# An accelerated frequency-independent solver for oscillatory differential equations

Tara Stojimirovic

*Department of Mathematics, University of Toronto*

James Bremer

*Department of Mathematics, University of Toronto*

**Abstract**

Oscillatory second order linear ordinary differential equations arise in many numerical and scientific calculations. Because the running times of standard ODE solvers increase roughly linearly with frequency when they are applied to such problems, a variety of specialized methods, most of them quite complicated, have been proposed. Here, we point out that one of the simplest conceivable approaches not only works, but yields a scheme for solving oscillatory second order linear ordinary differential equations which is significantly faster than current state-of-the-art techniques. Our method, which operates by constructing a slowly varying phase function representing a basis in the space of solutions of the differential equation, runs in time independent of the frequency of oscillations of the solutions and can be applied to second order equations whose solutions are oscillatory in some regions and slowly varying in others. In the high-frequency regime, our algorithm discretizes the nonlinear Riccati equation satisfied by the derivative of the phase function via a Chebyshev spectral collocation method and applies the Newton-Kantorovich method to the resulting system of nonlinear algebraic equations. We prove that the iterates converge quadratically to a nonoscillatory solution of the Riccati equation. The quadratic convergence of the Newton-Kantorovich method and the simple form of the linearized equations ensure that this procedure is extremely efficient. Our algorithm then extends the slowly varying phase function calculated in the high-frequency regime throughout the solution domain by solving a certain third order linear ordinary differential equation related to the Riccati equation. We describe the results of numerical experiments showing that our algorithm is orders of magnitude faster than existing schemes, including the modified Magnus method [18], the current state-of-the-art approach [7] and the recently introduced ARDC method [1].

*Keywords:* oscillatory problems, fast algorithms, ordinary differential equations

## 1. Introduction

Oscillatory second order linear ordinary differential equations are widely used in scientific and numerical calculations. They arise, for instance, in plasma physics [15, 11, 2], Hamiltonian dynamics [28], quantum mechanics [14] and cosmology [24, 2]. Moreover, numerous important families of special functions, such as the Jacobi polynomials, the Bessel functions and the spheroidal wave functions, satisfy such equations (see, for instance, [12]).

In many applications, the speed with which the differential equations are solved is of great importance. This is case for the calculations described in [2], which entail the numerical solution of billions of oscillatory ODEs, and there are numerous other such examples. Moreover, many calculations related to special functions require a large number of function evaluations. To give one example, it is necessary to rapidly evaluate associated Legendre functions of many different orders and degrees at a large number of points in order to apply the spherical harmonic transform via a butterfly algorithm [9].

The principal result of this paper applies to equations of the form

$$y''(t) + \omega^2 q(t, \omega) y(t) = 0, \quad -\infty < a < t < b < \infty, \tag{1}$$

where $q(t, \omega)$ is a smooth function such that there exist positive constants $q_{\min}$, $\omega_0$ and a positive integer $M$ with

$$\left| \left( \frac{d}{dt} \right)^j q(t, \omega) \right| \leq 1 \quad \text{for all} \quad 0 \leq j \leq 2M \tag{2}$$

and

$$q_{\min} \leq q(t, \omega) \tag{3}$$

for all $a \leq t \leq b$ and all $\omega \geq \omega_0$. It is well known that under these assumptions, there exists a basis $\{u, v\}$ in the space of solutions of (1) such that

$$u(t) = \frac{\exp\left( i\omega \int_a^t \sqrt{q(s, \omega)} \, ds \right)}{\sqrt{\omega} \, q(t, \omega)^{\frac{1}{4}}} + \mathcal{O}\left( \frac{1}{\omega} \right) \quad \text{as} \quad \omega \to \infty \tag{4}$$

and

$$v(t) = \frac{\exp\left( -i\omega \int_a^t \sqrt{q(s, \omega)} \, ds \right)}{\sqrt{\omega} \, q(t, \omega)^{\frac{1}{4}}} + \mathcal{O}\left( \frac{1}{\omega} \right) \quad \text{as} \quad \omega \to \infty. \tag{5}$$

These expressions are known as Liouville-Green or first-order WKB approximates for (1), and they are usually derived by analyzing the Riccati equation

$$r'(t) + (r(t))^2 + \omega^2 q(t, \omega) = 0 \tag{6}$$

satisfied by the logarithmic derivatives of solutions of (1). Throughout this paper, we will refer to any function $\psi$ such that $\exp(\psi(t))$ is a solution of (1) as an *exponential phase function* for (1), so that the solutions of the Riccati equation (6) are the derivatives of exponential phase functions. Among other things, (4) and (5) imply that the solutions of (1) are oscillatory, and that the parameter $\omega$ is a reasonable proxy for their frequency of their oscillations. It follows that when a standard ODE solver is applied to (1), its running time grows roughly linearly with $\omega$.

Here, we give an elementary proof that under the assumptions (2) and (3), there exists a nonoscillatory solution of the Riccati equation satisfied by the logarithmic derivatives of the solutions of (1). The solution is nonoscillatory in the sense that it can be approximated to a fixed accuracy via a Chebyshev expansion the number of terms of which is independent of the frequency parameter $\omega$, at least for sufficiently large $\omega$. Moreover, we show that when the Riccati equation is discretized using a standard Chebyshev spectral collocation method and the Newton-Kantorovich

method is applied to the resulting system of nonlinear algebraic equations, the iterates converge quadratically to a vector giving the values of this nonoscillatory phase function at the collocation nodes, again provided that the frequency parameter $\omega$ is sufficiently large. Because of the quadratic convergence of the iterates and the fact that the linearized equations which arise are of an extremely simple type that can be inverted rapidly, this procedure is considerable faster than existing methods for constructing nonoscillatory phase functions representing solutions of ordinary differential equations of the form (1).

We go on to describe a numerical algorithm for solving a large class of oscillatory differential equations. In particular, our algorithm applies in cases when the solutions are highly oscillatory in some regions and slowly varying in others. It runs in time independent of frequency, and obtains accuracy on the order of the condition number of evaluation of the solutions. Our method operates by constructing a phase function which is represented via a piecewise Chebyshev expansion. More explicitly, it adaptively subdivides the solution domain $[a, b]$ of the equation into a collection of discretization subintervals $[a_1, b_1], \ldots, [a_m, b_m]$ and, on each subinterval, the phase function is represented via a Chebyshev expansion of a fixed order $k$. The Newton-Kantorovich method is used to construct the polynomial expansion over each "high-frequency interval." The linearized equations which arise during the course of the Newton-Kantorovich iterations are not uniquely solvable in the low-frequency regime, however, and so our algorithm extends the phase function into the rest of the solution domain by solving initial and terminal value problems for Appell's equation. Appell's equation is a certain third order linear ordinary differential equation related to the Riccati equation (6). We prefer Appell's equation to the Riccati equation for two reasons. First, it is linear and this means that it can be solved more rapidly than the Riccati equation, at least in the low-frequency regime where the simple form of the linearized Riccati equation is difficult to exploit. Secondly, Riccati's equation becomes numerically degenerate when the coefficient $q$ is zero or close to it, whereas, as observed in [8], Appell's equation can be solved in a numerical stable fashion even when $q$ is close to 0 or negative and of large magnitude. As described here, our algorithm is limited to equations whose coefficients are nonnegative throughout the solution domain; however, with minor alterations, it could be applied in cases in which the coefficient becomes negative on some parts of its domain. Over such regions, the solutions of the differential equation behave like combinations of rapidly increasing and decreasing exponential functions, but they can nonetheless be accurately represented via phase functions (see [8]).

Almost all specialized techniques for solving oscillatory ordinary differential equations are based on the observation that such equations admit nonoscillatory exponential representations. One of the most widely used algorithms is the modified Magnus expansion method described in [18]. It exploits the existence of efficient exponential representations of the solutions of oscillatory differential equations by using an exponential integrator as the basis of a step method. Moreover, in each step, the equation is preconditioned by the solution of the constant coefficient equation obtained by freezing the coefficients at the midpoint. The running time of the modified Magnus method grows as $\mathcal{O}\left(\omega^{3/4}\right)$ when it is applied to a scalar equation of the form (1). It should be noted, though, that it applies to a wider class of differential equations than that considered here, including quite general systems of linear ordinary differential equations. Like the modified Magnus method, the scheme of [23] applies to a large class of systems of differential equations. It operates by constructing a preconditioner using the eigendecomposition of the system's coefficient matrix. When applied to scalar equations of the type appearing here, its running time

grows as $\mathcal{O}\left(\sqrt{\omega}\right)$. The WKB marching method of [4, 21], which is specific to second order linear ordinary differential equations, uses second order WKB expansions as a preconditioner in regions where the solutions are rapidly oscillating and applies a standard Runge-Kutta method in the low-frequency regime. The running time of the WKB marching method also grows as $\mathcal{O}\left(\sqrt{\omega}\right)$ in typical cases.

Another class of numerical methods uses asymptotic approximations to represent solutions of second order linear ordinary differential equations in the high-frequency regime directly, as opposed to using asymptotic approximations as preconditioners. The article [2] introduces a solver for second order linear ordinary differential equations that represents solutions via low-order WKB expansions in regions where the solutions are highly oscillatory, and applies a standard Runge-Kutta method in the low-frequency regime. The running time of this algorithm is independent of frequency in certain special cases, for instance when the solutions oscillate at extremely high frequency over the whole solution domain, but, in the general case, it grows as $\mathcal{O}\left(\omega\right)$. The ARDC scheme [1] improves on [2] by using higher order WKB-like approximations constructed numerically through an iterative method to represent solutions in the oscillatory regime. The running time of [1] is independent of frequency, but it is significantly slower and no more widely applicable than the earlier frequency-independent method introduced in [7] (see Section 6.2).

The iterative method used by [1] is strongly related to the approach of [29, 30, 31]. There, symbolic formulas which represent asymptotic approximations of solutions of the differential equation satisfied by the square of the imaginary part of solutions of the Riccati equation are constructed via an iterative scheme. The calculations are conducted symbolically rather than numerically because each iteration requires a further derivative of the coefficient. In [27], a method for solving a large class of oscillatory differential equations in time independent of frequency is described. It applies the "differential GMRES" procedure described in [26] to a auxiliary system derived from the differential equation under consideration. It also uses symbolic calculations because it requires high order derivatives of the entries of the coefficient matrix.

The method of [7] appears to be the current state-of-the-art algorithm for numerical solution of the class of equations considered here. It operates by solving the equation

$$\left(\alpha'(t)\right)^2 - \omega^2 q(\omega, t) - \frac{3}{4}\left(\frac{\alpha''(t)}{\alpha'(t)}\right)^2 + \frac{1}{2}\frac{\alpha'''(t)}{\alpha'(t)} = 0 \tag{7}$$

obtained by considering the real and imaginary parts of (6) separately in order to calculate a function $\alpha$ such that

$$\frac{\cos\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}} \quad \text{and} \quad \frac{\sin\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}} \tag{8}$$

constitute a basis in the space of solutions of (1). We refer any $\alpha$ such that the function $u$ and $v$ defined in (8) are a pair of independent solutions of (1) as a *trigonometric phase function* for (1). Equation (7) is satisfied by the derivatives of trigonometric phase functions, and we refer to it as Kummer's equation after E. E. Kummer who studied it in [22]. The algorithm of [7] runs in time independent of frequency and like the method of this paper, but unlike those proposed in [1, 4, 21, 2], the same exponential representations of solutions are used throughout the solution domain. This is highly conducive to certain calculations, such as the computation of the zeros of solutions [6] and the application of Strum-Liouville eigentransforms [9]. Almost

4

all of the solutions of the Riccati equation are oscillatory or singular and the main focus of [7] is a mechanism which selects the desired nonoscillatory solution. This is done by smoothly deforming $q$ so that it is constant on $[a, (3a+b)/4]$ and equal to the original coefficient $q$ on $[(a+3b)/4, b]$. The value of an appropriate nonoscillatory solution of the Riccati equation corresponding to the deformed coefficient is known at $a$, and by solving an initial value problem for the modified Riccati equation, the value at $b$ of a slowly-varying solution of the original Riccati equation corresponding to (1) is obtained. The derivative of the desired phase function is then constructed by solving a terminal value problem for the original Riccati equation. One of the main observations of this paper is that the machinery of [7] is unnecessary because discretizing the Riccati equation in one of the simplest ways possible and applying one of the most basic strategies for solving the resulting nonlinear system yields a faster method.

The remainder of this paper is structured as follows. Section 2 reviews the necessary mathematical and numerical preliminaries. Section 3 gives an elementary proof of the existence of a nonoscillatory solution of the Riccati equation. In Section 4, we prove that when the Riccati equation is discretized via a Chebyshev spectral method and the Newton-Kantorovich method method is applied to the resulting system of nonlinear algebraic equations, the iterates converge to a vector which represents the nonoscillatory solution whose existence is shown in Section 3. Section 5 details our numerical algorithm, and we go on to describe the results of numerical experiments conducted to demonstrate its properties in Section 6. We close with a few brief remarks regarding this work and future directions for research in Section 7.


## 2. Mathematical and Numerical Preliminaries

### 2.1. Notation and conventions

We denote the Fréchet derivative of a map $F : X \to Y$ between Banach spaces at the point $x$ via $F'(x)$. Assuming it exists, the Fréchet derivative is an element of the space of bounded linear functions $X \to Y$, for which we use the notation $L(X, Y)$. We will use $B_r(x)$ for the ball of radius $r > 0$ centered at the point $x \in X$. A function $F : \Omega \subset X \to Y$ given on an open subset $\Omega$ of a Banach space $X$ is said to be continuously differentiable on $\Omega$ provided the map $\Omega \to L(X, Y)$ given by $x \mapsto F'(x)$ is continuous.

In the analysis of Section 4, we will use vectors in $\mathbb{R}^n$ to represent certain functions and it is convenient to introduce a convention for distinguishing the two. Accordingly, throughout this paper we will display the names of vectors in the space $\mathbb{R}^n$ using a bold font. When working with $\mathbb{R}^n$, we will use the $l^\infty$ norm rather the Euclidean norm. We also adopt the conventions that $\|\mathbf{v}\|$ refers to the $l^\infty$ norm of the vector $\mathbf{v} \in \mathbb{R}^n$ and $\|T\|$ refers to the $l^\infty$ operator norm of the linear mapping $T : \mathbb{R}^n \to \mathbb{R}^m$. Moreover, we let $\text{diag}(\mathbf{x})$ denote the $n \times n$ diagonal matrix whose diagonal entries are the elements of the vector $\mathbf{x} \in \mathbb{R}^n$. We use the notation $\mathbf{v} \circ \mathbf{w}$ to denote the pointwise product of the vectors $\mathbf{v}$ and $\mathbf{w}$ — that is, the vector whose $j^{th}$ entry is the product of the $j^{th}$ entry of the vector $\mathbf{v}$ with the $j^{th}$ entry of the vector $\mathbf{w}$.

We use $T_j$ for the Chebyshev polynomial of degree $j$. We denote nodes of the $k$-point Chebyshev extremal grid on $[-1, 1]$ by

$$t_{1,k}^{\text{cheb}}, t_{2,k}^{\text{cheb}}, \ldots, t_{k,k}^{\text{cheb}}, \tag{9}$$

so that

$$t_{i,k}^{\text{cheb}} = \cos\left(\pi \frac{k-i}{k-1}\right). \tag{10}$$

We use $\mathscr{D}_k$ for the $k \times k$ spectral differential matrix which takes the vector

$$\begin{pmatrix} p\left(t_{1,k}^{\text{cheb}}\right) & p\left(t_{2,k}^{\text{cheb}}\right) & \cdots & p\left(t_{k,k}^{\text{cheb}}\right) \end{pmatrix} \tag{11}$$

of values of a polynomial $p$ of degree less than $k$ at the Chebyshev nodes to the vector

$$\begin{pmatrix} p'\left(t_{1,k}^{\text{cheb}}\right) & p'\left(t_{2,k}^{\text{cheb}}\right) & \cdots & p'\left(t_{k,k}^{\text{cheb}}\right) \end{pmatrix} \tag{12}$$

of the values of its derivatives at the same nodes. Moreover, we let $\mathscr{I}_k$ be the spectral integration matrix which takes the vector of values (11) to the vector

$$\begin{pmatrix} q\left(t_{1,k}^{\text{cheb}}\right) & q\left(t_{2,k}^{\text{cheb}}\right) & \cdots & q\left(t_{k,k}^{\text{cheb}}\right) \end{pmatrix}, \tag{13}$$

where

$$q(t) = \int_{-1}^{t} p(s) \ ds. \tag{14}$$

Finally, we let $\mathscr{C}_k$ denote the matrix which takes the vector of values (11) of a polynomial $p$ of degree less $k$ to the vector

$$\begin{pmatrix} a_0 & a_2 & \cdots & a_{k-1} \end{pmatrix}, \tag{15}$$

of coefficients in the Chebyshev expansion

$$p(t) = \sum_{j=0}^{k-1} a_j T_j(t). \tag{16}$$

### 2.2. The Newton-Kantorovich theorem

In [20], Kantorovich generalized Newton's method to the case of maps between Banach spaces and gave conditions for its convergence. Here, we state a simplified version of his theorem which can be found as Theorem 7.7-4 in Section 7.7 of [10].

**Theorem 1** (Newton-Kantorovich). *Suppose that $\Omega$ is an open subset of the Banach space $X$, that $Y$ is a Banach space and that $F : \Omega \subset X \to Y$ is continuously differentiable. Assume also that there exist a point $x_0 \in \Omega$ and constants $\lambda$ and $\eta$ such that*

1. *$F'(x_0)$ admits an inverse $F'(x_0)^{-1} \in L(Y, X)$,*
2. *$B_\eta(x_0) \subset \Omega$,*
3. *$0 < \lambda \leq \dfrac{\eta}{2}$,*
4. *$\left\| F'(x_0)^{-1} F(x_0) \right\| \leq \lambda$   and*
5. *$\left\| F'(x_0)^{-1} \left( F(x) - F(y) \right) \right\| \leq \dfrac{1}{\eta} \left\| x - y \right\|$  for all $x, y \in B_\eta(x_0)$.*

*Then, $F'(x)$ has a bounded inverse $F'(x)^{-1} \in L(Y, X)$ for each $x \in B_\eta(x_0)$, the sequence $\{x_n\}$ defined by*

$$x_{n+1} = x_n - F'(x_n)^{-1} F(x_n) \tag{17}$$

6

*is contained in the open ball*

$$B_{\eta^-}(x_0), \quad where \quad \eta^- = \eta\left(1 - \sqrt{1 - \frac{2\lambda}{\eta}}\right) \le \eta, \tag{18}$$

*and $x_n$ converges to a zero $x^*$ of $F$. Moreover, for each $n \ge 0$,*

$$\|x_n - x^*\| \le \frac{\eta}{2^n}\left(\frac{\eta^-}{\eta}\right)^{2^n} \tag{19}$$

*and $x^*$ is the only zero of $F$ in the ball $B_\eta(x_0)$.*

We will refer to the equation

$$F'(x_n)h = -F(x_n), \tag{20}$$

which arises when calculating the $(n+1)^{st}$ Newton-Kantorovich iterate as the linearization of the equation $F(x) = 0$ around the point $x_n$.

### 2.3. The Banach fixed point theorem

A proof of the following well-known theorem can be found in Section 3.7 of [10], among many other sources.

**Theorem 2** (Banach fixed point theorem). *Suppose that $F : X \to X$ is a map between Banach spaces, and that there exists a constant $0 \le \gamma < 1$ such that $\|F(x_1) - F(x_2)\| \le \gamma \|x_1 - x_2\|$ for all $x_1, x_2 \in X$. Then $F$ has a unique fixed point $x^*$. Moreover, given any $x_0 \in X$, the sequence $\{x_n\}$ defined via $x_{n+1} = F(x_n)$ converges to $x^*$, and the following estimate holds*

$$\|x_n - x^*\| \le \frac{\gamma^n \|F(x_0) - x_0\|}{1 - \gamma}. \tag{21}$$

An immediate consequence of Theorem 2 is that a system of linear algebraic equations of the form

$$(I + S)\mathbf{x} = \mathbf{y} \tag{22}$$

can be solved extremely rapidly via the iteration

$$\mathbf{x_0} = 0 \quad and \quad \mathbf{x_{n+1}} = -S\mathbf{x_n} + \mathbf{y} \tag{23}$$

as long as there is some operator norm $\|\cdot\|_0$ such that $\|S\|_0 \ll 1$. The spectral radius $\rho(A)$ of a matrix $A$ is the maximum of the magnitudes of its eigenvalues. It is well known (see, for instance, [17]) that given any $\epsilon > 0$, there exists a matrix norm $\|\cdot\|_\epsilon$ such that

$$\rho(S) \le \|S\|_\epsilon \le \rho(S) + \epsilon, \tag{24}$$

so the iteration (23) converges rapidly to a solution of (22) assuming $\rho(S) \ll 1$.

### 2.4. The Riccati equation and its variants

In this section, we briefly discuss the relationship between the solutions of the ordinary differential equations (1), (6) and (7). Moreover, we discuss a related third order linear ordinary differential equation.

It can be verified through direct computation that if

$$y(t) = \exp\left(\int r(s)\,ds\right) \tag{25}$$

satisfies (1), then $r$ is a solution of the Riccati equation (6). It follows by separating out the real and imaginary parts of the Riccati equation that its solutions are of the form

$$r(t) = i\alpha'(t) - \frac{1}{2}\frac{\alpha''(t)}{\alpha'(t)}, \tag{26}$$

where $\alpha$ satisfies Kummer's equation (7). In particular, the functions

$$u(t) = \frac{\sin\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}} \quad \text{and} \quad v(t) = \frac{\cos\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}} \tag{27}$$

constitute a basis of real-valued functions in the space of solutions of (1). In a similar vein, if $u$, $v$ is any pair of real-valued solutions of (1) whose Wronskian is 1, then

$$\alpha'(t) = \frac{1}{(u(t))^2 + (v(t))^2} \tag{28}$$

is a solution of Kummer's equation. Moreover, from (27) it is clear that the modulus function

$$m(t) = (u(t))^2 + (v(t))^2 \tag{29}$$

is equal to the reciprocal of $\alpha'(t)$. A somewhat lengthy, but straightforward, computation shows that $m(t)$ is a solution of the third order linear ordinary differential equation

$$m'''(t) + 4\omega^2 q(\omega, t)m'(t) + 2\omega^2 \frac{dq}{dt}(\omega, t)m(t) = 0, \tag{30}$$

which we refer to as Appell's equation in light of [3].

In this way, each "phase function" for (1) corresponds with a solution of (6), a solution of (7), a modulus function satisfying (30) and a pair of real-valued solutions of (1) whose Wronskian is 1. For the sake of clarity and precision, we refer to functions $\psi$ such that $\exp(\psi(t))$ is a solution of (1) as exponential phase functions, and we say that $\alpha$ is a trigonometric phase function if the functions defined in (27) constitute a basis in the space of solutions of (1). Given the close relationship (26) between the solutions of the Riccati equation and those of Kummer's equation, however, there is little real distinction between the two.

### 2.5. A bound on the magnitude of Chebyshev coefficients

The Chebyshev expansion of a function $f : [a, b] \to \mathbb{C}$ is the series

$$\sum_{n=0}^{\infty} a_n T_n \left(\frac{2}{b-a}t - \frac{b+a}{b-a}\right), \tag{31}$$

where

$$a_0 = \frac{1}{\pi}\int_{-1}^{1} f\left(\frac{b-a}{2}t + \frac{b+a}{2}\right)\frac{dt}{\sqrt{1-t^2}} \tag{32}$$

and

$$a_n = \frac{2}{\pi}\int_{-1}^{1} f\left(\frac{b-a}{2}t + \frac{b+a}{2}\right)T_n(t)\frac{dt}{\sqrt{1-t^2}} \quad \text{for all} \quad n > 0. \tag{33}$$

8

The following theorem, which can be found in a slightly different form in [32], gives an estimate on the size of the Chebyshev coefficients of $f$.

**Theorem 3.** *If $f, f', f'', \ldots, f^{(k-1)}$ are absolutely continuous on $[a, b]$ and*

$$V = \left(\frac{b-a}{2}\right)^k \int_{-1}^1 \left| \frac{f^{(k)}\left(\frac{b-a}{2}t + \frac{b+a}{2}\right)}{\sqrt{1-t^2}} \right| dt, \tag{34}$$

*then*

$$|a_n| \leq \frac{2V}{\pi n(n-1)(n-2)\cdots(n-k)} \quad \text{for all} \quad n > k. \tag{35}$$

## 3. Existence of Nonoscillatory Phase Functions

In this section, we first give an elementary argument showing that under the assumptions (2) and (3), the Riccati equation admits a solution which is nonoscillatory in the sense that it can be approximated via a polynomial expansion whose number of terms $k$ is independent of $\omega$, at least for large values of $\omega$. In many cases of interest, $q$ is nonoscillatory in more robust sense. Accordingly, we go on to discuss the principal theorem of [7], which shows the existence of a phase function which is nonoscillatory in a stronger sense under regularity conditions on $q$ which are more stringent than (2) and (3).

*3.1. Elementary proof of the existence of a nonoscillatory phase function*
We let $r_M$ be defined via the formula

$$r_M(t) = \omega u_0(t) + u_1(t) + \cdots + u_M(t)\omega^{1-M}, \tag{36}$$

where

$$u_0(t) = i\omega\sqrt{q(t,\omega)}, \quad u_1(t) = -\frac{1}{4}\frac{q'(t)}{q(t)} \tag{37}$$

and

$$u_n(t) = -\frac{1}{2u_0(t)}\left(u'_{n-1}(t) + \sum_{j=1}^{n-1} u_j(t)u_{n-j}(t)\right) \quad \text{for all} \quad j \geq 1. \tag{38}$$

We will first show that there exists a solution $r$ of (6) such that

$$\|r - r_M\|_{L^\infty([a,b])} = \mathcal{O}\left(\frac{1}{\omega^M}\right) \quad \text{as} \quad \omega \to \infty. \tag{39}$$

To that end, we insert the expression

$$r(t) = r_M(t) + \delta(t) \tag{40}$$

into (6), which yields the equation

$$\delta'(t) + 2r_M(t)\delta(t) + (\delta(t))^2 = -G(t), \tag{41}$$

where

$$G(t) = r'_M(t) + (r_M(t))^2 + \omega^2 q(t,\omega). \tag{42}$$

9

We observe that the assumptions (2) and (3) together with the fact that $u_j(t)$ involves only $q$ and the first $j$ derivatives of $q$, imply

$$\|G\|_{L^\infty([a,b])} = \mathcal{O}\left(\frac{1}{\omega^{M-1}}\right) \quad \text{and} \quad \|G'\|_{L^\infty([a,b])} = \mathcal{O}\left(\frac{1}{\omega^{M-1}}\right) \quad \text{as} \quad \omega \to \infty. \tag{43}$$

We proceed by setting

$$H(t) = 2\int_a^t r_M(s), \tag{44}$$

and multiplying both sides of (41) by $\exp(H(t))$ to obtain

$$\exp(H(t))\delta'(t) + 2\exp(H(t))r_M(t)\delta(t) = -\exp(H(t))\left(\delta(t)\right)^2 + \exp(H(t))G(t). \tag{45}$$

Since the left-hand side of (45) is the derivative of $\delta(t)\exp(H(t))$, we can reformulate it as the integral equation

$$\delta(t) = -\exp(-H(t))\int_a^t \exp(H(s))\left((\delta(s))^2 + G(s)\right) ds. \tag{46}$$

We will show that when $\omega$ is sufficiently large, the operator

$$S\left[\delta\right](t) = -\exp(-H(t))\int_a^t \exp(H(s))\left((\delta(s))^2 + G(s)\right) ds \tag{47}$$

is a contraction on a closed ball $B$ of whose radius $2C/\omega^M$ centered at $0$ in $L^\infty\left([a,b]\right)$, where $C$ is an appropriately chosen constant. It will then follow from Theorem 2 that there is a solution of (46) in this ball, and the desired bound (39) is a consequence of the existence of this solution and (40).

Our assumptions on $q$ and its derivatives imply that the functions $|\exp(\pm H(t))|$ are bounded independent of $\omega$. It follows that there exists a constant $C > 0$ which is independent of $\omega$ and such that

$$\left\| -\exp(-H(t))\int_a^t \exp(H(s))f(s)\,ds \right\|_{L^\infty([a,b])} \le C\|f\|_{L^\infty([a,b])} \tag{48}$$

for any $f \in L^\infty\left([a,b]\right)$. In particular, we have

$$\left\| -\exp(-H(t))\int_a^t \exp(H(s))\left(\delta(s)\right)^2 ds \right\|_{L^\infty([a,b])} \le C\|\delta\|_{L^\infty([a,b])}^2. \tag{49}$$

Now, integration by parts shows that

$$\begin{aligned}
&\exp(-H(t))\int_a^t \exp(H(s))G(s)\,ds \\
&= \exp(-H(t))\int_a^t \exp(H(s))H'(s)\frac{G(s)}{H'(s)}\,ds \\
&= \exp(-H(t))\left(\exp(H(t))\frac{G(t)}{H'(t)} - \exp(H(a))\frac{G(a)}{H'(a)}\right. \\
&\quad \left. - \int_a^t \exp(H(s))\frac{G'(s)}{H'(s)}\,ds + \int_a^t \exp(H(s))\frac{G(s)H''(s)}{(H'(s))^2}\,ds\right).
\end{aligned} \tag{50}$$

It is easy to see that

$$\frac{1}{H'(s)} = \mathcal{O}\left(\frac{1}{\omega}\right) \quad \text{and} \quad \frac{H''(s)}{(H'(s))^2} = \mathcal{O}\left(\frac{1}{\omega}\right), \tag{51}$$

and it follows from (51), the fact that the functions $\exp(\pm H(t))$ are bounded independent of $\omega$, (50) and (43) that we can adjust the constant $C$ so that

$$\left\| \exp(-H(t)) \int_a^t \exp(H(s)) G(s) \, ds \right\|_{L^\infty([a,b])} \leq \frac{C}{\omega^M} \tag{52}$$

holds for all sufficiently large $\omega$. Together (49) and (52) imply that

$$\| S\left[\delta\right] \|_{L^\infty([a,b])} \leq \frac{4C^3}{\omega^{2M}} + \frac{C}{\omega^M} \tag{53}$$

whenever $\|\delta\|_{L^\infty([a,b])} \leq 2C/\omega^M$. It follows that for all sufficiently large $\omega$ and all $\delta$ in the ball $B$ of radius $2C/\omega^M$ centered at $0$,

$$\| S\left[\delta\right] \|_{L^\infty([a,b])} \leq \frac{2C}{\omega^M}; \tag{54}$$

that is, the operator $S$ preserves the ball $B$. Next, we observe that for all $\delta$ in $B$, we have

$$\left| \exp(-H(t)) \int_a^t \exp(H(s)) \left(\delta_1(s) - \delta_2(s)\right) \left(\delta_1(s) + \delta_2(s)\right) \, ds \right|$$
$$\leq C \left\| \delta_1 - \delta_2 \right\|_{L^\infty([a,b])} \left\| \delta_1 + \delta_2 \right\|_{L^\infty([a,b])} \tag{55}$$
$$\leq \frac{4C^2}{\omega^M} \left\| \delta_1 - \delta_2 \right\|_{L^\infty([a,b])},$$

from which we see that $S$ is a contraction on the ball $B$ provided $\omega$ is sufficiently large. By applying Theorem 2, we conclude that $S$ admits a unique fixed point in $B$ and, as previously discussed, the desired bound (39) follows.

We now apply Theorem 3 to the function $r_M$. Since the first $M$ derivatives of $r_M$ depend on the first $2M$ derivatives of $q$, (2) and (3) imply that

$$V = \left(\frac{b-a}{2}\right)^M \int_a^b \left| \frac{r_M^{(M)}\left(\frac{b-a}{2}t + \frac{b+a}{2}\right)}{\sqrt{1-t^2}} \right| \, dt, \tag{56}$$

where $r_M^{(M)}$ is the $M^{th}$ derivative of the function $r_M$, is bounded independent of $\omega$. We invoke Theorem 3 to see that

$$r_M(t) = \sum_{n=0}^\infty a_n T_n \left(\frac{b-a}{2}t + \frac{b+a}{2}\right), \tag{57}$$

where

$$|a_n| \leq \frac{V}{\pi n(n-1)\cdots(n-M)} \quad \text{for all } n > M. \tag{58}$$

It follows that there exists a constant $D$ independent of $\omega$ such that

$$\left\| r_M(t) - \sum_0^{k-1} a_n T_n \left(\frac{b-a}{2}t + \frac{b+a}{2}\right) \right\|_{L^\infty([a,b])} \leq \frac{D}{k^M}. \tag{59}$$

11

Combining (59) and (39) shows that given $\epsilon > 0$ we can choose an integer $k$ and a solution $r$ of (6) such that

$$\left\| r(t) - \sum_0^{k-1} a_n T_n \left( \frac{b-a}{2} t + \frac{b+a}{2} \right) \right\|_{L^\infty([a,b])} < \epsilon \tag{60}$$

for all sufficiently large $\omega$. Moreover, the proceeding analysis implies that the necessary values of $k$ and $\omega$ are on the order of $\epsilon^{-\frac{1}{M}}$.

*3.2. A stronger results regarding nonoscillatory solutions of the Riccati equation*

We now reproduce a theorem proved in [16] which gives conditions under which the Riccati equation admits solutions that are nonoscillatory in a strong sense. It applies to equations given on the entire real line. However, it is easy to transform to an equation of the type (1) to one given on the real line; for instance, by letting

$$t = \frac{b-a}{2} \tanh(x) + \frac{b+a}{2}. \tag{61}$$

Moreover, in order to simplify notation, we will restrict attention to the case in which the coefficient $q$ does not depend on $\omega$. The theorem can be applied in the general case, assuming that $q$ satisfies the necessary conditions uniformly in $\omega$, for sufficiently large $\omega$. We let

$$p(t) = \frac{1}{q(t)} \left( \frac{5}{4} \left( \frac{q'(t)}{q(t)} \right)^2 - \frac{q''(t)}{q(t)} \right) \tag{62}$$

and define the new variable $x$ via

$$x(t) = \int_0^t \sqrt{q(s)}\, ds. \tag{63}$$

The theorem requires a condition on the function $p(t(x))$. At first glance, it might seem that the relationship between this function and the coefficient $q$ is quite complicated; however, when $p$ is viewed as a function of $x$, it is simply equal to a constant multiple of the Schwarzian derivative $\{t, x\}$ of $t$ with respect to $x$ (see, for instance, Chapter 6 of [25] for a discussion of the Schwarzian derivative and an illustration of the role it plays in the analysis of oscillatory differential equations).

**Theorem 4.** *Suppose that $q \in C^\infty(\mathbb{R})$ is strictly positive, that $x(t)$ is defined via (63) and that $p(x) = -2\{t, x\}$. Suppose further that there exist positive constants $\omega$, $\Gamma$ and $\mu$ with*

$$\omega > 2 \max \left\{ \frac{1}{\mu}, \Gamma \right\} \tag{64}$$

*and*

$$|\widehat{p}(\xi)| \leq \Gamma \exp(-\mu |\xi|) \quad \text{for all} \quad \xi \in \mathbb{R}. \tag{65}$$

*Then there exist smooth functions $\nu$ and $\delta$ such that:*

1. $|\nu(t)| \leq \dfrac{3\Gamma}{8\mu} \left( 1 + \dfrac{4\Gamma}{\omega} \right) \exp(-\mu\omega)$ *for all $t \in \mathbb{R}$,*

2. *the Fourier transform of $\delta$ is supported on $[-\sqrt{2}\omega, \sqrt{2}\omega]$,*

3. $\left|\widehat{\delta}(\xi)\right| \leq \left(1 + \dfrac{2\Gamma}{\omega}\right) \dfrac{\exp\left(-\mu\left|\xi\right|\right)}{4\omega^2 - \xi^2}$ *for all $\left|\xi\right| \leq \sqrt{2}\omega$,*

4. *the function $\alpha$ defined via*

$$\alpha(t) = \omega\sqrt{q(t)} \int_0^t \exp\left(\delta(x(s))\right) \, ds \tag{66}$$

*satisfies the equation*

$$\left(\alpha'(t)\right)^2 - \left(\omega^2 q(t) + \nu(t)\right) - \dfrac{3}{4}\left(\dfrac{\alpha''(t)}{\alpha'(t)}\right)^2 + \dfrac{1}{2}\dfrac{\alpha'''(t)}{\alpha'(t)} = 0, \tag{67}$$

5. *the function $r$ defined via*

$$r(t) = i\alpha'(t) - \dfrac{1}{2}\dfrac{\alpha''(t)}{\alpha'(t)} \tag{68}$$

*satisfies the equation*

$$r'(t) + (r(t))^2 + \left(\omega^2 q(t) + \nu(t)\right) = 0 \quad and \tag{69}$$

6. *both*

$$\left\{\dfrac{\cos\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}}, \ \dfrac{\sin\left(\alpha(t)\right)}{\sqrt{\alpha'(t)}}\right\} \tag{70}$$

*and*

$$\left\{\exp\left(\int_0^t r(s)\,ds\right), \ \exp\left(\int_0^t \overline{r(s)}\,ds\right)\right\} \tag{71}$$

*are bases in the space of solutions of the differential equation*

$$y''(t) + \omega^2\left(q(t) + \nu(t)\right)y(t) = 0, \quad -\infty < t < \infty. \tag{72}$$

Theorem 4 bounds a measure of the complexity of a solution of the perturbed Riccati equation (69) in terms of a bound on the complexity of the coefficient $q$. More explicitly, it states that the Fourier transform of the function $\delta$ is exponentially decaying with frequency assuming that the same is true of $p(t(x))$. Since $r$ is derived from $q$ and $\delta$, this has the effect of also bounding the complexity of $r$. The perturbation $\nu$ decays exponentially fast with the frequency parameter $\omega$, which means that (69) becomes indistinguishable from the Riccati equation for the original differential equation even at very modest values of $\omega$. Since all solutions of the Riccati equation are slowly varying when $\omega$ is small, it follows that for all intents and purposes we can always assume the Riccati equation has a nonoscillatory solution. The bound on the Fourier transform of $\delta$ implies that it can be well-approximated via a polynomial expansion whose number of terms is independent of $\omega$. Assuming the same is true of $q$, the solution $r$ of the perturbed Riccati equation, which is derived from $q$ and $r$, has this property as well. Finally, we note that while the conditions of Theorem 4 appear to be quite stringent, it can, in fact, be applied to a wide class of differential equations since it suffices to approximate the coefficient $q$ using one with the required properties.

## 4. Analysis of the Discretized Riccati Equation

In this section, we prove that when the Riccati equation (6) is discretized over an interval $[a, b]$ via a Chebyshev spectral collocation method and the Newton-Kantorovich method is applied to the resulting system of nonlinear algebraic equations, the iterates converge to a vector which represents a nonoscillatory solution of (6), provided that the frequency parameter $\omega$ is sufficiently large.

We begin by discretizing (6) via a standard Chebyshev spectral collocation method. To that end, we represent the solution $r$ of the Riccati equation by the vector

$$\mathbf{r} = \begin{pmatrix} r(t_1) & r(t_2) & \cdots & r(t_k) \end{pmatrix} \tag{73}$$

of its values at the nodes $t_1, \ldots, t_k$ of the Chebyshev extremal grid on $[a, b]$, replace the differential operator by the Chebyshev spectral differentiation matrix and require that the resulting semi-discrete equation holds at $t_1, \ldots, t_k$. This yields the system of nonlinear algebraic equations

$$0 = F(\mathbf{r}) := \frac{2}{b-a} \mathscr{D}_k \mathbf{r} + \mathbf{r} \cdot \mathbf{r} + \omega^2 \mathbf{q}, \tag{74}$$

where the vector $q$ is given by

$$\mathbf{q} = \begin{pmatrix} q(t_1, \omega) & q(t_2, \omega) & \cdots & q(t_k, \omega) \end{pmatrix}. \tag{75}$$

It can be readily seen that the Fréchet derivative of $F$ at $\mathbf{r}$ is the matrix

$$F'(\mathbf{r}) = \frac{2}{b-a} \mathscr{D}_k + \operatorname{diag}(2\mathbf{r}). \tag{76}$$

We now apply the Newton-Kantorovich theorem to the system (6), with the initial iterate given by the vector

$$\mathbf{r_0} = \begin{pmatrix} r_{\mathrm{LG}}(t_1) & r_{\mathrm{LG}}(t_2) & \cdots & r_{\mathrm{LG}}(t_k) \end{pmatrix}, \tag{77}$$

where $r_{\mathrm{LG}}$ is defined via

$$r_{\mathrm{LG}}(t) = i\omega\sqrt{q(t,\omega)} - \frac{1}{4}\frac{q'(t)}{q(t)}. \tag{78}$$

That is, $r_{\mathrm{LG}}$ is the first order asymptotic approximation of the desired nonoscillatory solution of Riccati's equation considered in Section 3. We use the notation $r_{\mathrm{LG}}$ because (78) is the logarithmic derivative of the Liouville-Green approximate appearing in (4). We could initialize the Newton-Kantorovich iterations with a higher order approximation in lieu of $r_{\mathrm{LG}}$; however, it is impractical to calculate the necessary high-order derivatives of the coefficient $q$ numerically, so this is only viable if symbolic formulas for these derivatives are known. Under our assumptions on $q$, we have the bound

$$\omega\sqrt{q_{\mathrm{cmin}}} \leq \|\mathbf{r_0}\| \leq \omega + \frac{1}{4q_{\min}}, \tag{79}$$

which holds for all $a \leq t \leq b$ and $\omega \geq \omega_0$.

As a next step, we establish a bound on the $l^\infty$ operator norm of the inverse of the Fréchet

14

derivative of $F$ at the starting point $\mathbf{r_0}$. To that end, we observe that

$$
\begin{aligned}
\left\| F'\left(\mathbf{r_0}\right)^{-1} \right\| &= \left\| \left( \frac{2}{b-a} \mathscr{D}_k + \mathrm{diag}\left(2\mathbf{r_0}\right) \right)^{-1} \right\| \\
&= \left\| \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \left( \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \frac{2}{b-a} \mathscr{D}_k + I \right)^{-1} \right\| \\
&= \left\| \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \left(I - S\right)^{-1} \right\| \\
&\leq \left\| \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \right\| \left\| I + S + S^2 + S^3 + \cdots \right\| \\
&\leq \left\| \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \right\| \frac{1}{1 - \|S\|},
\end{aligned}
\tag{80}
$$

where we have let

$$
S = -\frac{2}{b-a} \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \mathscr{D}_k
\tag{81}
$$

and assumed that Neumann series for $(I - S)$ converges. It is easy to see that this assumption is true when $\omega$ is sufficiently large; indeed, (79) gives us

$$
\|S\| \leq \frac{\|\mathscr{D}_k\|}{\omega(b-a)\sqrt{q_{\min}}},
\tag{82}
$$

and the right-hand side of the inequality converges to 0 as $\omega \to \infty$. So we can choose $\omega$ large enough such that

$$
\frac{1}{1 - \|S\|} \leq \frac{1}{2},
\tag{83}
$$

which ensures the convergence of the Neumann series and provides us with a bound that simplifies what follows. Now (79) also implies that

$$
\left\| \mathrm{diag}\left(2\mathbf{r_0}\right)^{-1} \right\| \leq \frac{1}{2\omega\sqrt{q_{\min}}},
\tag{84}
$$

and combining this with (83) yields the desired bound

$$
\left\| F'\left(\mathbf{r_0}\right)^{-1} \right\| \leq \frac{1}{\omega\sqrt{q_{\min}}}
\tag{85}
$$

on the operator norm of the inverse of the Fréchet derivative of $F\left(\mathbf{r_0}\right)$.

We next derive two bounds which will help us to show that fourth and fifth conditions in Theorem 1 are met. The assumptions (2) and (3) on $q$ imply that

$$
\sup_{a \leq t \leq b} \left| r'_{\mathrm{LG}}(t) + \left(r_{\mathrm{LG}}(t)\right)^2 + \omega^2 q(t, \omega) \right| \leq C_{\mathrm{LG}},
\tag{86}
$$

where

$$
C_{\mathrm{LG}} = \frac{5}{16} \frac{1}{q_{\min}^2} + \frac{1}{4} \frac{1}{q_{\min}}
\tag{87}
$$

is independent of $\omega$. In particular,

$$
\left\| F\left(\mathbf{r_0}\right) \right\| \leq C_{\mathrm{LG}}.
\tag{88}
$$

15

Combining this inequality with (85) gives us

$$\left\| F'(\mathbf{r_0})^{-1} F(\mathbf{r_0}) \right\| \leq \frac{C_{\mathrm{LG}}}{\omega \sqrt{q_{\min}}}. \tag{89}$$

Moreover, for any $\eta > 0$ and any $\mathbf{r}$ and $\mathbf{s}$ in $B_\eta(\mathbf{r_0})$, we have

$$\begin{aligned}
\|F(\mathbf{r}) - F(\mathbf{s})\| &= \left\| \frac{2}{b-a} \mathscr{D}_k(\mathbf{r}-\mathbf{s}) + (\mathbf{r}-\mathbf{s})(\mathbf{r}+\mathbf{s}) \right\| \\
&\leq \left( \frac{2}{b-a} \|\mathscr{D}_k\| + 2\|\mathbf{r_0}\| + 2\eta \right) \|\mathbf{r}-\mathbf{s}\| \\
&\leq \left( \frac{2}{b-a} \|\mathscr{D}_k\| + 2\omega + 2\eta + \frac{1}{2q_{\min}} \right) \|\mathbf{r}-\mathbf{s}\|.
\end{aligned} \tag{90}$$

From the combination of (85) and (90), we obtain

$$\left\| F'(\mathbf{r_0})^{-1}(F(\mathbf{r}) - F(\mathbf{s})) \right\| \leq \frac{\frac{2}{b-a}\|\mathscr{D}_k\| + 2\omega + 2\eta + \frac{1}{2q_{\min}}}{\omega\sqrt{q_{\min}}} \|\mathbf{r}-\mathbf{s}\|. \tag{91}$$

Having established the bounds (89) and (91), we are now in a position to choose the constants $\eta$ and $\lambda$ in Theorem 1 and show that all of its conditions are satisfied. We note first that the function $F$ is clearly continuously differentiable everywhere, so we can take $\Omega = \mathbb{R}^k$ and this implies the second condition of the theorem is satisfied regardless of our choice of $\eta$. Earlier, we showed that the Fréchet derivative of $F$ at $\mathbf{r_0}$ is invertible, so the first condition is also satisfied. We now choose the constants to be

$$\eta = \frac{\sqrt{q_{\min}}}{4} \quad \text{and} \quad \lambda = \frac{C_{\mathrm{LG}}}{\omega\sqrt{q_{\min}}}. \tag{92}$$

Since

$$\frac{\frac{2}{b-a}\|\mathscr{D}_k\| + 2\omega + 2\eta + \frac{1}{2q_{\min}}}{\omega\sqrt{q_{\min}}} \to \frac{2}{\sqrt{q_{\min}}} \quad \text{as} \quad \omega \to \infty, \tag{93}$$

it follows that

$$\frac{\frac{2}{b-a}\|\mathscr{D}_k\| + 2\omega + 2\eta + \frac{1}{2q_{\min}}}{\omega\sqrt{q_{\min}}} \leq \frac{1}{\eta} \tag{94}$$

for sufficiently large $\omega$. Combining this with our earlier bound (91) establishes the fifth condition in Theorem 1. For sufficiently large $\omega$, the third condition is met since $\lambda \to 0$ as $\omega \to \infty$ and $\eta$ has a positive limit. Finally, we observe that in light of our choice of $\lambda$, the fourth condition is simply (89).

Having shown that all of the conditions of Theorem 1 are satisfied, it now follows that the sequence of iterates $\{\mathbf{r_n}\}$ defined by (77) and

$$\mathbf{r_{n+1}} = \mathbf{r_n} - F'(\mathbf{r_n})^{-1} F(\mathbf{r_n}) \tag{95}$$

converges to a solution $\mathbf{r}^*$ of the discretized Riccati equation. The theorem also gives us the following bound on the rate of convergence of this sequence:

$$\|\mathbf{r}^* - \mathbf{r_n}\| \leq \frac{\eta}{2^n} \left( 1 - \sqrt{1 - \frac{2\lambda}{\eta}} \right)^{2^n} \leq \frac{\eta}{2^n} \left( \frac{\lambda}{\eta} \right)^{2^n} = \mathcal{O}\left( \frac{1}{\omega^{2^n}} \right). \tag{96}$$

16

Thus we have established that when $\omega$ is large enough, the iterates $\mathbf{r_n}$ converge rapidly to a solution of the *discretized* Riccati equation. Although their limit $\mathbf{r}^*$ corresponds to a polynomial $r^*$ of degree $(k-1)$ whose values at the discretization nodes $t_1, \ldots, t_k$ are given by the values of the vector $\mathbf{r}^*$, it is not immediately obvious that $r^*$ approximates a solution of the continuous Riccati equation (6). However, from the proof in Subsection 3.1, we know that the conditions (2) and (3) imply that there exists a nonoscillatory solution $r_{\mathrm{non}}$ of the Riccati equation which can be well-approximated by a polynomial expansion using a number of terms which is independent of $\omega$, at least for large enough frequencies. This means that if the number of collocation points $k$ suffices and the frequency $\omega$ is large enough, then

$$\mathbf{r}_{\mathrm{non}} = \begin{pmatrix} r_{\mathrm{non}}(t_1) & r_{\mathrm{non}}(t_2) & \cdots & r_{\mathrm{non}}(t_k) \end{pmatrix} \tag{97}$$

closely approximates a solution of the discretized Riccati equation. In other words, in this case, the nonoscillatory solution of the continuous Riccati equation yields an approximate solution of the discretized equation. Moreover, it is clear from (36) that $\mathbf{r}_{\mathrm{non}}$ is in the ball $B_\eta(\mathbf{r_0})$ for large enough values of $\omega$. Since $\mathbf{r}^*$ is the unique solution of the discretized Riccati equation in that ball, $\mathbf{r}^*$ must closely coincide with $\mathbf{r}_{\mathrm{non}}$. In particular, the polynomial $r^*$ represented by the vector $\mathbf{r}^*$ must closely approximate the nonoscillatory solution $r_{\mathrm{non}}$ of the Riccati equation (6).

It is difficult to derive an explicit lower bound on the value of $\omega$ which ensures that the Newton-Kantorovich iterates converge to a vector which represents a nonoscillatory solution of the Riccati equation. It is relatively easy, however, to develop a criterion which works well in practice. The procedure seems to succeed whenever the frequency is high enough to ensure that each linearized equation which arises has a unique solution. The linearization of the discretized Riccati operator around the vector $\mathbf{r}$ is

$$\left( \mathrm{diag}(2\mathbf{r}) + \frac{2}{b-a} \mathscr{D}_k \right) \mathbf{h} = -F(\mathbf{r}), \tag{98}$$

and multiplying both sides of this equation by the inverse of the diagonal matrix yields

$$(I + S_k) h = -\mathrm{diag}(2\mathbf{r})^{-1} F(\mathbf{r}), \tag{99}$$

where we have let

$$S_k = \mathrm{diag}(2\mathbf{r})^{-1} \frac{2}{b-a} \mathscr{D}_k. \tag{100}$$

Because we sample $r_{\mathrm{LG}}$ to initialize the Newton-Kantorovich iterations, $\mathbf{r} \sim i\omega\sqrt{\mathbf{q}}$, and this gives us the approximate bound

$$\|S_k\| \lesssim \frac{\|\mathscr{D}_k\|}{\omega \sqrt{q_{\min}}(b-a)}. \tag{101}$$

We have found (101) to be excellent estimate of the $l^\infty$ operator norm of $S_k$ and when $\|S_k\| < 1$, Theorem 2 ensures that the linearized equations which arise in the course of the applying the Newton-Kantorovich method to Riccati's equation can be rapidly solved via a fixed point iteration.

Accordingly, it is tempting to use

$$\frac{\|\mathscr{D}_k\|}{\omega \sqrt{q_{\min}}(b-a)} < 1 \tag{102}$$

as a criterion for deciding whether the interval $[a, b]$ is in the high-frequency regime or not.

However, the spectral radius $\rho(S_k)$ of $S_k$ is considerably smaller than its $l^\infty$ norm, a property inherited from the spectral differentiation matrix $\mathscr{D}_k$. Since given any $\epsilon > 0$, there exists a matrix norm $\|\cdot\|_\epsilon$ such that

$$\rho(S_k) \leq \|S_k\|_\epsilon \leq \rho(S_k) + \epsilon, \tag{103}$$

it is the spectral radius of $S_k$ which really controls the behavior of fixed point iterations for the linearized equation (99). Instead of (102), we use the criterion

$$\omega\sqrt{q_{\min}}(b-a) > \text{thresh}, \tag{104}$$

where thresh is a threshold parameter which must be adapted to the choice of $k$. Assuming thresh is properly set, when (104) holds, Theorem 2 implies the existence of a unique solution of (99) and the rapid convergence of the sequence $\{\mathbf{h_n}\}$ of fixed point iterates defined via

$$\mathbf{h_0} = 0 \quad \text{and} \quad \mathbf{h}_{n+1} = -\text{diag}\,(2\mathbf{r})^{-1}\frac{2}{b-a}\mathscr{D}_k\,\mathbf{h_n} - \text{diag}\,(2\mathbf{r})^{-1}F\,(\mathbf{r}) \tag{105}$$

to that solution.

## 5. Numerical Algorithm

In this section, we describe our algorithm for the construction of of a slowly-varying trigonometric phase function $\alpha$ such that

$$\left\{\frac{\cos{(\alpha(t))}}{\sqrt{\alpha'(t)}}, \quad \frac{\sin{(\alpha(t))}}{\sqrt{\alpha'(t)}}\right\} \tag{106}$$

in basis in the space of solutions of the oscillatory differential equation (1). The algorithm takes the following as inputs:

1. the endpoints $a$ and $b$ of the solution domain $[a, b]$,

2. an external subroutine for evaluating the coefficient $q$ at arbitrary points in $[a, b]$,

3. an integer parameter $k$ specifying the number of points in the Chebyshev collocation grids used to represent functions,

4. a real-valued parameter $\epsilon$ which controls the precision of the calculations and

5. a real-valued parameter thresh which is used to determine when an interval is in the high-frequency regime.

In all of the experiments of this paper, we let $k = 16$, $\epsilon = 1.0 \times 10^{-12}$ and thresh $= 10$. The parameter thresh might need to be adjusted if the value of $k$ is modified.

Our algorithm outputs a collection of discretization intervals $[a_1, b_1]$, $[a_2, b_2], \ldots, [a_m, b_m]$ such that

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < b_3 = \cdots = a_m < b_m = b, \tag{107}$$

and the values of $\alpha$, $\alpha'$ and $\alpha''$ at the nodes of the $k$-point Chebyshev extremal grid on each of these subintervals. Using this data, the functions $\alpha$, $\alpha'$ and $\alpha''$ can be easily evaluated at any point on the interval $[a, b]$. Moreover, any reasonable initial or boundary value problem for (1) can be solved by computing the appropriate linear combination of the basis functions (106).

We note that, as written, our algorithm will fail if there is no discretization interval in the high-frequency regime. This is a relatively simple problem to overcome, however, because in this event, all phase functions for (1) are slowly varying throughout the solution domain, and a suitable one can be constructed by solving the Riccati equation with essentially arbitrary initial conditions.

The algorithm operates in four stages, each of which is described in detail below. In the first stage, we form an initial set of discretization intervals which suffices to represent the coefficient $q$. In the second, we traverse traverse this initial collection of discretization intervals from left-to-right, solving Riccati's equation over each interval in the high-frequency regime and solving initial value problems for Appell's equation when possible in order to extend the solution of Riccati equations into the low-frequency regime. During the the second stage, the collection of discretization intervals is adaptively refined as necessary in order to represent the phase function accurately. In the third stage, we traverse the discretization intervals from right-to-left, solving terminal value problems for Appell's equation in order to extend the solution. Again, during this stage, the collection of discretization intervals is adaptively refined as needed. In the fourth stage, we integrate the obtained solution of Riccati's equation in order to obtain the desired trigonometric phase function $\alpha$.

In order to make the description of our algorithm simpler, we first define several subprocedures. The most often used of these is the following "goodness of fit" procedure for testing whether or not a function $f$ is well represented by a $(k-1)$-term Chebyshev expansion on an interval $[c, d]$. It consists of the following steps:

1. Form the vector

$$\mathbf{f} = (f(t_1), \ldots, f(t_k)). \tag{108}$$

   of values of the function $f$ at the nodes $t_1, \ldots, t_k$ of the Chebyshev extremal grid on the interval $[c, d]$.

2. Apply the matrix $\mathscr{C}_k$ to the vector $\mathbf{f}$ in order to compute coefficients $a_0, \ldots, a_{k-1}$ in the Chebyshev expansion

$$\sum_{j=0}^{k-1} a_j T_j \left( \frac{2}{d-c} t + \frac{d+c}{d-c} \right) \tag{109}$$

   which agrees with the function $f$ at the nodes $t_1, \ldots, t_k$.

3. If the quantity

$$\frac{\max \left\{ |a_{k-2}|, |a_{k-1}| \right\}}{\max\limits_{j=0,\ldots,k-1} |a_j|} \tag{110}$$

   is less than the input parameter $\epsilon$, we regard the expansion (109) as a good approximation of the function $f$. Otherwise, we regard it as poor approximation.

The next subprocedure we describe is our method for solving Riccati's equation over an interval [c,d] in the high-frequency regime. It is a fairly straightforward application of the Newton-Kantorovich method with the linearized equations solved via the fixed point iteration (105). The only unusual aspect of our implementation is that we always use the second iterate $\mathbf{h_2}$ in the sequence defined by (105) to approximate solutions of the linearized equations. This suffices

because the fixed point scheme converges rapidly and, when applying the Newton-Kantorovich scheme, it is only necessary to compute the solution of the linearized equation defining the $(i+1)^{st}$ iterate with accuracy on the order of $\|\mathbf{r_{i+1}} - \mathbf{r_i}\|$, where $\{\mathbf{r_i}\}$ is the sequence of iterates obtained by solving the linearized equations exactly. Here are the steps of the procedure in detail:

1. Form the vectors

$$\mathbf{q} = \begin{pmatrix} q(\omega, t_1) & q(\omega, t_2) & \cdots & q(\omega, t_k) \end{pmatrix} \tag{111}$$

   and

$$\mathbf{r_0} = \begin{pmatrix} r_{\mathrm{LG}}(t_1) & r_{\mathrm{LG}}(t_2) & \cdots & r_{\mathrm{LG}}(t_k) \end{pmatrix}, \tag{112}$$

   where $t_1, \ldots, t_k$ are the nodes of the Chebyshev extremal grid on the interval $[c, d]$ and $r_{\mathrm{LG}}$ is defined in (78). Also, set the integer $i$, which is the index of the current Newton-Kantorovich iteration, to 0.

2. Compute the residual

$$F(\mathbf{r_i}) = \frac{2}{d-c} \mathscr{D}_k \, \mathbf{r_i} + \mathbf{r_i} \cdot \mathbf{r_i} + \omega^2 \mathbf{q}. \tag{113}$$

3. Approximate the solution of the linearized problem

$$\left( \frac{2}{d-c} \mathscr{D}_k + \mathrm{diag}\left(2\mathbf{r_i}\right) \right) \mathbf{h} = -F(\mathbf{r_i}) \tag{114}$$

   via

$$\mathbf{h} = \mathrm{diag}\left(2\mathbf{r_i}\right)^{-1} \left( \mathrm{diag}\left(2\mathbf{r_i}\right)^{-1} \frac{2}{d-c} \mathscr{D}_k - I \right) F(\mathbf{r_i}). \tag{115}$$

   This is the second iterate $\mathbf{h_2}$ in the fixed point scheme (105).

4. Form the $(i+1)^{st}$ iterate $\mathbf{r_{i+1}} = \mathbf{r_i} + \mathbf{h}$.

5. If $\|\mathbf{h}\| > \epsilon \|\mathbf{r_i}\|$ then continue the Newton-Kantorovich iterations by incrementing $i$ and going to Step 2. If $\|\mathbf{h}\| \le \epsilon \|\mathbf{r_i}\|$, then terminate the iterative procedure and regard $\mathbf{r} := \mathbf{r_{i+1}}$ as the obtained solution of the discretized Riccati equation.

The next subprocedure we describe is an integral equation method for solving an initial value problem for Appell's equation (30) on the interval $[c, d]$. When using Chebyshev spectral techniques to solve initial value problems for differential equations, we prefer integral formulations to differential formulations because the former provide a natural way to incorporate the initial conditions into the problem, while differential formulations require various ad hoc procedures for enforcing the initial conditions. Letting

$$m(t) = m(c) + m'(c)(t-c) + m''(c)\frac{(t-c)^2}{2} + \int_c^t \frac{(t-s)^2}{2} \sigma(s) \, ds \tag{116}$$

20

in (30) yields the integral equation

$$
\begin{aligned}
\sigma(t) + 4\omega^2 q(t,\omega) \int_c^t \int_c^{s_2} \sigma(s_1)\, ds_1\, ds_2 &+ 2\omega^2 q'(t) \int_c^t \int_c^{s_3} \int_c^{s_2} \sigma(s_1)\, ds_1\, ds_2\, ds_3 \\
&= -4\omega^2 q(t,\omega) \left( m'(c) + m''(c)(t-c) \right) \\
&- 2\omega^2 q'(t,\omega) \left( m(c) + m'(c)(t-c) + m''(c)\frac{(t-c)^2}{2} \right).
\end{aligned}
\tag{117}
$$

We discretize and solve it as follows:

1. Form the vector

$$
\mathbf{q} = \begin{pmatrix} q(t_1,\omega) & q(t_2,\omega) & \cdots & q(t_k,\omega) \end{pmatrix},
\tag{118}
$$

   where $t_1,\ldots,t_k$ are the nodes of the Chebyshev extremal grid on the interval $[c,d]$.

2. Apply the spectral differentiation matrix $2/(d-c)\mathscr{D}_k$ to $\mathbf{q}$ to form the vector $\mathbf{qp}$ of values of the derivatives of $q$ at the nodes $t_1,\ldots,t_k$.

3. Form the $k \times k$ coefficient matrix for the discretized version of (117) by letting

$$
A = I_k + \operatorname{diag}\left(4\omega^2 \mathbf{q}\right) \left(\frac{d-c}{2}\mathscr{I}_k\right)^2 + \operatorname{diag}\left(2\omega^2 \mathbf{qp}\right) \left(\frac{d-c}{2}\mathscr{I}_k\right)^3
\tag{119}
$$

   where $I_k$ is the $k \times k$ identity matrix.

4. Form the vector

$$
\begin{aligned}
\mathbf{y} = -\, & 4\omega^2 \left(m'(c)\mathbf{q} + m''(c)\mathbf{q}\circ\mathbf{tc}\right) \\
& - 2\omega^2 \left(m(c)\mathbf{qp} + m'(c)\mathbf{qp}\circ\mathbf{tc} + \frac{1}{2}m''(c)\mathbf{qp}\circ\mathbf{tc}\circ\mathbf{tc}\right),
\end{aligned}
\tag{120}
$$

   where

$$
\mathbf{tc} = \begin{pmatrix} (t_1 - c) & (t_2 - c) & \cdots & (t_k - c) \end{pmatrix}
\tag{121}
$$

   and $\mathbf{v}\circ\mathbf{w}$ represents pointwise product of the vectors $\mathbf{v}$ and $\mathbf{w}$.

5. Use a standard method to solve the system of linear equations $A\boldsymbol{\sigma} = \mathbf{y}$.

6. Compute vectors $\mathbf{m}$ and $\mathbf{mp}$ representing the solution of Appell's equation and its derivative via the formulas

$$
\begin{aligned}
\mathbf{m} &= m(c)\mathbf{1} + m'(c)\mathbf{tc} + \frac{1}{2}m''(c)\mathbf{tc}\circ\mathbf{tc} + \left(\frac{d-c}{2}\mathscr{I}_k\right)^3 \boldsymbol{\sigma} \quad \text{and} \\
\mathbf{mp} &= m'(c)\mathbf{1} + m''(c)\mathbf{tc} + \left(\frac{d-c}{2}\mathscr{I}_k\right)^2 \boldsymbol{\sigma},
\end{aligned}
\tag{122}
$$

   where $\mathbf{1}$ is the vector whose entries are all 1's and $\mathbf{tc}$ is as before.

We use a completely analogous procedure to solve terminal boundary value problems for Appell's equation. Because of the great similarities between that procedure and the one just described, we omit a detailed description of it.

We are now in a position to describe each stage of our algorithm.

*5.1. Stage one: adaptive discretization of the coefficient q*

In this stage, we construct an initial list of intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_m, b_m]$ which suffice to discretize the coefficient $q$. In addition to this list of discretization intervals, a list of intervals to process is maintained. In the first instance, the list of discretization intervals is empty and the list of intervals to process contains only $[a, b]$. This stage proceeds by executing the following steps until the list of intervals to process is empty:

1. Remove an interval $[c, d]$ form the list of intervals to process.

2. Check the accuracy with which the $q$ is represented via a $k$-term Chebyshev expansion over $[c, d]$ using the goodness of fit procedure. If it is deemed to be well represented, put $[c, d]$ into the list of discretization intervals. Otherwise, put the intervals $[c, (c + d)/2]$ and $[(c + d)/2, d]$ into the list of intervals to process.

*5.2. Stage two: solution of the Riccati equation and left-to-right sweep*

In this stage, we traverse the list $[a_1, b_1]$, ..., $[a_m, b_m]$ of discretization intervals formed in the previous stage from left-to-right. We solve Riccati's equation on any interval in the high-frequency regime, and solve an initial value problem for Appell's equation in order to extend the phase function whenever possible. During this stage, we adaptively refine the list of discretization intervals as needed to ensure the phase function is accurately represented. To this end, we maintain two lists, one containing the intervals which need to be processed and the other specifying the new collection of discretization intervals created during this stage. The list of intervals to process is initialized with all of the discretization intervals constructed in the preceding phase, and the following sequence of steps is executed until the list of intervals to process is empty:

1. Remove from the list of intervals to process the left-most interval $[c, d]$.

2. Form the vector $\mathbf{q}$ of the values of the coefficient at the nodes $t_1, \ldots, t_k$ of the Chebyshev extremal grid on $[c, d]$ and let $q_{\min}$ denote the minimum value of $q$ at those nodes.

3. Compute the quantity $\gamma = \omega \sqrt{q_{\min}}(b - a)$.

4. If $\gamma > $ thresh, execute the following sequence of steps:

    (a) Compute a solution $\mathbf{r}$ of the Riccati equation using the subprocedure described above.

    (b) Let $\boldsymbol{ap}$ be the real-valued vector whose entries are the imaginary parts of the entries of $\mathbf{r}$, and let $\boldsymbol{app}$ be the pointwise produce of the vector $-2\boldsymbol{ap}$ and the real part of $\mathbf{r}$. According to Formula (26), these are the computed values of the derivative $\alpha'$ and second derivative $\alpha''$ of the desired trigonometric phase function $\alpha$.

    (c) Perform the goodness of fit procedure on $\boldsymbol{ap}$. If it is well-approximated by $(k - 1)$-term Chebyshev expansion, then add the interval $[c, d]$ to the list of output intervals. Otherwise, add the intervals $[c, (c + d)/2]$ and $[(c + d)/2, d]$ to the list of intervals to process. In either case, goto Step 1 of the procedure of this stage.

5. If $\gamma \geq $ thresh and the functions $\alpha'$ and $\alpha''$ have already been computed over the interval immediately to the left of $[c, d]$, then we solve an initial value problem for Appell's equation to construct the vectors $\mathbf{ap}$ and $\mathbf{app}$ of values of $\alpha'$ and $\alpha''$ on $[c, d]$ using the following sequence of steps:

(a) Let *apval* and *appval* denote the values of $\alpha'(c)$ and $\alpha''(c)$ (these are already known since $c$ is the right-hand endpoint of the interval to the left of $[c, d]$) and let *qval* be the value of the coefficient $q$ at the point $c$. Use Kummer's equation (7) to form an estimate *apppval* of the value of $\alpha'''(c)$ via the formula

$$apppval = \frac{4\, qval \times apval^2 - 4apval^4 + 3\, appval^2}{2\, apval}.$$  (123)

(b) Calculate the initial values of the solution $m$ of Appell's equation such that $1/m$ extends the derivative $\alpha'$ of the trigonometric phase function from the preceding interval to the current one. More explicitly, let

$$m(c) = \frac{1}{apval}, \quad m'(c) = -\frac{appval}{apval^2} \quad \text{and} \quad m''(c) = 2\frac{appval^2}{apval^3} - \frac{apppval}{apval^2}.$$  (124)

(c) Solve the initial value problem for Appell's equation using the procedure described above. This results in vectors **m** and **mp** which give the values of the solution of Appell's equation $m(t)$ and its derivative $m'(t)$ at the Chebyshev nodes $t_1, \ldots, t_k$.

(d) Compute the vectors **ap** and **app**, whose entries give the values of $\alpha'$ and $\alpha''$ at the Chebyshev nodes $t_1, \ldots, t_k$ on $[c, d]$, using vectors **m** and **mp**. To be more explicit, **ap** is the vector whose entries are the reciprocals of those of **m**, and

$$\mathbf{app} = -\mathbf{ap} \circ \mathbf{ap} \circ \mathbf{mp}.$$  (125)

(e) Perform the goodness of fit procedure on **ap**. If it is well-approximated by $(k - 1)$-term Chebyshev expansion, then add the interval $[c, d]$ to the list of output intervals. Otherwise, add the intervals $[c, (c + d)/2]$ and $[(c + d)/2, d]$ to the list of intervals to process. In either case, goto Step 1 of the procedure of this stage.

*5.3. Stage three: right-to-left sweep*

At the conclusion of the second state of the procedure, we have a refined list $[a_1, b_1], \ldots, [a_m, b_m]$ of discretization intervals and the vectors **ap** and **app** have been constructed over all intervals which lie to the right of an interval in the high-frequency regime. In this stage, we sweep from right-to-left, solving terminal value problems for Appell's equation over any interval for which **ap** and **app** have not yet been constructed.

As before, we maintain two lists, one containing the intervals which need to be processed and the other specifying the new collection of discretization intervals created during this stage. The list of intervals to process is initialized with all of the discretization intervals constructed in the preceding phase, and the following sequence of steps is executed until the list of intervals to process is empty:

1. Remove from the list of discretization intervals to process the right-most interval $[c, d]$ for which the vectors **ap** and **app** have not yet been constructed.

2. Solve a terminal value problem for Appell's equation in order to construct the vectors **ap** and **app** We omit the details because the procedure is entirely analogous to that used in

the second stage of the procedure.

3. Perform the goodness of fit procedure on **ap**. If it is well-approximated by $(k-1)$-term Chebyshev expansion, then add the interval $[c, d]$ to the list of output intervals. Otherwise, add the intervals $[c, (c+d)/2]$ and $[(c+d)/2, d]$ to the list of intervals to process. In either case, goto Step 1 of the procedure of this stage.

### 5.4. Stage four: spectral integration

At the conclusion of the third stage, we have the final list $[a_1, b_1], \ldots, [a_m, b_m]$ of discretization intervals and the vectors **ap** and **app** have been constructed over each interval. In this final stage, we use spectral integration to compute the values of the desired nonoscillatory trigonometric phase function $\alpha$ at the Chebyshev nodes of each discretization interval. More explicitly, we initial set $aval = 0$ and then, for each $j = 1, \ldots, m$ we execute the following steps:

1. Compute the vector **a** giving the values of $\alpha$ at the Chebyshev nodes on $[a_j, b_j]$ via the formula

$$\mathbf{a} = aval + \frac{b_j - a_j}{2} \mathscr{I}_k \mathbf{ap} \tag{126}$$

2. We let aval be equal to the last entry of the vector **a**.

## 6. Numerical Experiments

In this section, we present the results of numerical experiments which were conducted to illustrate the properties of the method of this paper and to compare it with other solvers for oscillatory differential equations. We implemented our algorithm in Fortran and compiled our code with version 14.2.1 of the GNU Fortran compiler. All experiments were performed on a desktop computer equipped with an AMD 9950X processor and 64GB of RAM. This processor has 16 cores, but only one core was utilized in our experiments.

In all of our experiments, the value of the parameter $k$, which determines the order of the Chebyshev expansions used to represent phase functions, was taken to be 16, the parameter $\epsilon$ that controls the accuracy of the obtained phase functions was set to $10^{-12}$ and the parameter thres was set to 10.

In almost all of our experiments, we tested the accuracy of our method and those we compare it to by using it to calculate solutions to initial and boundary value problems for second order equations. The experiment of Section 6.3, in which the accuracy of the phase functions produced by our algorithm was measured directly, is the sole exception. Because the condition numbers of initial and boundary value problems for equations of the form (1) grow with $\omega$, the accuracy of any numerical method used to solve them is expected to deteriorate with increasing frequency. In the case of our algorithm, the phase functions themselves are calculated to high precision, but the magnitudes of the phase functions increase with frequency and accuracy is lost when the phase functions are exponentiated. One implication is that calculations which involve only the phase functions and not the solutions of the scalar equation can be performed to high accuracy. The article [6], for example, describes a phase function method for rapidly computing the zeros of solutions of second order linear ordinary differential equations to extremely high accuracy.

To account for the vagaries of modern computing environments, all reported times were obtained by averaging the cost of each calculation over 100 runs.

## 6.1. Comparison with the Modified Magnus expansion method

In this first experiment, we compared the performance of the modified Magnus method of [18], which is one of the most widely used methods for solving oscillatory differential equations, with the algorithm of this paper. The former is a step method that combines an exponential integrator with preconditioning via the solutions of a constant coefficient equation obtained by freezing the coefficient matrix of the oscillatory equation. We used a fourth order exponential integrator and equispaced step sizes in our implementation of the modified Magnus method, which was written in Fortran and closely follows the description provided in [18] and [19].

For each $n = 2^6, 2^7, 2^8, \ldots, 2^{20}$, we used both methods to evaluate the solution

$$L_n(x) = P_n(x) + i\frac{2}{\pi}Q_n(x) \tag{127}$$

of Legendre's differential equation

$$(1 - t^2)y''(t) - 2ty'(t) + n(n+1)y(t) = 0 \tag{128}$$

at a collection of points on the interval $[0, 0.9]$. Because Equation (127) is not in the normal form (1), we actually applied the algorithm of this paper to the differential equation

$$y''(t) + \left(\frac{1}{(1-t^2)^2} + \frac{n(n+1)}{1-t^2}\right)y(t) = 0, \tag{129}$$

which has $L_n(t)\sqrt{1-t^2}$ as a solution. We choose to evaluate the Legendre function $L_n$ because its logarithmic derivative is nonoscillatory. We discuss this further in Section 6.3, and it can be seen from the graph of the accuracy predicted by its condition number of evaluation in Figure 1.

For each value of $n$ considered, we executed the modified Magnus method twice, once with the step size $h$ taken to be a constant multiple of $n^{-1/2}$ and again with the step size taken to be the same constant multiple of $n^{-3/4}$. The constant was set so as to ensure $10^{-10}$ relative accuracy for the smallest value of $n$ considered. We struggled to obtain higher accuracy with the modified Magnus method at high frequencies because it becomes numerically unstable when the step size is extremely small. In the case of our algorithm, for each value of $n$ considered, we evaluated $L_n$ at a collection of 1,000 equispaced points on the interval $[0, 0.9]$, including 0 and 0.9, and recorded the largest relative error which was observed. For the modified Magnus method, we recorded the largest relative error which was observed at the steps taken by the solver. Figure 1 gives the results. The plot on the left shows the time taken by each method as a function of $n$, while that on the right gives the largest observed relative errors, again as a function of $n$. The plot on the right also shows the relative accuracy predicted by the condition number of evaluation of the function $L_n$. More explicitly, it displays a graph of

$$\kappa(n) = \max_{j=1,\ldots,1000} \epsilon_0 \left| \frac{t_j L_n'(t_j)}{L_n(t_j)} \right|, \tag{130}$$

where $t_1, \ldots, t_{1000}$ are the equispaced nodes on $[0, 0.9]$ at which we evaluated $L_n$ using the algorithm of this paper and $\epsilon_0 \approx 2.220446049250313 \times 10^{-16}$ is machine zero.

From Figure 1, we see that the modified Magnus method maintains roughly constant accuracy when the number of steps scales as $n^{-3/4}$, but loses accuracy when the step size scales as $n^{-1/2}$.

This is consistent with the theoretical estimates presented in [18]. We note that the method of this paper is several orders of magnitude faster than the modified Magnus method, even at low frequencies, and it is more than 4 orders of magnitude faster at the highest frequency considered.
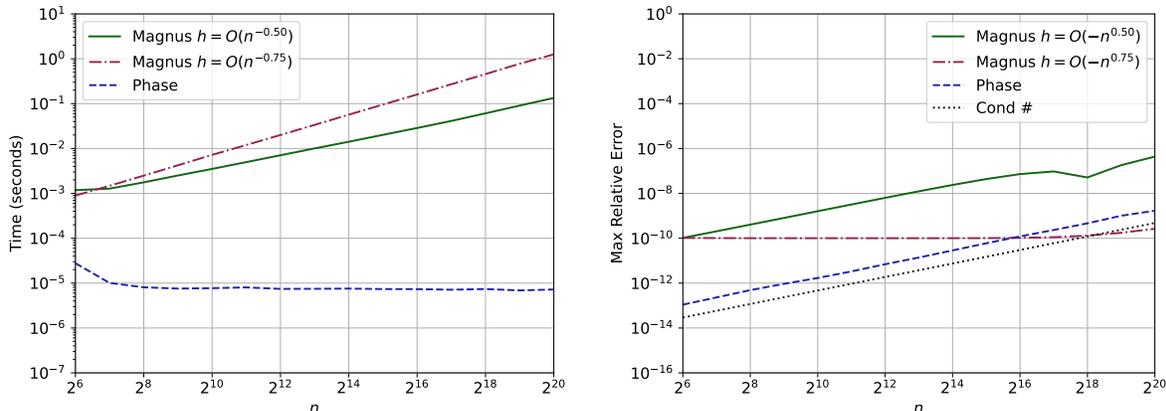


Figure 1: The results of the experiment of Section 6.1 in which the modified Magnus method of [18] and the algorithm of this paper were used to evaluate the Legendre function $L_n(t)$ defined in (127) on the interval $[0, 0.9]$. The plot on the left gives the time in seconds required by each method as a function of the degree $n$ of the Legendre function, while the plot on the right gives the maximum relative error observed in the course of evaluating the Legendre function using each method. The plot on the right also shows the accuracy predicted by the condition number of evaluation of the Legendre function. We give the results for the modified Magnus method when the step size $h$ was taken to be a constant multiple of $n^{-1/2}$, and when it was taken to be a constant multiple of $n^{-3/4}$.

### 6.2. Comparison with two frequency-independent solvers

In this experiment, we compared the algorithm of this paper with the ARDC method of [1] and the frequency-independent solver [7] previously developed by one of this paper's authors by once again evaluating the solution $L_n$ of Legendre's differential equation (128). More explicitly, for each $n = 2^6, 2^7, 2^8, \ldots, 2^{20}$, we used each method to compute the Legendre function $L_n$ defined in (127) at 100 equispaced nodes on $[0, 0.999]$, including the points 0 and 0.999. We evaluated $L_n$ at a relatively small number of nodes since the method of [1] is a step scheme which only returns the values of the solution at a sparse set of nodes sampled well below the Nyquist frequency. This means that the solution returned by the ARDC method cannot be accurately interpolated at arbitrary points on the solution domain. In order to evaluate it at the 100 equispaced nodes we considered, we had to specify them as inputs to ARDC and doing so increased the cost of the algorithm. Adding more points would increase the cost further, and we regard 100 points as sufficient to measure the accuracy of the solution. Both the algorithm of this paper and that of [7] allow for the solution to be evaluated at any point on the solution domain.

Figure 2 gives the results of this experiment. The plot on the left gives the time required by each method as a function of $n$, while that on the right shows the maximum relative accuracy achieved by each method as a function of $n$, as well as the relative accuracy predicted by the condition number of evaluation of $L_n$. We note that the three methods achieve very similar levels of accuracy for large values of $n$, while the the ARDC method loses some accuracy at small values of $n$. ARDC is noticeably slower than [7], and it is several orders of magnitude slower than the method of this paper.
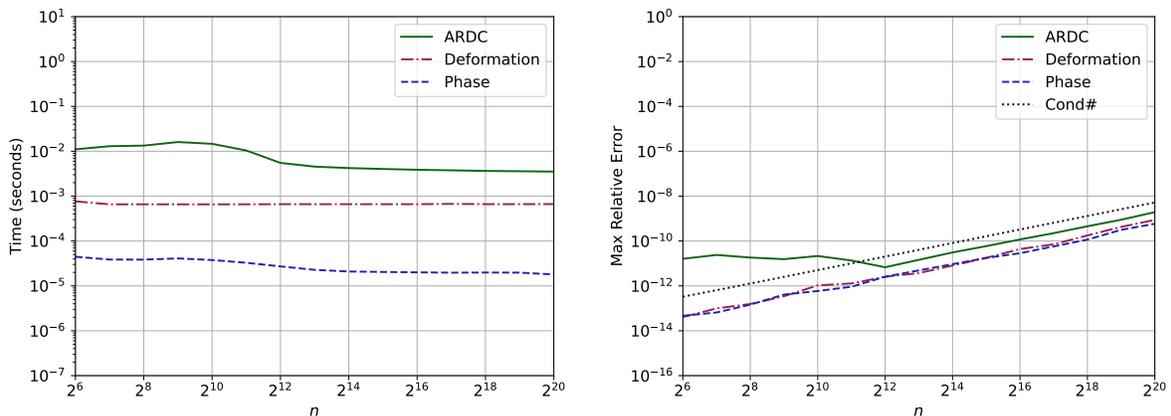
26

Figure 2: The results of the experiment of Subsection 6.2 in which the method of this paper, the smooth deformation scheme of [7] and the ARDC method of [1] were used to evaluate the Legendre function $L_n$ defined in (127) on the interval $[0, 0.999]$. The plot on the left gives the time required by each algorithm as a function of the degree $n$ of the Legendre function, while the plot on the right shows the maximum relative error observed when $L_n$ was evaluated at 100 equispaced points on the interval $[0, .999]$ using each approach. The plot on the right also gives the maximum relative error predicted by the condition number of evaluation of this function.

### 6.3. Accuracy of the computed phase functions

In this experiment, we measured the accuracy with which our method computes phase functions. For each $n = 2^7, 2^8, \ldots, 2^{21}$, we used our algorithm to construct the derivative of a nonoscillatory trigonometric phase function for the normal form (129) of Legendre's differential equation on the interval $\left[0, 1 - 10^{-7}\right]$.

There is an explicit expression for the desired solution of Kummer's equation in this case. Indeed, it can be shown using the formula

$$(P_n(t))^2 + \frac{4}{\pi^2} (Q_n(t))^2 = \frac{4}{\pi^2} \int_1^\infty Q_n\left(t^2 + (1 - t^2)s\right) \frac{ds}{\sqrt{s^2 - 1}}, \quad -1 < t < 1, \tag{131}$$

which appears in [13], that the function on the left-hand side of (131) is absolutely monotone on $(0, 1)$. Since the functions

$$\sqrt{\frac{\pi}{2}} P_n(t) \sqrt{1 - t^2}, \quad \frac{2}{\pi} Q_n(t) \sqrt{1 - t^2} \tag{132}$$

are a pair of solutions of (129) whose Wronskian is 1 (see, for instance, Chapter 3 of [5]), the function

$$\alpha'(t) = \frac{1}{(1 - t^2)\left(\frac{\pi}{2} P_n^2(t) + \frac{2}{\pi} Q_n^2(t)\right)}. \tag{133}$$

is the derivative of the desired nonoscillatory trigonometric phase function for the normal form of Legendre's differential equation.

For each value of $n$ considered in this experiment, we measured the accuracy of the obtained solution of Kummer's equation by comparing it with the reference solution (133) at at 1,000 equispaced points on the interval $\left[0, 1 - 10^{-7}\right]$, including the points 0 and $1 - 10^{-7}$. Figure 3 gives the largest observed relative error in the computed value of the phase function as a function of $n$, as well as the time required to construct the phase function. From these results, we see

that the phase function is always computed with relative accuracy which is greater than the requested precision $1.0 \times 10^{-12}$.
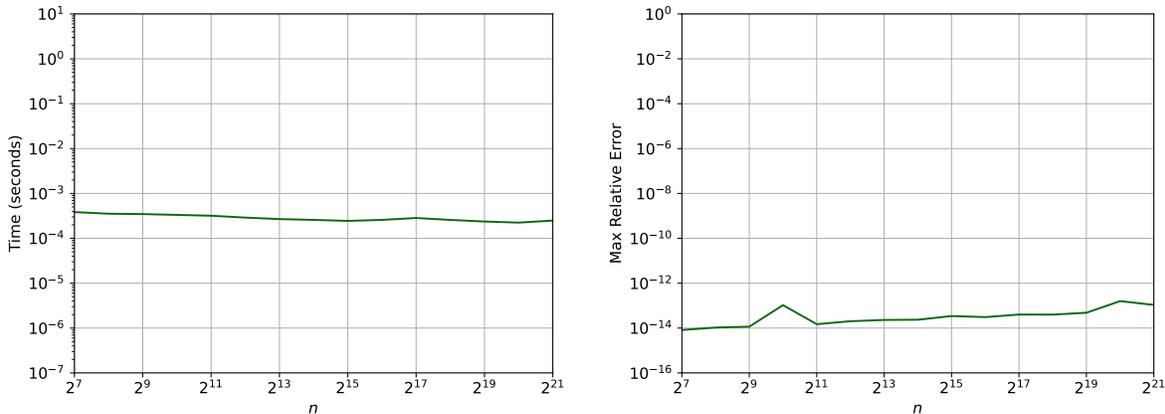


Figure 3: The results of the experiment of Subsection 6.3 in which the accuracy of the phase functions computed by our algorithm was measured. The plot on the left gives the time taken by our algorithm as a function of the frequency parameter $n$. The plot on the right gives the maximum relative error observed while evaluating the solution of Kummer's equation calculated by our algorithm at 1,000 points in the interval $\left[0, 1 - 10^{-7}\right]$, including the points 0 and $1 - 10^{-7}$.

### 6.4. Evaluation of Gegenbauer polynomials

In the experiment of this section, we used the scheme of this paper to evaluate the Gegenbauer polynomial $C_n^\alpha(t)$, which is the unique solution of the differential equation

$$(1 - t^2)y''(t) - 2(\alpha + 1)y(t) + n(n + 2\alpha)y(t) = 0 \tag{134}$$

that is continuous on $[-1, 1]$ and such that

$$C_n^\alpha(1) = \frac{\Gamma(2\alpha + 1)}{\Gamma(2\alpha)\Gamma(n + 1)}. \tag{135}$$

For each $n = 2^6, 2^7, 2^8, \ldots, 2^{20}$ and $\alpha = -0.499, 0.25, 1.00$, we computed a phase function for the normal form

$$y''(t) + \left( \frac{\alpha - \alpha^2 + \frac{3}{4}}{(1 - t^2)^2} + \frac{\left(n + \alpha - \frac{1}{2}\right)\left(n + \alpha + \frac{1}{2}\right)}{1 - t^2} \right) y(t) = 0 \tag{136}$$

of (134) satisfied by $C_n^\alpha(t)\left(1 - t^2\right)^{(2\alpha+1)/4}$ via the algorithm of this paper and used it to evaluate $C_n^\alpha(x)$ at 1,000 equispaced points on the interval $(0, 0.999)$. We compared the obtained values to those computed using the well-known three-term recurrence relation satisfied by the Gegenbauer polynomials. The results are shown in Figure 4. There, we give the largest observed absolute error as a function of $n$, as well as the time required to construct each phase function. We measured absolute error rather than relative error in these experiments because $C_n^\alpha$ has zeros on the interval $(0, 1)$.
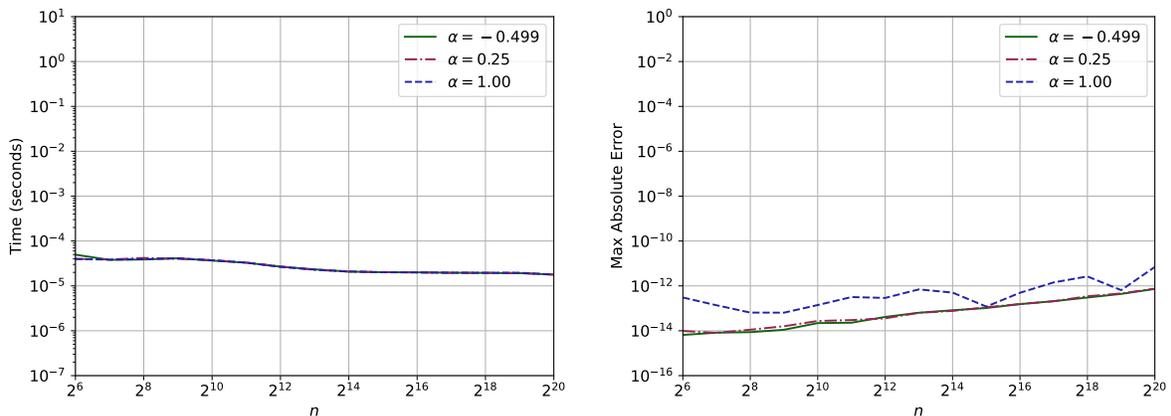
Figure 4: The results of the experiment of Subsection 6.4 in which our algorithm was used to evaluate Gegenbauer polynomials of orders $\alpha = -0.499, 0.25, 1.0$. The plot on the left gives the time required to construct each phase function. The plot on the right gives the maximum absolute accuracy observed while evaluating the Gegenbauer polynomials at 1,000 equispaced points on the interval $[0, 0.999]$.

*6.5. A boundary value problem*

In this final experiment, for each $\omega = 2^6, 2^7, \ldots, 2^{20}$, we constructed a phase function for the differential equation

$$y''(t) + \omega^2 \left( \frac{3t^2\omega^2 + t^2\omega + 1}{-(t^2+1)\,\omega + \omega^2 + 1} + \frac{2e^{-t}}{t^2 + \frac{1}{10}} \right) y''(t) = 0, \quad -1 < t < 1, \tag{137}$$

and used it to evaluate the solution $y$ which satisfied the boundary conditions $y(-1) = y(1) = 1$ at 1,000 equispaced points on $[-1, 1]$. We then applied a standard adaptive Chebyshev spectral method to these boundary value problems in order to construct reference solutions. Figure 5 gives the results. The time required by our algorithm and by the standard solver are plotted as functions of the frequency $\omega$ on the left, while the maximum observed absolute error is plotted as a function of $\omega$ on the right.

## 7. Conclusion

We have introduced a novel method for solving oscillatory second order linear ordinary differential equations. The running time of our scheme is independent of frequency, and it is considerably faster than previous methods with this property. In the high-frequency regime, it applies a remarkably simple approach, namely, discretizing the Riccati equation via a standard Chebyshev spectral method and using the Newton-Kantorovich algorithm to invert the resulting linear system. Moreover, the results of Section 3 and (4) rigorously establish the validity of this method.

The machinery necessary to handle the low-frequency regime is somewhat more complicated, and this is a weakness of the algorithm of this paper which the authors hope to address shortly. In most cases of interest, the low-frequency regime is close to a turning point of the differential equation. Although phase functions can be used to represent solutions of second order linear ordinary differential equations near a turning point, their derivatives behave like steep errors functions there, which makes them expensive to approximate via polynomial expansions. The
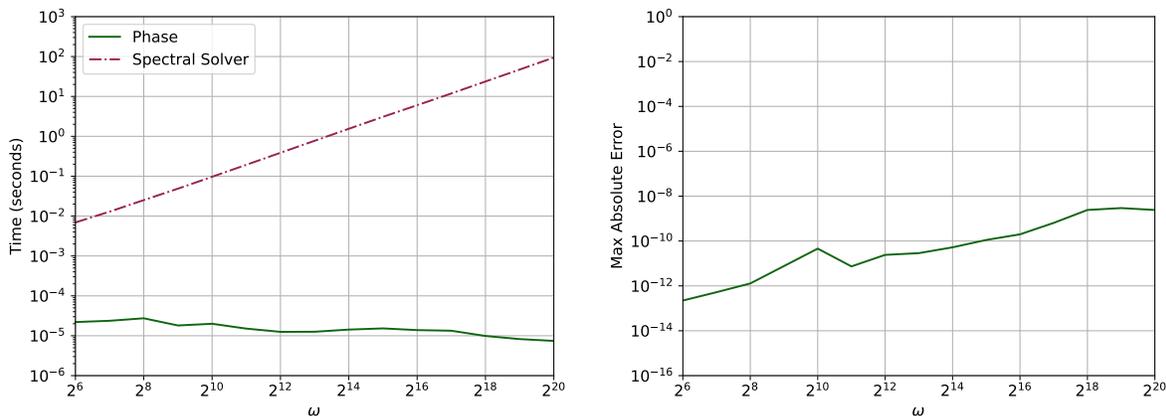
29

Figure 5: The results of the experiment of Subsection 6.5 in which our algorithm was applied to boundary value problem for the differential equation (137) and the accuracy of the resulting solution was tested using a standard adaptive Chebyshev spectral solver. The plot on the left gives the time required to solve the problem using each method as a function of the frequency $\omega$, while the plot on the right gives the maximum absolute error observed while comparing the solution obtained using our method with the reference solution computed via the standard solver.

authors are currently developing an algorithm based on an alternate approach to representing solutions of second order linear ordinary differential equations near turning points. This should allow for the further acceleration of the method discussed here, as well as making it simpler and more robust.

We also observe that the bulk of the calculations performed by the algorithm of this paper can be parallelized. There are two problems that must be overcome to do so, but both admit obvious solutions. First, the solution of Appell's equation is carried out sequentially, after the Riccati equation has been solved over each high-frequency interval. Rather than doing this, though, we could construct a basis in the space of solutions of Appell's equation for each low-frequency interval and construct the desired phase function from these basis functions after either Appell's equation or the Riccati equation has been solved in each interval. This would allow the bulk of the computations to be parallelized. However, another problem remains, namely, we adaptively discretize the phase function as we compute it and such calculations are difficult to parallelize. It would behoove us to have a set of discretization intervals determined in advance, before solving the Riccati equation or Appell's equation. Fortunately, there is a relatively straightforward mechanism for determining a suitable set of discretization intervals *a priori*, namely, we can adaptively discretize the asymptotic approximate $r_{\mathrm{LG}}$. The authors have found that this provides a collection of discretization intervals which suffices to represent the phase function in almost all cases of interest.

Finally, we note that because many special functions satisfy second order equations, there are numerous applications of this work to special functions. For instance, it should allow for the extremely rapid and high-accuracy computation of classical Gaussian quadrature rules. The authors are also interested in using the techniques discussed here to solve Sturm-Liouville problems, and to rapidly apply the associated eigentransforms.

## 8. Acknowledgments

## References

[1] Agocs, F. J., and Barnett, A. H. An adaptive spectral method for oscillatory second-order linear odes with frequency-independent cost. *SIAM Journal on Numerical Analysis 62*, 1 (2024), 295–321.

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

[3] Appell, P. Sur la transformation des équations différentielles linéaires. *Comptes Rendus 91* (1880), 211–214.

[4] Arnold, A., Abdallah, N. B., and Negulescu, C. WKB-based schemes for the oscillatory 1D Schrödinger equation in the semiclassical limit. *SIAM Journal on Numerical Analysis 49*, 4 (2011), 1436–1460.

[5] Bateman, H., and Erdélyi, A. *Higher Transcendental Functions*, vol. I. McGraw-Hill, New York, New York, 1953.

[6] Bremer, J. On the numerical calculation of the roots of special functions satisfying second order ordinary differential equations. *SIAM Journal on Scientific Computing 39* (2017), A55–A82.

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

[8] Bremer, J. Phase function methods for second order linear ordinary differential equations with turning points. *Applied and Computational Harmonic Analysis 65* (2023), 137–169.

[9] Bremer, J., Chen, Z., and Yang, H. Rapid application of the spherical harmonic transform via interpolative decomposition butterfly factorization. *SIAM Journal on Scientific Computing 43*, 6 (2021), A3789–A3808.

[10] Ciarlet, P. *Linear and Nonlinear Functional Analysis with Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[11] Davidson, R., and Hong, Q. *Physics of intense charged particle beams in high energy accelerators*. World Scientific, Singapore, 2001.

[12] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.1.0 of 2020-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[13] Durand, L. Nicholson-type integrals for products of Gegenbauer functions and related topics. In *Theory and Application of Special Functions*, R. A. Askey, Ed. Academic Press, 1975, pp. 353–374.

[14] EINAUDI, F., AND HINES, C. WKB approximation in application to acoustic-gravity waves. *Canadian Journal of Physics 48* (02 2011), 1458–1471.

[15] HAZELTINE, R. D., AND MEISS, J. D. *Plasma confinement.* Courier Corporation, North Chelmsford, Massachusett, 2003.

[16] HEITMAN, Z., BREMER, J., AND ROKHLIN, V. On the existence of nonoscillatory phase functions for second order ordinary differential equations in the high-frequency regime. *Journal of Computational Physics 290* (2015), 1–27.

[17] HORN, R. A., AND JOHNSON, C. R. *Matrix Analysis.* Cambridge University Press, Cambridge, England, 1990.

[18] ISERLES, A. On the global error of discretization methods for highly-oscillatory ordinary differential equations. *BIT Numerical Mathematics 32* (2002), 561–599.

[19] ISERLES, A. Think globally, act locally: solving highly-oscillatory ordinary differential equations. *Applied Numerical Mathematics 43* (2002), 145–160.

[20] KANTOROVICH, L. Functional analysis and applied mathematics. *Uspehi Matematiceskii Nauk 3* (1948), 89–185.

[21] KŘNER, J., ARNOLD, A., AND DÖPFNER, K. WKB-based scheme with adaptive step size control for the Schrödinger equation in the highly oscillatory regime. *Journal of Computational and Applied Mathematics 404* (2022), 113905.

[22] KUMMER, E. De generali quadam aequatione differentiali tertti ordinis. *Progr. Evang. Köngil. Stadtgymnasium Liegnitz* (1834).

[23] LORENZ, K., JAHNKE, T., AND LUBICH, C. Adiabatic integrators for highly oscillatory second-order linear differential equations with time-varying eigendecomposition. *BIT Numerical Mathematics* (2005), 91–115.

[24] MARTIN, J., AND SCHWARZ, D. WKB approximation for inflationary cosmological perturbations. *Physical Review D 67* (10 2002).

[25] OLVER, F. W. *Asymptotics and Special Functions.* A.K. Peters, Wellesley, Massachusetts, 1997.

[26] OLVER, S. GMRES for the differentiation operator. *SIAM Journal on Numerical Analysis 47*, 5 (2009), 3359–3373.

[27] OLVER, S. GMRES for oscillatory matrix-valued differential equations. In *Spectral and High Order Methods for Partial Differential Equations* (Berlin, Heidelberg, 2011), J. S. Hesthaven and E. M. Rønquist, Eds., Springer Berlin Heidelberg, pp. 267–274.

[28] PRITULA, G. M., PETRENKO, E. V., AND USATENKO, O. V. Adiabatic dynamics of one-dimensional classical Hamiltonian dissipative systems. *Physics Letters, Section A: General, Atomic and Solid State Physics 382*, 8 (feb 2018), 548–553.

[29] SPIGLER, R. Asymptotic-numerical approximations for highly oscillatory second-order differential equations by the phase function method. *Journal of Mathematical Analysis and Applications 463* (2018), 318–344.

[30] Spigler, R., and Vianello, M. A numerical method for evaluating the zeros of solutions of second-order linear differential equations. *Mathematics of Computation 55* (1990), 591–612.

[31] Spigler, R., and Vianello, M. The phase function method to solve second-order asymptotically polynomial differential equations. *Numerische Mathematik 121* (2012), 565–586.

[32] Trefethen, L. Is Gauss quadrature better than Clenshaw–Curtis? *SIAM Review 50* (2008), 67–87.