Wave Propagation in Quadratic-Finite-Element Approximations to Hyperbolic Equations *

Dale R. Durran †
University of Washington, USA

October 8, 1999

SUMMARY
Eigenmodes for the quadratic-finite-element method are expressed as a linear combination of two conventional semi-discrete Fourier modes. Each of these Fourier modes moves at a different phase speed, but both modes have the same group velocity. This representation of the QFEM eigenmodes clarifies the significance of the negative phase speeds that naturally arise as part of the conventional analysis.

KEYWORDS: 65P25 Finite Elements   76M10 Finite Element Methods

1 Introduction

Finite-element approximations to hyperbolic partial differential equations usually employ piecewise linear elements. The use of piecewise quadratics and higher-order elements has been discouraged by their algorithmic complexity and by confusion about the susceptibility of such methods to errors in the smallest scales resolvable on the numerical mesh. This note attempts to clarify some of the confusion involving the wave propagation characteristics of solutions to semi-discrete quadratic-finite-element approximations to the one-way wave equation

$$\frac{\partial \psi}{\partial t} + c\frac{\partial \psi}{\partial x} = 0 \tag{1}$$

on the periodic domain $x \in [0, L]$.

As in [1], let the x-dependence of the solution to (1) be approximated by quadratic Lagrange interpolating polynomials such that

$$\phi(x,t) = \sum_{n=0}^{N} a_{2n}(t)\varphi_{2n}^{e} \ + \sum_{n=1}^{N} b_{2n-1}(t)\varphi_{2n-1}^{m}(x), \tag{2}$$

*To be submitted to the *Journal of Computational Physics*
†Corresponding author: University of Washington, Atmospheric Sciences, Box 351640, Seattle, WA 98195-1640, USA.

Here φ_j^e is an "endpoint" expansion function defined as

$$\varphi_j^e(x) = \begin{cases} 1 - 3\dfrac{|x-x_j|}{2\Delta x} + \dfrac{1}{2}\left(\dfrac{x-x_j}{\Delta x}\right)^2, & \text{if } |x-x_j| \le 2\Delta x, \\ 0, & \text{otherwise,} \end{cases}$$

φ_j^m is the "midpoint" expansion function

$$\varphi_j^m(x) = \begin{cases} \left(\dfrac{x-x_{j-1}}{\Delta x}\right)\left(2 - \dfrac{x-x_{j-1}}{\Delta x}\right), & \text{if } |x-x_j| \le \Delta x, \\ 0, & \text{otherwise,} \end{cases}$$

and $x_j = j\Delta x$ for $j = 0, 1, \ldots, 2N$. The standard Galerkin finite-element approximation to (1) yields two families of equations for the evolution of the expansion coefficients. The equations centered at the endpoint nodes are

$$\frac{-\dot{a}_{j-2} + 2\dot{b}_{j-1} + 8\dot{a}_j + 2\dot{b}_{j+1} - \dot{a}_{j+2}}{10} + c\left(\frac{b_{j+1} - b_{j-1}}{\Delta x} - \frac{a_{j+2} - a_{j-2}}{4\Delta x}\right) = 0, \tag{3}$$

whereas those centered on the midpoints are

$$\frac{\dot{a}_{\ell-1} + 8\dot{b}_\ell + \dot{a}_{\ell+1}}{10} + c\left(\frac{a_{\ell+1} - a_{\ell-1}}{2\Delta x}\right) = 0. \tag{4}$$

As is widely appreciated, no single semi-discrete Fourier mode of the form $\phi_j(t) = e^{i(kj\Delta x - \omega t)}$ will simultaneously satisfy (3) and (4). The solution for the nodal values may, however, be expressed in the form

$$a_j(t) = e^{ik(j\Delta x - c^* t)} \qquad b_\ell(t) = r_a e^{ik(\ell\Delta x - c^* t)}, \tag{5}$$

where r_a is the ratio of the amplitude at a midpoint node to the amplitude at an endpoint node and, loosely speaking, c^* is a phase-speed [1, 2, 3, 4]. Substitution of (5) into (3) and (4) yields

$$\frac{1}{5}\left(8 - 2\cos 2\theta + 4r_a\cos\theta\right)c^* - \frac{c}{\theta}\left(4r_a\sin\theta - \sin 2\theta\right) = 0.$$

and

$$\frac{1}{5}\left(\cos\theta + 4r_a\right)c^* - \frac{c}{\theta}\sin\theta = 0,$$

where $\theta = k\Delta x$. Solutions to this system of two equations in the two unknowns c^* and r_a may be expressed in the form

$$\frac{c^*}{c} = \frac{\sin\theta}{\theta\left(1+\sin^2\theta\right)}\left(-2\cos\theta \pm \left(9 + \sin^2\theta\right)^{1/2}\right) \tag{6}$$

and

$$r_a = \frac{1}{4}\left(5\frac{c}{c^*}\frac{\sin\theta}{\theta} - \cos\theta\right). \tag{7}$$

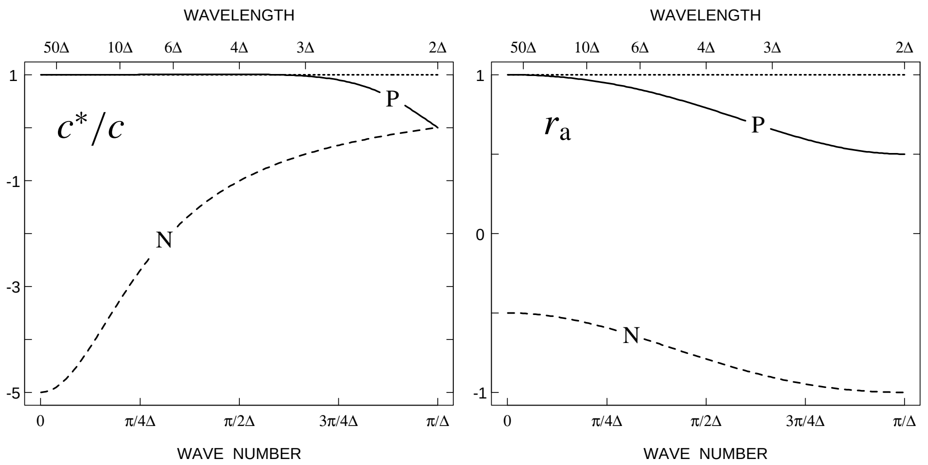

Figure 1: Normalized phase speed, $c_{\mathrm{p}}^*(k)/c$, and amplitude ratio, r_{a}, plotted as a function of horizontal wavenumber for the positive (P) and negative (N) roots of (6).

The relative phase speed error, c^*/c, and the corresponding values of r_{a} are plotted as a function of horizontal wavenumber in Fig. 1.

Two difficulties arise in connection with the interpretation of these results. The first, which mislead this author [1] as well as previous researchers [5, 6], is the apparent presence of two distinct sets of solutions: "physical" and "computational" modes associated, respectively, with the positive and negative roots in (6). A more correct description was provided by Gresho and Lee [3] (see also [4]), who recognized that only a single set of modes was required to span the solution space defined by the linear system of ordinary differential equations (3)–(4). Unfortunately, [3, 4] also incorrectly suggest that those modes associated with the negative root can be dismissed as extraneous. An essentially correct interpretation was offered by Cullen [2], who recognized that none of the solutions are extraneous, because the set of solutions given by the positive root in (6) is identical to the set of solutions given by the negative root.

As a preliminary step to the subsequent analysis, we derive the relation between the two sets of solutions stated in [2, 3]. Let those quantities related to the positive root in (6) be denoted with a subscript "p" and those related to the negative root be denoted by a subscript "n." Also let $\omega^* = kc^*$ be the frequency of the quadratic-finite-element method (QFEM) solution. Elementary trigonometric identities imply that for $k \in [0, \pi/\Delta x]$, i.e., for the positive wave numbers resolvable on the numerical grid,

$$\omega_{\mathrm{p}}^*(k) = \omega_{\mathrm{n}}^*(k - \pi/\Delta x) \quad \text{and} \quad r_{\mathrm{a_p}}(k) = -r_{\mathrm{a_n}}(k - \pi/\Delta x), \tag{8}$$

and for $k \in [-\pi/\Delta x, 0]$

$$\omega_{\mathrm{p}}^*(k) = \omega_{\mathrm{n}}^*(k + \pi/\Delta x) \quad \text{and} \quad r_{\mathrm{a_p}}(k) = -r_{\mathrm{a_n}}(k + \pi/\Delta x). \tag{9}$$

Now consider a negative-root mode with endpoint values a_j and midpoint values

b_ℓ. Consistent with (2), the indices j are even and ℓ are odd. Thus, using (8) and (9)

$$a_j = e^{i\left[(k-\pi/\Delta x)j\Delta x-\omega_{\mathrm{n}}^*(k-\pi/\Delta x)t\right]} = (-1)^j e^{ikj\Delta x} e^{-i\omega_{\mathrm{p}}^*(k)t} = e^{i\left[kj\Delta x-\omega_{\mathrm{p}}^*(k)t\right]},$$

and

$$\begin{aligned} b_\ell &= r_{\mathrm{a_n}}(k-\pi/\Delta x)e^{i\left[(k-\pi/\Delta x)\ell\Delta x-\omega_{\mathrm{n}}^*(k-\pi/\Delta x)t\right]} \\ &= -r_{\mathrm{a_p}}(k)(-1)^\ell e^{ik\ell\Delta x} e^{-i\omega^*{}_{\mathrm{p}}(k)t} \\ &= r_{\mathrm{a_p}}(k)e^{i\left[k\ell\Delta x-\omega^*{}_{\mathrm{p}}(k)t\right]}, \end{aligned}$$

which demonstrates that for $k \in [0, \pi/\Delta x]$ the negative-root mode with wavenumber $k - \pi/\Delta x$ is identical to the positive-root mode with wavenumber k. Similarly, for $k \in [-\pi/\Delta x, 0]$, the negative-root mode with wavenumber $k + \pi/\Delta x$ is identical to the positive-root mode with wavenumber k.

We now arrive at the second difficulty in interpreting our results: how can both of the phase speeds given by the two curves in Fig. 1 simultaneously describe the behavior of individual solutions to (3)–(4)? The answer lies in the fact that solutions of the form (5) are not conventional semi-discrete Fourier modes that propagate without changing shape between each pair of adjacent nodes. Instead, (5) is a semi-discrete approximation to a function of the form $g(x)e^{i(kx-\omega t)}$, where the factor $g(x)$ is introduced to account for the extra spatial dependence associated with the midpoint-node coefficient r_{a}.

The reason that two nonspurious phase speeds are obtained from (6) can be made more apparent by re-expressing the QFEM eigenmodes as the sum of two *conventional* semi-discrete Fourier modes traveling at different speeds and moving in opposite directions. As a first step, (5) is written in the alternative form

$$\hat{q}_n(t) = \tfrac{1}{2}\left[(1+r_{\mathrm{a}}) + (-1)^n(1-r_{\mathrm{a}})\right] e^{i[kn\Delta x-\omega^*(k)t]}, \qquad n = 0, 1, \ldots 2N.$$

Even values of n give the nodal values a_j; odd values yield the b_ℓ. Then for a positive-root eigenmode,

$$\begin{aligned} \hat{q}_n(t) &= \left(\frac{1+r_{\mathrm{a_p}}}{2}\right) e^{i\left[kn\Delta x-\omega_{\mathrm{p}}^*(k)t\right]} + \left(\frac{1-r_{\mathrm{a_p}}}{2}\right) e^{i\left[(k-\pi/\Delta x)n\Delta x-\omega_{\mathrm{p}}^*(k)t\right]} \\ &= \left(\frac{1+r_{\mathrm{a_p}}}{2}\right)\left[e^{i\left[kn\Delta x-\omega_{\mathrm{p}}^*(k)t\right]} + \left(\frac{1-r_{\mathrm{a_p}}}{1+r_{\mathrm{a_p}}}\right) e^{i\left[(k-\pi/\Delta x)n\Delta x-\omega_{\mathrm{n}}^*(k-\pi/\Delta x)t\right]}\right], \end{aligned}$$

where (8) is used to obtain the last equality and without loss of generality it is assumed that $k \in [0, \pi/\Delta x]$. Defining $q = 2\hat{q}/(1 + r_{\mathrm{a_p}})$ and $\beta = (1 - r_{\mathrm{a_p}})/(1 + r_{\mathrm{a_p}})$, the preceding may be expressed in the form

$$q_n(t) = e^{ik\left[n\Delta x-c_{\mathrm{p}}^*(k)t\right]} + \beta e^{i(k-\pi/\Delta x)\left[n\Delta x-c_{\mathrm{n}}^*(k-\pi/\Delta x)t\right]}. \tag{10}$$

Here $c_{\mathrm{p}}^*(k)$ and $c_{\mathrm{n}}^*(k-\pi/\Delta x)$ are given, respectively, by the positive and negative roots in (6), and the value of $r_{\mathrm{a_p}}$ used to evaluate β is obtained by substituting $c_{\mathrm{p}}^*(k)/c$ into (7). If $-\pi/\Delta x \le k \le 0$, the preceding expression for q_n is modified by replacing $k - \pi/\Delta x$ with $k + \pi/\Delta x$.

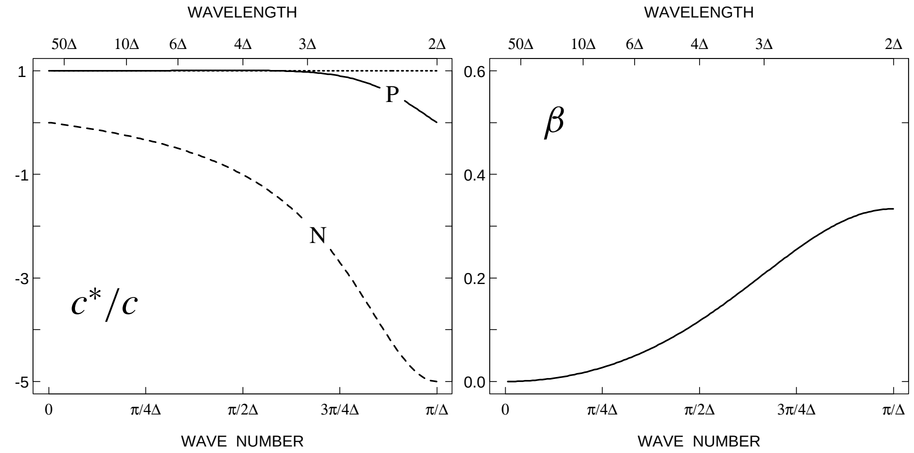

Figure 2: Normalized phase speeds, $c_\mathrm{p}^*(k)/c$, for the positive-root (physical) component (P) and $c_\mathrm{n}^*(k - \pi/\Delta x)/c$ for the negative-root (nonphysical) component (N) of a quadratic finite-element approximation to the constant-wind-speed advection equation plotted as a function of the wavenumber of the correct solution (k). Also shown is β, the ratio of the amplitude of the negative-root component to the amplitude of the positive-root component.

According to (10), *the eigenmodes of the QFEM approximation to* (1) *are the superposition of a wave moving downstream at speed c_p^* and a second nonphysical wave moving upstream at speed c_n^**. Neither the upstream moving wave, nor the downstream moving wave can satisfy (3)–(4) without the simultaneous presence of the second wave whose amplitude, relative to the first wave, is determined by β. Values of $c_\mathrm{p}^*(k)/c$, $c_\mathrm{n}^*(k - \pi/\Delta x)/c$, and $\beta(k)$ are plotted as a function of k in Fig. 2.

Each of the two waves that compose the QFEM eigenmode is a conventional semi-discrete Fourier mode with a well-defined phase speed and group velocity. The phase speeds given by the two roots of (6) are the phase speeds of these individual Fourier modes. No single phase speed precisely describes the motion of a QFEM eigenmode, although since β is small for eigenmodes longer than about $4\Delta x$, well-resolved eigenmodes appear to translate with almost no change in form at speed c_p.

The group velocity, $c_g^* = \partial\omega/\partial k$, for each individual Fourier mode in (10) satisfies

$$\frac{c_g^*}{c} = \frac{\pm\cos\theta\left(2 + 7\cos^2\theta\right) + 2\left(2 - 3\cos^2\theta\right)\sqrt{10 - \cos^2\theta}}{\left(2 - \cos^2\theta\right)^2\sqrt{10 - \cos^2\theta}}.,$$

where the positive root is associated with $\partial\omega_\mathrm{p}/\partial k$ and the negative root with $\partial\omega_\mathrm{n}/\partial k$.[1] Note that $c_{g_\mathrm{p}}(k) = c_{g_\mathrm{n}}(k \pm \pi/\Delta x)$, since $\cos\theta = -\cos(\theta \pm \pi)$, implying that both of the individual modes in (10) propagate at the same group velocity. Thus, although there is some ambiguity in the precise determination of the phase speed of a single

[1] The preceding expression for c_{g_p} is mathematically equivalent to (2.6-52) in [4].

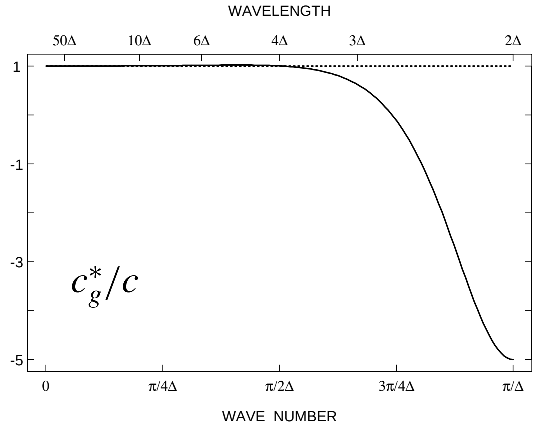

Figure 3: Normalized group velocity, $c_{\mathrm{p}}^*(k)/c$, for the QFEM eigenmodes plotted as a function of k.

QFEM eigenmode, each eigenmode does have a well-defined group velocity. This group velocity is plotted as a function of wavenumber in Fig. 3.

The data in Figs. 2 and 3 suggest that, in many respects, a QFEM approximation to (1) yields qualitatively similar results to those obtained with linear finite elements or finite differences: long, well-resolved waves will be treated accurately while disturbances with wavelengths near $2\Delta x$ will be subject to substantial error. This is supported by many examples in [4], which show qualitatively similar behaviors in both linear- and quadratic-finite element approximations to poorly resolved waves. As the numerical resolution improves, the QFEM can, nevertheless, yield much better solutions to (1) than those obtained using linear elements [1, 4]. The advantage of the QFEM method seems to be attributable to its low phase-speed error and relative freedom from numerical dispersion. The phase-speed error in the downstream moving (physical) component of the QFEM solution is very small, both for relatively short waves, such as a $3\Delta x$ wave, and in the limit of good spatial resolution, for which

$$\frac{c_{\mathrm{p}}^*}{c} \approx 1 + \frac{(k\Delta x)^4}{270}.$$

The most serious inaccuracies in QFEM approximations to (1) are associated with the nonphysical upstream propagating mode, whose amplitude is at least 10% that of the physical mode for all wavelengths shorter than $4\Delta x$. Moreover, the amplitude of the nonphysical mode decays rather slowly with increasing numerical resolution. In the limit $k\Delta x \to 0$,

$$\beta \approx \frac{(k\Delta x)^2}{24},$$

which is consistent with the well known fact [7] that unless the initial data is appropriately filtered to remove the backward propagating wave, the truncation error in the QFEM approximation to (1) is only $O\left[(\Delta x)^2\right]$.

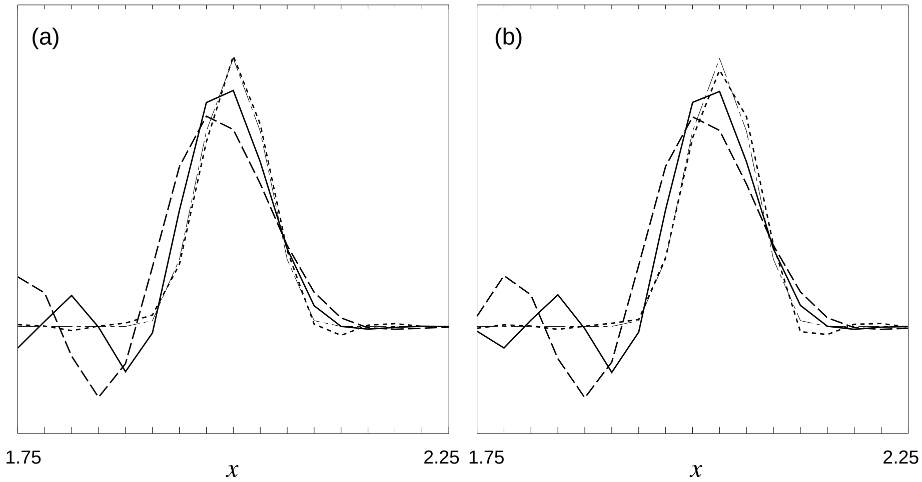

Figure 4: Comparison of solutions to the constant-wind-speed advection equation at (a) $t = 10$ and (b) $t = 10\frac{5}{16}$: quadratic finite element (short dashed), linear finite element (solid), fourth-order centered explicit finite difference (long dashed) and exact (thin dot-dashed). After [1].

How is the difference in the eigenmode structure to linear- and quadratic-finite-element approximations to (1) made manifest in practical applications? One example is provided by the comparison of finite-difference, linear- and quadratic-finite-element solutions to (1) shown in Fig. 4. Numerical solutions were computed on the periodic domain $0 \leq x \leq 3$ subject to the initial condition

$$\psi(x,0) = \begin{cases} \frac{1}{4}(\cos(8\pi(x-1)) + 1)^2 & \text{if } |x-1| \leq \frac{1}{8}; \\ 0, & \text{otherwise} . \end{cases}$$

The wind speed is $c = 0.1$. The horizontal mesh spacing is $\Delta x = 1/32$, implying that the total width of the initial spike is $8\Delta x$, which is sufficiently narrow to reveal short-wavelength errors without allowing the solution to be completely dominated by very short-wavelength disturbances. Further numerical details are given in [1].

The solution in the central portion of the domain is shown at $t = 10$ in Fig. 4a, at which time the peak in the true solution is centered at $x = 2$. For simplicity, the quadratic-finite-element solution is plotted as a piecewise-linear function between the nodes. The superiority of the QFEM result over those obtained using the linear finite element method and using finite differences is clearly evident.

A distinct difference between the amplitude error in the QFEM solution and the other solutions can be seen by comparing panels (a) and (b) of Fig. 4. The true solution propagates exactly one grid interval between the times shown in panels (a) and (b). There are essentially no changes in the shapes of the linear-finite-element and the finite-difference solutions over this short period of time, but the peak in the QFEM solution is noticeably damped. This damping is followed by re-amplification as the solution translates another Δx. The amplitude of the peak in the QFEM solution continues to oscillate as it moves alternatively over the midpoint and endpoint nodes. This oscillation is produced by the alternate constructive and destructive superposition of the pairs of semi-discrete Fourier modes (10) that make up each QFEM eigenmode.

Although the amplitude of the nonphysical component of each QFEM eigenmode remains small in this linear constant-coefficient test problem, there is no guarantee that it will not be amplified by wave-wave interactions and contribute to aliasing error in nonlinear problems. For example, Cullen [2] reports that, in those regions where the solution is smooth, quadratic elements give worse results than linear elements in FEM approximations to the inviscid Burgers' equation.

Acknowledgments. This research was stimulated by several productive conversations with Phil Gresho. This research was sponsored by NSF grant ATM-9817728.

References

[1] D. R. DURRAN, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Springer-Verlag, New York, 1999, 465 p.

[2] M. J. P. CULLEN, *J. Comp. Phys.* **45**, 221 (1982).

[3] P. M. GRESHO and R. L. LEE, *Int. J. Numer. Methods Fluids* **7**, 1357 (1987).

[4] P. M. GRESHO and R. L. SANI, *Incompressible Flow and the Finite Element Method*, John Wiley and Sons, New York, 1998, 1040 p.

[5] P. M. GRESHO, R. L. LEE, and R. L. SANI, in *Finite Elements in Fluids*, volume 3, pp. 335–350, John Wiley and Sons, New York, 1978.

[6] B. CATHERS and B. A. O'CONNOR, *Int. J. Numer. Methods Fluids* **5**, 201 (1985).

[7] G. HEDSTROM, *J. Comp. Phys.* **30**, 222 (1979).