An International Journal
computers & mathematics
with applications

PERGAMON Computers and Mathematics with Applications 41 (2001) 1293–1315
www.elsevier.nl/locate/camwa

Rational Chebyshev Spectral Methods for Unbounded Solutions on an Infinite Interval Using Polynomial-Growth Special Basis Functions

J. P. Boyd
Department of Atmospheric, Oceanic, and Space Science
and Laboratory for Scientific Computation
University of Michigan, 2455 Hayward Avenue
Ann Arbor, MI 48109, U.S.A.
jpboyd@engin.umich.edu
http://www.engin.umich.edu:/~jpboyd/

(Received August 1999; revised and accepted September 2000)

Abstract—In the method of matched asymptotic expansions, one is often forced to compute solutions which grow as a polynomial in y as $|y| \to \infty$. Similarly, the integral or repeated integral of a bounded function $f(y)$ is generally unbounded also. The k^{th} integral of a function $f(y)$ solves $\frac{d^k u}{dy^k} = f(y)$. We describe a two-part algorithm for solving linear differential equations on $y \in [-\infty, \infty]$ where $u(y)$ grows as a polynomial as $|y| \to \infty$. First, perform an explicit, analytic transformation to a new unknown v so that v is bounded. Second, expand v as a rational Chebyshev series and apply a pseudospectral or Galerkin discretization. (For our examples, it is convenient to perform a preliminary step of splitting the problem into uncoupled equations for the parts of u which are symmetric and antisymmetric with respect to $y = 0$, but although this is very helpful when applicable, it is not necessary.) For the integral and iterated integrals and for constant coefficient differential equations in general, the Galerkin matrices are banded with very low bandwidth. We derive an improvement on the "last coefficient error estimate" of the author's book which applies to series with a subgeometric rate of convergence, as is normally true of rational Chebyshev expansions. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords—Rational Chebyshev functions, Spectral method, Quadrature, Unbounded domain, Matched asymptotic expansions.

1. INTRODUCTION

Physics and engineering problems are often solved on an unbounded domain, but the solution is almost always required to be bounded at infinity. In the perturbation method known as "matched asymptotic expansions" [1,2], the domain is divided into two (or more) subdomains and different (approximate) differential equations are solved on each subdomain. In the limit

This work was supported by the National Science Foundation through Grant OCE9521133.
I thank my former students, G. Chen and N. Flyer, whose thesis work using matched asymptotic expansions required inner solutions with polynomial growth.
I thank the reviewer for helpful comments.

that the perturbation parameter $\epsilon \to 0$, the width of the "inner" domain is unbounded. The "inner" subproblem must therefore be solved on a semi-infinite or infinite interval in the "inner" coordinate even though the inner approximation will be used, for finite ϵ, only on part of the physical domain. It follows that the usual physical constraint of boundedness at infinity in the coordinate x does not apply to the inner problem.

Our own interest in unbounded solutions to differential equations arose from applying matched asymptotics to nonlinear waves of the rotationally-modified Korteweg-deVries (RMKdV) equation [3]. The j^{th}-order inner approximation grows as a polynomial of degree $(j-1)$ in the inner coordinate y as $|y| \to \infty$. However, the waves are not unbounded, but rather oscillate sinusoidally as the physical coordinate tends to infinity [4]. The growing-in-y term in the $O(\epsilon^2)$ inner approximation matches smoothly to the inner limit of the outer approximation, $\epsilon \sin(\epsilon y) \approx \epsilon^2 y$, $y \to 0$. Similarly, at all higher orders, polynomial growth in y matches to sinusoidal oscillations in the outer region also.

Neither finite differences nor spectral methods are at all happy with unbounded solutions. Our central theme is that linear problems can be easily transformed into a new differential equation with a new unknown $v(y)$ such that v is bounded. The key strategy is to subtract a set of smooth but unbounded functions, $\phi_j(y)$, from the original unknown u. The unbounded growth of $u(y)$ is completely captured by a weighted sum of the ϕ_j. One can then apply any infinite interval numerical method to compute the bounded, modified unknown $v(y)$.

A convenient spectral basis is the set of rational Chebyshev functions, $TB_j(y)$ [5–8]. These are images of a Fourier cosine basis under a change-of-coordinate. It is not necessary that the transformed solution decay to zero as $|y| \to \infty$. The usual "spectral accuracy", that is, an error that decreases exponentially fast with the number of basis functions N, is obtained even if $v(y)$ merely asymptotes to a constant.

The general problem attacked here is a linear, inhomogeneous boundary value problem on an unbounded domain of the form

$$\mathcal{L}u = f(y), \tag{1}$$

where $u(y)$ and $f(y)$ asymptote to polynomials in y as $|y| \to \infty$ and the differential operator is of the form

$$\mathcal{L}u \equiv \sum_{k=0}^{\nu} b_k(y) \frac{d^k}{dy^k}. \tag{2}$$

We shall only explicitly discuss first- and second-order examples, but the methods apply to differential equations of any order.

An important special case is the integral or iterated integral of a bounded function $f(y)$ on an unbounded interval,

$$\begin{aligned} I^{(1)} &\equiv \int_0^y f(z)\,dz, \\ I^{(2)} &\equiv \int_0^y dz \int_0^z f(w)\,dw, \end{aligned} \tag{3}$$

and so on. The k^{th} such integral is equivalent to the differential equation $\frac{d^k u}{dy^k} = f(y)$, with

$$\frac{d^{k-1}u}{dy^{k-1}(0)} = \frac{d^{k-2}u}{dy^{k-2}(0)} = \cdots = u(0) = 0.$$

Because of its simplicity, we shall use this to give concrete illustrations of more general and abstract methods that are needed for variable coefficient differential equations.

Both the analytic subtractions and the rational Chebyshev spectral method apply to nonlinear differential equations also. However, because the asymptotic behavior of the solution $u(y)$ must be analyzed on a case-by-case basis for nonlinear problems, we shall not discuss them explicitly in our article.

2. SUBTRACTIONS

The crucial step in taming an unbounded $u(y)$ is to transform it. The original unknown is assumed to be unbounded through a polynomial as $|y| \to \infty$, i.e.,

$$u(y) \sim \begin{cases} \sum_{j=1}^{M} \alpha_j^{(+)} y^j, & y \to \infty, \\ \sum_{j=1}^{M} \alpha_j^{(-)} y^j, & y \to -\infty, \end{cases} \tag{4}$$

where in $\alpha_j^{(+)} \neq \alpha_j^{(-)}$ in general. The transformation replaces the unbounded function $u(y)$ with the new unknown v defined by

$$u = v + \sum_{j=1}^{2M} \Omega_j \phi_j(y), \tag{5}$$

where the ϕ_j will be defined below. The inhomogeneous term in the differential equation is replaced by

$$g(y) = f(y) - \sum_{j=1}^{2M} \Omega_j \mathcal{L}\phi_j(y). \tag{6}$$

The transformed differential equation is

$$\mathcal{L}v = g. \tag{7}$$

The challenge is to choose the "subtraction functions" $\phi_j(y)$ and weights Ω_j so that the transformed problem has a bounded solution.

There are many possible choices for basis functions, but our preference is the following. Note that because we shall split the differential equation into two subproblems whose solutions are of definite parity with respect to $y = 0$, we shall define two basis sets accordingly.

$$\phi_j(y) \equiv \begin{cases} y^j, & j = \text{even integer}, \\ y^j \operatorname{erf}(y), & j = \text{odd integer}, \end{cases} \qquad [\text{Symmetric}], \tag{8}$$

$$\phi_j(y) \equiv \begin{cases} y^j, & j = \text{ odd integer}, \\ y^j \operatorname{erf}(y), & j = \text{ even integer}, \end{cases} \qquad [\text{Antisymmetric}]. \tag{9}$$

Why the error function? A good subtraction function must be explicitly integrable, analytically simple, and preferably introduce no poles or branch points which are not already present in the original integrand. (Singularities of the subtracted function, even if off the real axis, could degrade the rate of convergence of the rational Chebyshev series.) Powers of y are the best, but they are not enough because the even powers of y are always symmetric. The error function factor allows us to also mimic symmetric functions that asymptote to odd powers of $|y|$ as $y \to \infty$, and to similarly imitate antisymmetric functions whose magnitude grows proportionally to even powers of $|y|$. The error function can be generalized to include a scale factor, i.e., $\operatorname{erf}(\lambda y)$, where λ is a constant chosen to match the scale of the error function to the length scale of $f(y)$. For simplicity, we set $\lambda = 1$ in what follows.

Assume

$$f(y) \sim \begin{cases} \sum_{j=1}^{M'} \beta_j^{(+)} y^j, & y \to \infty, \\ \sum_{j=1}^{M'} \beta_j^{(-)} y^j, & y \to -\infty, \end{cases} \tag{10}$$

where $\beta_j^{(+)} \neq \beta_j^{(-)}$. Similarly, assume that the coefficients of the differential equation $b_k(y)$ asymptote to constants or polynomials as $|y| \to \infty$. The constants $\alpha_j^{(\pm)}$ in u can then be found, merely by matching powers of y, in terms of the known constants $\beta_j^{(\pm)}$ and the asymptotic expansion coefficients of the $b_k(y)$. This matching is a small linear algebra problem.

If, for example, the coefficients of the differential equation asymptote to constants where those for b_0 and b_ν are nonzero where ν is the order of the differential equation, then the number of terms in the large-$|y|$ polynomial for $u(y)$ matches the number of terms in the similar representation of $f(y)$, that is, $M' = M$. This is the simplest case because the coefficients of the polynomial growth in u are completely determined by the large-$|y|$ behavior of the inhomogeneous term f.

The iterated integral is more complicated. For the second integral, for example, b_2 asymptotes to one, while all the other b_k are identically zero. The degree of the asymptotic polynomial part of $u(y)$ is then larger by two than the degree of the asymptotic polynomial in $f(y)$, that is, $M = M' + 2$. Because of this discrepancy, it is necessary to employ subtraction functions which asymptote to a constant and zero, as well as additional unbounded functions if the integrand $f(y)$ is unbounded. For this reason, we shall give this important special case a separate, detailed treatment in later sections. However, the general principle is not changed: transform the problem by subtracting known functions $\phi_j(y)$ with calculable weights Ω_j from the unbounded unknown $u(y)$ to obtain a new unknown $v(y)$ which can be computed by standard infinite interval algorithms.

3. PARITY DECOMPOSITION: THE FIRST STEP

For many problems, the subtraction step and the computation of the transformed unknown $v(y)$ are simplified by first splitting the problem into two uncoupled subproblems of definite parity. A function $f(y)$ is said to be "symmetric with respect to the origin" or to have "even parity" if $f(y) = f(-y)$ for all y. Similarly, an "antisymmetric" or "odd parity" function has the property $f(y) = -f(-y)$ for all y. An arbitrary function can always be split into its symmetric part u_S and antisymmetric part u_A through the following:

$$u(y) = u_S(y) + u_A(y), \qquad u_S = \frac{1}{2}\left(u(y) + u(-y)\right), \qquad u_A = \frac{1}{2}\left(u(y) - u(-y)\right). \tag{11}$$

Differentiation is a parity-reversing operation: the first derivative of a symmetric function is antisymmetric, but its second derivative is symmetric and so on. The rational Chebyshev functions have the property that even degree functions (TB_{2j}) are symmetric, whereas the odd degree functions are of odd parity.

The book [8] describes parity in more detail, but the salient point is that a differential equation can *always* be split into symmetric and antisymmetric problems. In the general case, these problems are *coupled*, and then the parity decomposition is not useful. For integrals and iterated integrals and also for the inner problems of the RMKdV equation, the two subproblems are *uncoupled*. The parity decomposition then reduces the computational cost as shown below.

4. RATIONAL CHEBYSHEV SPECTRAL METHODS

The transformed differential equation for the bounded unknown $v(y)$ still must be solved. One can combine the strategies of earlier sections with finite difference, finite element, or spectral algorithms for solving the transformed differential equation. In this article, we chose to use spectral methods with a basis set of the "rational Chebyshev" functions.

There are two reasons for this choice. First, the spectral method yields an error that decreases exponentially fast with the size N of the truncated basis set (after transformation so that the expansion is applied only to bounded functions) [8,9]. Second, no special procedures are needed for solutions that asymptote to a constant, rather than to zero, for large $|y|$.

The rational Chebyshev functions are defined on the interval $y \in [-\infty, \infty]$ by

$$TB_n(y) \equiv \cos(nt), \tag{12}$$

where the coordinates are related via

$$y = L\cot(t), \qquad t = \operatorname{arccot}\left(\frac{y}{L}\right). \tag{13}$$

The constant L is a user-choosable map parameter; strategies for optimizing L are given in [6–8], but the most fundamental idea is to choose L to be roughly equal to the length scale of the desired solution. The actual basis functions are given by (for $L = 1$),

$$\begin{gathered} TB_0(y) \equiv 1, \qquad TB_1(y) = \frac{y}{(y^2+1)^{1/2}}, \qquad TB_2(y) = \frac{(y^2-1)}{(y^2+1)}, \\ TB_3 = \frac{y\,(y^2-3)}{(y^2+1)^{3/2}}, \qquad TB_4(y) = \frac{(y^4-6y^2+1)}{(y^2+1)^2}, \end{gathered} \tag{14}$$

and so on; the functions for general L are obtained by replacing y by y/L in the formulas above. The odd degree Chebyshev functions are not rational functions because of the square root in the denominator, but in a minor abuse of terminology, we shall apply the label "rational" to all members of the basis anyway. A Matlab function for computing these basis functions and their derivatives is given in [10, p. 147].

A differential equation can be solved in two different ways. Either way, a truncated series of rational Chebyshev functions is substituted into the differential equation to define the "residual" function R,

$$v(y) \approx v_N(y) \equiv \sum_{j=0}^{N} a_j TB_j(y; L), \tag{15}$$

$$R(y; a_0, a_1, \dots, a_N) \equiv \mathcal{L}v_N - g(y). \tag{16}$$

The coefficients $a_0, \dots, a_N$ are determined by solving a matrix problem which results from imposing $(N+1)$ constraints that minimize the residual function. (If v_N were the exact solution, the residual function would be identically zero.) The two algorithms differ only in the form of the smallness-of-residual constraints.

In the Galerkin method, the constraints are that the first $(N+1)$ coefficients of the spectral series for R are zero. This is equivalent, after expressing the coefficient integrals in terms of the trigonometric coordinate, to

$$\int_0^{\pi} R(y[t])\cos(jt)\,dt = 0, \qquad j = 0, 1, \dots, N. \tag{17}$$

In the pseudospectral method, the constraints are that the residual is zero at each of N points which are evenly spaced in t:

$$R(L\cot(t_i)) = 0, \quad t_i = \frac{\pi(2i-1)}{(2N+2)}, \qquad i = 1, 2, \dots, (N+1). \tag{18}$$

To exploit parity, the basis is restricted to *even* degree basis functions, $TB_{2j}(y)$, for *symmetric* solutions $v(y)$, and to *odd* degree to compute solutions that are antisymmetric in y. The Galerkin constraints are similarly restricted to products of only $\cos(2jt)$ or $\cos([2j-1]t)$ with the residual, respectively. The pseudospectral collocation points are restricted to $t_i < \pi/2$.

For general differential equations, both Galerkin and pseudospectral discretization matrices are dense. The pseudospectral method is preferable because it is simpler to program.

For constant coefficient differential equations, which includes computing the integral or iterated integral of a function $f(y)$, Galerkin's method is preferable. The reason is that it gives a *banded* matrix. The Galerkin first-derivative matrix, for either a basis of definite parity, is a tridiagonal matrix. The second derivative Galerkin-TB matrices are similarly pentadiagonal with five nonzero elements in each row. These banded matrices can be factored and solved in $O(N)$ operations.

This is much cheaper than the $O(N^3)$ cost of the LU factorization of a dense matrix. Unfortunately, the Galerkin matrices are banded only when the differential equation has constant coefficients or other special cases.

5. FUNCTIONS WITH ALGEBRAIC DECAY AT INFINITY

If the transformed unknown $g(y)$ decays *exponentially* fast to either zero or a constant as $|y| \to$, as illustrated by examples such as $g = \exp(-y^2)$ and $\tanh(y)$, respectively, then the TB_j basis described above is *always sufficient*. However, a function like

$$g(y) \equiv \frac{1}{(1+y^2)^{3/2}} \leftrightarrow g(\cot(t)) = \sin^3(t) \tag{19}$$

is equivalent under the mapping $y = \cot(t)$ to the cube of the *sine* function. However, $TB_j \equiv \cos(jt)$, so expanding this particular $g(y)$ as a series of TB_j is equivalent to approximating $\sin^3(t)$ by a cosine series, which converges very slowly. A far better strategy is to write

$$g(y) = \sum_{j=1}^{\infty} b_j \sin(jt) = \sum_{j=1}^{\infty} b_j SB_{j-1}(y; L=1), \tag{20}$$

where the new basis functions are defined by $SB_{j-1}(y; L) \equiv \sin(jt)$, where $t = \operatorname{acot}(y/L)$.

This difficulty arises because a cosine series is always symmetric about $t = 0$, that is, its sum $\tilde{g}(t)$ always has the property that $\tilde{g}(t) = \tilde{g}(-t)$ for all t. It follows that the cosine approximation to $g(t)$ is not to the function itself, but rather to

$$\tilde{g}(t) \equiv \begin{cases} \sin^3(t), & t \in [0, \pi], \\ -\sin^3(t), & t \in [-\pi, 0]. \end{cases} \tag{21}$$

If a function $g(y)$ decays algebraically fast to its limit, then its symmetrization $\tilde{g}$ may have discontinuities.

This difficulty does not arise for functions that decay exponentially fast as $|y| \to \infty$ because then $g(y[t])$ is a function whose derivatives to *all orders* are zero at $t = 0$ and $t = \pi$. Such a function can be extended across $t = 0$ in either a symmetric or antisymmetric way without inducing discontinuities of derivatives of any order in the extended function. Consequently, it is sufficient to use *either* a cosine series or a sine series to represent such a function on $t \in [0, \pi]$. The sum of the sine series will be the negative of the sum of the cosine series for $t \in [-\pi, 0]$, but the two expansions will agree for $t \in [0, \pi]$, the interval that is the image of the entire real y-axis under the map $y = L\cot(t)$.

It is possible to contrive examples where a general Fourier series in t also fails to converge rapidly. An illustration is

$$g(y) = \frac{1}{(1+y^2)^{5/4}} \leftrightarrow g(y[t]) = \sin^{5/2}(t). \tag{22}$$

Because the sine is raised to a fractional power, $g(y[t])$ has a square root singularity at $t = 0$ and its Fourier series in t converges poorly. Its rational Chebyshev expansion in y, which has the same coefficients, must converge poorly also. For this case, the only remedy is make a preliminary change of the coordinate y to a new variable z, chosen so that $g(y[z(t)])$ will have a rapidly convergent Fourier series.

In summary, the change of coordinate $y = L\cot(t)$ will yield an efficient spectral method if and only if $g(y[t])$ has a rapidly convergent Fourier series in t. The marvel of the mapping is that a problem on an unbounded domain is converted to a problem of ordinary Fourier analysis on a finite interval.

For functions that decay exponentially to their limits as $|y| \to \infty$, and furthermore, have no singularities on the real y-axis (except at infinity), the rapid convergence of the TB_j series is guaranteed. For functions that decay as *powers* of y or as some other nonexponential, algebraic functions, then a more careful analysis is needed. For some problems, adding the Fourier sine terms is sufficient to remove all difficulties. This is the case for the problem of integrating the rational Chebyshev functions TB_j, where the SB_j functions play an essential role, as explained in Appendix A.

6. THE MAP PARAMETER L

There is no simple way to choose the map parameter L. The first tactic is to choose L to equal the dominant length scale of the solution. The second is to apply the simple formulas given in [6], assuming that one has some information about

(i) the rate of asymptotic decay of the solution with $|y|$, and

(ii) the singularities of the solution in the complex y-plane.

Since this information is usually unavailable, the third strategy—a little experimentation with different L for a single moderate value of N—is usually the fastest and most effective way to optimize L.

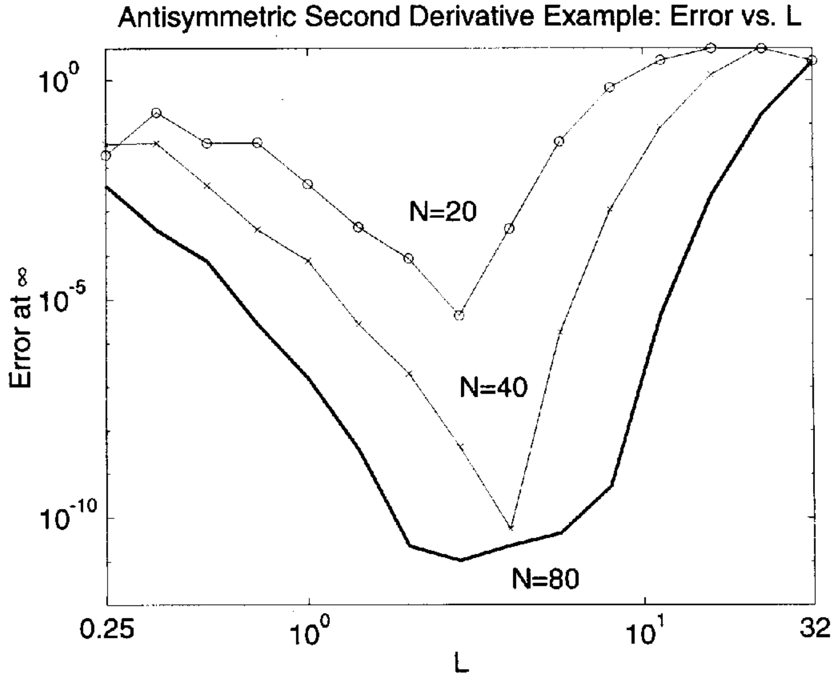

Figure 1. A graph of the maximum absolute error versus L for three different N for the antisymmetric second derivative example discussed later. For a given L, the best L is that which minimizes the error. In agreement with the theory of [6], the optimum L increases slowly with N.

Typically, the curve of error versus L for fixed N has a "V-shape" (Figure 1) for reasons explained in more detail in [6,7]. In brief, off-the-real-axis poles and branch points of u are best resolved by *small* L; the exponential decay of $u(y)$ towards its asymptotic value is best resolved by *large* L. The total error is thus the sum of two independent contributions which vary oppositely with L. Note that $L = 1$, which was used for the two numerical examples, is decidedly *not* optimum for any N for the example of Figure 1. Nevertheless, one can obtain very accurate solutions as illustrated in Figures 3 and 4.

7. NUMERICAL EXAMPLE FROM MATCHED ASYMPTOTIC EXPANSIONS

As noted earlier, the theory of matched asymptotic expansions for the rotation-modified Korteweg-deVries (RMKdV) equation requires solving a sequence of inner problems on an infinite domain which are identical in form to those studied here [3,4,11]. The unapproximated "linear" RMKdV problem [11] is

$$w_{4y} - w_{yy} - \epsilon^2 u = -18\,\mathrm{sech}^4\left(\frac{y}{2}\right) + \frac{45}{2}\,\mathrm{sech}^6\left(\frac{y}{2}\right). \tag{23}$$

From this, perturbation theory gives the second-order inner problem as, after two formal integrations in y,

$$u_{yy} - u = 12\log\left(\cosh\left(\frac{y}{2}\right)\right), \tag{24}$$

This has the exact solution, symmetrical with respect to $y = 0$,

$$\begin{aligned} u = {} & \cosh(y)\left\{6y\mathrm{th} + 3\mathrm{sh}^2 - 12\log\left(\cosh\left(\frac{y}{2}\right)\right) - 6\log(2) + \left(1 + \mathrm{th}^2\right)\right\} \\ & + 6y\mathrm{th} - 12\log\left(\cosh\left(\frac{y}{2}\right)\right) + 3\mathrm{sh}^2 - 6\log(2)\left(1 + \mathrm{th}^2\right), \end{aligned} \tag{25}$$

where $\mathrm{th} \equiv \tanh(y/2)$ and $\mathrm{sh} \equiv \mathrm{sech}(y/2)$. To avoid cancellation errors that are bigger than the (very tiny!) errors of the rational Chebyshev series, we computed the exact solution for $|y| > 7$ through

$$u \sim -6y + 12\log(2) - (6y + 9)\exp(-y) - 2\exp(-2y) + \frac{1}{2}\exp(-3y) - \frac{1}{5}\exp(-4y), \tag{26}$$

which shows explicitly that the solution grows linearly with y for large y. The solution and forcing are both *symmetric* with respect to $y = 0$. The only initial or boundary conditions are those of no exponential growth as $y \to \infty$, which are automatically and implicitly satisfied by every member of the TB basis set.

Only one special basis function is needed because u grows only linearly:

$$\phi_1 \equiv y\,\mathrm{erf}(y). \tag{27}$$

The transformed problem is

$$v = u + \sigma\phi_1, \tag{28}$$

where σ is chosen so that the transformed problem

$$v_{yy} - v = g(y) \tag{29}$$

has an inhomogeneous term $g(y)$, which is bounded at infinity, and where

$$g(y) \equiv 12\log\left(\cosh\left(\frac{y}{2}\right)\right) + \sigma\left\{\phi_{1,yy} - \phi_1\right\}. \tag{30}$$

The asymptotic relation

$$12\log\left(\cosh\left(\frac{y}{2}\right)\right) \sim 6y, \qquad y \gg 1 \tag{31}$$

implies that $\sigma = 6$, and therefore,

$$g(y) = 12\left(\cosh\left(\frac{y}{2}\right)\right) + 6\left\{\frac{4}{\sqrt{\pi}}\left(1-y^2\right)\exp\left(-y^2\right) - y\,\mathrm{erf}\,(y)\right\}. \tag{32}$$

Because this equation is *constant coefficient*, it is efficient to solve it by Galerkin's method. Because the inhomogeneous term in the differential equation decays *exponentially* fast to its limits, a rational Chebyshev series in the TB_j functions is sufficient. Because the solution is symmetric with respect to $y = 0$, the basis can be restricted to functions of *even degree*, that is, TB_{2j}. The nonzero elements of the Galerkin matrix are

$$\begin{aligned} G_{jj} &\equiv -1 - \frac{1}{L^2}\frac{3}{8}(2j-2)^2, \qquad j = 1,2,\dots N, \\ G_{j,j\pm1} &\equiv \frac{1}{L^2}\left\{\frac{1}{4}(2j-2)^2 \pm \frac{3}{4}(2j-2) + \frac{1}{2}\right\}, \\ G_{j,j\pm2} &\equiv \frac{1}{L^2}\left\{-\frac{1}{16}(2j-2)^2 \mp \frac{3}{8}(2j-2) - \frac{1}{2}\right\}. \end{aligned} \tag{33}$$

The elements of the inhomogeneous term in the matrix equation can be calculated using the quadrature approximation

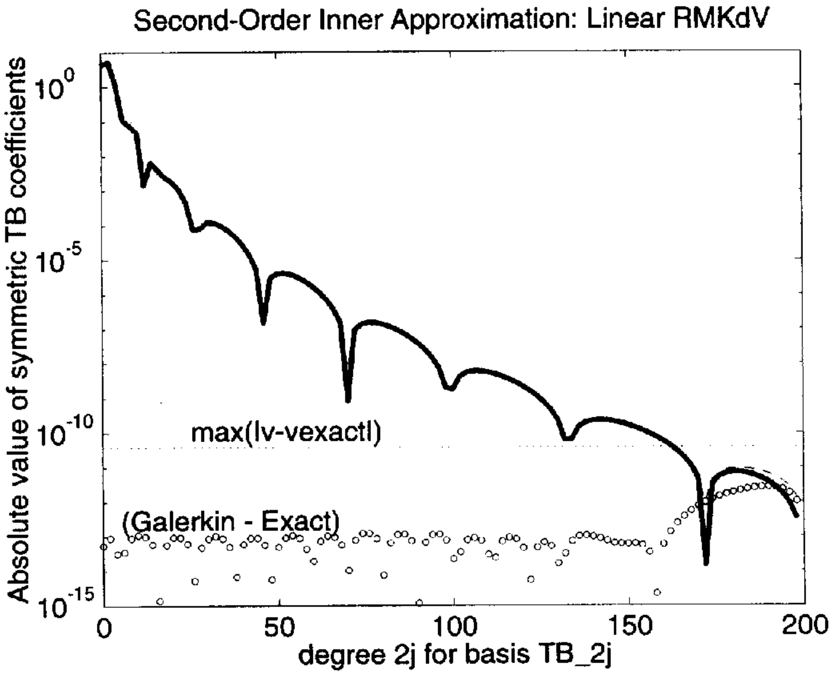

Figure 2. Thick solid line: numerically-computed TB coefficients for the solution of $v_{yy} - v = 12\log(\cosh(y/2)) + 6\{(4/\sqrt{\pi})(1-y^2)\exp(-y^2) - y\,\mathrm{erf}\,(y)\}$, which is transformed from the second-order inner approximation for the linear RMKdV equation. The dashed line, almost hidden by the thick curve, shows the exact TB coefficients. The circles are the difference between the exact and Galerkin spectral coefficients. The horizontal dotted line marks the maximum pointwise error, $\max_{y\in[-\infty,\infty]}|v_{\mathrm{Galerkin}}(y) - v(y)|$, which is 3.8E−11. The limit of the envelope is roughly $\mathcal{E}(198) \approx 2.E - 12$. The map parameter $L = 2$; the basis was $TB_0(y), TB_2(y), \dots, TB_{198}(y)$.

$$\begin{aligned} g_j &\equiv \frac{4}{\pi}\int_0^{\pi/2} g(L\cot(t))\cos((2j-2)t)\,dt \\ &\approx \frac{1}{c_{j-1}}\frac{2}{N}\sum_{k=1}^{N} g\left(L\cot\left\{\frac{\pi(2k-1)}{(4N)}\right\}\right)\cos\left((2j-2)\left\{\frac{\pi(2k-1)}{(4N)}\right\}\right), \end{aligned} \tag{34}$$

where $c_0 = 2$, $c_k = 1$, $\forall\, k > 0$.

Figure 2 shows the result of solving the tridiagonal matrix equation

$$\vec{\vec{G}}\begin{vmatrix} a_0 \\ a_2 \\ a_4 \\ \dots \\ a_{2j} \end{vmatrix} = \begin{vmatrix} g_0 \\ g_2 \\ g_4 \\ \dots \\ g_{2j} \end{vmatrix}. \tag{35}$$

8. UNBOUNDED INTEGRALS ON AN INFINITE DOMAIN

An integral such as

$$I^{(1)}(y) \equiv \int_0^y f(z)\,dz, \qquad y \in [-\infty, \infty] \tag{36}$$

can be evaluated in two ways if the integrand $f(y)$ is bounded. First, one can expand $f(y)$ as a series of rational Chebyshev functions and integrate term-by-term. Second, one can solve the differential equation

$$I_y^{(1)} = f(y), \qquad I^{(1)}(0) = 0. \tag{37}$$

To implement the first strategy, compute the coefficient f_j of TB_j in the Chebyshev series of the integrand. This coefficient is given by the usual inner product integral which is most conveniently evaluated by converting to the trigonometric coordinate

$$f_j \equiv \frac{c_j}{\pi}\int_0^{\pi} f(L\cot(t))\cos(jt)\,dt, \tag{38}$$

where $c_0 = 2, c_k = 1$ otherwise. One can then integrate the series term-by-term.

It is straightforward to integrate *individual* rational Chebyshev functions. For example, for $L = 1$,

$$\int_0^y TB_0\,dz = y, \qquad \int_0^y TB_1(z)\,dz = \sqrt{1+y^2} - 1, \tag{39}$$

$$\int_0^y TB_2(z)\,dz = y - 2\arctan(y), \qquad \int_0^y TB_3(z)\,dz = \frac{y^2+5}{\sqrt{y^2+1}} - 5, \tag{40}$$

$$\int_0^y TB_4(z)\,dz = \frac{y^3+5y}{y^2+1} - 4\arctan(y). \tag{41}$$

Unfortunately, these formulas become increasingly complicated as the degree j increases. Appendix A gives a simple recurrence to compute these integrals to arbitrarily high order.

The alternative differential equation strategy requires subtractions because $u(y)$ is usually unbounded even if $f(y)$ is finite for all real y. However, if $f(y)$ itself is unbounded, or if the integral is iterated, then subtractions are needed even for the term-by-term approach. We shall therefore leave the integrals of individual rational Chebyshev functions to the appendix. In the next few sections, we concentrate on integrating or iteratively-integrating a function $f(y)$ by solving the differential equation.

9. UNBOUNDED INTEGRALS: THEORY

The evaluation of an integral on an unbounded domain is more difficult than the general differential equation discussed earlier. The problem can be stated either as the integral or as the equivalent first-order differential equation

$$u(y) \equiv \int_0^y f(z)\,dz \leftrightarrow u_y = f(y), \qquad u(0) = 0. \tag{42}$$

The difficulty is that $f(y) \equiv 1$ implies $u(y) = y$, and therefore, the integral can be unbounded as $|y| \to \infty$ even when the integrand $f(y)$ is bounded. If f is bounded, then the following theorem shows that it is sufficient to subtract one symmetric and one antisymmetric function from f so that the transformed integrand g asymptotes to zero as $y \to \infty$. The functions $\phi_0(y) \equiv 1$ (for the symmetric part of f) and $\phi_1 \equiv y\,\mathrm{erf}\,(y)$ (for the antisymmetric part) are good choices.

Theorem 1. Boundedness of Integral. *Let $g(y)$ be a function, nonsingular for all finite real y, which satisfies either of the two (equivalent) conditions.*

1. *g asymptotes to zero as $|y| \to \infty$ with sufficient rapidity that the following bounds apply for some constants A, B:*

$$|g(y)| < \frac{A}{(B+y^2)}, \qquad \forall y, \tag{43}$$

or

2. $$|h(t)| \le D, \qquad \forall t \in [0,\pi], \qquad h(t) \equiv \frac{g(L\cot(t))}{\sin^2(t)}, \tag{44}$$

for some constant D where $t(y) \equiv \mathrm{acot}\,(y/L)$.

Then

$$\int_0^y g(z)\,dz < C, \qquad \forall y, \tag{45}$$

for some constant C. This is equivalent to the statement that the solution u to

$$v_y = g(y), \qquad v(0) = 0 \tag{46}$$

is bounded as $|y| \to \infty$.

Proof. When the inequalities apply, the integrand is bounded by a rational function for all y. Then

$$\int_0^y g(z)\,dz < \frac{A}{\sqrt{B}} \arctan\left(\frac{y}{\sqrt{B}}\right) < C \equiv \frac{\pi}{2}\frac{A}{\sqrt{B}}, \tag{47}$$

by explicit integration of this bounding integral. The necessity of the second condition follows by rewriting the integral definition of $v(y)$ in terms of the trigonometric coordinate t as

$$v \equiv L \int_{\mathrm{acot}\,(y/L)}^{\pi/2} g(L\cot(t)) \frac{1}{\sin^2(t)}\,dt. \tag{48}$$

Unless the integrand is bounded, v itself will be unbounded. ∎

The transformed differential equation is given in the summary below.

10. UNBOUNDED INTEGRALS: NUMERICAL TECHNOLOGY

To solve $v_y = g$, it is convenient to rewrite it in the equivalent form in the trigonometric coordinate t as

$$v_t = -L\frac{g(L\cot(t))}{\sin^2(t)} = -Lh(t), \tag{49}$$

where $h \equiv g(L\cot(t))/\sin^2(t)$, the same as defined in equation (44) in Theorem 1. The theorem demands that $h(t)$ be bounded for the boundedness of v. Assume $h(t)$ has a convergent expansion as a general Fourier series,

$$h(t) = a_0 + \sum_{j=1}^{\infty} a_j \cos(jt) + \sum_{j=1}^{\infty} b_j \sin(jt), \tag{50}$$

where the coefficient integrals are defined in the usual way as

$$a_0 = \frac{1}{\pi}\int_0^{\pi} h(t)\,dt, \qquad a_j = \frac{2}{\pi}\int_0^{\pi} h(t)\cos(jt)\,dt, \qquad b_j = \frac{2}{\pi}\int_0^{\pi} \sin(jt)h(t)\,dt. \tag{51}$$

These can be approximated (with exponential accuracy, if $h(t)$ is periodic and infinitely differentiable on $t \in [0,\pi]$) by trapezoidal rule quadrature as in equation (34).

The general solution is then

$$\begin{aligned} v &= C - La_0\left(t - \frac{\pi}{2}\right) - L\sum_{j=1}^{\infty}\frac{1}{j}a_j\sin(jt) + L\sum_{j=1}^{\infty}\frac{1}{j}b_j\cos(jt) \\ &= C + L\arctan\left(\frac{y}{L}\right) - L\sum_{j=1}^{\infty}\frac{1}{j}a_j SB_{j-1}(y;L) + L\sum_{j=1}^{\infty}\frac{1}{j}b_j TB_j(y;L), \end{aligned} \tag{52}$$

using the identity $\operatorname{acot}(y/L) - \pi/2 = -\arctan(y/L)$. The constant C, which enforces the initial condition $v = 0$ at $y = 0$ ($\leftrightarrow t = \pi/2$), is

$$C \equiv L\sum_{j=1}^{\infty}\frac{1}{j}a_j\sin\left(\frac{j\pi}{2}\right) - L\sum_{j=1}^{\infty}\frac{1}{j}b_j\cos\left(\frac{j}{pi/2}\right). \tag{53}$$

Note that unlike the other cases described here, $v_y = g$ can be solved explicitly without the need to solve a banded Galerkin's matrix.

The inhomogeneous term of the transformed differential equation, $v_y = g$, is

$$g(y) \equiv f(y) - \frac{1}{2}\{f(\infty) - f(-\infty)\}\operatorname{erf}(y) - \frac{1}{2}\{f(\infty) + f(-\infty)\}. \tag{54}$$

The solution to

$$u_y = f(y), \qquad u(0) = 0, \leftrightarrow u(y) = \int_0^y f(z)\,dz \tag{55}$$

is

$$\begin{aligned} u(y) &= v(y) + \frac{1}{2}(f(\infty) + f(-\infty))\,y \\ &+\frac{1}{2}(f(\infty) - f(-\infty))\left\{y\operatorname{erf}(y) + \frac{1}{\sqrt{\pi}}\left(\exp(-y^2) - 1\right)\right\}. \end{aligned} \tag{56}$$

11. INTEGRAL-OF-INTEGRAL

The problem is to find $u(y)$ where

$$u(y) \equiv \int_0^y dz \int_0^z f(w)\,dw \leftrightarrow u_{yy} = f(y), \qquad u(0) = u_y(0) = 0. \tag{57}$$

The once-iterated integral is harder because now we need *two* subtractions for each parity, versus only one for integral (42). The reason is illustrated by this example.

$$\int_0^y dz \int_0^z \operatorname{sech}^2(w)\,dw = \int_0^y dz \tanh(z)\,dz = \log(\cosh(y)) + \log(2). \tag{58}$$

Even though the integrand $\operatorname{sech}^2(w)$ asymptotes exponentially fast to zero, and thus, has a bounded integral by the previous theorem, the integral-of-its-integral asymptotes to $|y|$ as $|y| \to \infty$.

The following theorem provides sufficient conditions for removing the unboundedness of u.

Theorem 2. Boundedness of Integral-of-Integral. *Let $g(y)$ be a function such that the following bound applies for some constants A, B:*

$$\left|\int_0^y g(z)\,dz\right| < \frac{A}{(B+y^2)}, \qquad \forall y. \tag{59}$$

Then

$$\left|\int_0^y dz \int_0^z g(w)\,dw\right| < C, \qquad \forall y, \tag{60}$$

for some constant C. For the boundedness of the integral of g, it is also necessary that

$$|g(y)| < \frac{A'}{(B'+y^2)}, \qquad \forall y, \tag{61}$$

for some constants A', B'. Equivalently, it is sufficient that

$$h(t) \equiv \frac{g(L\cot(t))}{\sin^2(t)} \qquad \text{and} \qquad k(t) \equiv \frac{1}{\sin^2(t)} \int_t^{\pi/2} h(s)\,ds \tag{62}$$

are bounded.

Proof. Apply Theorem 1 twice, once to the integral of g and then again to g itself. ∎

To impose these conditions, it is convenient to choose functions with simple second derivatives so that $f(y)$ is transformed easily; the corresponding ϕ_j are the iterated integrals of these. The second derivatives of the subtraction functions are the constant one and $\exp(-y^2)$ for the symmetric part of f and $\operatorname{erf}(y)$ and $2y\exp(-y^2)$ for the antisymmetric part. These functions and the weights that enforce both conditions of the theorem, (60) and (61), are given in the summary (75).

12. INTEGRAL-OF-AN-INTEGRAL: NUMERICAL TECHNOLOGY

The second derivative, after separation into two subproblems of definite parity, gives pentadiagonal Galerkin matrices. The technical complication of expanding the inhomogeneous term as a sine series but the unknown as a cosine series does not arise here. As in the previous section, the constant in the spectral series must be determined from the initial condition $u(0) = 0$ rather than from the residual of the differential equation.

Iterating the chain rule gives

$$\frac{d^2}{dy^2} = \frac{\sin^4(t)}{L^2}\frac{d^2}{dt^2} + 2\frac{\cos(t)\sin^3(t)}{L^2}\frac{d}{dt}. \tag{63}$$

The Galerkin matrix elements are, therefore,

$$G^{(2)}_{jk} = \frac{4}{\pi}\int_0^{\pi/2} dt \cos(jt)\left\{\frac{\sin^4(t)}{L^2}\left(-k^2\right)\cos(kt) + 2\frac{\cos(t)\sin^3(t)}{L^2}(-k)\sin(kt)\right\}. \tag{64}$$

Analytical evaluation in Maple shows that the nonzero elements are

$$G_{jj} = -\frac{(3/8)\,j^2}{L^2}, \tag{65}$$

$$G_{j,j\pm2} = \frac{\left\{j^2/4 \pm (3/4)j + 1/2 - (3/16)\delta_{1j}\right\}}{L^2}, \tag{66}$$

$$G_{j,j\pm4} = \frac{\left\{-j^2/16 \mp (3/8)j - 1/2\right\}}{L^2}, \tag{67}$$

where δ_{1j} is zero unless $j = 1$, in which case, $\delta_{11} = 1$.

The column vector on the right-hand side of $\vec{\vec{G}}\vec{a} = \vec{g}$ has elements

$$g_j \equiv \frac{4}{\pi}\int_0^{\pi/2} g(L\cot(t))\cos(jt)\, dt. \tag{68}$$

In summary, define $v_S(y)$ to denote the sum of the even degree basis functions, *omitting the constant*, and similarly for v_A:

$$v_S(y) \equiv \sum_{j=1}^{N} a_{2j} TB_{2j}(y), \qquad v_A(y) \equiv \sum_{j=1}^{N} a_{2j-1} TB_{2j-1}(y), \tag{69}$$

where the coefficients a_j are determined by solving $\vec{\vec{G}}\vec{a} = \vec{g}$ for each parity. The transformed differential equation pair is

$$v_{S,yy} = g_S(y), \qquad v_{A,yy} = g_A(y), \tag{70}$$

where, defining $f_S \equiv \{f(y) + f(-y)\}/2$ and $f_A \equiv \{f(y) - f(-y)\}/2$,

$$g_S(y) \equiv f_S(y) - f_S(\infty) - \sigma_S \frac{2}{\sqrt{\pi}}\exp\left(-y^2\right), \tag{71}$$

$$g_A(y) \equiv f_A(y) - f_A(\infty)\operatorname{erf}(y) - \sigma_A 2y \exp\left(-y^2\right), \tag{72}$$

$$\sigma_S \equiv \int_0^{\infty}\{f_S(y) - f_S(\infty)\}\, dy, \qquad \sigma_A \equiv \int_0^{\infty}\{f_A(x) - f_A(\infty)\operatorname{erf}(y)\}\, dy. \tag{73}$$

The solution to

$$u_{yy} = f(y) \tag{74}$$

is

$$\begin{aligned} u(y) = {} & v_S(y) - v_S(0) + f_S(\infty)\frac{1}{2}y^2 + \sigma_S\left\{y\operatorname{erf}(y) + \frac{1}{\sqrt{\pi}}\left(\exp\left(-y^2\right) - 1\right)\right\} \\ & + v_A(y) + f_A(\infty)\left\{\left(\frac{1}{2}y^2 + \frac{1}{4}\right)\operatorname{erf}(y) + \frac{y}{2\sqrt{\pi}}\left(\exp\left(-y^2\right) - 2\right)\right\} \\ & + \sigma_A\left\{y - \frac{\sqrt{\pi}}{2}\operatorname{erf}(y)\right\}, \end{aligned} \tag{75}$$

where

$$v_S(0) = \sum_{j=1}^{N} a_{2j}(-1)^j. \tag{76}$$

13. INTEGRAL-OF-AN-INTEGRAL: SYMMETRIC NUMERICAL EXAMPLE

The symmetric example is

$$u_S = \frac{y^2}{2} - \log\left(\cosh\left(y\right)\right), \qquad f_S(y) \equiv \tanh^2(y). \tag{77}$$

The two parameters that determine the subtractions are

$$f_S(\infty) = 1, \qquad \sigma_S = \int_0^\infty \left\{f_S(y) - f_S(\infty)\right\} dy = -1. \tag{78}$$

The transformed symmetric problem is

$$v_{S,yy} = \tanh^2(y) - 1 + \frac{2}{\sqrt{\pi}} \exp\left(-y^2\right), \tag{79}$$

with the exact solution

$$v_S = -\log\left(\cosh\left(y\right)\right) + \left\{y \operatorname{erf}(y) + \frac{1}{\sqrt{\pi}}\left(\exp\left(-y^2\right) - 1\right)\right\}, \tag{80}$$

which asymptotes to the constant $\log(2) - 1/\sqrt{\pi}$ and is zero at $y = 0$.

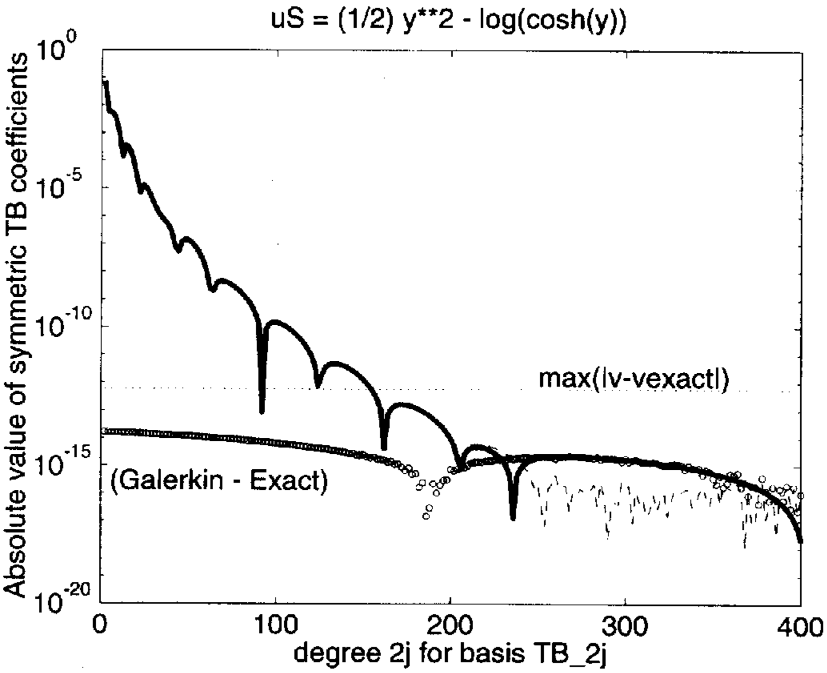

Figure 3. Solid: absolute value of Galerkin-computed coefficients as calculated using even degree rational Chebyshev functions up to and including TB_{400} for a map parameter $L = 1$. Thin dashed: exact coefficients of the TB expansion of v_S. These two curves are indistinguishable until the coefficients have decreased to $O(10^{-15})$. Circles: error in the coefficients, which is the difference between the other two curves. Note that the error curve is almost independent of degree j, at a magnitude controlled by roundoff error, roughly $O(10^{-14})$. The maximum pointwise error (horizontal dotted line) is about 6.7 x 10^{-13}; this, too, is controlled by roundoff error, and would be much smaller (for this value of N) in a computation with higher-precision arithmetic.

Figure 3 is a log-linear plot of the rational Chebyshev coefficients of v_S. For comparison, the dashed curve—indistinguishable from the Galerkin coefficients until the degree is very large—gives the magnitudes of the exact coefficients of the function v_S as computed through the usual Fourier integrals, $a_{2j}^{(\text{exact})} \equiv (2/\pi) \int_0^{\pi} \cos(2jt) v_S(L \cot(t))\, dt$. There are two sources of error in the solution to the differential equation. First, the Chebyshev series must be truncated at $j = 2N$ for some finite N, which gives a "truncation error" that is the sum of all the neglected, higher-degree terms. Second, the Galerkin method (or any of the alternatives like collocation) invariably computes coefficients of low degree which are slightly different from those of the exact expansion. The graph shows that this "discretization error" is roughly the same order of magnitude for all computed coefficients.

It can be proved [8] that both sources of error decrease exponentially fast with N. On a log-linear plot, the coefficients (and error) would asymptote to a straight line if the error decreased geometrically, that is, $\log(\text{error}) \sim -qN$ for sufficiently large N and some positive constant q. Unfortunately, as explained in [6–8], the convergence rate on an unbounded domain is "subgeometric" with $\log(\text{error}) \sim -qN^r$ where $r < 1$, typically in the range of $1/2$ to $2/3$, depending on the problem. (For our examples, Boyd [6] has shown that $r = 1/2$.) The figure shows that with sufficient basis functions, one can obtain full machine precision.

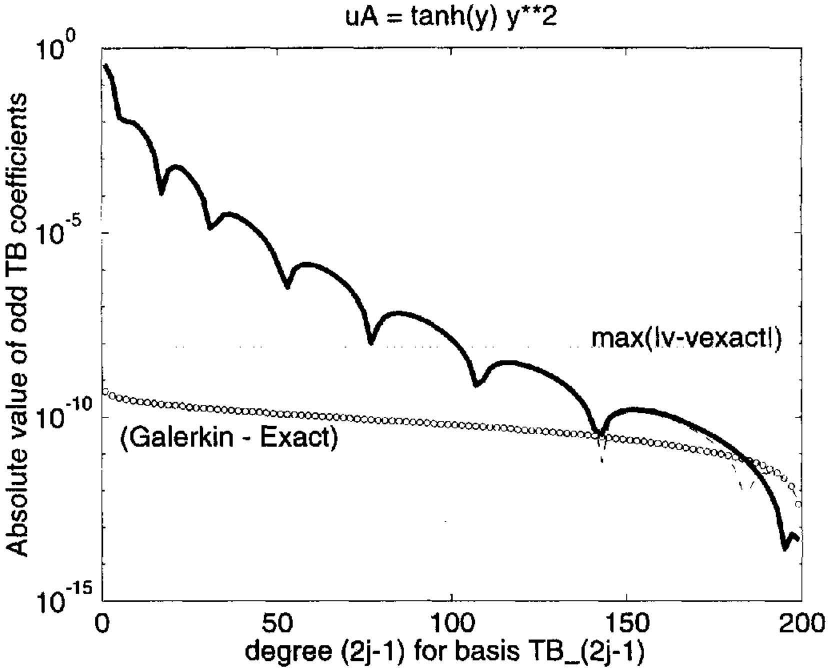

Figure 4. The same as previous figure, but for the antisymmetric example. Note that the coefficients graphed (solid and dashed curves) are those of Galerkin and exact expansion of v_A; an appropriate combination of the error functions and various Gaussians must be added to v_A to obtain the solution to the original problem, $u_A = y^2 \tanh(y)$. The maximum pointwise error is 7.7E−9, which is marked by the thin horizontal dividing line.

14. INTEGRAL-OF-AN-INTEGRAL: ANTISYMMETRIC NUMERICAL EXAMPLE

The problem is

$$u_A = y^2 \tanh(y), \qquad f_A(y) \equiv 4y\,\text{sech}^2(y) + 2\tanh(y) - 2y^2 \tanh(y)\,\text{sech}^2(y). \tag{81}$$

The two parameters that determine the subtractions are

$$f_A(\infty) = 2, \qquad \sigma_A = \int_0^{\infty} \{f_A(y) - f_A(\infty)\,\text{erf}(y)\}\, dy = \frac{2}{\sqrt{\pi}}. \tag{82}$$

The transformed antisymmetric problem is

$$v_{A,yy} = 4y\,\mathrm{sech}^2(y) + 2\tanh(y) - 2y^2\tanh(y)\,\mathrm{sech}^2(y) - 2\,\mathrm{erf}(y) - \frac{4}{\sqrt{\pi}}y\exp\left(-y^2\right), \tag{83}$$

with the exact solution

$$\begin{aligned} v_A = y^2\tanh(y) - 2\left\{\left(\frac{1}{2}y^2 + \frac{1}{4}\right)\mathrm{erf}(y) + \frac{y}{2\sqrt{\pi}}\left(\exp\left(-y^2\right) - 2\right)\right\} \\ - \frac{2}{\sqrt{\pi}}\left\{y - \frac{\sqrt{\pi}}{2}\mathrm{erf}(y)\right\}, \end{aligned} \tag{84}$$

which asymptotes to the constant 1/2.

Figure 4 shows that exponential-but-subgeometric convergence is obtained for this example also. The number of basis functions was halved from the symmetric example to show that the discretization error for the Galerkin-computed coefficients is again roughly independent of the degree j even when N is small enough so that the error (now about 10^{-10} for each coefficient) is less than machine precision. The maximum pointwise error (L_∞ error) is about 7.7×10^{-9}, which is marked on the graph by the thin horizontal dotted line. We see that the average discretization error in a coefficient is roughly the maximum pointwise error divided by the truncation N. Because these errors can accumulate, and the truncation error must be added also, the maximum difference between the sum of the truncated Chebyshev series and the exact $v_A(y)$ is considerably larger than the error in any individual Chebyshev coefficient.

15. MYSTERY: WHY IS THE MAXIMUM POINTWISE ERROR SO LARGE?

A good check on the accuracy of a spectral calculation is to graph the magnitude of the coefficients. In Appendix B, we show that the truncation error, made by chopping a spectral series after the N^{th} term, is

$$E_T(N) \equiv \max_{y\in[-\infty,\infty]}\left|\sum_{j=N+1}^{\infty} a_j TB_j(y)\right| \sim O\left(N^{1-r}\right)\mathcal{E}(N), \tag{85}$$

where $\mathcal{E}(N)$, the "envelope" of the spectral coefficients, is a monotonically-decreasing function which provides an upper bound on the spectral coefficients. Thus, we can conservatively estimate the truncation error from a graph of the computed spectral coefficients as illustrated in Figure 5, where the "envelope" is the slanting dashed line and the lower of the two horizontal dashed lines indicates the estimated truncated error E_T.

The discretization error E_D of a spectral calculation is the sum of the differences between the first $N+1$ exact coefficients and those same coefficients as computed using a Galerkin or pseudospectral algorithm; this error is the small circles in Figures 3 and 4. The total error in solving a differential equation is the sum of E_T and E_D. As explained in the book [8], estimating the truncation error is fairly easy but estimating the discretization error is hard.

Empirically, however, the discretization error is *usually* the *same* order-of-magnitude as the truncation error as formalized as Rule-of-Thumb 1 [8, p. 31]. The reason is obvious. If the truncation error is zero, which is equivalent to the statement that an N-term spectral series exactly represents the solution $u(y)$, then the discretization error E_D will be zero also. For any well-behaved spectral algorithm, the truncation and discretization errors are handcuffed together in the sense that both must go to zero simultaneously as N increases.

Unfortunately, Figure 5 shows that the maximum pointwise error, $\max_{y\in[-\infty,\infty]}|v_S(y) - v_S^{(\text{Galerkin})}|$, is about SIXTY times the truncation error! This implies that the discretization error E_D is roughly sixty times larger than E_T.

We have no explanation for why the discretization error is so large for this example.

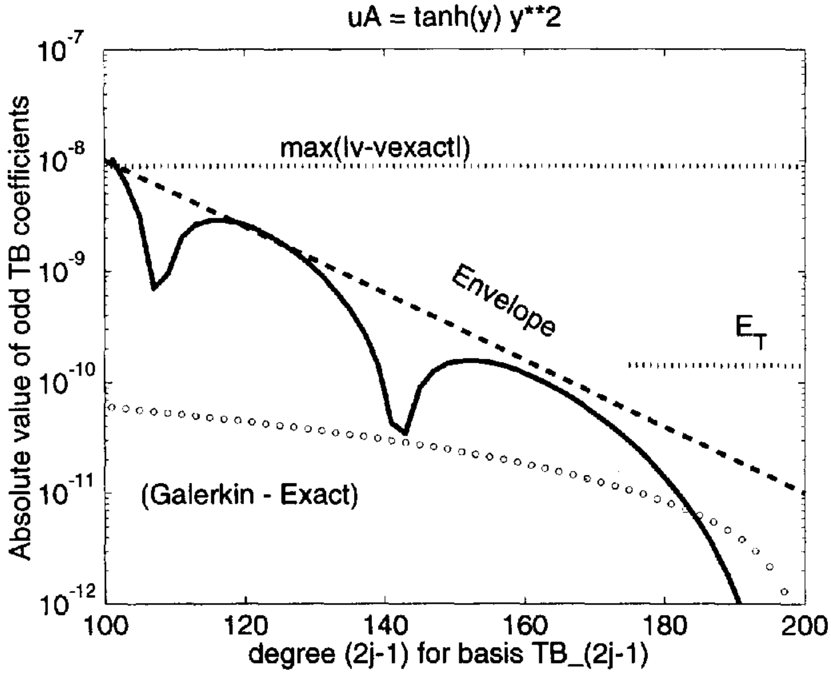

Figure 5. The same as the previous graph, but showing only half the range in degree j and with a compressed vertical scale as well ("zoom" version of the previous figure). The slanting dashed line is the "envelope" $\mathcal{E}(j)$ of the spectral coefficients, which are the heavy solid line; the estimated truncation error is $E_T = \sqrt{N}\mathcal{E}(N)$ as marked by the lower of the two horizontal dotted lines. The maximum pointwise error (upper dotted line) is expected to be the same order of magnitude as the truncation error E_T, but is actually about 60 times larger.

16. SUMMARY

In this work, we have shown that it is straightforward to use rational Chebyshev expansions to solve differential equations or to compute indefinite iterated integrals on an unbounded domain. This is true even though such solutions or integrals may asymptote to polynomials as the coordinate $y \to \infty$. The key trick is to construct simple, explicit special basis functions which have polynomial unboundedness. For linear differential equations and for indefinite integrals, the coefficients of these special functions can be found through an asymptotic analysis for large y. The differential equation can then be transformed so that the new unknown $v(y)$ is bounded, and therefore, has a rapidly-convergent series in rational Chebyshev functions TB_j.

When the differential equation has constant coefficients, a Galerkin discretization is very efficient because it generates a banded matrix. Furthermore, when the differential equation is parity-preserving, symmetry can be exploited to

(i) split the matrix problem into two subproblems of half the size and a smaller bandwidth, and also

(ii) to simplify the unboundedness-removing transformation.

Problems with polynomial-unboundedness arise very naturally in the method of matched asymptotic expansions as noted earlier. The composite matched asymptotics approximation is bounded because the outer approximation, which does not grow with $|y|$, replaces the growing inner approximation for large $|y|$. In the RMKdV problem for nonlinear waves in a channel [4], the inner differential equations are linear. However, nonlinear inner problems can arise in matched asymptotics also.

In principle, the strategies described here can be extended to nonlinear differential equations. However, the Galerkin matrix is almost always dense, and it is probably easier to apply the collocation or pseudospectral method [9,12].

In Appendix A, we give a trio of three-term recurrence relations which collectively compute the integrals of the rational Chebyshev functions. In Appendix B, we derive a bound/estimate of the truncation error which applies to any spectral series. This generalizes the bounds of [8] to series with an exponential but subgeometric rate of convergence, which is the usual rate for expansions on an unbounded spatial interval.

Our examples have been confined to first- and second-order linear ordinary differential equations. However, the underlying ideas have a much broader applicability. Special basis functions that depend *nonlinearly* on the unknown have been applied to nonlinear differential equations [10,13], for example.

Clearly, unbounded solutions on an unbounded integral are not necessarily difficult. With simple tricks, it is possible to achieve spectral accuracy.

APPENDIX A
RECURRENCE AND LEMMA FOR INTEGRALS OF THE RATIONAL CHEBYSHEV FUNCTIONS

The desired integrals are defined by

$$\mathcal{I}_n \equiv \int_0^y TB_n(x)\,dy \tag{86}$$

$$= \int_{\operatorname{acot}(y/L)}^{\pi/2} \cos(n\tau)\frac{L}{\sin^2(\tau)}\,d\tau. \tag{87}$$

A recurrence relation for these integrals can be derived by using the recurrence relation for the cosine functions

$$\cos([n+1]\tau) = 2\cos(\tau)\cos(n\tau) - \cos([n-1]\tau), \qquad n = 1, 2, \dots. \tag{88}$$

(Parenthetically, it may be noted that this, after a change-of-coordinate, is the usual three-term recurrence for the Chebyshev polynomials.) Integrating both sides of this recurrence and invoking the definition of $\mathcal{I}_n$ gives a formula for $\mathcal{I}_{n+1}$ in terms of $\mathcal{I}_{n-1}$ and the quantity defined by

$$\mathcal{J}_n \equiv \int_0^y TB_1(y)TB_n(y)\,dy = \int_{\operatorname{acot}(y/L)}^{\pi/2} \cos(\tau)\cos(n\tau)\frac{L}{\sin^2(\tau)}\,d\tau. \tag{89}$$

We can derive a recurrence for the auxiliary integrals $\mathcal{J}_n$ by again applying the cosine identity (88), and then splitting the integral whose integrand is proportional to $\cos^2(\tau)\cos(n\tau)$ by the identity $\cos^2(\tau) = 1 - \sin^2(\tau)$. The result can be written in terms of the $\mathcal{J}_n$ and $\mathcal{I}_n$ plus the integral of $\cos(n\tau)$, which, of course, can be integrated explicitly to give $\sin(nt)$.

The images of $\sin(nt)$ under the mapping $y = L\cot(t)$ have been discussed in [8] using the notation

$$SB_{n-1}(y; L) \equiv \sin\left(n \operatorname{acot}\left(\frac{y}{L}\right)\right), \qquad n = 1, 2, \dots. \tag{90}$$

These basis functions satisfy the same recurrence relation as their cousins, the TB_n, which are the images of $\cos(nt)$ under the same map. One finds that the integrals of the TB functions can thus be computed by the following two-step procedure. (For simplicity, we set $L = 1$ in the rest of this appendix, but the general case follows by making the elementary change-of-variable $y \to y/L$.) The first is to initialize the recurrences through

$$\mathcal{I}_0 = y, \qquad \mathcal{I}_1 = \sqrt{1+y^2} - 1, \tag{91}$$

$$\mathcal{J}_0 = \sqrt{1+y^2} - 1, \qquad \mathcal{J}_1 = y - \arctan(y), \tag{92}$$

$$SB_0 = \frac{1}{\sqrt{1+y^2}}, \qquad SB_1 = 2\frac{y}{1+y^2}. \tag{93}$$

The second step is to simultaneously advance three recurrences through a single loop:

$$SB_{n+1} = 2\frac{y}{\sqrt{1+y^2}}SB_n - SB_{n-1}, \tag{94}$$

$$\mathcal{I}_{n+1} = 2\mathcal{J}_n - \mathcal{I}_{n-1}, \tag{95}$$

$$\mathcal{J}_{n+1} = 2\mathcal{I}_n + \frac{2}{n}\left(SB_{n-1} - \sin\left(n\frac{\pi}{2}\right)\right) - \mathcal{J}_{n-1}. \tag{96}$$

If one wishes to compute all the $\mathcal{I}_n$ for $n \leq N$, then the cost (in floating point arithmetic) is directly proportional to N.

In principle, this formalism can be extended to iterated integrals also. However, after the first integration, one must subtract the unbounded, growing-linearly-with-$|y|$ factors and re-expand the difference as a series of TB functions. Thus, the strategy of subtractions is necessary for iterated integrals even when using recurrence relations.

Although suitable for numerical evaluation, the recurrence leaves unanswered an important question. How do the integrals behave as $y \to \infty$? The following provides an answer.

Theorem 3. Asymptotics of Integrals of TB Functions.

$$\int_0^y TB_n(x)\,dy \sim (\operatorname{sign}(y))^n\, y + O\left(n^2\right), \qquad |y| \gg \frac{1}{n}. \tag{97}$$

Proof. The trigonometric definition of the rational Chebyshev functions is

$$TB_n(y;L) \equiv \cos(nt[y]), \qquad t = \operatorname{acot}\left(\frac{y}{L}\right). \tag{98}$$

When $y \gg L$, power series for the trigonometric functions show that $t \approx L/y$. Similarly, when y is large and negative, $t \to \pi$. Combining both limits gives

$$TB_n(y;L) \sim (\operatorname{sign}(y))^n - \frac{n^2L^2}{2}\frac{1}{y^2} + O\left(y^{-4}\right). \tag{99}$$

Next, split the range of integration, $x \in [0,y]$, into two parts. On the subinterval $x \in [0, n/\epsilon]$ for some $0 < \epsilon$, the asymptotic approximation (99) does not apply. However, since $|TB_n(y;L)| \leq 1$ for all n and all real y as follows from its definition in terms of the cosine, it follows that the integral of TB_n on this subinterval is bounded for all n and $\epsilon > 0$. If $\epsilon \ll 1$, then the asymptotic approximation will be accurate on the rest of the integration range, $x \in [n/\epsilon, y]$. The sign function integrates to $(\operatorname{sign}(y))^n y$; the integral of the error terms is bounded as $y \to \infty$. ∎

APPENDIX B
IMPROVED TRUNCATION ERROR BOUND

In this appendix, we derive a bound on the truncation error for a spectral series which has "subgeometric" convergence as defined in [8]:

$$a_j \sim \{\,\}\exp\left(-qj^r\right), \qquad j \to \infty, \tag{100}$$

where the empty braces denote factors that vary slower-than-exponentially with j or are oscillatory with degree j. Although the rest of this article has focused solely on rational Chebyshev functions, the theorem derived here applies equally to Chebyshev and Legendre polynomials, Fourier series, and to all other spectral series for which the basis functions have been normalized to maximum values of one on the expansion interval.

Theorem 4. Truncation Error Bound. *Let $\phi_j(y)$ denote the elements of a spectral basis set which have been normalized to a maximum value of one on the expansion interval. Define the* *truncation error $E_T(N)$ as the sum of all terms in the infinite series which are neglected when the series is truncated after the term of degree $j = N - 1$:*

$$E_T(N) \equiv \sum_{j=N+1}^{\infty} a_j \phi_j(y). \tag{101}$$

Suppose that the spectral coefficients satisfy a bound of the form, with $r \geq 0$,

$$|a_j| \leq C \exp\left(-q j^r\right) \equiv \mathcal{E}(j), \qquad j > N, \tag{102}$$

where the bounding function $\mathcal{E}(j)$ is said to be an "envelope of the spectral coefficients". Then

$$\begin{aligned} |E_T(N)| &\leq C \frac{1}{r q^{1/r}} \Gamma\left(\frac{1}{r}; q N^r\right), \qquad N \gg 1 \\ &\leq C \frac{1}{rq} N^{1-r} \exp\left(-q N^r\right) \left\{ 1 + \frac{1/r - 1}{q N^r} + \frac{(1/r - 1)(1/r - 2)}{q^2 N^{2r}} + \cdots \right\}, \end{aligned} \tag{103}$$

where $\Gamma(\alpha; z)$ is the usual incomplete Γ function defined by

$$\Gamma(\alpha; z) \equiv \int_z^{\infty} \exp\left(-t\right) t^{\alpha - 1}\, dt. \tag{104}$$

Proof. Because each basis function has a maximum value of one, it follows that each term in the spectral series is individually bounded by $|a_j|$:

$$E_T(N) \leq \sum_{j=N+1}^{\infty} |a_j|. \tag{105}$$

Replacing each coefficient by its upper bound as specified in equation (102) gives

$$E_T(N) \leq C \sum_{j=N+1}^{\infty} \exp\left(-q j^r\right). \tag{106}$$

It is easier to analyze integrals than sums, so note that without approximation,

$$s \equiv \sum_{j=N+1}^{\infty} \exp\left(-q j^r\right) = N \int_1^{\infty} \sigma(x)\, dx, \tag{107}$$

where $\sigma(x)$ is the piecewise-constant function

$$\sigma(x) \equiv \exp\left(-q N^r \left(\frac{j+1}{N}\right)^r\right), \qquad x \in \left[\frac{j}{N}, \frac{(j+1)}{N}\right]. \tag{108}$$

Note further that σ is bounded from above by the integrand of

$$I(r, Q) = \int_1^{\infty} dx \exp\left(-Q x^r\right), \tag{109}$$

where $Q \equiv q N^r$. It follows that $s \leq I(r, Q)$ and, in fact, this is a very tight bound in the sense that $s \approx I(r, Q)$ within a relative error of $O(1/N)$ as $N \rightarrow \infty$.

By the change of coordinate $Q x^r \equiv z$, the integral $I(r, Q)$ can be transformed into the usual definition of the incomplete gamma function given in the theorem. The last line of the theorem follows by replacing the gamma function by its large-Q asymptotic approximation. ∎

To make practical use of this theorem, three further observations are necessary. First, the integral that defines the upper bound becomes an increasingly good approximation (with a relative error $O(1/N)$) as N increases. If we relax the certainty of a bound for the explicitness of an approximation, then the theorem can be restated as: the truncation error in the sum of a series whose j^{th} term is $\exp(-qj^r)$ is approximately

$$\sum_{j=N+1}^{\infty} \exp\left(-qj^r\right) \sim \frac{1}{rq} N^{1-r} \exp\left(-qN^r\right). \tag{110}$$

Second, the last retained coefficient in the series is $\exp(-qN^r)$. Therefore, equation (110) can be restated as: the truncation error of the subgeometrically-converging sum is the magnitude of the N^{th} term multiplied by $N^{1-r}/(rq)$. This implies that the last coefficient that we *keep* provides us useful information about the sum of all the higher-degree terms that we *drop* in the truncation. Similar reasoning led to the "last coefficient error estimate", Rule-of-Thumb 2 [8, p. 51]. The book restricted itself to the special case $r = 1$ (geometric convergence); we have here generalized the argument to subgeometric convergence ($r < 1$) also.

Third, spectral coefficients usually *oscillate* as well as *decay* with increasing degree j. (In the author's experience, this seems to be almost universal for rational Chebyshev series.) Thus, one has to be careful: a truncation error based on the size of a_N could be wildly optimistic if degree N happened to be a zero or near-zero of the oscillations of the spectral coefficients with degree. For this reason, Boyd [8] and Flyer [14] have introduced the concept of the "envelope of the spectral coefficients", $\mathcal{E}(j)$, as defined above in the body of the theorem, equation (102). When the coefficients are oscillatory-in-degree, the truncation error is

$$E_T(N) \sim O\left(N^{1-r}\right) \mathcal{E}(N). \tag{111}$$

An algebraically-converging series, i.e., one where the best bound one can establish is of the form

$$|a_j| \leq C \frac{1}{j^k}, \tag{112}$$

for some constant k, is the limit $r \to 0$ of an exponentially-convergent series. A geometrically-converging series is the limit $r = 1$. The theorem above interpolates between the two cases $r = 1$ (geometric convergence) and $r = 0$ (algebraic convergence) given in the last coefficient error estimate, Rule-of-Thumb 2 of [8, p. 51].

REFERENCES

1. J. Kevorkian and J.D. Cole, *Multiple Scale and Singular Perturbation Methods*, Springer-Verlag, New York, (1996).
2. M. Van Dyke, *Perturbation Methods in Fluid Mechanics*, Second Edition, Parabolic Press, Stanford, CA, (1975).
3. M. Van Dyke, Radiative decay of weakly nonlocal solitary waves, *Wave Motion* **27**, 211–221 (1998).
4. J.P. Boyd and G.-Y. Chen, Analytical and numerical studies of weakly nonlocal solitary waves of the Rotation-Modified Korteweg-deVries equation, *Physica D* (submitted).
5. C.E. Grosch and S.A. Orszag, Numerical solution of problems in unbounded regions: Coordinate transforms, *Journal of Computational Physics* **25**, 273–296 (1977).
6. J.P. Boyd, The optimization of convergence for Chebyshev polynomial methods in an unbounded domain, *Journal of Computational Physics* **45**, 43–79 (1982).
7. J.P. Boyd, Spectral methods using rational basis functions on an infinite interval, *Journal of Computational Physics* **69**, 112–142 (1987).
8. J.P. Boyd, *Chebyshev and Fourier Spectral Methods*, Second Edition, Dover, Mineola, NY, (2000).
9. D. Gottlieb and S.A. Orszag, *Numerical Analysis of Spectral Methods*, SIAM, Philadelphia, PA, (1977).
10. J.P. Boyd, *Weakly Nonlocal Solitary Waves and Beyond-All-Orders Asymptotics: Generalized Solitons and Hyperasymptotic Perturbation Theory, Volume 442, Mathematics and Its Applications*, Kluwer, Amsterdam, (1998).
11. G.-Y. Chen, Application of weak nonlinearity in ocean waves, Ph.D. Dissertation, University of Michigan, Department of Atmospheric, Oceanic and Space Science (July 1998).
12. B.A. Finlayson, *The Method of Weighted Residuals and Variational Principles*, Academic, New York, (1973).

13. B.A. Finlayson, Weakly nonlocal solitons for capillary-gravity waves: Fifth-degree Korteweg-de Vries equation, *Physica D* **48**, 129–146 (1991).
14. N. Flyer, Asymptotic upper bounds for the coefficients in the Chebyshev series expansion for a general order integral of a function, *Math. Comput.* **67**, 1601–1616 (1998).
15. B.A. Finlayson, The asymptotic Chebyshev coefficients for functions with logarithmic endpoint singularities, *Applied Mathematics and Computation* **29**, 49–67 (1989).