Boise State University
ScholarWorks

Mathematics Faculty Publications and Presentations | Department of Mathematics

1-1-2002

Accuracy, Resolution and Stability Properties of a Modified Chebyshev Method

Jodi Mead
Boise State University

Rosemary A. Renaut
Arizona State University, Tempe Campus

This document was originally published by Society for Industrial and Applied Mathematics (SIAM) in *SIAM Journal on Scientific Computing*. Copyright restrictions may apply. DOI: 10.1137/S1064827500381501

SIAM J. SCI. COMPUT.
Vol. 24, No. 1, pp. 143–160
© 2002 Society for Industrial and Applied Mathematics

ACCURACY, RESOLUTION, AND STABILITY PROPERTIES OF A MODIFIED CHEBYSHEV METHOD*

JODI L. MEAD† AND ROSEMARY A. RENAUT‡

Abstract. While the Chebyshev pseudospectral method provides a spectrally accurate method, integration of partial differential equations with spatial derivatives of order M requires time steps of approximately $O(N^{-2M})$ for stable explicit solvers. Theoretically, time steps may be increased to $O(N^{-M})$ with the use of a parameter, α-dependent mapped method introduced by Kosloff and Tal-Ezer [*J. Comput. Phys.*, 104 (1993), pp. 457–469]. Our analysis focuses on the utilization of this method for reasonable practical choices for N, namely $N \lesssim 30$, as may be needed for two- or three-dimensional modeling. Results presented confirm that spectral accuracy with increasing N is possible both for constant α (Hesthaven, Dinesen, and Lynov [*J. Comput. Phys.*, 155 (1999), pp. 287–306]) and for α scaled with N, α sufficiently different from 1 (Don and Solomonoff [*SIAM J. Sci. Comput.*, 18 (1997), pp. 1040–1055]). Theoretical bounds, however, show that any realistic choice for α, in which both resolution and accuracy considerations are imposed, permits no more than a doubling of the time step for a stable explicit integrator in time, much less than the $O(N)$ improvement claimed by Kosloff and Tal-Ezer. On the other hand, by choosing α carefully, it is possible to improve on the resolution of the Chebyshev method; in particular, one may achieve satisfactory resolution with fewer than π points per wavelength. Moreover, this improvement is noted not only for waves with the minimal resolution but also for waves sampled up to about 8 points per wavelength. Our conclusions are verified by calculation of phase and amplitude errors for numerical solutions of first and second order one-dimensional wave equations. Specifically, while α can be chosen such that the mapped method improves the accuracy and resolution of the Chebyshev method, for practical choices of N, it is not possible to achieve both single precision accuracy and gain the advantage of an $O(N^{-M})$ time step.

Key words. Chebyshev collocation, accuracy, stability, partial differential equations

AMS subject classifications. 65M70, 65M12

PII. S1064827500381501

1. Introduction.

Spectral methods are based on global approximations,

$$u(x) \approx \sum_{j=0}^{N} a_j \phi_j(x),$$

where the basis functions $\phi_j(x)$ are assumed to be infinitely differentiable global functions, typically eigenfunctions of singular Sturm–Liouville problems [4]. The method of calculation for the coefficients $\{a_j\}$ determines the type of spectral method as Galerkin, Galerkin–Tau, or collocation [4].

The study presented here concerns collocation methods, in particular the Chebyshev pseudospectral (CPS) collocation method, for which the basis functions are the set of orthogonal polynomials, the Chebyshev polynomials,

$$\phi_j(x) = T_j(x) = \cos(j \cos^{-1} x),$$

and collocation is at the Chebyshev points $x_k = \cos(\pi k/N)$, $k = 0 : N$. This method gives highly accurate approximations for the solution of partial differential equations

*Received by the editors November 21, 2000; accepted for publication (in revised form) November 14, 2001; published electronically May 20, 2002.
http://www.siam.org/journals/sisc/24-1/38150.html
†Department of Mathematics, Boise State University, Boise, ID 83725-1555 (mead@math.boisestate.edu).
‡Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804 (renaut@asu.edu).

[4, 8] and has been widely used and studied [3, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. On the other hand, although the low to mid end of the spectrum of the differential operator is well approximated by the eigenvalues of the CPS differential operator as N tends to infinity, the CPS operator overestimates the larger values of the spectrum [20]. These so-called outliers are of $O(N^2)$ and restrict the time step that may be used to integrate a partial differential equation with an explicit ordinary differential equation solver to be $O(N^{-2})$. By suitable transformations the Chebyshev grid becomes more evenly spaced [3, 13] and the outliers are reduced to $O(N)$, hence potentially permitting larger time steps, $O(N^{-1})$.

In section 2 we review the properties of the parameter dependent grid transformation introduced by Kosloff and Tal-Ezer [13], and, contrary to the conclusion of Hesthaven, Dinesen, and Lynov [11], we determine that the optimal choice of α with regard to both accuracy and resolution theoretically returns the method to the Chebyshev case. On the other hand, it is possible to achieve good results in relation to practical levels of precision, single or double, for fixed α, provided that α is not chosen near 1. The emphasis of this work concerns the impact of the mapped method for a choice of N that may be realistically used in three-dimensional simulations; we assume then that N is taken to be rather small, $N \lesssim 30$. In this case, α must be taken to be rather small such that the choice suggested by Hesthaven, Dinesen, and Lynov [11] is indeed good. However, we also show that for small choices of α the accuracy is very sensitive to small changes in α so that, for example, in practice the choice $\sin(1.0)$ is much better than $\cos(0.5)$. In addition, for such practical choices of both α and N, the anticipated $O(N^{-1})$ time step is replaced by at most a doubling of the time step used in the Chebyshev method; otherwise minimal single precision accuracy is not achievable. In sections 3 and 4 we confirm these observations by calculation of amplitude and phase errors introduced by numerical solutions of first and second order wave equations in one dimension. Moreover, we demonstrate the sensitivity of the accuracy to the choice of the parameter. We find that choosing α for resolution, as suggested in [13], minimizes the phase and amplitude errors for the first order case. For the second order wave, the optimal parameter choice does lead to minimal phase and amplitude errors for large N. Finally, since our focus is the study of these methods for practical N our results complement and extend the analysis in Don and Solomonoff [7], in which the focus was strictly on accuracy, without regard to a need to keep N relatively small.

2. Grid transformations. We consider the grid transformations in which the grid for $-1 \le y \le 1$ is obtained by the transformation $y = g(x)$ for some, possibly parameter dependent, continuous transformation $g : [-1, 1] \to [-1, 1]$. The resulting grid in y is *stretched* with respect to the original grid in x. Under the mapping, interpolation of function $u(x)$ by the order N interpolant

$$u(x) \approx P_N(x) = \sum_{j=0}^{N} u(x_j)\phi_j(x),$$

for a set of polynomial basis functions $\phi_j(x)$, $j = 0 : N$, is replaced by an interpolant in the variable $y = g(x)$,

$$u(y) \approx q_N(y) = \sum_{j=0}^{N} u(y_j)\psi_j(y), \tag{2.1}$$

for a new set of basis functions $\psi_j(y)$, $j = 0 : N$, $\psi_j(y) = \phi_j(g^{-1}(y))$, not necessarily polynomial in y.

In order to solve a given partial differential equation, in variable $u(y)$, on the transformed grid, the grid values $u(y_j)$ must be used to obtain values for the derivatives of u of any order. This is accomplished via repeated application of the chain rule to reexpress derivatives of $u(y)$ in terms of derivatives of $u(x)$ with respect to x, yielding, for example, for the first order derivative

$$\frac{du}{dy}(y_k) \cong \frac{1}{g'(x_k)} \sum_{k=0}^{N} D_{kj} u(y_j) \tag{2.2}$$

[3, 13]. Here the entries D_{kj} are the entries of the matrix, D, which approximates the first order differential operator such that the derivative approximations on the original grid are given by $u'(x_k) \cong (Du)_k$. Thus for the derivative on the stretched grid, the operator D is replaced by AD, where A is a diagonal matrix with entries $A_{kk} = 1/g'(x_k)$. We note that the approximations to derivatives on the new grid are not necessarily polynomial. This contrasts with the mapping method introduced by Carpenter and Gottlieb [5] in which the derivative is provided with respect to polynomial basis functions in the original variable. For this method it is necessary that the initial function is expressed with respect to a grid in which its accuracy is high at all points.

The operators for higher order derivatives are obtained in a similar fashion [13, 16]. Here we will need the second order derivative operator

$$\frac{d^2}{dy^2} \approx A^2 D_2 - A_2 D, \tag{2.3}$$

where D_2 is the second order operator for the original grid and A_2 is the diagonal matrix with entries $g''(x_k)/(g'(x_k))^3$. We note that this operator is not equivalent to operation by $(AD)^2$. Likewise, the third order derivative operator should be obtained similarly, rather than by $(AD)^3$, contrary to the suggestion in [7].

2.1. The Kosloff–Tal-Ezer transformation. In this report our focus is on a study of the parameter dependent transformation of the Chebyshev grid,

$$y(x) = g(x) = \frac{\sin^{-1}(\alpha x)}{\sin^{-1}(\alpha)}, \qquad 0 < \alpha < 1, \tag{2.4}$$

introduced by Kosloff and Tal-Ezer [13]. The choice of the parameter α determines the degree to which the grid is stretched under the mapping and the degree to which the elements of the matrix D are scaled by the diagonal matrix A. The scaling of D for $\alpha = .99999$ and $N = 32$ is illustrated in Figures 2.1 and 2.2. We see that the larger entries in D are considerably damped by A. While the entries of D are of $O(N)$ for the middle rows, those entries in the bordering rows are of $O(N^2)$; see, for example, [4]. Obviously, the entries of $O(N^2)$ have been damped to size $O(N)$ by the scaling with A,

$$A_{kk} = \frac{1}{g'(\alpha, x_k)} = \frac{\sin^{-1}(\alpha)\sqrt{1-\alpha^2 x_k^2}}{\alpha}.$$

Noting the approximation $\alpha_l = \cos(l\pi/N) \approx 1 - (1/2)(l\pi/N)^2 + O((l\pi/N)^4)$, for l small relative to N, this scaling can be verified. In particular, A_{kk} increases from $O(1/N)$ at $k = 0$, to $O(1)$ for k at the middle of the matrix, and then decreases back

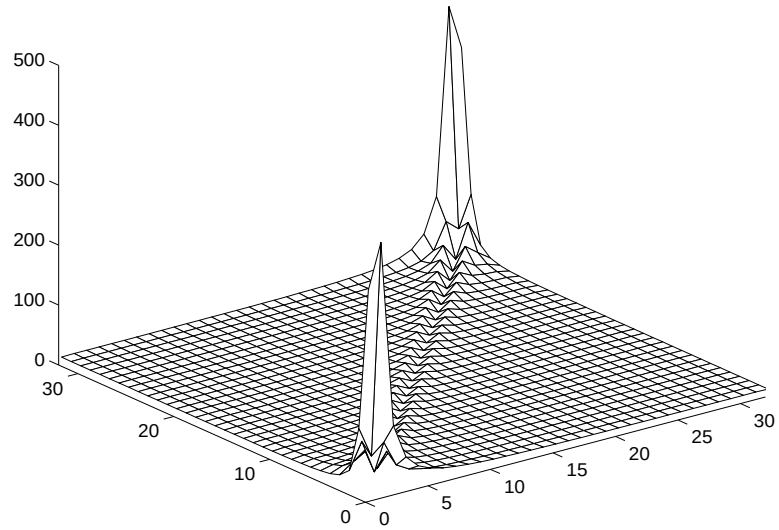

FIG. 2.1. *Magnitude of elements of D for $N = 32$.*

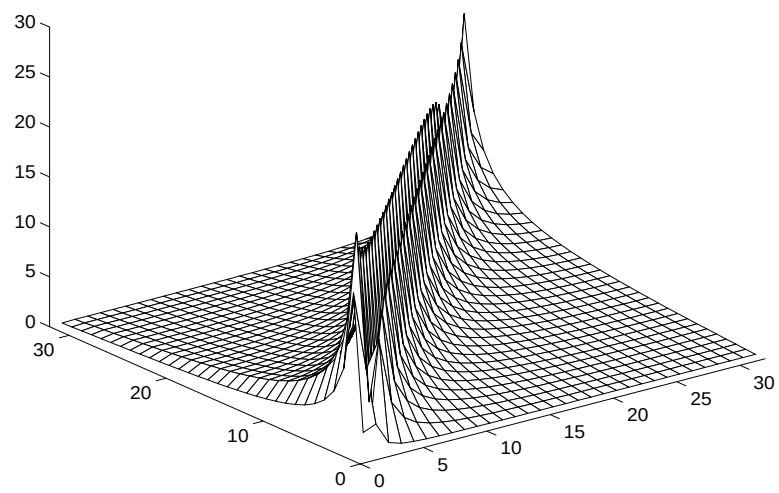

FIG. 2.2. *Magnitude of elements of AD for $N = 32$, $\alpha = .99999$.*

to $O(1/N)$ for $k = N$. While the more severe scaling with α tending to 1 permits the use of larger time steps for stability, in the limit $\alpha = 1$ the mapping (2.4) is singular. In this case, the matrix A is singular, $A_{00} = A_{NN} = 0$, and the Dirichlet boundary conditions

$$u'(-1) = u'(1) = 0$$

are imposed on the solution u.

2.2. The parameter α.

2.2.1. Accuracy. In order for the mapped method to maintain the high accuracy of the Chebyshev method, it is essential that the singularity in the mapping is controlled. Kosloff and Tal-Ezer [13] demonstrated that this can be accomplished by specifying α through the relationship

$$\epsilon_N = \epsilon^N = \left(\frac{1-\sqrt{1-\alpha^2}}{\alpha}\right)^N, \tag{2.5}$$

namely,

$$\alpha_a = \left(\cosh\frac{|\ln\epsilon_N|}{N}\right)^{-1}. \tag{2.6}$$

If ϵ_N is taken to be machine precision the error due to the mapping can be ignored.

2.2.2. Resolution. *Chebyshev.* The truncation of the series approximation to $u(y)$ also introduces an error, ϵ_A, which for the unmapped Chebyshev method controls the resolution of waves on the grid. Specifically, we consider the resolution of waves $\sin(k\pi x)$, $k = 1, 2, \ldots, k_{\max}$, for a series approximation of order N. Gottlieb and Orszag [9] noted that ϵ_A, which can be expressed as a sum of terms involving Bessel functions, $J_m(k\pi)$, $m \geq N+1$, is dominated by the first ignored term of the series,

$$\epsilon_A \approx 2(-1)^p J_{2p+1}(k\pi)\cos((2p+1)\theta),$$

$N = 2p-1$ (see equation (3.41) in [9]) because $|J_m(k\pi)|$ tends to zero very quickly with increasing m. The extent to which ϵ_A can be taken as negligible for given k depends on the speed with which $|J_N(k\pi)|$ tends to zero with increasing N. Gottlieb and Orszag [9] found that ϵ_A decreases very quickly for $N/k > \pi$. Hesthaven, Dinesen, and Lynov [11] deduced this requirement by consideration of the asymptotic limit for ϵ_A from

$$\lim_{N\to\infty} J_N(k\pi) \simeq \frac{1}{\sqrt{2\pi N}}\left(\frac{ek\pi}{N}\right)^N;$$

see equation (9.3.1) in [1]. However, it is readily verified computationally that this estimate is not valid for practical choices of N, $N \leq 200$. On the contrary, equation (9.1.63) in [1] may be used to obtain the much better estimate

$$\begin{aligned} |J_N(k\pi)| &= \left|J_N\left(\frac{Nk\pi}{N}\right)\right| \\ &\leq \left|\frac{(k\pi/N)^N \exp(N\sqrt{1-(k\pi/N)^2}}{1+\sqrt{1-(k\pi/N)^2}}\right|, \end{aligned}$$

from which we also deduce that $\epsilon_A \to 0$, provided $N/k > \pi$. Moreover, for $N/k = \pi$, by equation (9.1.61) in [1], we find that

$$0 < J_N(N) < \frac{\sqrt{2}}{3^{2/3}\Gamma(2/3)N^{1/3}}$$

decreases very slowly with N. Hence the requirement of π points per wavelength for the Chebyshev method is strict.

Mapped Chebyshev. Kosloff and Tal-Ezer [13] extended these ideas for the mapped method and were able to demonstrate a potential improvement in resolution as compared to the Chebyshev method. In particular, by analysis of the dominant term in the truncated series they found that the maximal wave number which can be resolved is

$$k_{\max} = \frac{N \sin^{-1}(\alpha)}{\pi\alpha}. \tag{2.7}$$

In the limit as $\alpha \to 1$, this provides the Fourier result, $k_{\max} = N/2$, but as $\alpha \to 0$ the Chebyshev conclusion, $k_{\max} = N/\pi$, is obtained. We may therefore anticipate that, while the resolution requirement does indeed conflict with the accuracy requirement, the mapped method potentially offers improved resolution as compared to the Chebyshev method, provided $(N/\pi) < k_{\max} < N/2$ can be chosen such that accuracy is not compromised. This contradicts the assumption that $k_{\max} < (N/\pi)$ in [11].

Hesthaven, Dinesen, and Lynov [11] rightly noted that an N dependent choice for α in which α becomes closer to one with increasing N will not permit exponential convergence with N. Specifically, the choice $\alpha = \cos(l\pi/N)$ for small l, when inserted in (2.5), yields the fixed limit

$$\lim_{N\to\infty} \epsilon_N = e^{-l\pi}.$$

On the other hand, taking $l = N/(2\pi)$, hence $\alpha = \cos(0.5)$, does provide exponential convergence, $\epsilon_N \sim e^{-N/2}$ [11]. However, inserting $\alpha = \cos(0.5)$ in (2.7) we find $k_{\max} \sim (1.22)(N/\pi)$, some 20% larger than the suggested maximal N/π in [11]. The discrepancy arises from the utilization of the approximation of (2.7) for $\alpha \approx 1$, $\alpha = \cos(l\pi/N)$, $k_{\max} = (N/2) - l$, for α sufficiently different from 1. Alternatively, using the approximation $\alpha \approx 1$ in (2.7) directly, and $k_{\max}/N = 1/\pi$, yields $\alpha = \sin(1.0)$, rather than $\cos(0.5)$. However, neither of these choices actually enforces the resolution requirement of π points per wavelength. Rather, on inserting the choices $\alpha = \cos(0.5)$, $\sin(1.0)$ in (2.7), without any approximation of α, yields 2.57 and 2.64 points per wavelength, respectively. Moreover, any choice of $\alpha > 0$ predicts the mapped resolution to be better than that of the Chebyshev method, while the limit case $k_{\max} = N/\pi$ inserted in (2.7) yields $\alpha = 0.0$.

Suppose, then, that we seek to resolve all waves up to a certain percentage of the maximum possible wavenumber $N/2$, $k_{\max} = (p/100)(N/2)$, $(200/\pi) < p < 100$; equivalently, we seek resolution, r, $\pi < r < 2$, where r is the number of points per wavelength. The parameter α, independent of N, can then be obtained as the solution of

$$\sin\left(\frac{\pi\alpha k_{\max}}{N}\right) - \alpha = 0, \tag{2.8}$$

from which the value for ϵ in (2.5) may also be calculated. We illustrate these calculations in Figure 2.3 in which we plot $\alpha(p)$ and $\epsilon(p)$, asterisks and pluses, respectively. For comparison, we also indicate the values of ϵ for $\alpha = \cos(0.5)$, $\sin(1.0)$ and an additional choice $\alpha = .91901$, to be explained shortly. We note that α and ϵ are both increasing functions of p, which illustrates the competition between the requirements of resolution and accuracy. In particular, for higher percentages p resolution improves and the range of wavenumbers resolved is larger, but concurrently ϵ in-

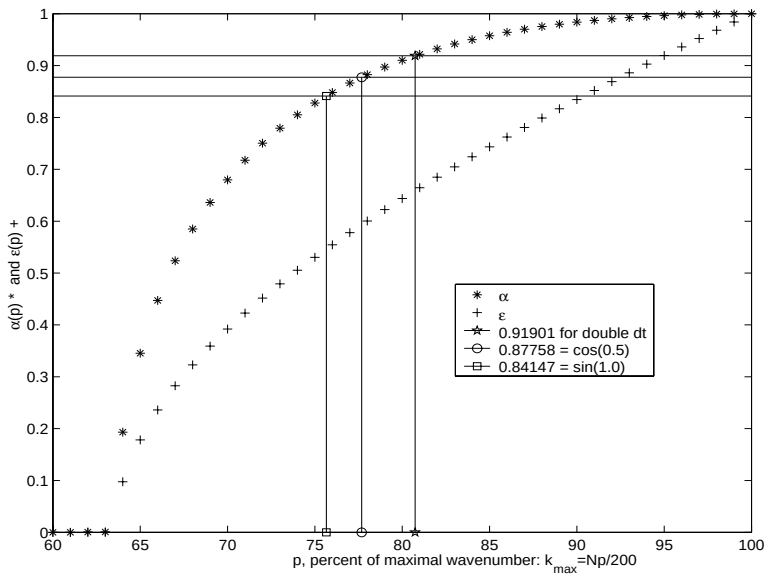

FIG. 2.3. α, (2.8), *and* ϵ, (2.5), *as a function of percent of modes resolved up to* $N/2$.

creases and the accuracy is reduced. On the other hand, while α rapidly increases from 0 to roughly .8, for resolution of waves up to roughly $.7N/2$, ϵ increases to 1 more gradually. Moreover, while the choices for α, $\sin(1.0)$ and $\cos(0.5)$, are close, roughly .84 and .88, respectively, the resulting small change in ϵ, also in the second decimal place, becomes significant for the accuracy, dependent on ϵ^N, as illustrated in Figure 2.4. For comparison a selection of values $\epsilon(p)^N$, $p = 75, 85, 95$, is also shown. We deduce that even single precision accuracy, $\epsilon^N \sim 10^{-6}$, cannot be achieved with a resolution of fewer than π points per wavelength ($p \approx 64$) unless $N > 30$ is taken. On the other hand, for larger but still marginally practical N, a small change in α is significant for the attainment of double precision; in particular, for $N = 60$, we find $\epsilon^N \sim 10^{-16}$, 10^{-14} for $\alpha = \sin(1.0)$ and $\cos(0.5)$, respectively. While these choices both yield spectral accuracy in the limit as $N \to \infty$, this limiting case is not relevant for any practical range of N.

In summary our results further complement the analyses presented in [7] and [11]. Specifically Don and Solomonoff [7] deduced that α must scale with N to provide spectral accuracy, while Hesthaven, Dinesen, and Lynov [11] deduced that a fixed choice of α is sufficient, specifically $\alpha = \cos(0.5)$. While these results are indeed valid, we have shown that the dependence of the accuracy on N for a fixed choice of α severely limits accuracy for small N and that for larger N the accuracy is very sensitive to small changes in α. Moreover, in considering the range of waves $\sin k\pi x$, $0 < k_{\max} < N/2$, which can be resolved on a grid, it is more appropriate to consider the choice of α from percentage, p, of waves up to the maximum $N/2$, $k_{\max} = pN/200$ that are resolved, rather than the number of modes, l, that are not resolved $k_{\max} = N/2 - l$.

2.2.3. Stability. Our final consideration in the choice of α is the size of the time step that might be expected in order to stably integrate a first order equation. Again, using the approach in [11], which is derived from that in [13], we note that

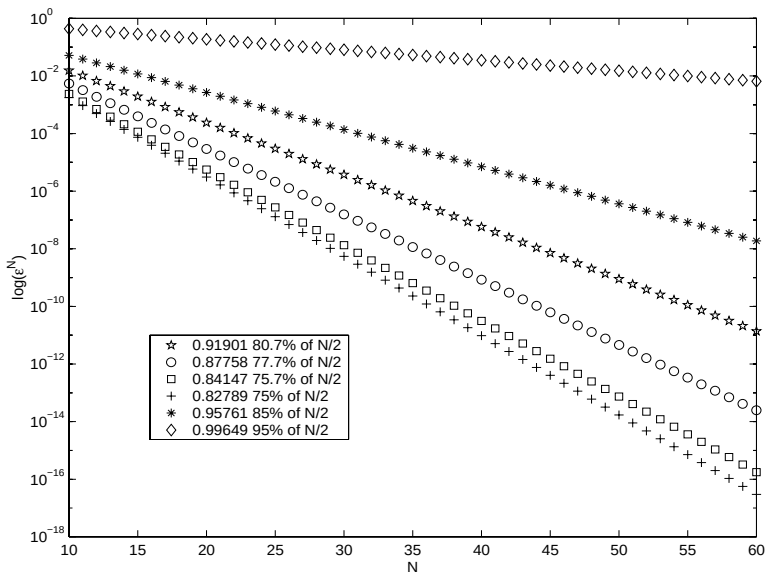

FIG. 2.4. ϵ^N, (2.5), *as a function of percent of modes resolved up to* $N/2$.

asymptotically with N the time step should scale inversely with the minimum grid size, such that

$$\Delta t_{\text{map}} = \frac{\alpha}{\sin^{-1}(\alpha)\sqrt{1-\alpha^2}}\Delta t_{\text{cheb}} = \beta(\alpha)\Delta t_{\text{cheb}},$$

where Δt_{map} and Δt_{cheb} are the time steps associated with the mapped and unmapped methods, respectively. Here it is the potential that $\beta(\alpha) \cong N$, near $\alpha = 1$, that makes the mapped method attractive. However, it is clear from Figure 2.5 that the maximal values of β are possible only when the choice for α compromises accuracy. Still, theoretically, a rough doubling of time step, particularly for a large simulation, would still be attractive. This is achieved for $\alpha = .91901 > \cos(0.5)$, the root of $\beta(\alpha) = 2$, for which accuracy $\epsilon^N \sim 10^{-10}$, for $N = 60$ is achievable, but for which accuracy does not become acceptable until N is at least 35. Hence by this analysis one must compromise either accuracy or the speed of integration if the mapped method is to be used to advantage over the unmapped method, unless a single large domain is used in the simulation. To confirm these estimates as a predictor for the maximal time step we also calculated the ratio $\rho(D)/\rho(AD)$, ρ the spectral radius, for a range of values for α and N; see Table 2.1. The results in Table 2.1 show that the theoretical ratio increases with N to a maximum value as predicted by $\beta(\alpha)$, illustrated in Figure 2.5. This maximum value is approached very quickly for small N and α, discouraging a move to larger N in order to gain relative increase in stable time step. Moreover, these results do not predict that the potential gains for increase in the time step are close to $O(N)$. We conclude that there is limited potential to significantly increase the stable time step for a given choice of N. While these results are negative regarding usage of the mapped method, Don and Solomonoff [7] included in their analysis the sensitivity of these eigenvalue calculations, for which the mapped method is far less sensitive. Hence the achievable improvement may be anticipated to

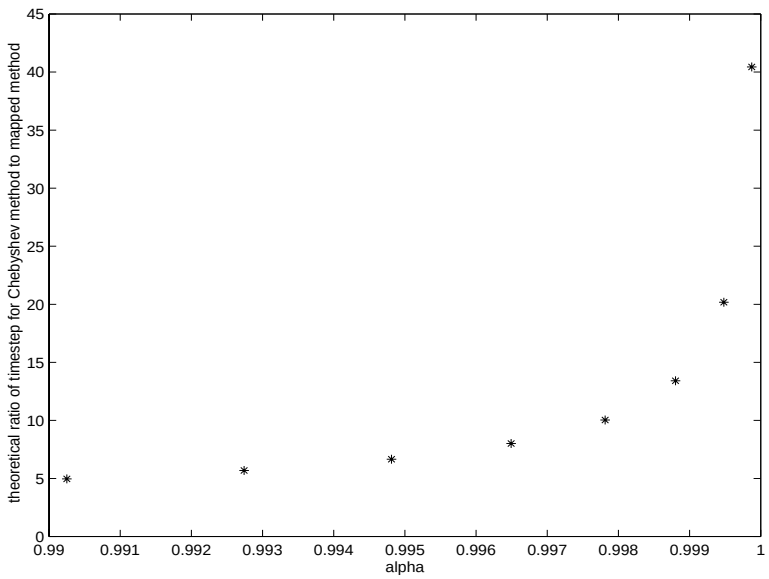

FIG. 2.5. *Theoretical scaling in timestep $\beta(\alpha)$ for optimal α illustrated in Figures* 2.3 *and* 2.4.

TABLE 2.1
Estimate of increase in time step for the mapped method from the ratio $\rho(D)/\rho(AD)$.

	alpha									
N	0.3090	0.3827	0.5000	0.7071	0.8415	0.8660	0.8776	0.9190	0.9900	1.0000
8	1.0003	0.9999	0.9980	0.9828	0.9508	0.9412	0.9361	0.9147	0.8630	0.8278
16	1.0240	1.0379	1.0689	1.1647	1.2747	1.2983	1.3082	1.3161	1.2231	1.1487
32	1.0314	1.0501	1.0933	1.2416	1.4682	1.5390	1.5782	1.7656	2.1737	2.0264
64	1.0335	1.0536	1.1002	1.2648	1.5329	1.6220	1.6727	1.9318	3.7901	3.8292
128	1.0341	1.0545	1.1020	1.2711	1.5510	1.6456	1.6999	1.9819	4.5522	7.4462
256	1.0342	1.0547	1.1025	1.2727	1.5558	1.6519	1.7070	1.9954	4.8118	14.6780
512	1.0342	1.0548	1.1026	1.2731	1.5570	1.6535	1.7088	1.9988	4.8847	29.1330
1024	1.0342	1.0548	1.1026	1.2732	1.5573	1.6539	1.7093	1.9997	4.9037	58.0315

be somewhat higher than these estimates because $\rho(D)$ typically underestimates the actual value of the spectral radius.

3. Numerical evaluation: First order equation.

3.1. Phase and amplitude errors. To verify the discussion in section 2 we investigate the accuracy of the modified method from the calculation of the phase and amplitude errors introduced in the solution of the one-dimensional wave equation. The approach follows the method of Kopriva [12] for the equivalent investigation of the Chebyshev method and assumes that the error introduced by time integration can be made negligible by taking a sufficiently small time step.

For the initial value problem

$$\begin{aligned} &u_t + u_x = 0, \quad -1 \le x \le 1, \quad 0 \le t < 8, \\ &u(x,0) = e^{ik\pi x}, \quad -1 \le x \le 1, \\ &u(-1,t) = e^{-ik\pi(1+t)}, \quad 0 \le t < 8, \end{aligned} \tag{3.1}$$

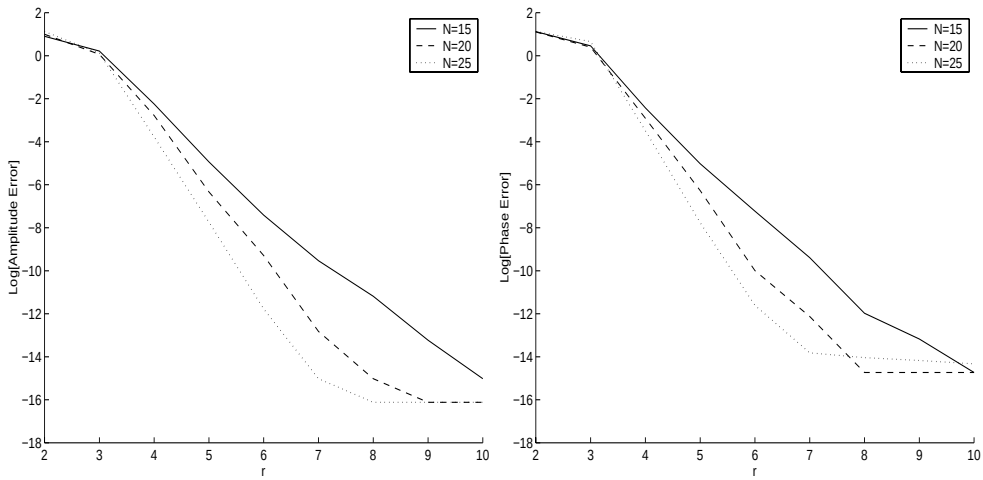

FIG. 3.1. *Phase and amplitude errors Chebyshev,* $\alpha_c = 0$.

the analytic solution is given by

$$u(x,t) = e^{ik\pi(x-t)},$$

for which the amplitude is one, the phase $k\pi(x-t)$, and the number of points per wavelength, for grid with $N+1$ points, is $r = N/k$. Then, assuming that the computed solution is of the form

$$\hat{u}(x_j,t) = e^{i\theta_j}, \qquad \theta_j = a_j + ib_j,$$

the amplitude error is $1 - e^{-b_j}$ and the phase error $k\pi(x_j - t) - a_j$. Kopriva [12] used this calculation for the Chebyshev method to demonstrate that these errors decrease exponentially as the number of points per wavelength, r, increases. This is illustrated in Figure 3.1.

3.2. Implementation. In our implementation the time integration is accomplished using the standard fourth order four stage Runge–Kutta (RK) solver. To maintain the formal accuracy in time, the time dependent boundary condition, $u(-1,t) = e^{-ik\pi(1+t)} = h(t)$, is imposed at each stage of the RK integration using the method in [6]. Specifically, the boundary condition and its derivatives are used to obtain the intermediate stage values, $u^{(m)}(-1,t)$, $m = 1:3$, as

$$\begin{aligned} u^{(1)}(-1,t) &= h(t) + \frac{\delta t}{2}h'(t), \\ u^{(2)}(-1,t) &= u^{(1)}(-1,t) + \frac{\delta t^2}{4}h''(t), \\ u^{(3)}(-1,t) &= h(t) + \delta t h'(t) + \frac{\delta t^2}{2}h''(t) + \frac{\delta t^3}{4}h'''(t). \end{aligned} \tag{3.2}$$

The calculation of entries of the derivative operator, D, can be accomplished via a number of different algorithms; see, e.g., [2, 7, 8, 19]. In our implementation Fornberg's algorithm, [8], was used. While this operator can be used to evaluate the derivative at the grid points using a matrix-vector product, the derivative may also be

TABLE 3.1
Relative error in the solution of (3.1) *with* 4 *points per wavelength.*

	Matrix-vector		FFT	
N	Chebyshev	$\alpha_s = \sin(1.0)$	Chebyshev	$\alpha_s = \sin(1.0)$
8	3.006480E−01	5.502124E−02	3.006480E−01	5.502132E−02
16	1.007892E−01	9.352306E−04	1.007892E−01	9.352304E−04
32	1.511967E−02	1.473285E−06	1.511970E−02	1.516212E−06

TABLE 3.2
Relative error in the solution of (3.1) *with* 8 *points per wavelength.*

	Matrix-vector		FFT	
N	Chebyshev	$\alpha_s = \sin(1.0)$	Chebyshev	$\alpha_s = \sin(1.0)$
8	2.005779E−03	4.793553E−03	2.005823E−03	4.793477E−03
16	6.898087E−06	7.127883E−05	6.905703E−06	7.127340E−05
32	1.984337E−07	1.176971E−07	4.119758E−08	3.412978E−08

obtained using the discrete fast Fourier transform; see [4]. This approach immediately extends also for the mapped method. Specifically, suppose that $u(y(x))$ is given by

$$u(y_j) = \sum_{k=0}^{N} \bar{a}_k T_k(x_j), \qquad j = 0 : N,$$

where

$$\bar{a}_k = \frac{1}{\gamma_k} \sum_{j=0}^{N} u(y_j) T_k(x_j) w_j,$$

with

$$w_j = \begin{cases} \pi/2N, & j = 0, N, \\ \pi/N, & 1 \le j \le N-1, \end{cases} \qquad \gamma_k = \begin{cases} \frac{\pi}{2} c_k, & k < N, \\ \pi, & k = N, \end{cases} \quad \text{and} \quad c_k = \begin{cases} 2k = 0, \\ 1 \quad \text{otherwise.} \end{cases}$$

Then the derivative values are given by

$$u'(y_j) = \frac{1}{g'(x_j)} \sum_{k=0}^{N} \bar{b}_k T_k(x_j),$$

where the $\bar{b}_k$ are defined by the recurrence

$$c_k \bar{b}_k = \bar{b}_{k+2} + 2(k+1)\bar{a}_k, \quad 0 \le k \le N-1,$$

equivalent to the recurrence for the series coefficients a_k for $u(x)$.

In all numerical tests the conclusions are valid regardless of the method for the implementation of the derivative, matrix-vector (MV) or fast Fourier transform (FFT). This is illustrated in Tables 3.1 and 3.2 for two choices, the Chebyshev and the case with $\alpha_s = \sin(1.0)$, for N up to $N = 32$. Results when $r = 4$ are identical up to 6 decimal places and up to 5 decimal places for $r = 8$. For larger, but not practical, N, one should anticipate reduced accuracy of the MV method; see Canuto et al. [4]. Hence, for practical choices, the method which can be implemented most efficiently for the given architecture may be chosen. While Canuto et al. [4] demonstrated that the MV method is most likely most efficient for practical choices of N, this may not always be the case if a well-tuned FFT routine is available.

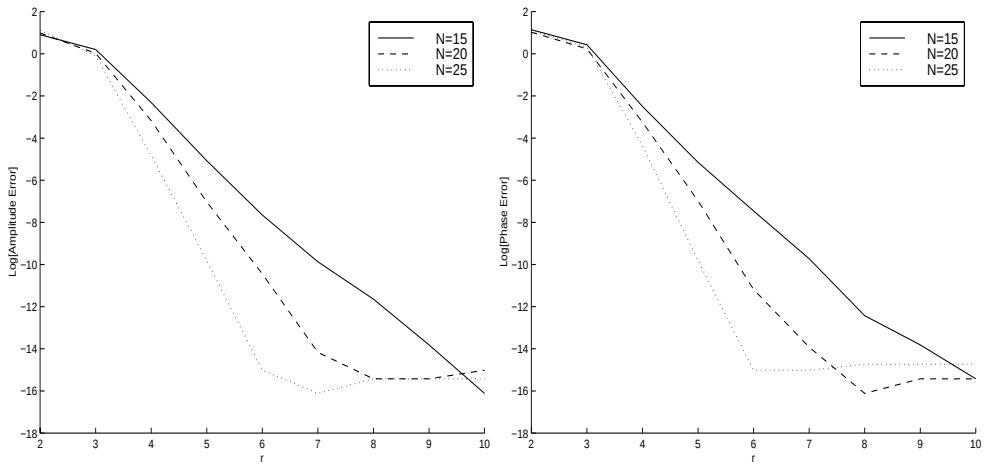

FIG. 3.2. *Phase and amplitude errors for high accuracy using α_a.*

3.3. Numerical experiments. Numerical tests were designed to evaluate the spectral accuracy of the mapped Chebyshev method with certain choices of the parameter α, as compared to the Chebyshev method, for realistic practical choices of N. To evaluate the impact of the choice for α, the following choices were made:

(i) α chosen for maximal theoretical accuracy, i.e., α_a is chosen according to (2.6), with $\epsilon \approx 2^{-53}$.

(ii) α chosen to show sensitivity due to small change in α as compared to α_h:

$$\alpha_s = \sin(1.0) \simeq .84. \tag{3.3}$$

(iii) α chosen for maximal theoretical resolution, where $\alpha \simeq 1$ is assumed, as in [13]:

$$\alpha_r = \sin\left(\frac{\pi}{r}\right). \tag{3.4}$$

(iv) α chosen to study impact of fixed choice for α near 1, for which boundary conditions will become Dirichlet in limit $\alpha = 1$:

$$\alpha_{db} = .99. \tag{3.5}$$

(v) α chosen as per Hesthaven, Dinesen, and Lynov [11]:

$$\alpha_h = \cos(0.5) \simeq .88.$$

These choices, and the Chebyshev method $\alpha_c = 0$, were evaluated for a range of values for N and the number of points per wavelength. Finally, we evaluate the number of points per wavelength required to achieve minimal relative error for these methods.

3.4. Numerical results. The results of the experiments are illustrated in Figures 3.2–3.7. It is immediately confirmed from Figure 3.1 that the Chebyshev method is not accurate for less than 3 points per wavelength. Moreover, the phase error dominates the amplitude error for small r, but at roughly 5 points per wavelength the errors are comparable.

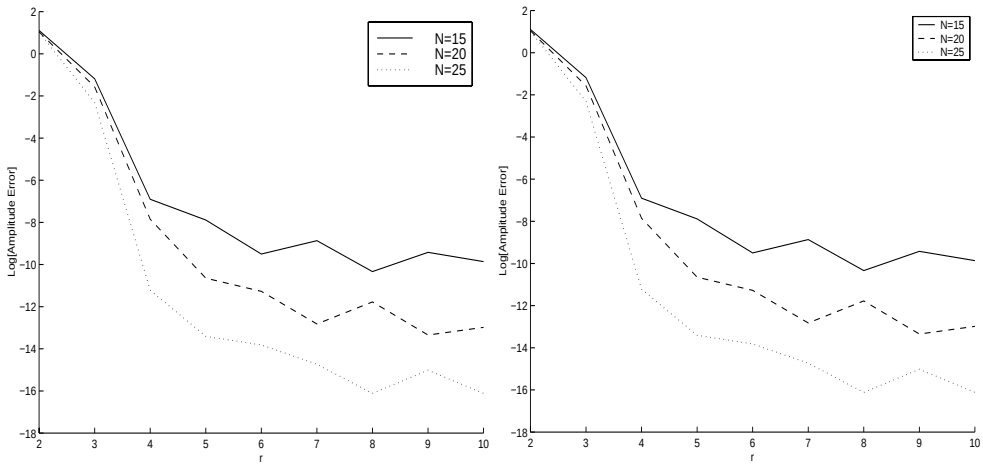

FIG. 3.3. *Phase and amplitude errors for* $\alpha_s = \sin(1.0)$.

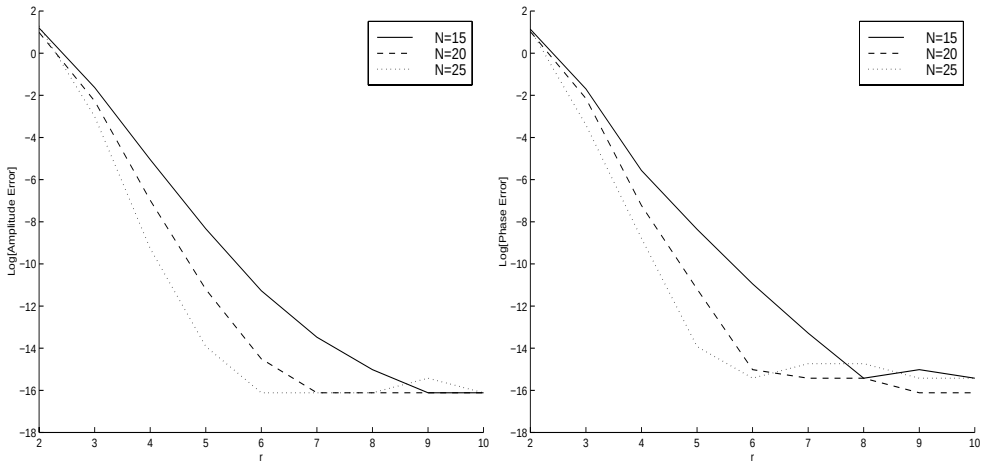

FIG. 3.4. *Phase and amplitude errors for resolution,* $\alpha_r = \sin(\pi/r)$.

While the modest increases in N yield only modest improvements in accuracy, the impact is more pronounced as r increases, until maximum achievable accuracy in double precision is attained. In particular, for $N = 25$ all modes with $r \gtrsim 7$ are resolved equally well in terms of both phase and amplitude accuracy. Not surprisingly, the high accuracy choice for α, α_a (see Figure 3.2), for which α_a is relatively small, .171460, .310752, and .436969 for $N = 15, 20, 25$, respectively, yields conclusions similar to that for the $\alpha_c = 0$ case. With $N = 25$ all nodes with $r \gtrsim 6$ yield equivalent accuracy in phase and amplitude, while with $N = 15$ the maximal accuracy is still not achieved until $r = 10$. The positive impact of picking α to provide improved resolution, e.g., $\alpha_r = 1$, .587785, .309017, for $r = 2, 5, 10$, respectively, particularly for small r, is clearly seen in Figure 3.4, but the impact as r increases is not significant. Note the large errors for small r and large α, which decrease substantially as r increases,

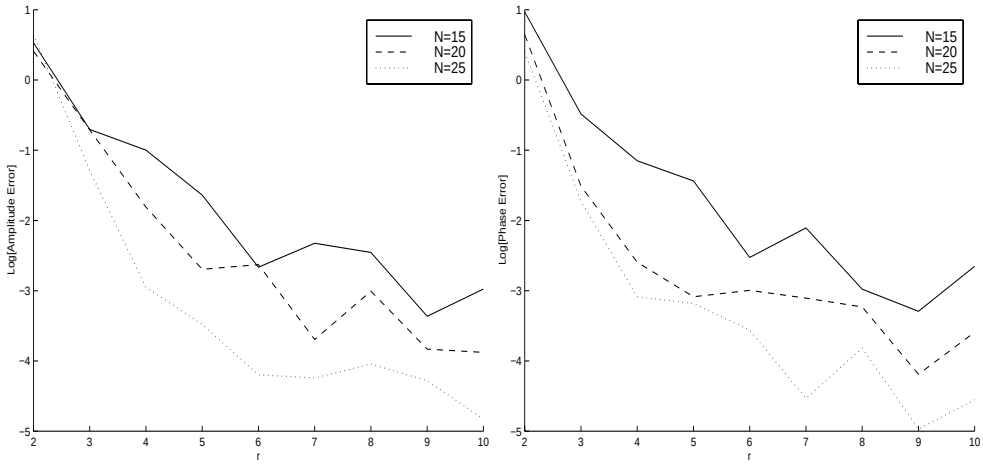

FIG. 3.5. *Phase and amplitude errors for* $\alpha_{db} = .99$.

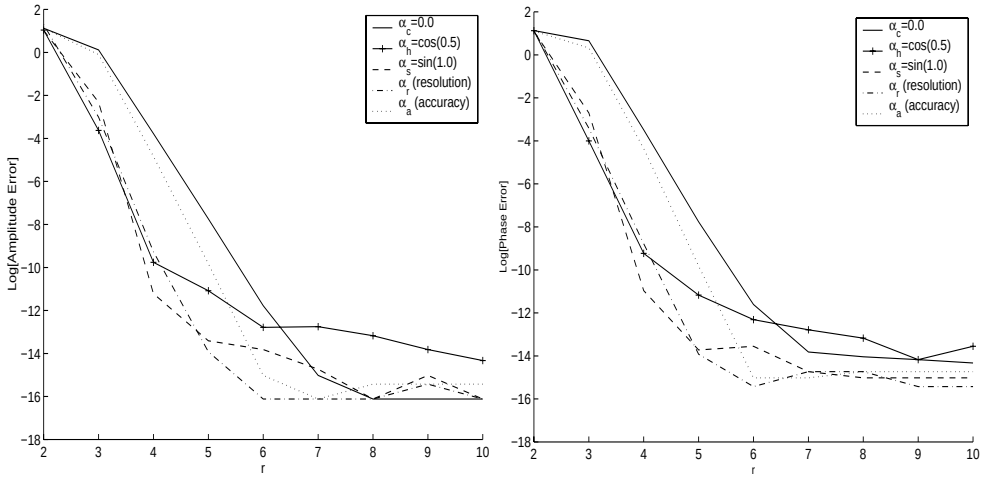

FIG. 3.6. *Comparison of phase and amplitude errors in first order derivative approximation for different* α *and fixed* $N = 25$.

and hence α decreases. On the other hand, the very limited achievable accuracy for $\alpha_{db} = .99$ is apparent in Figure 3.5. Still, fixing α but increasing N, corresponding to increasing the maximal resolution for this α, improves the results.

In Figure 3.6 we compare the previous results, ignoring the worst case, $\alpha_{db} = .99$, and introducing the comparison of the choices α_h and α_s. The improved accuracy of α_s over α_h confirms the theoretical study in section 2. Moreover, the mapped methods outperform the Chebyshev method, not only for small r, hence confirming the resolution analysis in section 2, but also at larger r showing the gains that are realized by emphasizing the theoretical ability of the mapped method to offer accuracy for fewer than π points per wavelength. Moreover, while these results suggest that the choice α_r is optimal, it must also be noted that these results pick α_r to match the ratio N/k to the given initial condition. Thus these results for α_r are falsely

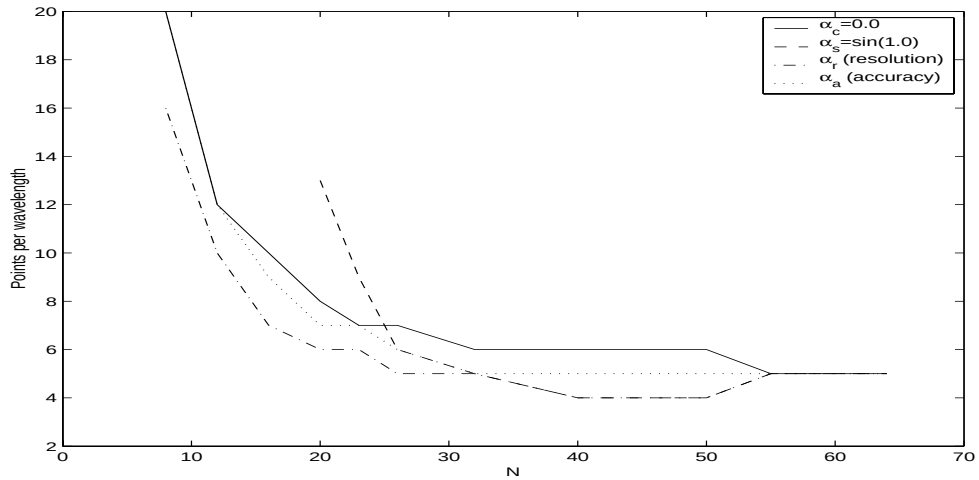

FIG. 3.7. *For given N, r needed to achieve relative error less than* 10^{-6}.

optimistic. However, almost comparable accuracy is achievable over a range of initial conditions for α_s, while the accuracy choice α_a is too severe.

Finally, in Figure 3.7 we illustrate the number of points per wavelength that are needed, for choices of α, as a function of N. We conclude that single precision accuracy requires in all cases tested up to $N = 32$, a minimum of 5 points per wavelength. Moreover, the mapped method marginally outperforms the Chebyshev method for N small, $N \lesssim 30$, provided α is far from 1.0. Additionally, it is not possible to achieve single precision accuracy with the fixed choice $\alpha_s = \sin(1.0)$ until N approaches 20. Similarly, single precision accuracy cannot be achieved if $\alpha_{db} = 0.99$ and $N \lesssim 60$.

4. Numerical evaluation: Second order equation. In order to fully evaluate the impact of the modified method on accuracy we also considered the one-dimensional two way wave equation,

$$u_{tt} = u_{xx}, \quad -1 \le x \le 1, \quad 0 \le t < 8, \tag{4.1}$$

$$\left.\begin{aligned} u(x,0) &= e^{ik\pi x} \\ u_t(x,0) &= -ik\pi e^{ik\pi x} \end{aligned}\right\}, \qquad -1 \le x \le 1, \tag{4.2}$$

$$\left.\begin{aligned} u(-1,t) &= e^{-ik\pi(1+t)} \\ u(1,t) &= e^{ik\pi(1-t)} \end{aligned}\right\}, \qquad 0 \le t < 8, \tag{4.3}$$

for which $u(x,t) = e^{ik\pi(x-t)}$. For implementation we extended the first order implementation described in section 3. We note that the second order derivative can again be calculated by either MV multiply or the FFT. In order to use the RK method we reformulate the second order equation as a system of first order equations in $v = [u, w]^T$, $w = u_t$. Formal accuracy of the RK method is imposed by applying the method in [6], (4.1), for both boundaries, and to both components of $v(x,t)$.

In Tables 4.1 and 4.2, we evaluate the accuracy on the implementation of the derivative operation for the second order wave equation. While the results again agree, this time to 5 decimal places, the FFT is in this case more reliable, confirming the discussion of Don and Solomonoff [7] on the sensitivity of the calculation of the higher derivative operators, particularly with large N and for the mapped method.

TABLE 4.1
Relative error in the solution of (4.1)–(4.3) *with* 4 *points per wavelength.*

	Matrix		FFT	
N	Chebyshev	$\alpha_s = \sin(1.0)$	Chebyshev	$\alpha_s = \sin(1.0)$
8	7.293058E−01	3.768539E−02	7.293058E−01	3.768472E−02
16	1.357839E−01	2.859397E−04	1.357852E−01	2.852290E−04
32	5.458733E−03	7.434244E−07	5.459800E−03	3.233326E−07

TABLE 4.2
Relative error in the solution of (4.1)–(4.3) *with* 8 *points per wavelength.*

	Matrix		FFT	
N	Chebyshev	$\alpha = \sin(1.0)$	Chebyshev	$\alpha = \sin(1.0)$
8	6.585752E−04	1.403180E−03	6.596680E−04	1.401453E−03
16	2.239754E−06	1.833787E−05	9.982809E−07	1.998241E−05
32	2.100210E−06	1.201101E−06	1.455820E−07	3.843321E−08

Moreover, with 4 points per wavelength the mapped method using $\alpha_s = \sin(1.0)$ outperforms the Chebyshev method, but for $r = 8$ the benefit is not realized for small N.

A comparison of the accuracy for different choices of α similar to that done in the first derivative approximation was done. Individual results for each choice of α yield similar behavior as for the first order equation. In particular, for fixed α with N, $\alpha_s = \sin(1.0)$ provides comparable accuracy for $r \geq 4$, but errors decrease exponentially when α is chosen either for resolution, α dependent on r not N, or for accuracy, α dependent on N and not r. There is again an advantage to use of the mapped method for a small number of grid points per wavelength.

5. Conclusions. Don and Solomonoff [7] have shown that α must be scaled with N to provide spectral accuracy, while Hesthaven, Dinesen, and Lynov [11] have shown that spectral accuracy can be achieved with N fixed. In this work, we calculated the phase and amplitude errors in the modified method and illustrate that for small $N \lesssim 30$, high accuracy is achievable only if α is small. The cases with small α (namely when α is chosen to optimize either accuracy, α_a, or resolution, α_r) illustrate similar properties as the Chebyshev case, namely that the phase and amplitude errors decay exponentially as the number of points per wavelength increases.

There are two main conclusions of this work. First, the $O(N)$ increase in the stable time step for the mapped method cannot be obtained for realistic N. One of the apparent major benefits of choosing $\alpha \neq 0$ is that the time step for stable solutions may be increased by as much as N. Theoretically, however, if we hope to double the time step for stable implementations and still achieve high accuracy, N needs to be large; for example, for accuracy 10^{-10}, N must exceed 60.

Second, the theoretical estimates indicate that a choice of $\alpha \neq 0$ provides good resolution with fewer than π points per wavelength. Since the Chebyshev method requires at least π points per wavelength for resolution, more accurate solutions can be found by the modified method under the right circumstances, as described in the next paragraph.

The choice of α for a given N depends on the accuracy required, and we have shown that the accuracy for a given N, N relatively large, $N > 30$, is very sensitive to small changes in α. In addition, it is not always possible to reach single precision accuracy with the modified method. For example, the choice of $\alpha_s = \sin(1.0)$ must

be accompanied by $N > 20$ for single precision accuracy; see Figure 3.7. However, if N is large enough, or α chosen small enough, the modified method can be more accurate than the Chebyshev method even when more than the minimum number of points per wavelength for resolution are used; see Table 3.1 and Figure 3.6, where the same number of points per wavelength are used for both methods. In Table 3.1, 4 points per wavelength were used, while in Figure 3.6 the mapped method has smaller phase and amplitude errors when up to 8 points per wavelength are used. Of course, the Chebyshev method is more accurate when there are a large number of points per wavelength, which for a wide range of wavelengths requires large N, but with 8 points per wavelength (Table 3.2) both the mapped and unmapped methods have errors with the same order of magnitude. With 4 points per wavelength (Table 3.1) the error of the mapped method is up to four orders of magnitude smaller than that of the unmapped Chebyshev method. The better accuracy is a result of the fact that $\alpha \neq 0$ requires fewer points per wavelength for resolution than $\alpha = 0$. It is significant that the mapped method is able to improve accuracy with a smaller sampling rate because this is crucial in keeping the size of N within a reasonable range computationally, both with regards to memory and computation requirements, the latter through the size of time step that can be used for the time integration.

REFERENCES

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1972.

[2] A. Bayliss, A. Class, and B. Matkowsky, *Roundoff error in computing derivatives using the Chebyshev differentiation matrix*, J. Comput. Phys., 116 (1994), pp. 380–383.

[3] A. Bayliss and A. Turkel, *Mappings and accuracy for Chebyshev pseudo-spectral approximations*, J. Comput. Phys., 101 (1992), pp. 349–359.

[4] C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang, *Spectral Methods in Fluid Dynamics*, Springer-Verlag, New York, 1988.

[5] M.H. Carpenter and D. Gottlieb, *Spectral methods on arbitrary grids*, J. Comput. Phys., 129 (1996), pp. 74–86.

[6] M.H. Carpenter, D. Gottlieb, S. Abarbanel, and W.-S. Don, *The theoretical accuracy of Runge–Kutta time discretizations for the initial boundary value problem: A study of the boundary error*, SIAM J. Sci. Comput., 16 (1995), pp. 1241–1252.

[7] W.S. Don and A. Solomonoff, *Accuracy enhancement for higher derivatives using Chebyshev collocation and a mapping technique*, SIAM J. Sci. Comput., 18 (1997), pp. 1040–1055.

[8] B. Fornberg, *A Practical Guide to Pseudospectral Methods*, 1st ed., Cambridge University Press, New York, 1996.

[9] D. Gottlieb and S.A. Orszag, *Numerical analysis of spectral methods: Theory and applications*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 26, SIAM, Philadelphia, 1977.

[10] J.C. Hardin, J.R. Ristorcelli, and C.K.W. Tam, eds., *ICASE/LaRC Workshop on Benchmark Problems in Computational Aeroacoustics*, NASA CP 3300, Hampton, VA, 1995.

[11] J.S. Hesthaven, P.G. Dinesen, and J.P. Lynov, *Spectral collocation time-domain modelling of diffractive optical elements*, J. Comput. Phys., 155 (1999), pp. 287–306.

[12] D.A. Kopriva, *Spectral solution of acoustic wave propagation problems*, in Proceedings of the AIAA 13th Aeroacoustics Conference, 1990.

[13] D. Kosloff and H. Tal-Ezer, *A modified Chebyshev pseudospectral method with an $O(N^{-1})$ time step restriction*, J. Comput. Phys., 104 (1993), pp. 457–469.

[14] J.L. Mead, *Numerical Methods for Problems in Computational Aeroacoustics*, Ph.D. thesis, Arizona State University, Tempe, AZ, 1998.

[15] R. Renaut and J. Frohlich, *A pseudospectral Chebychev method for the 2D wave equation with domain stretching and absorbing boundary conditions*, J. Comput. Phys., 124 (1996), pp. 324–336.

[16] R.A. Renaut and Y. Su, *Evaluation of Chebychev pseudospectral methods for third order differential equations*, Numer. Algorithms, 16 (1997), pp. 255–281.

[17] R. Renaut, *Stability of a Chebychev pseudospectral solution of the wave equation with absorbing boundaries*, J. Comput. Appl. Math., 87 (1997), pp. 243–259.

[18] C.K.W. Tam, *Computational aeroacoustics: Issues and methods*, AIAA J., 33 (1995), pp. 1788–1795.

[19] J.A.C. Weideman and S.C. Reddy, *A Matlab differentiation matrix suite*, ACM Trans. Math. Software, 26 (2000), pp. 465–519.

[20] J.A.C. Weideman and L.N. Trefethen, *The eigenvalues of second-order spectral differentiation matrices*, SIAM J. Numer. Anal., 25 (1988), pp. 1279–1298.

[21] B.D. Welfert, *Generation of pseudospectral differentiation matrices* I, SIAM J. Numer. Anal., 34 (1997), pp. 1640–1657.