left[ \frac{1}{\varepsilon} A_0(x) + \varepsilon A_1(x) \right] U(x), & 0 < x < 1, \\ U(0) = U_I, \end{cases}$$

with the two matrices

$$A_0(x) = \sqrt{a(x)} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad A_1(x) = \begin{pmatrix} 0 & 0 \\ 2\beta(x) & 0 \end{pmatrix}.$$

Next the dominant part (w.r.t.  $\varepsilon$ ) of the resulting system matrix in (2.2), i.e.  $\frac{1}{\varepsilon}A_0$ , is diagonalized by the following change of variable:

$$Y(x) := PU(x),$$

with the unitary matrix

$$P := \frac{1}{\sqrt{2}} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}; \quad P^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix}.$$

The final transformation step eliminates the leading oscillations by using the diagonal matrix

$$\Phi(x) := \begin{pmatrix} \phi & 0 \\ 0 & -\phi \end{pmatrix},$$

with the  $\varepsilon$ -dependent, (real valued) phase function  $\phi(x)$ . The change of unknown

$$Z(x) = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} := e^{-\frac{i}{\varepsilon}\Phi(x)} Y(x),$$

finally leads to the system

$$(2.3) \quad \begin{cases} \frac{dZ}{dx} = \varepsilon BZ, & 0 < x < 1, \\ Z(0) = Z_I = Y_I = P U_I. \end{cases}$$

Here, the matrix

$$(2.4) \quad B(x) = \beta(x) \begin{pmatrix} 0 & e^{-\frac{2i}{\varepsilon}\phi(x)} \\ e^{\frac{2i}{\varepsilon}\phi(x)} & 0 \end{pmatrix}$$

is off-diagonal,  $\varepsilon$ -dependent – in fact highly oscillatory, but bounded independently of  $\varepsilon$ .

The principal idea of the WKB-method for  $\varepsilon \ll 1$  is as follows: Instead of solving the highly oscillatory problem (2.1) on a fine mesh, one solves numerically the smooth problem (2.3) on a coarse mesh, possibly with  $h > \varepsilon$ . Then the original solution is recovered by

$$(2.5) \quad U(x) = P^{-1} e^{\frac{i}{\varepsilon} \Phi(x)} Z(x).$$

As in [1], the above transformation assumes that the phase  $\phi(x)$  is known exactly on the considered interval  $[0, 1]$ . In special cases (like piecewise linear coefficient functions  $a(x)$ ) this is indeed possible (since  $\sqrt{a}$  and  $\beta$  are then explicitly integrable). Then, the crucial first error terms in (1.6), (1.7) would be absent. But in general,  $\phi$  has to be obtained by a numerical quadrature, yielding an approximation  $\tilde{\phi}(x)$ . This approximate phase will be used only in the final numerical method and in our error analysis, which generalizes the analysis from [1].

**2.2. Numerical discretization of the transformed problem.** In this section we review the discretizations of (2.3), as developed in [1]. Let  $0 = x_1 < \dots < x_n < \dots < x_N = 1$  be a discretization of the interval  $(0, 1)$  and  $h := \max_{n=1, \dots, N-1} |x_{n+1} - x_n|$ . The marching method from [1] is based on two steps, firstly a truncated Picard iteration for (2.3) with  $\bar{p} = 1, 2$  (corresponding to first and second order schemes, respectively):

$$(2.6) \quad Z(x_{n+1}) = Z(x_n) + \sum_{p=1}^{\bar{p}} \varepsilon^p M_p(x_{n+1}; x_n) Z(x_n),$$

with the matrices

$$(2.7) \quad M_1(x_{n+1}; x_n) = \int_{x_n}^{x_{n+1}} B(y) dy, \quad M_2(x_{n+1}; x_n) = \int_{x_n}^{x_{n+1}} \int_{x_n}^{y_1} B(y_1) B(y_2) dy_2 dy_1, \dots$$

In the second step numerical approximations for  $M_1$  and  $M_2$  have to be found. Since the matrix  $B$  is highly oscillatory w.r.t.  $\varepsilon$ , the key issue is to find an  $\varepsilon$ -uniform discretization of the resulting oscillatory integrals. Since  $B \in \mathbb{C}^{2 \times 2}$  is an off-diagonal matrix, the matrices  $M_p$  are alternately off-diagonal (Hermitian) and diagonal (with complex conjugate entries). Their entries are (iterated) oscillatory integrals (without stationary points). We remark that there is a vast literature on the numerical treatment of oscillatory integrals (see [17, 26], e.g.). But there, the standard setting is to treat an integral on a fixed interval, where the integrand involves a small parameter (corresponding to  $\varepsilon$  in our case). But our application actually involves two small parameters,  $\varepsilon$  and  $h$ , since (2.6), (2.7) constitute a stepping method for the ODE (2.1). Hence, the numerical errors for the oscillatory integrals (2.7) must also be small in  $h$ , in fact of the order  $\mathcal{O}(h^2)$  and  $\mathcal{O}(h^3)$  (for a first and second order scheme, resp.). Due to this additional requirement, standard approaches like the *asymptotic method* [17] do not apply here (directly). In §2.2 of [1] a variant of this latter method was presented that allows to “trade in”  $h$ -powers in the error for  $\varepsilon$ -powers. Next we summarize the resulting method, and refer the reader to [1] for a detailed derivation.

To present the discrete analogs of (2.6)-(2.7) we need some notation. We define the following functions

$$H_1(x) := e^{ix} - 1, \quad H_2(x) := e^{ix} - 1 - ix,$$

$$(2.8) \quad \beta_0(y) := \frac{\beta}{2\phi'}(y) = \frac{\beta}{2(\sqrt{a} - \varepsilon^2\beta)}(y), \quad \beta_k(y) := \frac{1}{2\phi'(y)} \frac{d}{dy} (\beta_{k-1})(y), \quad k = 1, 2, 3,$$

and the phase increments

$$S_n := \phi(x_{n+1}) - \phi(x_n) = \int_{x_n}^{x_{n+1}} \left( \sqrt{a(\tau)} - \varepsilon^2\beta(\tau) \right) d\tau.$$

Now we recall from [1] the two marching methods for the vector  $Z$ .

**First order scheme.** Let  $Z_1 := Z_I$  be the initial condition and let  $n = 1, \dots, N-1$ . Then we define

$$(2.9) \quad Z_{n+1} := (I + A_n^1)Z_n,$$

with the  $2 \times 2$  matrix

$$(2.10) \quad A_n^1 := \varepsilon^3 \beta_1(x_{n+1}) \begin{pmatrix} 0 & e^{-\frac{2i}{\varepsilon}\phi(x_n)} H_1(-\frac{2}{\varepsilon}S_n) \\ e^{\frac{2i}{\varepsilon}\phi(x_n)} H_1(\frac{2}{\varepsilon}S_n) & 0 \end{pmatrix} - i\varepsilon^2 \begin{pmatrix} 0 & \beta_0(x_n)e^{-\frac{2i}{\varepsilon}\phi(x_n)} - \beta_0(x_{n+1})e^{-\frac{2i}{\varepsilon}\phi(x_{n+1})} \\ \beta_0(x_{n+1})e^{\frac{2i}{\varepsilon}\phi(x_{n+1})} - \beta_0(x_n)e^{\frac{2i}{\varepsilon}\phi(x_n)} & 0 \end{pmatrix}.$$

**Second order scheme.** Let  $Z_1 := Z_I$  be the initial condition and let  $n = 1, \dots, N-1$ . Then we define

$$(2.11) \quad Z_{n+1} := (I + A_n^2 + A_n^3)Z_n,$$

with the matrices

$$(2.12) \quad A_n^2 := -i\varepsilon^2 \begin{pmatrix} 0 & \beta_0(x_n)e^{-\frac{2i}{\varepsilon}\phi(x_n)} - \beta_0(x_{n+1})e^{-\frac{2i}{\varepsilon}\phi(x_{n+1})} \\ \beta_0(x_{n+1})e^{\frac{2i}{\varepsilon}\phi(x_{n+1})} - \beta_0(x_n)e^{\frac{2i}{\varepsilon}\phi(x_n)} & 0 \end{pmatrix} + \varepsilon^3 \begin{pmatrix} 0 & \beta_1(x_{n+1})e^{-\frac{2i}{\varepsilon}\phi(x_{n+1})} - \beta_1(x_n)e^{-\frac{2i}{\varepsilon}\phi(x_n)} \\ \beta_1(x_{n+1})e^{\frac{2i}{\varepsilon}\phi(x_{n+1})} - \beta_1(x_n)e^{\frac{2i}{\varepsilon}\phi(x_n)} & 0 \end{pmatrix} + i\varepsilon^4 \beta_2(x_{n+1}) \begin{pmatrix} 0 & -e^{-\frac{2i}{\varepsilon}\phi(x_n)} H_1(-\frac{2}{\varepsilon}S_n) \\ e^{\frac{2i}{\varepsilon}\phi(x_n)} H_1(\frac{2}{\varepsilon}S_n) & 0 \end{pmatrix} - \varepsilon^5 \beta_3(x_{n+1}) \begin{pmatrix} 0 & e^{-\frac{2i}{\varepsilon}\phi(x_n)} H_2(-\frac{2}{\varepsilon}S_n) \\ e^{\frac{2i}{\varepsilon}\phi(x_n)} H_2(\frac{2}{\varepsilon}S_n) & 0 \end{pmatrix},$$

$$\begin{aligned}
(2.13) \quad A_n^3 &:= -i\varepsilon^3(x_{n+1} - x_n) \frac{\beta(x_{n+1})\beta_0(x_{n+1}) + \beta(x_n)\beta_0(x_n)}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\
&\quad -\varepsilon^4\beta_0(x_n)\beta_0(x_{n+1}) \begin{pmatrix} H_1(-\frac{2}{\varepsilon}S_n) & 0 \\ 0 & H_1(\frac{2}{\varepsilon}S_n) \end{pmatrix} \\
&\quad +i\varepsilon^5\beta_1(x_{n+1})[\beta_0(x_n) - \beta_0(x_{n+1})] \begin{pmatrix} H_2(-\frac{2}{\varepsilon}S_n) & 0 \\ 0 & -H_2(\frac{2}{\varepsilon}S_n) \end{pmatrix}.
\end{aligned}$$

In order to compute now the numerical approximation of (2.2) and thus to obtain the wave function  $\varphi$  as well as its derivative  $\varphi'$ , we have to transform back via

$$(2.14) \quad U_n = P^{-1}e^{\frac{i}{\varepsilon}\Phi(x_n)}Z_n, \quad n = 1, \dots, N.$$

In these schemes we assumed so far that the phase  $\phi$ , and the functions  $\beta$  and  $\beta_k$ ,  $k = 0, 1, 2, 3$  (which involve up to five derivatives of  $a$ ) are explicitly “available”. For  $\beta$  and  $\beta_k$  this is feasible, as they involve only derivatives of  $a$ . But typically the phase and phase increment have to be replaced by their numerical approximates  $\tilde{\phi}$  and  $\tilde{S}_n := \tilde{\phi}(x_{n+1}) - \tilde{\phi}(x_n)$ . At this point we have two options: On the one hand we could use the approximate  $\tilde{\phi}$  along with the exact functions  $a$ ,  $\beta$ , and  $\beta_k$ . On the other hand it appears more consistent to combine  $\tilde{\phi}$  with its exact derivatives that lead to numerical approximates  $\tilde{a}$ ,  $\tilde{\beta}$ ,  $\tilde{\beta}_k$ . Since we shall use a spectral integration of the phase (cf. §4), the exact derivatives of  $\tilde{\phi}$  are readily available. We shall use the second option here, as this will simplify the error analysis in §3. Moreover, the error plots of these two versions are almost indistinguishable.

With the replacements  $\tilde{\phi}$ ,  $\tilde{S}_n$ ,  $\tilde{\beta}$ , and  $\tilde{\beta}_k$ , the above matrices (2.10), (2.12), and (2.13) shall be called  $\tilde{A}_n^1$ ,  $\tilde{A}_n^2$ , and  $\tilde{A}_n^3$ . Hence, the first and second order methods with approximate phase read

$$(2.15) \quad \tilde{Z}_{n+1} := (I + \tilde{A}_n^1)\tilde{Z}_n, \quad n = 1, \dots, N-1,$$

and

$$(2.16) \quad \tilde{Z}_{n+1} := (I + \tilde{A}_n^2 + \tilde{A}_n^3)\tilde{Z}_n, \quad n = 1, \dots, N-1.$$

For both methods, the corresponding back-transformation to the variable  $U$  then reads

$$(2.17) \quad \tilde{U}_n = P^{-1}e^{\frac{i}{\varepsilon}\tilde{\Phi}(x_n)}\tilde{Z}_n, \quad n = 1, \dots, N,$$

with  $\tilde{\Phi} := \text{diag}(\tilde{\phi}, -\tilde{\phi})$ . Note that these two WKB-schemes were introduced as a numerical approximation of the schemes (2.9), (2.11) for the (transformed) Schrödinger equation (2.2) with the (unperturbed) coefficients  $a$ ,  $\beta$  — just because the exact phase  $\phi$  is typically not available. But, in parallel, the two schemes (2.15), (2.16) can also be considered as WKB-schemes *without a numerically perturbed phase* for the perturbed Schrödinger equation (2.2), i.e. with the perturbed coefficients  $\tilde{a}$ ,  $\tilde{\beta}$ . The second point of view will be helpful in the subsequent numerical analysis.



In Section 3 we shall give a complete error analysis of the first order scheme (2.15), (2.17) and of the second order scheme (2.16), (2.17). This is a completion of the error estimates given in [1], since that paper used the approximate phase  $\tilde{\phi}$  only in the back-transformation (2.17) but not in the matrices  $A_n^1$ ,  $A_n^2$ , and  $A_n^3$ .

**3. Error analysis including phase errors.** First we decompose the exact phase from (1.8) as  $\phi(x) = \phi_1(x) - \varepsilon^2 \phi_2(x)$  with

$$(3.1) \quad \phi_1(x) := \int_0^x \sqrt{a(\tau)} d\tau, \quad \phi_2(x) := \int_0^x \beta(\tau) d\tau.$$

For the approximate phase  $\tilde{\phi} = \tilde{\phi}_1 - \varepsilon^2 \tilde{\phi}_2$  we shall make the following assumptions:

**Hypothesis B** Let  $\tilde{\phi}_1, \tilde{\phi}_2 \in C^\infty[0, 1]$  with  $\tilde{\phi}_1'(x) \geq C_2 > 0$ .

This Hypothesis implies that there are positive constants  $C_3, C_4, C_5$  such that

$$\begin{aligned} 0 < C_3 \leq \tilde{\phi}'(x) = \tilde{\phi}_1'(x) - \varepsilon^2 \tilde{\phi}_2'(x) &\leq C_4 \quad \text{for any } 0 < \varepsilon \leq \tilde{\varepsilon}_0, \\ \tilde{\phi}''(x) = \tilde{\phi}_1''(x) - \varepsilon^2 \tilde{\phi}_2''(x) &\leq C_5 \quad \text{for any } 0 < \varepsilon \leq \tilde{\varepsilon}_0, \end{aligned}$$

with some  $0 < \tilde{\varepsilon}_0 \leq \varepsilon_0$ . Hence, Hypothesis **B** implies that there are positive numbers  $E, E', E''$ , such that the following error bounds on  $\tilde{\phi}$  hold uniformly in  $\varepsilon \in (0, \tilde{\varepsilon}_0]$ :

$$(3.2) \quad \begin{aligned} \|\phi - \tilde{\phi}\|_{L^\infty(0,1)} &\leq E, \\ \|\phi' - \tilde{\phi}'\|_{L^\infty(0,1)} &\leq \|\phi_1' - \tilde{\phi}_1'\|_{L^\infty(0,1)} + \varepsilon^2 \|\phi_2' - \tilde{\phi}_2'\|_{L^\infty(0,1)} \leq E', \\ \|\phi'' - \tilde{\phi}''\|_{L^\infty(0,1)} &\leq \|\phi_1'' - \tilde{\phi}_1''\|_{L^\infty(0,1)} + \varepsilon^2 \|\phi_2'' - \tilde{\phi}_2''\|_{L^\infty(0,1)} \leq E'', \end{aligned}$$

These error bounds will be important ingredients for the subsequent error estimates. In accordance with (3.1) we also define  $\tilde{a}(x) := (\phi_1'(x))^2$ ,  $\tilde{\beta}(x) := \phi_2'(x)$ , but we do *not require* that  $\tilde{\beta} = -\frac{1}{2\tilde{a}^{1/4}}(\tilde{a}^{-1/4})''$  holds (cp. with (1.4)). We remark that this equality was also *not used* in the error analysis of [1].

Note that we assume here that  $\tilde{\phi}$  is a continuous (and smooth) function on  $[0, 1]$ , and it is not only defined on the grid points  $x_n$ . In particular, this is satisfied for the spectral approximation constructed in §4 below.

As a first step of the error analysis we shall estimate the error between the (continuous) solution  $Z(x)$  to (2.3) and its perturbed analog  $\tilde{Z}(x)$ , which is the exact solution to

$$(3.3) \quad \begin{cases} \frac{d\tilde{Z}}{dx} = \varepsilon \tilde{B} \tilde{Z}, & 0 < x < 1, \\ \tilde{Z}(0) = Z_I = P U_I, \end{cases}$$

with the matrix

$$\tilde{B}(x) := \tilde{\beta}(x) \begin{pmatrix} 0 & e^{-\frac{2i}{\varepsilon}\tilde{\phi}(x)} \\ e^{\frac{2i}{\varepsilon}\tilde{\phi}(x)} & 0 \end{pmatrix}.$$

**LEMMA 3.1.** *Let the coefficient function  $a$  satisfy Hypothesis **A** and let  $\tilde{\phi}$  satisfy Hypothesis **B**. Then we have*

$$(3.4) \quad \|Z - \tilde{Z}\|_{L^\infty(0,1)} \leq c\varepsilon[\min(\varepsilon, E) + \varepsilon(E' + E'')],$$

with some generic constant  $c$  independent of  $\varepsilon \in (0, \tilde{\varepsilon}_0]$ .

*Proof.* Step 1 (bound on the solution propagator:) We define the *propagator* pertaining to the ODE in (2.3) as the matrix  $S(x, y) \in \mathbb{C}^{2 \times 2}$  that satisfies  $Z(x) = S(x, y)Z(y)$ , and  $\tilde{S}(x, y)$  is the propagator for (3.3). To estimate the growth of the solution  $Z(x) \in \mathbb{C}^2$  we compute

$$\begin{aligned} \frac{d}{dx} \|Z\|^2 &= 2\varepsilon\beta(x)\bar{Z}^T \begin{pmatrix} 0 & e^{-\frac{2i}{\varepsilon}\phi(x)} \\ e^{\frac{2i}{\varepsilon}\phi(x)} & 0 \end{pmatrix} Z = 2\varepsilon\beta(x) \left( e^{\frac{2i}{\varepsilon}\phi(x)} z_1 \bar{z}_2 + e^{-\frac{2i}{\varepsilon}\phi(x)} \bar{z}_1 z_2 \right) \\ &\leq 2\varepsilon\|\beta\|_\infty \|Z\|^2. \end{aligned}$$

Hence  $\frac{d}{dx} \|Z\| \leq \varepsilon\|\beta\|_\infty \|Z\|$ , where we used the abbreviation  $\|\cdot\|_\infty$  for  $\|\cdot\|_{L^\infty(0,1)}$ . Gronwall's lemma then implies for the solution propagator of (2.3):

$$(3.5) \quad \|S(x, y)\| \leq e^{\varepsilon\|\beta\|_\infty|x-y|}, \quad 0 \leq x, y \leq 1,$$

and analogously for the solution propagator of (3.3):

$$(3.6) \quad \|\tilde{S}(x, y)\| \leq e^{\varepsilon\|\tilde{\beta}\|_\infty|x-y|} \quad 0 \leq x, y \leq 1.$$

Step 2 (bound on  $\Delta Z$ ): We denote  $\Delta Z := Z - \tilde{Z}$ , which satisfies

$$\Delta Z' = \varepsilon B(x)\Delta Z + \varepsilon(B(x) - \tilde{B}(x))\tilde{Z}(x), \quad \Delta Z_I = 0.$$

The solution of this inhomogeneous equation reads

$$\begin{aligned} \Delta Z &= \varepsilon \int_0^x S(x, y) [(B(y) - \tilde{B}(y))\tilde{Z}(y)] dy \\ &= i\varepsilon^2 \int_0^x S(x, y) \begin{pmatrix} [\beta_0(e^{-\frac{2i}{\varepsilon}\phi})' - \tilde{\beta}_0(e^{-\frac{2i}{\varepsilon}\tilde{\phi}})'] \tilde{z}_2 \\ [-\beta_0(e^{\frac{2i}{\varepsilon}\phi})' + \tilde{\beta}_0(e^{\frac{2i}{\varepsilon}\tilde{\phi}})'] \tilde{z}_1 \end{pmatrix} dy \\ (3.7) \quad &= i\varepsilon^2 S(x, y) \begin{pmatrix} [\beta_0 e^{-\frac{2i}{\varepsilon}\phi} - \tilde{\beta}_0 e^{-\frac{2i}{\varepsilon}\tilde{\phi}}] \tilde{z}_2 \\ [-\beta_0 e^{\frac{2i}{\varepsilon}\phi} + \tilde{\beta}_0 e^{\frac{2i}{\varepsilon}\tilde{\phi}}] \tilde{z}_1 \end{pmatrix} \Big|_{y=0}^{y=x} \\ &\quad - i\varepsilon^2 \int_0^x S(x, y) \begin{pmatrix} (\beta_0 \tilde{z}_2)' e^{-\frac{2i}{\varepsilon}\phi} - (\tilde{\beta}_0 \tilde{z}_2)' e^{-\frac{2i}{\varepsilon}\tilde{\phi}} \\ -(\beta_0 \tilde{z}_1)' e^{\frac{2i}{\varepsilon}\phi} + (\tilde{\beta}_0 \tilde{z}_1)' e^{\frac{2i}{\varepsilon}\tilde{\phi}} \end{pmatrix} dy \\ &\quad + i\varepsilon^3 \int_0^x S(x, y) \left[ B(y) \begin{pmatrix} [\beta_0 e^{-\frac{2i}{\varepsilon}\phi} - \tilde{\beta}_0 e^{-\frac{2i}{\varepsilon}\tilde{\phi}}] \tilde{z}_2 \\ [-\beta_0 e^{\frac{2i}{\varepsilon}\phi} + \tilde{\beta}_0 e^{\frac{2i}{\varepsilon}\tilde{\phi}}] \tilde{z}_1 \end{pmatrix} \right] dy, \end{aligned}$$

where we skipped in the integrands the argument " $(y)$ " for brevity. In (3.7) we used the following integration by parts formula involving the propagator  $T(x, y)$  for some linear evolution equation  $u' = A(x)u$ :

$$\begin{aligned} \int_0^x T(x, y) [f'(y)g(y)] dy &= T(x, y) [f(y)g(y)] \Big|_{y=0}^{y=x} \\ &\quad - \int_0^x T(x, y) [f(y)g'(y)] dy + \int_0^x T(x, y) [A(y)\{f(y)g(y)\}] dy, \end{aligned}$$

which can be verified easily by using  $\frac{\partial}{\partial y} T(x, y) = -T(x, y)A(y)$ . Note that the integration by parts in the oscillatory integral of (3.7) will allow to recover one more

$\varepsilon$ -power in the estimate of  $\Delta Z$ . This strategy was already used in Proposition 2.2 of [1]. In the last line of (3.7) we used also

$$(3.8) \quad \frac{\partial}{\partial y} S(x, y) = -\varepsilon S(x, y) B(y).$$

On the r.h.s. of (3.7) we have to consider two types of differences: First we estimate

$$(3.9) \quad \begin{aligned} & \left| \beta_0(y) e^{\frac{2i}{\varepsilon} \phi(y)} - \tilde{\beta}_0(y) e^{\frac{2i}{\varepsilon} \tilde{\phi}(y)} \right| \\ &= \left| \beta_0(y) \left[ e^{\frac{2i}{\varepsilon} \phi(y)} - e^{\frac{2i}{\varepsilon} \tilde{\phi}(y)} \right] + \left[ -\frac{\beta(y)}{2} \frac{\phi'(y) - \tilde{\phi}'(y)}{\phi'(y) \tilde{\phi}'(y)} + \frac{\beta - \tilde{\beta}}{2\tilde{\phi}'} \right] e^{\frac{2i}{\varepsilon} \tilde{\phi}(y)} \right| \\ &\leq 2\|\beta_0\|_\infty \min\left(1, \frac{E}{\varepsilon}\right) + \|\beta\|_\infty \frac{E'}{2C_0 C_3} + \frac{E'}{2C_3}, \end{aligned}$$

where we used for the first term in (3.9) both the trivial estimate  $|e^{\frac{2i}{\varepsilon} \phi(y)} - e^{\frac{2i}{\varepsilon} \tilde{\phi}(y)}| \leq 2$  and the mean value theorem for vector functions. We also used  $\beta - \tilde{\beta} = \phi'_2 - \tilde{\phi}'_2$ , and we recall the definitions  $\beta_0(y) := \beta(y)/(2\phi'(y))$ ,  $\tilde{\beta}_0(y) := \tilde{\beta}(y)/(2\tilde{\phi}'(y))$ .

Secondly we estimate:

$$\begin{aligned} & \left| \beta'_0(y) e^{\frac{2i}{\varepsilon} \phi(y)} - \tilde{\beta}'_0(y) e^{\frac{2i}{\varepsilon} \tilde{\phi}(y)} \right| = \left| \beta'_0 \left[ e^{\frac{2i}{\varepsilon} \phi} - e^{\frac{2i}{\varepsilon} \tilde{\phi}} \right] - \right. \\ & \left. - \left[ \frac{\beta'}{2} \frac{\phi' - \tilde{\phi}'}{\phi' \tilde{\phi}'} + \frac{\beta}{2} \left\{ \frac{\phi''(\tilde{\phi}' - \phi')(\tilde{\phi}' + \phi')}{(\phi' \tilde{\phi}')^2} + \frac{\phi'' - \tilde{\phi}''}{(\tilde{\phi}')^2} \right\} - \frac{\beta' - \tilde{\beta}'}{2\tilde{\phi}'} + \frac{(\beta - \tilde{\beta})\tilde{\phi}''}{2(\tilde{\phi}')^2} \right] e^{\frac{2i}{\varepsilon} \tilde{\phi}} \right| \\ &\leq 2\|\beta'_0\|_\infty \min\left(1, \frac{E}{\varepsilon}\right) + \|\beta'\|_\infty \frac{E'}{2C_0 C_3} + \|\beta\|_\infty \frac{\|\phi''\|_\infty (C_1 + C_4) E'}{2C_0^2 C_3^2} + \|\beta\|_\infty \frac{E''}{2C_3^2} + \frac{E''}{2C_3} + \frac{C_5 E'}{2C_3^2}. \end{aligned}$$

From (3.5), (3.6), and (3.3) we recall that the propagator  $S(x, y)$ , the solution  $\tilde{Z}(x)$ , and  $\tilde{Z}'(x)$  are uniformly bounded in  $x, y$ , and  $\varepsilon$ . Hence, the estimates (3.9) and (3.10) yield the result (3.4) with a constant  $c$  that depends only on  $\|\beta\|_\infty$ ,  $\|\beta_0\|_{W^{1,\infty}}$ ,  $\|\phi''\|_\infty$ ,  $C_0$ ,  $C_1$ ,  $C_3$ ,  $C_4$ , and  $C_5$ .  $\square$

This lemma allows to derive the main result of this section:

**THEOREM 3.2.** *Let the coefficient function  $a$  satisfy Hypothesis **A** and let  $\tilde{\phi}$  satisfy Hypothesis **B**. Then the first order scheme (2.15), (2.17) and the second order scheme (2.16), (2.17) satisfy the following error estimates:*

$$(3.10) \quad \|U(x_n) - \tilde{U}_n\| \leq C \frac{E}{\varepsilon} + C\varepsilon^2 \min(\varepsilon, h) + C\varepsilon[\min(\varepsilon, E) + \varepsilon(E' + E'')], \quad 1 \leq n \leq N,$$

and

$$(3.11) \quad \|U(x_n) - \tilde{U}_n\| \leq C \frac{E}{\varepsilon} + C\varepsilon^3 h^2 + C\varepsilon[\min(\varepsilon, E) + \varepsilon(E' + E'')], \quad 1 \leq n \leq N,$$

with some generic constant  $C$  independent of  $\varepsilon \in (0, \tilde{\varepsilon}_0]$ ,  $n$ , and  $h$ .

Let us compare this result with the estimates (1.6), (1.7) that are due to [1, Th. 3.1]: The first error terms on the r.h.s. of (3.10) and (3.11) are generalizations to  $h$ -independent numerical integrations of the phase integral. The new (additional) third terms are due to using the perturbed phase  $\tilde{\phi}$  in the WKB-method.

*Proof.* [of Theorem 3.2] First we estimate the error in the  $Z$ -variable, for  $1 \leq n \leq N$ :

$$(3.12) \quad \begin{aligned} \|Z(x_n) - \tilde{Z}_n\| &\leq \|Z - \tilde{Z}\|_\infty + \|\tilde{Z}(x_n) - \tilde{Z}_n\| \\ &\leq C\varepsilon[\min(\varepsilon, E) + \varepsilon(E' + E'')] + \begin{cases} C\varepsilon^2 \min(\varepsilon, h) & \text{for first order,} \\ C\varepsilon^3 h^2 & \text{for second order,} \end{cases} \end{aligned}$$

where we used Lemma 3.1 and the WKB-error estimate for  $\tilde{Z}$  that is analogous to (1.6), (1.7) (skipping the first term, cf. [1, Th. 3.1]).

Next we estimate the error in the  $U$ -variable, using (2.14), (2.17):

$$\begin{aligned} \|U(x_n) - \tilde{U}_n\| &\leq \|P^{-1}(e^{\frac{i}{\varepsilon}\Phi(x_n)} - e^{\frac{i}{\varepsilon}\tilde{\Phi}(x_n)})Z(x_n)\| + \|P^{-1}e^{\frac{i}{\varepsilon}\tilde{\Phi}(x_n)}(Z(x_n) - \tilde{Z}_n)\| \\ &\leq C \min(1, \frac{E}{\varepsilon}) + \|Z(x_n) - \tilde{Z}_n\|, \end{aligned}$$

where we used an estimate like for the first term of (3.9), the  $\varepsilon$ -uniform boundedness of  $Z(x)$  (see (3.5)), and the unitarity of the matrices  $P^{-1}$ ,  $e^{\frac{i}{\varepsilon}\tilde{\Phi}(x_n)}$ . Combining the two estimates yields the result.  $\square$

**4. Spectral integration of the phase.** The estimation of the numerical errors (1.6) and (1.7) in the computation of a solution to the Schrödinger equation in the semi-classical limit via the approach of [1] indicates that the problematic term for small  $\varepsilon$  is the first on the right hand sides of these expressions. It arises from the numerical computation of the phase (1.8) and is not present if the latter can be computed exactly. In cases where this is not possible, a high order method is recommended to reduce as much as possible the role of the term proportional to  $1/\varepsilon$ . We use here spectral methods which are known to approximate analytic functions with spectral accuracy, i.e., an error decreasing exponentially with the number of modes. The numerical error for  $C^\infty$  functions in accordance with Hypothesis **A** is known to decrease faster than any power of  $h$ , which means in practice an exponential decrease, too, see e.g. [28]. Concretely we apply a Chebychev collocation method and use the Clenshaw-Curtis [6] algorithm for the integral in (1.8). For points in between collocation points of the Clenshaw-Curtis algorithm, we use barycentric interpolation, see [4].

The basic idea of spectral methods is to approximate a function  $f$  on the interval  $[a, b]$  via functions which are globally smooth on the considered interval. We will use here Chebychev polynomials since Chebychev series are related to Fourier series for which efficient numerical algorithms exist. Since any finite interval  $x \in [a, b]$  can be mapped via  $x = b(1+l)/2 + a(1-l)/2$  to the interval  $l \in [-1, 1]$ , we present all algorithms for the latter interval. We approximate  $f(l)$  via

$$(4.1) \quad f(l) \approx \sum_{n=0}^N a_n T_n(l), \quad l \in [-1, 1],$$

where the Chebychev polynomials  $T_n(l)$  are defined as

$$(4.2) \quad T_n(l) = \cos[n \arccos(l)], \quad n = 0, 1, \dots$$

The idea of a collocation method is to introduce collocation points  $l_j$ ,  $j = 0, \dots, N$  on  $[-1, 1]$  and to impose in (4.1) equality at the collocation points,

$$(4.3) \quad f(l_j) = \sum_{n=0}^N a_n T_n(l_j), \quad j = 0, \dots, N.$$

These are  $N + 1$  equations to determine the spectral coefficients  $a_n$ ,  $n = 0, \dots, N$ . Choosing the  $l_j$  as the Chebychev collocation points  $l_j = \cos(j\pi/N)$ ,  $j = 0, \dots, N$  (note in particular that these points avoid the Runge phenomenon in interpolation on equidistant points and allow a uniform accuracy in the interpolation, see e.g., the discussion in Chap. 5 of [28]), the equations (4.3) take the form

$$(4.4) \quad f(l_j) = \sum_{n=0}^N a_n \cos\left(\frac{nj\pi}{N}\right), \quad j = 0, \dots, N.$$

Thus the spectral coefficients are given by the discrete cosine transformation (DCT) of the function  $f$  at the collocation points. Since the DCT is related to the discrete Fourier transform, it can be computed with the fast Fourier transform algorithm after some preprocessing, see for instance Chap. 8 of [28]. Thus one advantage of a Chebychev collocation method is that a fast algorithm to compute the spectral coefficients exists.

To integrate a function approximated by the Chebychev sum (4.1), a very efficient algorithm exists due to Clenshaw and Curtis [6]. The basis of the algorithm is the well known identity for Chebychev polynomials (simply a consequence of the addition theorems for trigonometric functions)

$$(4.5) \quad \frac{T'_{n+1}(l)}{n+1} - \frac{T'_{n-1}(l)}{n-1} = 2T_n(l), \quad n > 1.$$

The antiderivative of a function  $f$  approximated as a Chebychev sum (4.1) can itself be approximated by such a sum,

$$(4.6) \quad \int_{-1}^l f(l') dl' \approx \sum_{n=0}^N \int_{-1}^l a_n T_n$$

$$= a_0(l+1) + \sum_{n=1}^N a_n \left( \frac{T_{n+1}(l) - (-1)^n}{n+1} - \frac{T_{n-1}(l) - (-1)^n}{n-1} \right) \approx \sum_{n=0}^N b_n T_n(l),$$

where the  $b_n$  follow from the  $a_n$  via (4.5),

$$(4.7) \quad b_n = \frac{1}{2(n-1)}(a_{n-1} - a_{n+1}), \quad n = 2, 3, \dots, N-1,$$

$$b_1 = a_0 - \frac{1}{2}a_2,$$

$$b_N = \frac{a_{N-1}}{2(N-1)},$$

$$b_0 = - \sum_{n=1}^N (-1)^n b_n.$$

In [5] the numerical error of the Clenshaw-Curtis algorithm was discussed showing that it is a spectral method. Identity (4.5) can obviously also be used to approximate derivatives of functions in coefficient space. Alternatively and with the same numerical accuracy, one can use the differentiation matrices of Chap. 6 of [28] following from Lagrangian interpolation on Chebychev collocation points. We use these matrices to compute the derivatives appearing in the definition of the  $\beta_k$ ,  $k = 0, 1, 2, 3$  (2.8). Thus these derivatives are also computed with spectral accuracy. For  $L^\infty$ -error bounds of

the Chebychev spectral approximation (and its derivatives) we refer to Theorem 5 and 6 in [28]. We recall that such estimates were assumed for the error analysis in §3.

As an example we consider the function  $f(x) = \exp(-x^2/2)$  appearing in the examples in the following section for  $x \in [0, 1]$ . The difference between the Clenshaw-Curtis approximation of  $\int_0^1 f(x)dx$  and the exact value  $\sqrt{\pi/2}\operatorname{erf}(1/\sqrt{2})$  (the well known *error function* computed in Matlab to machine precision via  $\operatorname{erf}(x)$ ) in dependence of the number  $N$  of collocation points is shown on the left of Fig. 4.1. It can be seen in the semilogarithmic plot that the numerical error decreases exponentially with the number  $N$  of collocation points up to  $N = 14$  where the numerical error reaches the saturation level (we work here in double precision, thus the accuracy is limited in practice to the order of  $10^{-16}$  because of rounding errors).

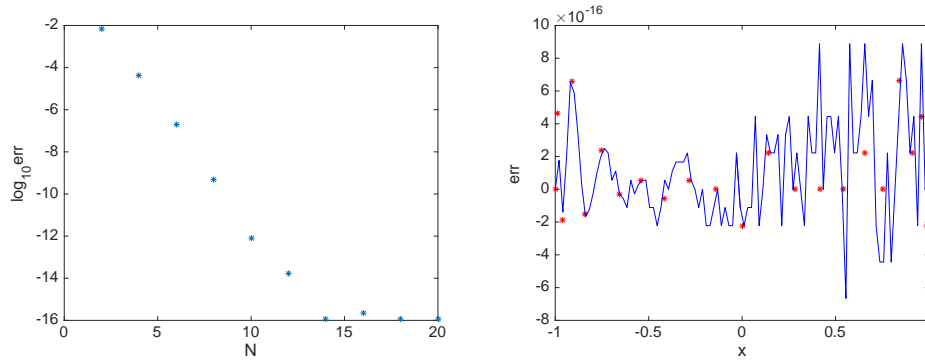


FIG. 4.1. Antiderivative of  $f = \exp(-x^2/2)$  on the interval  $[0, 1]$  as approximated by the Clenshaw-Curtis algorithm; on the left the difference between the approximation of  $\int_0^1 f(x)dx$  and  $\sqrt{\pi/2}\operatorname{erf}(1/\sqrt{2})$  in dependence of  $N$ ; on the right the difference between the antiderivative and  $\sqrt{\pi/2}\operatorname{erf}(x/\sqrt{2})$  for  $N = 20$  at the collocation points (marked with \* in red) and for intermediate values after barycentric interpolation.

The Clenshaw-Curtis algorithm gives in principle only the antiderivative of a function at the collocation points. However in the present context, the former will be needed on more general values of  $l$ . Since the basis of the approach is a sum of Chebychev polynomials, intermediate values can be obtained in principle from formula (4.1). A numerically stable and very efficient way to interpolate is to use Lagrange interpolation in the barycentric form, see [4] and references therein. For Chebychev collocation points, the interpolation weights can be given explicitly, and a Matlab code for this case can be found in [4]. The difference between the antiderivative of the function  $f(x) = \exp(-x^2/2)$  and  $\sqrt{\pi/2}\operatorname{erf}(x/\sqrt{2})$  in dependence of  $x$  can be seen for  $N = 20$  in Fig. 4.1 on the right. The difference on the collocation points is marked with red ‘\*s’. It can be seen that the error introduced by interpolation at intermediate points is also smaller than  $10^{-15}$ .

As mentioned above, the exponential decay of the numerical error with  $N$  in the Clenshaw-Curtis algorithm for functions analytic in a strip around the real axis in the complex plane is a general feature of Chebychev series for such functions. This can be seen on the left of Fig. 4.2 where the Chebychev coefficients of the antiderivative of  $\exp(-x^2/2)$  are shown. They decrease exponentially, and the numerical error due to truncation of the series at  $N + 1$  terms is actually due to the highest order spectral coefficient. Thus the Chebychev coefficients also provide an approach to estimate the numerical error due to a spectral method by studying the spectral coefficients. For the example  $a(x) = \exp(-x^2)$ , we just considered in Fig. 4.1 the term  $\int_0^x \sqrt{a(\tau)} d\tau$  in

the integral of (1.3), since we had an independent way to compute the exact integral via the error function. This is not the case for the terms proportional to  $\varepsilon^2$  in the integral in (1.3). But the spectral coefficients of the antiderivative of these terms ( $(1 + x^2/2) \exp(x^2/2)/4$  for the example  $a(x) = \exp(-x^2)$ ) on the right of Fig. 4.2 indicate a similar behavior of the error as in Fig. 4.1: For  $N \sim 20$  the Chebychev coefficients are of the order of the rounding error, and further increase of the number of coefficients no longer leads to higher accuracy.

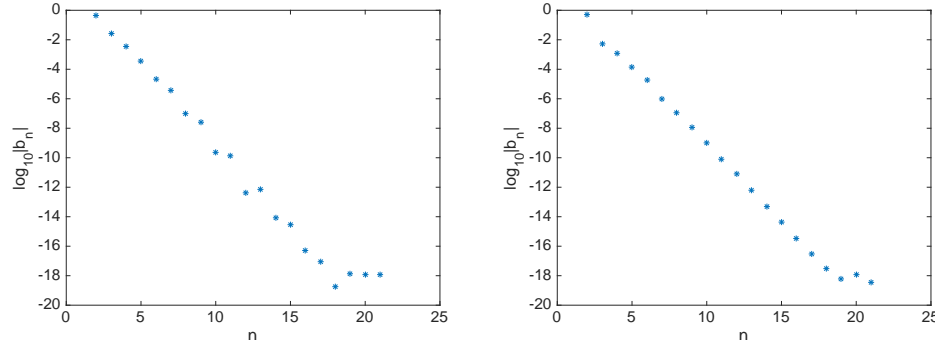


FIG. 4.2. Spectral coefficients for two antiderivatives in dependence of the number of collocation points: on the left for the function  $\exp(-x^2/2)$ , on the right for  $(1 + x^2/2) \exp(x^2/2)/4$ .

#### Remark:

The example in Fig. 4.1 shows that a spectral approach for  $C^\infty$  functions allows in practice to reach machine precision with low resolution, here with just 14 Chebychev polynomials. Thus the numerical error reaches a plateau which itself will increase with  $N$ . The latter is due to the fact that rounding errors pile up with larger values of  $N$ . With finite difference methods, considerably larger values of  $N$  (or equivalently smaller values of  $h$ ) are needed to reach the saturation of the numerical errors, and because of this, the plateau is reached in practice at much higher values than here, of the order of  $10^{-10}$  (see for instance examples in [28]).

Note that in [1], the saturation level of the numerical errors was not reached since quadruple precision was used, and since the values of  $h$  were not small enough to get there with the used precision. Here, we work in double precision and will reach the saturation level in most cases.

**5. Numerical results.** We shall present now numerical results obtained with the first and second order WKB-schemes from Section 2. For our numerical tests we chose  $a(x) = \exp(-x^2)$  on the spatial interval  $[0, 1]$  with a uniform grid, and the initial condition  $U_I = (1, -i)^\top$ . For both schemes we shall compare the results obtained with two versions of the numerical phase computation: on the one hand by the composite Simpson rule (with error order  $\gamma = 4$ ) on the WKB-grid  $\{x_n\}$ , and on the other hand by the spectral method from §4 along with barycentric interpolation at the WKB-grid points  $x_n$ . Note that we use a Chebychev grid with  $N = 20$  points as in the previous section for the computation of the phase  $\phi$  in (1.3) and then interpolate to the equidistant grid for the WKB-scheme. This is necessary since Chebychev collocation points are not equidistant. A consequence of this approach is that the spectral method computes the phase always to machine precision. In fact, the absolute and relative errors of  $\tilde{\phi}_j^{(k)}$ ;  $j = 1, 2$ ;  $k = 0, \dots, 4$  are always of the order  $10^{-16} - 4 \cdot 10^{-15}$  for this example.

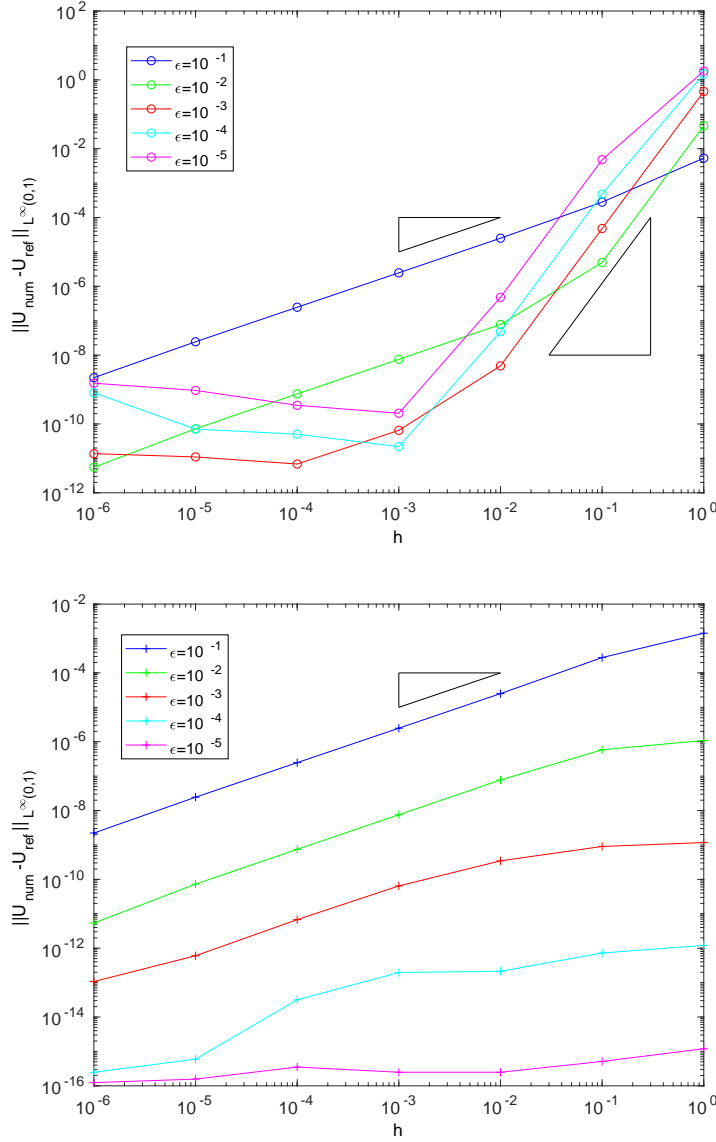


FIG. 5.1. Error of the first order WKB-method on the interval  $[0, 1]$  as a function of  $h$ , for 5 values of  $\varepsilon$ ; at the top the results with the phase  $\tilde{\phi}$  computed via Simpson's rule; at the bottom the analogous results with  $\tilde{\phi}$  computed via Clenshaw-Curtis algorithm and barycentric interpolation.

Fig. 5.1 shows the results for the first order method. Plotted are the  $L^\infty(0, 1)$ -errors of the numerical solution  $\{U_n\}$  as a function of  $h$  for 5 values of  $\varepsilon$ . The reference solutions were obtained with the same method, but on a much finer grid. For simplicity we shall refer in our discussion to the error estimate (1.6). The top plot is obtained with the Simpson rule to compute  $\tilde{\phi}$ . For  $\varepsilon = 10^{-1}$  the second term (i.e. the WKB-error) in (1.6) dominates and the method is clearly first order in  $h$ , as indicated by the upper slope triangle. For  $\varepsilon = 10^{-2}$  one observes this behavior for  $h \leq 10^{-2}$ . For smaller values of  $\varepsilon$  and large step sizes (e.g.  $h = 1$ ) the error behaves



like the first error term in (1.6), i.e.  $h^4/\varepsilon$  due to the Simpson rule, as visualized by the lower slope triangle. This leads to an inversion of the 5 error curves at  $h = 1$  ( $\varepsilon = 10^{-5}$  at the top,  $\varepsilon = 10^{-1}$  at the bottom). However, for small step sizes (e.g.  $h = 10^{-6}$ ) the error term  $\varepsilon^2 \min(\varepsilon, h)$  dominates, such that the inversion of the 5 error curves w.r.t.  $\varepsilon$  disappears. But for small values of  $h$  the error curves are also polluted by round-off errors (due to the double precision computations in Matlab). For the phase computation, the composite Simpson rule has a worse conditioning than the spectral method. Hence the former increases the effect of round-off errors here, see the remark in the previous section. Thus the error reaches the saturation level for the Simpson rule at higher values of the error than for the spectral method. This is also the reason why smaller errors can be reached for small  $\varepsilon$  and small  $h$  in the bottom figure of Fig. 5.1. In the top figure, it can be recognized that the errors in the phase computation lead to an effective increase of the numerical error with decreasing  $h$ , and the  $\varepsilon$  dependence of the related term on (1.6) implies that this is mainly visible for values of  $\varepsilon < 10^{-2}$ .

The bottom plot in Fig. 5.1 is obtained with the spectral method for  $\tilde{\phi}$ , and it reveals that the problematic first term in (1.6) has been essentially eliminated. As shown by the slope triangle, the method is first order in  $h$ . For large step sizes (e.g.  $h = 1$ ) the error behaves like  $\mathcal{O}(\varepsilon^3)$ , and for small step sizes (e.g.  $h = 10^{-6}$ ) roughly like  $\mathcal{O}(\varepsilon^2)$ , as predicted by (1.6).

Fig. 5.2 shows the results for the second order method, again for 5 values of  $\varepsilon$ . Grosso modo the error behavior is similar to the first order method, and we shall compare it to the error estimate (1.7). Of course the errors of the second order method are smaller than for the first order method. Therefore, the WKB-error reaches the saturation level for small step sizes (usually for  $h \leq 10^{-4} - 10^{-3}$ ) (note that the WKB-error did not reach the saturation level for the first order method in Fig. 5.1).

The top plot in Fig. 5.2 is obtained with the Simpson rule to compute  $\tilde{\phi}$ . For  $\varepsilon = 10^{-1}$  the second term (i.e. the WKB-error) in (1.7) dominates and the method is second order for  $h \geq 10^{-3}$ , as indicated by the upper slope triangle. Again the fact that the maximally achievable accuracy with the Simpson method is reached at higher values than with the spectral method leads to a slight increase of the error for very small values of  $h$ . Since, in this case, both errors of the WKB method and the Simpson integration of the phase are due to rounding errors, there is no simple dependence of  $\varepsilon$  and  $h$  and the errors are of the order of  $10^{-10}$ .

The bottom plot is obtained with the spectral method for  $\tilde{\phi}$ , and it reveals that the problematic first term in (1.7) has been eliminated again. As shown by the slope triangle, the method is second order in  $h$  (for  $h$  large, i.e. before the WKB-error reaches the saturation level). Again the better conditioning of the spectral method compared to the Simpson method allows to achieve smaller numerical errors depending on  $\varepsilon$ . For  $\varepsilon = 10^{-5}$  the error dropped below the relative machine precision for the values  $h = 10^{-5} - 10^{-2}$ . Therefore, Matlab's double precision could not compute a positive error value in these cases. Hence, these points are omitted in Fig. 5.2, bottom.

To round off the presentation of the second order method, we present in Fig. 5.3 the error behavior of the transformed variable  $Z$ . It satisfies the error bounds from (3.12), which does *not* include the problematic term  $\frac{E}{\varepsilon}$ . Note that the bottom plot includes all  $h$ -values in the error curve for  $\varepsilon = 10^{-5}$ .

**6. Conclusion.** In this paper we have reviewed the numerical approach of [1] to efficiently compute highly oscillatory solutions to the 1D Schrödinger equation (1.1)

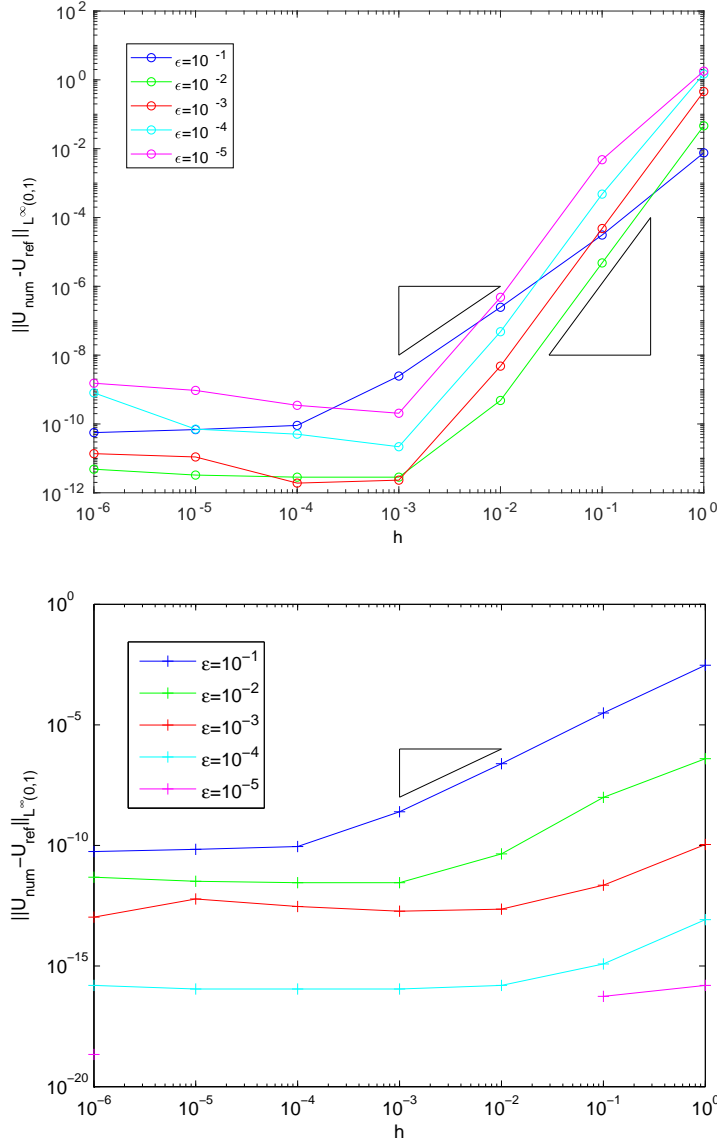


FIG. 5.2. Error of the second order WKB-method on the interval  $[0, 1]$  as a function of  $h$ , for 5 values of  $\epsilon$ ; at the top the results with the phase  $\tilde{\phi}$  computed via Simpson's rule; at the bottom the analogous results with  $\phi$  computed via Clenshaw-Curtis algorithm and barycentric interpolation.

for small values of the semiclassical parameter  $\epsilon$ . The method presented in [1] uses the leading terms of the WKB approximation to solutions of (1.1) to reformulate the problem in terms of less oscillatory functions. Two methods were given in [1], one of first order and one of second order. An error analysis showed that the error (1.6) has a term due to the numerical computation of the phase which is proportional to  $1/\epsilon$  and thus problematic in the limit  $\epsilon \rightarrow 0$ . In this paper we have first refined the error analysis by taking into account numerical errors in the phase computation also in the transformations (2.5) between the solution to the Schrödinger equation and

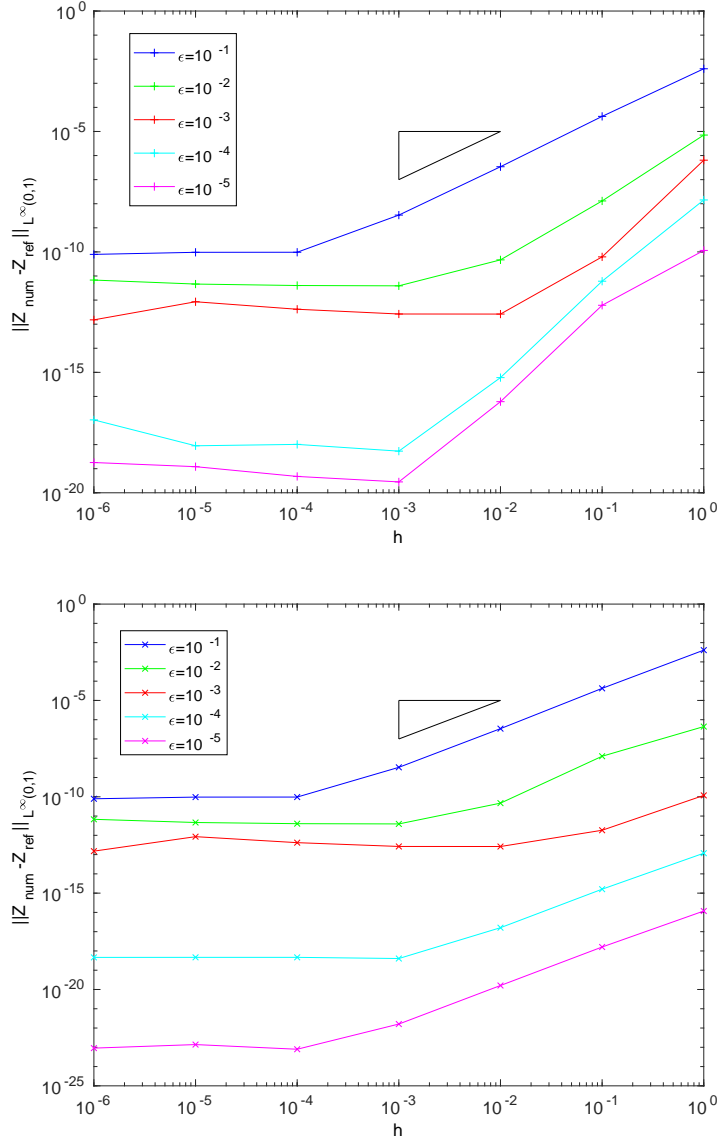


FIG. 5.3. Error of the second order WKB-method on the interval  $[0, 1]$  as a function of  $h$ , for 5 values of  $\epsilon$ ; at the top the results with the phase  $\tilde{\phi}$  computed via Simpson's rule; at the bottom the analogous results with  $\tilde{\phi}$  computed via Clenshaw-Curtis algorithm and barycentric interpolation.

the reduced system. In addition the phase is now computed with a spectral method showing an exponential decrease of the numerical error with the number of modes for analytical functions. The advantages of this approach with respect to a Simpson method are illustrated for several examples.

An interesting question as a consequence of this work is whether the spectral approach to the numerical computation of the phase can be extended to the complete reduced system (2.3), i.e., whether this system can be itself efficiently treated with a spectral method. This is not obvious since the function  $Z$  is known to have oscillations

of twice the frequency as the solution  $\varphi$  of the Schrödinger equation, but at much smaller amplitude. To check whether spectral methods can be efficient in this context will be the subject of further work.

**Acknowledgments.** The first author (AA) was supported by the FWF-doctoral school “Dissipation and dispersion in non-linear partial differential equations”, the SFB F65, the bi-national FWF-project I3538-N32, and a sponsorship by *Clear Sky Ventures*. This work was supported by the ANR-FWF project ANuI. CK thanks for support by the visite BFC project NAANoD, the ANR-17-EURE-0002 EIPHI and by the European Union Horizon 2020 research and innovation program under the Marie Skłodowska-Curie RISE 2017 grant agreement no. 778010 IPaDEGAN.

## REFERENCES

- [1] A. Arnold, N. Ben Abdallah, C. Negulescu, *WKB-based schemes for the oscillatory 1D Schrödinger equation in the semi-classical limit*, SIAM J. Numer. Anal. 49 (2011), no. 4 pp. 1436–1460.
- [2] A. Arnold, K. Döpfner, *Stationary Schrödinger equation in the semi-classical limit: WKB-based scheme coupled to a turning point*, Calcolo 57 (2020), no. 1, article no. 3.
- [3] A. Arnold, C. Negulescu, *Stationary Schrödinger equation in the semi-classical limit: numerical coupling of oscillatory and evanescent regions*, Numer. Math. 138 (2018), no. 2, pp. 501–536.
- [4] J.-P. Berrut, L. N. Trefethen, *Barycentric Lagrange Interpolation*, SIAM Review, Vol. 46, 3 (2004), pp. 501-517.
- [5] M.M. Chawla, *Error Estimates for the Clenshaw-Curtis Quadrature*, Math. Comp. Vol 22, 103 (1968), pp. 651-656.
- [6] C. W. Clenshaw, A. R.Curtis, *A method for numerical integration on an automatic computer*, Numerische Mathematik 2, 1 (1960), pp. 197-205.
- [7] D. Cohen, E. Hairer, C. Lubich, *Modulated Fourier Expansions of Highly Oscillatory Differential Equations*, Found. Comput. Math. 3 (2003) pp. 327–345.
- [8] P. Degond, S. Gallego, F. Méhats, *An asymptotic preserving scheme for the Schrödinger equation in the semiclassical limit*, C.R. Acad. Sci. Paris, Ser. I, 345 (2007), no. 9, pp. 531–536.
- [9] W. E, B. Engquist, X. Li, W. Ren, E. Vanden-Eijnden, *Heterogeneous multiscale methods: a review*, Commun. Comput. Phys. 2 (2007), no. 3, pp. 367–450.
- [10] J. Geier, *Efficient integrators for linear highly oscillatory ODEs based on asymptotic expansions*, PhD-dissertation at TU Wien, 2011.
- [11] E. Hairer, C. Lubich, G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd Ed., Springer-Verlag, Berlin Heidelberg, 2006.
- [12] W.J. Handley, A.N. Lasenby, M.P. Hobson, *The Runge-Kutta-Wentzel-Kramers-Brillouin method*, preprint, 2016. <https://arxiv.org/abs/1612.02288>
- [13] F. Ihlenburg, I. Babuška, *Finite element solution of the Helmholtz equation with high wave number. I. The h-version of the FEM*, Comput. Math. Appl. 30 (1995), no. 9, pp. 9–37.
- [14] F. Ihlenburg, I. Babuška, *Finite element solution of the Helmholtz equation with high wave number. II. The h-p version of the FEM*, SIAM J. Numer. Anal. 34 (1997), no. 1, pp. 315–358.
- [15] A. Iserles, *On the Global Error of Discretization Methods for Highly-Oscillatory Ordinary Differential Equations*, BIT, 42 (2002), no. 3, pp. 561-599.
- [16] A. Iserles, H.Z. Munthe-Kaas, S.P. Nørsett, A. Zanna, *Lie-group methods*, Acta Numerica 9 (2000) pp. 215–365.
- [17] A. Iserles, S.P. Nørsett, S. Olver, *Highly oscillatory quadrature: The story so far*, in A. Bermudez de Castro, ed., *Proceeding of ENuMath, Santiago de Compostella (2006)*, Springer Verlag, 2006, pp. 97–118.
- [18] T. Jahnke, *Long-time-step integrators for almost-adiabatic quantum dynamics*, SIAM J. Sci. Comp., 25 (2004), pp. 2145–2164.
- [19] T. Jahnke, C. Lubich, *Numerical integrators for quantum dynamics close to the adiabatic limit*, Numerische Mathematik, 94 (2003), pp. 289–314.
- [20] L.D. Landau, E.M. Lifschitz, *Quantenmechanik*, Akademie-Verlag, Berlin, 1985.
- [21] C.S. Lent, D.J. Kirkner, *The Quantum Transmitting Boundary Method*, J. Appl. Phys., 67 (1990), pp. 6353–6359.

- [22] K. Lorenz, T. Jahnke, C. Lubich, *Adiabatic integrators for highly oscillatory second-order linear differential equations with time-varying eigendecomposition*, BIT, 45 (2005), no. 1, pp. 91–115.
- [23] J.-F. Mennemann, A. Jüngel, H. Kosina, *Transient Schrödinger–Poisson simulations of a high-frequency resonant tunneling diode oscillator*, J. Computat. Phys. 239 (2013) pp. 187–205.
- [24] P.C. Moan, J. Niesen, *Convergence of the Magnus Series*, Found. Comput. Math. 8 (2008) pp. 291–301.
- [25] C. Negulescu, *Numerical analysis of a multiscale finite element scheme for the resolution of the stationary Schrödinger equation*, Numerische Mathematik, 108 (2008), no. 4, pp. 625–652.
- [26] S. Olver, *Moment-free numerical integration of highly oscillatory functions*, IMA J. Numer. Anal., 26 (2006), pp. 213–227.
- [27] J.P. Sun, G.I. Haddad, P. Mazumder, J.N. Schulman, *Resonant Tunneling Diodes: Models and Properties*, Proc. of the IEEE, 86 (1998), no. 4, pp. 641–661.
- [28] L. N. Trefethen, *Spectral Methods in MATLAB*, vol. 10 of Software, Environments, and Tools, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.