JOURNAL OF COMPUTATIONAL PHYSICS 122, 118–128 (1995)

Wavelet–Galerkin Discretization of Hyperbolic Equations

JUAN MARIO RESTREPO AND GARY K. LEAF

Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439

Received July 12, 1994; revised February 6, 1995

The relative merits of the wavelet–Galerkin solution of hyperbolic partial differential equations, typical of geophysical problems, are quantitatively and qualitatively compared to traditional finite difference and Fourier-pseudo-spectral methods. The wavelet–Galerkin solution presented here is found to be a viable alternative to the two conventional techniques. © 1995 Academic Press, Inc.

1. INTRODUCTION

In the past two decades interest in wavelets has been nothing short of remarkable. In the areas of time series analysis, matrix compression, and approximation theory, wavelets have carved out a practical niche. In the solution of differential equations, however, wavelets have not, thus far, been able to replace other more traditional techniques such as polynomial finite-element methods, except when nonlocal operators are involved. This is because (a) at present, wavelets are capable of dealing in a general way only with the simplest of boundary conditions; (b) until recently there were no techniques to compute the inner products of a wavelet Galerkin approximation easily and inexpensively; (c) the advent of more powerful computers has enabled researchers to stretch the computational usefulness of more traditional methods; (d) wavelet multiresolution analysis can, in most instances, be part of a postprocessing stage in the solution of the differential equation; and (e) adaptive and multigrid solvers are available for finite-difference and finite-element techniques. In our estimation, the usefulness of wavelets in the solution of differential equations is still a matter to be completely established. This study sheds some light on the practical use of wavelets in the solution of hyperbolic equations.

Recent developments in wavelet techniques [1] have made the wavelet–Galerkin procedure a viable option for the solution of some classes of partial differential equations. In this study we compare a Galerkin procedure based on the use of periodized Daubechies scaling functions with standard numerical methods such as finite difference and Fourier pseudo-spectral methods. Other studies that compare the wavelet–Galerkin are [2–4] and, in particular, [5]. In this last paper, Weiss compares wavelet–Galerkin methods with Fourier pseudo-spectral methods and concludes that the wavelet Galerkin method is faster than the dealiased Fourier pseudo-spectral solution of a two-dimensional Euler system and is capable of holding onto the exact solution for a considerably longer time span than the Fourier solution.

The specific hyperbolic problem to be considered is a variant of the Boussinesq system [6]. This system was chosen because it has many of the ingredients of hyperbolic equations that arise in geophysical problems. In scaled variables the Boussinesq system (BQS) is

$$\begin{aligned} \eta_t &= -(hu)_x - \alpha(u\eta)_x \\ v_t &= -\eta_x - \alpha\left(\frac{u^2}{2}\right)_x \\ v &= u - h^2\beta^2 u_{xx}, \end{aligned} \tag{1}$$

to be solved on the interval $x \in [0, 1]$ for $t > 0$ subject to periodic boundary conditions, and initial conditions $\eta(x, t = 0) = E^0(x)$ and $u(x, t = 0) = U^0(x)$. In the geophysical context the $O(1)$ variables $u(x, t)$ and $z = \eta(x, t)$ are thought of as the depth-averaged first-order velocity and wave displacement over $z = 0$, respectively, for weakly nonlinear shallow water dispersive waves traveling over a bottom topography $z = -h(x)$ that is periodic in x.

Equation (1) admits bidirectional, dispersive, weakly nonlinear wave solutions. The degree of nonlinearity is controlled by the parameter $\alpha \ll 1$ and the dispersiveness by parameter $\beta \ll 1$. By setting both parameters to zero, Eq. (1) becomes the linear wave equation (WE). The shallow water wave equation (SWWE) is obtained by letting $\beta = 0$ and $\alpha \neq 0$. The bottom topography $h(x)$ is $O(1)$; but when $\alpha \neq 0$ and $\beta \neq 0$, the additional restriction on the bottom topography is that its derivatives with respect to x have size comparable to α. Aside from a dissipative term, the model is seen to cover a variety of geophysically relevant phenomena.

To make the discretization comparison reasonable and understandable, we have not fine-tuned each of the approximations under consideration. For example, we did not use high order methods, preconditioning techniques for the inversion of local operators, etc. Furthermore, we employed the same time discretization technique for all three methods. We have chosen the leap-frog method [7], owing to its simplicity; its wide use, such as in applications in climate and weather dynamics [8–10]; and

0021-9991/95 $12.00

its nondissipative properties. The first time step is accomplished with a backwards Euler step. Since the above scheme is prone to exhibit growth of the so-called leap-frog computational mode [8], two time-consecutive sets of solutions are averaged periodically.

Application of the leap-frog scheme to Eq. (1) yields the semi-discrete system

$$\begin{aligned}
\tilde{\eta}^{n+1} &= \tilde{\eta}^{n-1} - 2\,\Delta t[(h\tilde{u})_x + \alpha(\tilde{u}\tilde{\eta})_x]^n \\
\tilde{v}^{n+1} &= \tilde{v}^{n-1} - 2\,\Delta t\left[\tilde{\eta}_x + \alpha\left(\frac{\tilde{u}^2}{2}\right)_x\right]^n \qquad (2)\\
\tilde{u}^n &= \mathbf{L}^{-1}\tilde{v}^n,
\end{aligned}$$

with $\mathbf{L} = (\mathbf{I} - h^2\beta^2\partial_{xx})$, where $t = n\,\Delta t$, Δt is taken as fixed during the integration, $n = 0, 1, 2, \ldots$, and the tilde variables $\tilde{f}^n(x) \equiv f(x, n\,\Delta t)$.

In Section 2 we briefly present the full discretization of Eq. (2) using finite difference (FD) and the Fourier pseudo-spectral (FS) schemes. Section 3 presents the wavelet–Galerkin (WG) method in full detail. Qualitative and quantitative comparisons are presented in Section 4. Section 5 summarizes what we have been able to learn about the merits and pitfalls of the WG scheme, as applied to hyperbolic problems, and sets the stage for a future paper on the use of the WG scheme to explore orographic effects on shallow water waves. Appendix A discusses some aspects of the evaluation of the inner products that arise in the wavelet–Galerkin approximation. Technical details related to the execution of this study appear in the Appendix B.

2. FINITE DIFFERENCE AND FOURIER PSEUDO-SPECTRAL DISCRETIZATION

The FD spatial discretization of Eq. (2) for $x \in [0, 1]$, subject to periodic boundary conditions on u and η, will be performed on a uniform spatial grid. Let $x_j = j\,\Delta x$, where $\Delta x = 1/N$ and $j = 0, 1, \ldots, N - 1$. Defining the fully discrete variable, in terms of the tilde variables $f_j^n \equiv \tilde{f}^n(j\,\Delta x)$, the discrete FD system is

$$\begin{aligned}
\eta_j^{n+1} &= \eta_j^{n-1} - \frac{2\,\Delta t}{\Delta x}[h_{j+1/2}u_{j+1/2}^n - h_{j-1/2}u_{j-1/2}^n] \\
&\quad - \frac{2\alpha\,\Delta t}{\Delta x}[\eta_{j+1/2}^n u_{j+1/2}^n - \eta_{j-1/2}^n u_{j-1/2}^n] \\
v_j^{n+1} &= v_j^{n-1} - \frac{2\,\Delta t}{\Delta x}[\eta_{j+1/2}^n - \eta_{j-1/2}^n] \qquad (3)\\
&\quad - \frac{\alpha\,\Delta t}{\Delta x}[u_{j+1/2}^n u_{j+1/2}^n - u_{j-1/2}^n u_{j-1/2}^n] \\
\mathbf{u}^n &= \mathbf{L}^{-1}\mathbf{v}^n,
\end{aligned}$$

where

$$\mathbf{L}\mathbf{v}^n = u_j^n - \frac{h^2\beta^2}{\Delta x^2}[u_{j+1}^n - 2u_j^n + u_{j-1}^n],$$

with boundary conditions and intial data

$$\begin{aligned}
u_0^n &= u_N^n \\
\eta_0^n &= \eta_N^n \\
h_0 &= h_N \\
u_j^0 &= U_j^0 \\
\eta_j^0 &= E_j^0.
\end{aligned}$$

The Fourier approximation of Eq. (2) will be performed pseudo-spectrally [11]. Define the discrete Fourier transform pair

$$\begin{aligned}
\hat{f}^n(k) &= \sum_{j=0}^{n-1} \tilde{f}^n(x_j)\exp(-ikx_j) \\
\tilde{f}^n(x_j) &= \sum_{k=-N/2}^{N/2-1} \frac{\hat{f}^n(k)}{N}\exp(ikx_j),
\end{aligned}$$

where $x_j = 2\pi j/N$, with $j = 0, 1, \ldots, N - 1$. Projecting Eq. (2) into Fourier space and exploiting orthogonality, we obtain

$$\begin{aligned}
\hat{\eta}^{n+1}(k) &= \hat{\eta}^{n-1}(k) - ik2\,\Delta t\widehat{h\tilde{u}^n}(k) - ik2\,\Delta t\widehat{\tilde{\eta}^n\tilde{u}^n}(k) \\
\hat{v}^{n+1}(k) &= \hat{v}^{n-1}(k) - ik2\,\Delta t\hat{\eta}^n(k) - ik\,\Delta t\widehat{\tilde{u}^n\tilde{u}^n}(k) \qquad (4)\\
\hat{v}^n(k) &= \hat{u}^n(k) - \beta^2\widehat{h^2\tilde{u}_{xx}^n}(k),
\end{aligned}$$

with $-N/2 \leq k < N/2$. For a flat bottom, the last equation in Eq. (4) reduces to

$$\hat{v}^n(k) = (1 + \beta^2k^2)\hat{u}^n(k). \qquad (5)$$

Hence, in this special case the operator $\mathbf{L}$ is easily invertible in the FS approximation. The initial data is

$$\begin{aligned}
\hat{\eta}^0(k) &= \hat{E}^0(k) \\
\hat{u}^0(k) &= \hat{U}^0(k).
\end{aligned}$$

Since the dependent variables are real, the discrete Fourier transforms are performed using real FFTs. Possible aliasing that may arise from the evaluation of the nonlinear terms was minimized by zero-padding the upper half of the spectrum since the nonlinear terms are quadratic.

3. WAVELET–GALERKIN DISCRETIZATION

Two discretization alternatives exist. The system can be treated either as a fully Galerkin procedure or as a mixed Galerkin–collocation problem. The presentation will be limited to the full Galerkin implementation; however, a few remarks on the mixed procedure are in order. In the mixed method, nonlinear terms as well as linear terms with spatially varying coefficients, are evaluated by collocation in a manner analogous to FS. Namely, one projects the appropriate variables back to real space, forms the nonlinear terms or the terms involving products of field variables and space-dependent coefficients and then projects these back to the trial space, thus preparing the system for the next time integration. The advantages of this technique are twofold: (a) simplicity of the resulting equations, since these invariably involve simpler inner products as compared with the full Galerkin procedure; (b) the mixed procedure has little or no aliasing problems as compared to the FS. The main disadvantage of the Galerkin–collocation method is that the operation count per time step is significantly higher than its Galerkin counterpart, which is especially troublesome in hyperbolic problems.

Our Galerkin procedure uses a class of compactly supported scaling functions introduced by Daubechies [12]. The scaling functions are determined by a genus index DN and a set of scaling parameters $\{c_k : 0 \le k \le DN\}$ that define the generator function $\phi(x)$ through the scaling relation

$$\phi(x) = \sum_{k=0}^{DN-1} c_k \phi(2x - k).$$

For each $0 \le j$ we set

$$\phi^j_k(x) = 2^{j/2}\phi(2^j x - k) \quad \text{for } 0 \le k < 2^j.$$

If one sets $V^j = \text{span}\{\phi^j_k : 0 \le k < 2^j\}$, in [13] it is shown that $\{\phi_k\}$ can be periodized and made to form an orthonormal basis for $V^j \in L^2[0, 1]$, with $\overline{\bigcup V^j} = L^2[0, 1]$ and $\bigcap V^j = 0$. Moreover, the subspaces V^j are nested, so that $V^j \subset V^{j+1}$. If one lets W^j denote the orthogonal complement of V^j in V^{j+1}, it is shown in [13] that W^j is spanned by an orthonormal set of wavelet functions $\psi^j_k = 2^{j/2}\psi(2^j x - k)$, where the generator wavelet $\psi(x)$ is defined by

$$\psi(x) = \sum_{k=-1}^{DN-2} (-1)^k c_{k+1} \phi(2x + k).$$

The base generators $\phi(x)$ and $\psi(x)$ have support $[0, DN - 1]$ and every polynomial of degree $K \le DN/2$ lies in the space V^0, which is equivalent to $\psi(x)$ having $DN/2$ vanishing moments. The Daubechies class is distinguished by having this interpolation property and the smallest possible support. Thus, from the interpolation property, we see that $\phi(x)$ has at least $DN/2$ continuous derivatives. As mentioned in Qian and Weiss [14], $\phi(x)$ is in the class C^γ with γ at least $0.55DN$.

Consider a set $\{\phi^p_k\}$ that spans the space $V^p[0, 1] \subset L^2[0, 1]$. A multiresolution is effected by noting that the space $V^p \supset V^{p-1} \supset \cdots \supset V^1 \supset V^0$. For the Galerkin approximation of the hyperbolic problem, the field variables are projected into the space of trial functions belonging to V^p. When we use test functions from the same space, a system of differential equations in time for the coefficients of the field variable results when the inner products $\langle\cdot, \cdot\rangle$ are evaluated and orthogonality among the elements of V^p is used. In this study the evolution equations are solved at scale p determined by the resolution of the space V^p. If, at any time, a multiresolution is desired, this can be performed as a postprocessing step or as an adjunct calculation.

In what follows, we project the semi-discrete real variable $\tilde{f}$, say, into V^p so that

$$\tilde{f}^n(x) = \sum_{l=0}^{N-1} f^n_l \phi_l(x_j), \tag{6}$$

where $f^n_l = \langle \tilde{f}^n, \phi_l \rangle$. For simplicity of notation it will be assumed in the remainder of this study that the ϕ_l's are of resolution $N \equiv 2^p$ and genus DN.

The weak formulation of the semi-discrete system is obtained by substituting Eq. (6) into Eq. (2), multiplying by a test function $\phi_k \in V^p$, and integrating

$$\begin{aligned}
\langle \tilde{\eta}^{n+1}, \phi_k \rangle &= \{\tilde{\eta}^{n-1}, \phi_k \rangle - 2\,\Delta t\,\langle (h\tilde{u}^n)_x, \phi_k \rangle \\
&\quad - 2\,\Delta t \alpha \langle (\tilde{u}^n \tilde{\eta}^n)_x, \phi_k \rangle \\
\langle \tilde{v}^{n+1}, \phi_k \rangle &= \langle \tilde{v}^{n-1}, \phi_k \rangle - 2\,\Delta t \langle \tilde{\eta}^n_x, \phi_k \rangle - \Delta t \alpha \langle (\tilde{u}^n \tilde{u}^n)_x, \phi_k \rangle \\
\langle \tilde{v}^n, \phi_k \rangle &= \langle \tilde{u}^n, \phi_k \rangle - \beta^2 \langle h^2 \tilde{u}^n_{xx}, \phi_k \rangle.
\end{aligned} \tag{7}$$

Following the convention in [1, 13], we refer to the inner products as connection coefficients:

$$\begin{aligned}
\Omega^{0,1}_{k,l} &= \langle \phi_k, \phi'_l \rangle \\
\Omega^{1,1}_{k,l} &= \langle \phi'_k, \phi'_l \rangle \\
\Omega^{0,1,1}_{k,j,l} &= \langle \phi_k, \phi'_j \phi'_l \rangle \\
\Omega^{1,0,0}_{k,j,l} &= \langle \phi'_k, \phi_j \phi_l \rangle \\
\Omega^{1,0,1}_{k,j,l} &= \langle \phi'_k, \phi_j \phi'_l \rangle.
\end{aligned}$$

The most expedient strategy available for the evaluation of these connection coefficients is given in [1]. A brief summary on the computational procedure is provided in the appendix of this paper. The connection coefficients should be precomputed, and the resulting tables are then read in the time-marching procedure.

After integrating by parts and exploiting periodicity, the full Galerkin implementation is

$$\begin{aligned} b_k^{n+1} &= b_k^{n-1} + 2\,\Delta t \sum_{l=0}^{N-1} a_l^n \sum_{j=0}^{N-1} h_j \Omega_{j-k,l-k}^{1,0,0} \\ &\quad + 2\alpha\,\Delta t \sum_{l=0}^{N-1} a_l^n \sum_{j=0}^{N-1} b_j^n \Omega_{j-k,l-k}^{1,0,0} \\ c_k^{n+1} &= c_k^{n-1} + 2\,\Delta t \sum_{l=0}^{N-1} b_l^n \Omega_{l-k}^{1,0} + 2\alpha\,\Delta t \sum_{l=0}^{N-1} a_l^n \sum_{j=0}^{N-1} a_j^n \Omega_{j-k,l-k}^{1,0,0} \\ c_k^n &= a_k^n + \beta^2 \sum_{l=0}^{N-1} a_l^n \sum_{j=0}^{N-1} (h^2)_j [\Omega_{j-k,l-k}^{1,0,1} + \Omega_{j-k,l-k}^{0,1,1}], \end{aligned} \tag{8}$$

with $0 \le k \le N - 1$, where b_k, c_k, and a_k, are the expansion coefficients associated with η, v, and u; while h_k and $(h^2)_k$ are associated with the bottom topography h, and its square. The initial data for the wavelet–Galerkin scheme is

$$b_k^0 = \mathcal{P}^p(E^0(x))$$

$$a_k^0 = \mathcal{P}^p(U^0(x)),$$

where $\mathcal{P}^p$ is the orthogonal projection operator to the space V^p.

By a change of variables the last two connection coefficients in Eq. (8) can be expressed in terms of elements of the same connection coefficient array [13], so that the last expression in Eq. (8) is transformed into

$$c_k^n = a_k^n + \beta^2 \sum_{l=0}^{N-1} a_l^n \sum_{j=0}^{N-1} h_j [\Omega_{h-j,l-j}^{0,1,1} + \Omega_{j-k,l-k}^{0,1,1}].$$

When $h(x) = 1$, the above equation can be further simplified to

$$c_k^n = a_k^n + \beta^2 \sum_{l=0}^{N-1} a_l^n \Omega_{l-k}^{1,1}. \tag{9}$$

4. COMPARISON STUDY

We compare the methods on three types of hyperbolic equations: the wave equation (WE), the shallow water wave equation (SWWE), and the Boussinesq system (BQS). To effect a comparison, we define a merit value based on two factors: the memory resources M and the wall-clock time T. In making a comparison we first establish a desired level of accuracy as follows: for a given N and Δt we monitor three norms of the solution at some time t_f, the final integration time. Our criterion for accuracy is established by demanding that each of the three norms l_1, l_2, and l_∞ of the solution agree, to three decimal places for the WE, and to four decimal places for the BQS. For each method, T is the time required to obtain a solution to this level of accuracy and will require storage M. Thus, we define the computational efficiency merit value

$$C_{\text{eff}} \equiv \frac{1}{T \times M}.$$

TABLE I

Storage Requirements

Problem	FD	FS	WG
WE	$5N$	$5N$	$5N$
SWWE	$5N$	$9N$	$7N$
BQS	$5N + 3N$	$9N + 0.75N^2$	$7N + 2N(DN - 1)$

Our determination of an acceptable solution was based on searching among the parameter values $\Delta t = 0.001/2^r$ and $N = 1/2^q$. We report the largest Δt and the smallest N encountered in meeting the accuracy criteria. This determines T and the corresponding M.

The storage requirements M depends on N. For the three methods as a function of the type of problem, the relation between M and N is given in Table I.

The numbers reflect "common" storage requirements as opposed to optimal requirements. The second number in the BQS row represents the memory requirements for the operator **L** for each method.

In order to simplify the comparison, the bottom topography will be set, for the remainder of this study, to $h(x) = 1$. However, although the inversion of **L** when $h = 1$ is trivial and exact in the FS case as shown in Eq. (5) and simpler for the WG using Eq. (9), neither of these advantages will be invoked in the comparison of the three implementations.

4.1. *The Wave Equation*

Table II shows a comparison of the computational efficiency and the energy E of the three methods on the WE problem. The last four entries correspond to the WG of genus DN. The initial data for this experiment was the cubic pulse

$$E^0 = \begin{cases} A\left(1 - 3\left|\dfrac{x-0.5}{\sigma}\right|^2 + 2\left|\dfrac{x-0.5}{\sigma}\right|^3\right) & \text{for } |x - 0.5| > \sigma \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

$$U^0 = E^0/2$$

TABLE II

Computational Efficiency for the Solution of the Wave Equation

Method	N	Δt	T	C_{eff}	E
FD	512	1.0(−3)	41.40	9.4354(−6)	1.000178
FS	32	1.0(−3)	7.28	8.5852(−4)	0.999972
*DN*4	128	1.0(−3)	54.67	2.8581(−5)	1.000388
*DN*6	128	1.0(−3)	88.24	1.7707(−5)	1.000470
*DN*8	128	1.0(−3)	115.32	1.3549(−5)	1.000478
*DN*16	64	2.0(−3)	90.59	3.4496(−5)	1.000472
*DN*20	64	1.0(−3)	272.05	1.1487(−5)	1.000478

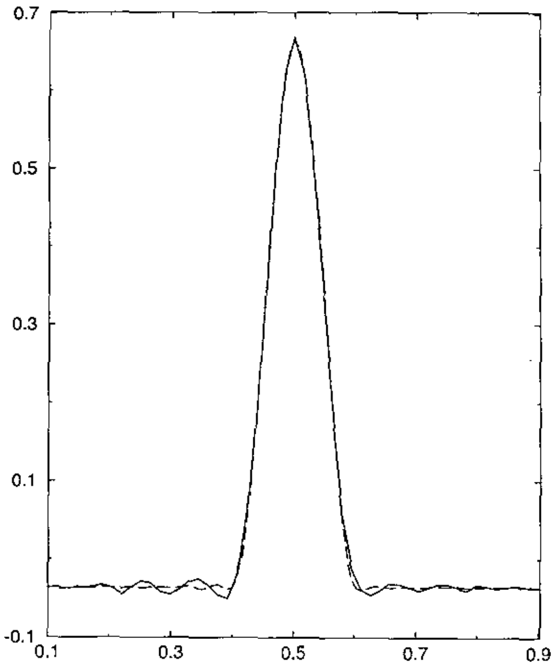

FIG. 1. WG *DN*6 solution of the wave equation at $t_f = 2$; $N = 32$ (solid, with oscillatory outskirts), and $N = 128$ (dashed).

with $A = 0.7$ and $\sigma = 0.1$. The integration is carried out to $t_f = 2$, at which time the solution should be an exact replica of the initial conditions.

For this particular initial data we found that the three methods were most successful in reaching first the l_2 norm, second in reaching the sup norm, and last in reaching the l_1.

To within the discretization size, all methods were capable of predicting correctly the location at which the sup norm is expected to be. It is also noted that conservation of the total energy is easily achieved, even when the computed solution looks unacceptable, namely when the solution has been underdiscretized. The most salient feature of an underdiscretized solution is the appearance of dispersive effects. Figure 1 illustrates the WG *DN*6 solution at $t = 2$ in the underdiscretized case: $\Delta t = 0.001$, $N = 32$. Superimposed on the underdiscretized solution in Fig. 1 is the converged solution reported in Table II.

Figure 2 shows the time evolution of the bidirectinal linear wave with a numerically induced dispersive tail resulting from underdiscretization. In this figure $t_f = 2.2$, $\Delta t = 0.001$, $N = 32$, and *DN*6. The FD, as is well known, will exhibit a very similar behavior when underdiscretized. The cost comparison, which is $1/C_{\text{eff}}$, of the three methods for the WE problem is shown in Fig. 3, as a function of N. In this cost comparison we do not consider the accuracy of the solution.

4.2. *The Shallow Water Wave Equation*

For the shallow water wave equation with $\alpha = 0.1$, the initial data is given by Eq. (10), with $A = 1.0$ and $\sigma = 0.1$. The integration time was $t_f = 0.64$, which was sufficient to make the nonlinear effects very obvious in the solution. The solution is a bidirectional steepening wave. Table III displays the results of the timing experiment. The last two columns show the location x_{sup}, to within $1/N$, of the sup norm and the value of the norm. For the SWWE we did not attempt to achieve similar norms in all methods, but rather monitored the quality of the shape of the solution and the size of the l_2 error.

Figure 4 shows the qualitative differences between the three

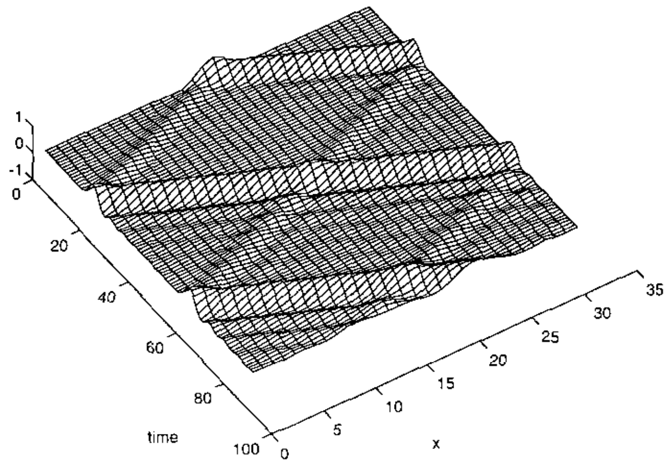

FIG. 2. WG *DN*6 solution of the wave equation; $\Delta t = 0.001$, $N = 32$.

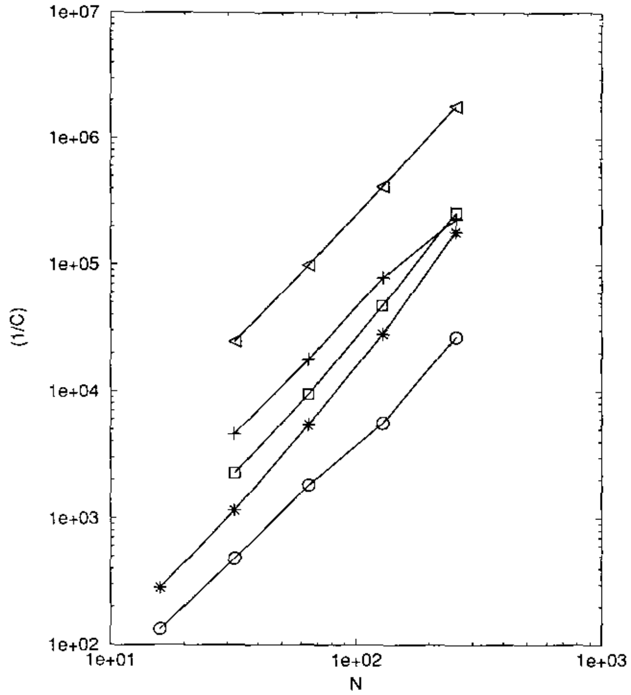

FIG. 3. Cost comparison of the three methods for the WE: FD (circles), FS (stars), *DN*6 (squares), *DN*8 (crosses), *DN*16 (triangles); $N = 32, 64, 128, 256$.

TABLE III

Computational Efficiency for the Solution of the Shallow Water Wave Equation

Method	N	Δt	T	C_{eff}	E	x_{sup}	l_∞
FD	2048	1.0(−4)	8360.2	1.1681(−8)	0.98840670	0.2217	0.887690
FS	1024	1.0(−4)	846.68	1.2826(−7)	0.98967046	0.2090	0.731428
*DN*6	2048	5.0(−5)	57521.2	1.2127(−9)	0.974146	0.2094	0.745228
*DN*16	2048	1.0(−4)	42324.4	1.6481(−9)	0.990746	0.2183	0.774571

methods in the calculation of the shocks at $t = 0.64$. At the end of the computations the solutions were filtered twice using a 1–2–1 spatial low-pass filter. The filtered solutions appear in Fig. 4. The noise could have been curbed significantly in all the solutions if an appropriate implicit time integrator replaced the leap-frog technique used in this study. The parameters for each of these curves appears in Table III. As expected, we found that the smaller wave (not shown) is very well captured by all three methods, but they handled poorly the high amplitude portion of the solution which is featured in Fig. 4. The phases of the FD and the FS are the same, whereas the phase of the WG solution is ahead of the aforementioned solutions. The shape of the unfiltered solutions is quite different: high frequency oscillations are significant in the WG case but limited to the neighborhood of the shock front, and they are smaller in magnitude in the FS solution, but present, throughout the domain. The second-order FD solution, on the other hand, shows large oscillations, but these are only present in the immediate vicinity of the shock front. As shown in Fig. 4, the filter has virtually eliminated the high frequency oscillations of the FS and significantly improved the situation for the WG solution. We found that the oscillations in the WG solutions could be eliminated to the same degree as the FS solution shown in the figure if the data is filtered once more. The FS method is clearly most efficient and the FD best able to capture the shape of the solution.

For the same problem Fig. 5 illustrates the differences between the methods when the same values of N and Δt are used in all three methods. The plot was obtained using $\Delta t = 10^{-4}$, with $N = 1024$. The WG solutions do not have the oscillations

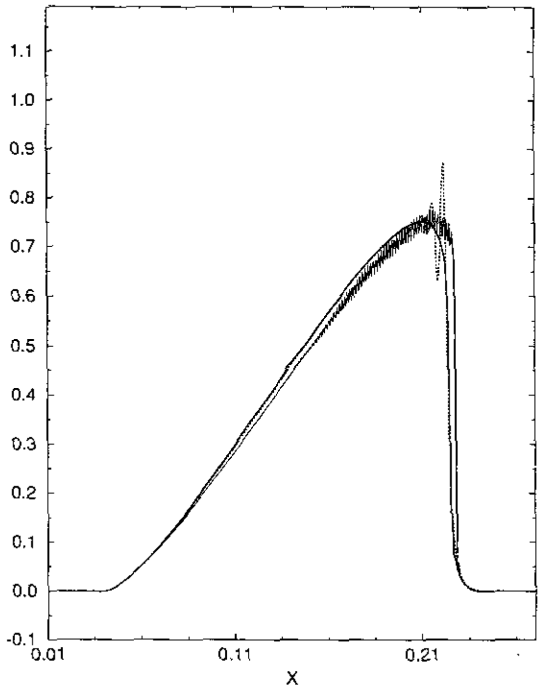

FIG. 4. Comparison of the three methods in the solution of the shallow water wave equation. Portion of the profile at $t = 0.64$: FS (left-most, solid), FD (dashed), *DN*16 (right-most, solid). Execution parameters are given in Table III.

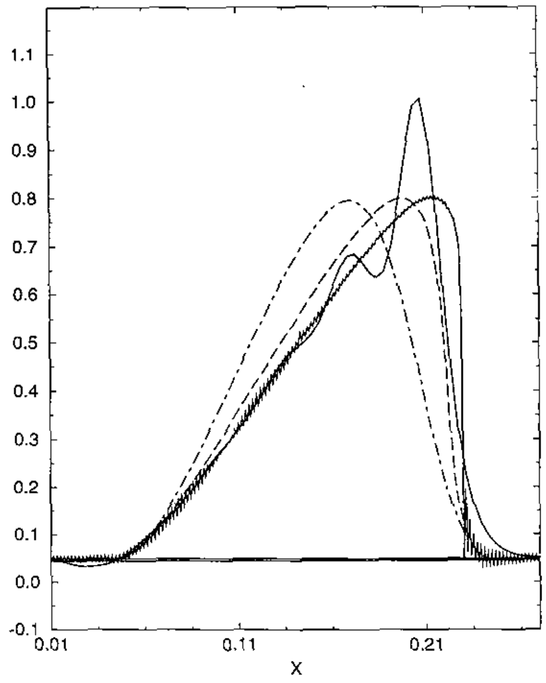

FIG. 5. Comparison of the three methods in the solution of the shallow water wave equation. Portion of the profile at $t = 0.64$: FS (solid, small oscillations), FD (solid, large oscillations), *DN*6 (dash), *DN*16 (dash-dot); $N = 1048$, $\Delta t = 10^{-4}$.

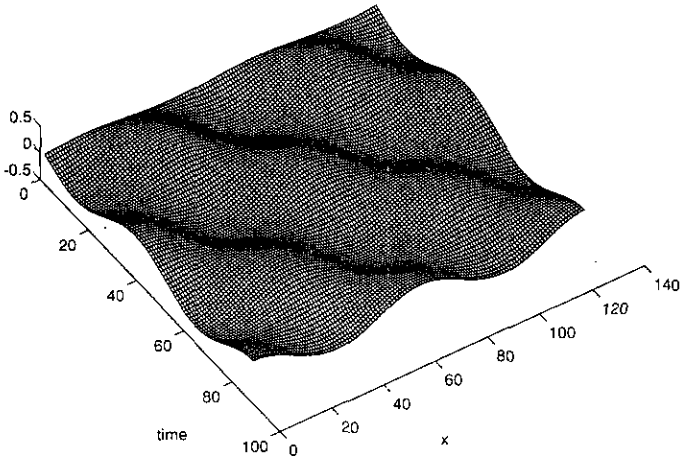

FIG. 6. WG *DN*6 Boussinesq solution for η: $N = 128$, $\Delta t = 0.002$. The time axis, in arbitrary units, increases toward the viewer.

present in the FS; however, the shock is not as steep. The steepness in the WG solution was less severe in the case $DN = 16$. The milder steepness of the WG method means that the location of the x_{sup} is very poorly predicted. The FD is next in getting this location; however, it suffers from poor shape-capturing characteristics. The FS is best, overall; however, the solution has a great deal of high frequency oscillations which propagate away from the shock and are present throughout the whole solution. Since the energy was slightly smaller in magnitude in the WG case than in the other methods, it may indicate that the dissipation was significant enough to affect the amplitude of the solution and thereby the velocity of the solution. This could account for the significant phase error.

We performed experiments with initial data with noncompact support. We found that the FD method had a significantly worse phase lag than reported in the above experiments. In fact, this phase lag was also significant if the solution, for some intitial data, eventually loses its compact support. This phase problem was absent in the FS and very minimally present in the WG experiments for noncompact solutions.

4.3. *The Boussinesq System*

For the computation of the Boussinesq system solution with $\alpha = 0.1$ and $\beta^2 = 0.03333$, we compared the solutions of the three methods at $t = 0.5$ for initial data,

$$\begin{aligned} U^0 &= 0.1\sin(4\pi x) \\ E^0 &= 0.5U^0. \end{aligned} \tag{11}$$

The solution, up to $t_f = 2.2$ with $\alpha = 0.1$ and $\beta^2 = 0.1$, is shown in Fig. 6 for the WG method with $DN = 6$, $\Delta t = 0.002$, and $N = 128$.

TABLE IV

Computational Efficiency for the Solution of the Boussinesq System

Method	N	Δt	T	C_{eff}
FD	256	1.0(−3)	14.75	3.3104(−5)
FS	128	2.0(−3)	21.30	3.4932(−6)
*DN*6	128	2.0(−3)	12.29	3.7393(−5)
*DN*8	128	2.0(−3)	16.77	2.2184(−5)
*DN*16	128	2.0(−3)	100.49	2.1012(−6)

The computational efficiency for the Boussinesq system is shown in Table IV. In this case T reflects the fact that the operator $\mathbf{L}$ needs to be inverted at each time step to find $\mathbf{u}$ from $\mathbf{v}$. We observe in this case that the WG *DN*6 is not only computationally more efficient but also has the least wall-clock time. For partial differential equations that generate systems of the form

$$A(t, y)\frac{dy}{dt} = f(t, y)$$

the WG approach appears viable. In particular, equations such as the Boussinesq system, the Benjamin–Bona–Mahony equation, the regularized Benjamin–Ono equation, and the regularized Korteweg–de Vries–Burger equation provide examples of such systems.

The cost comparison of the three methods for the BQS problem is shown in Fig. 7, as a function of N. The graph shows

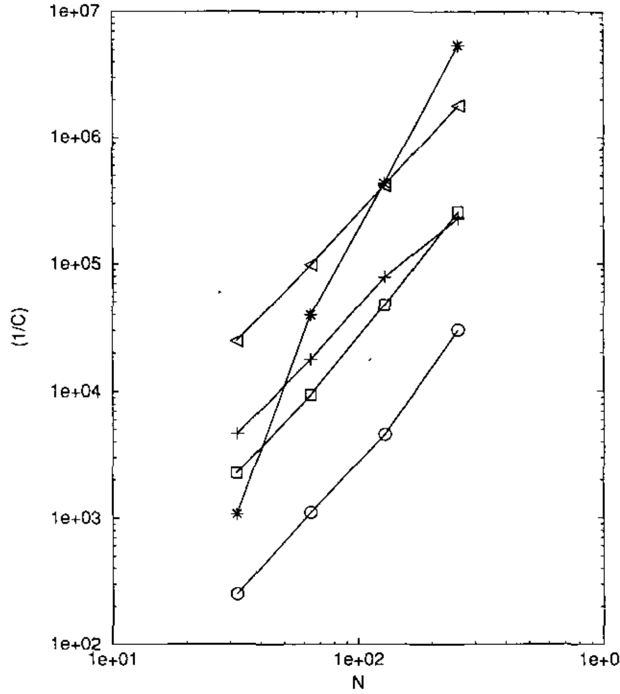

FIG. 7. BQS cost comparison of the three methods: FD (circles), FS (stars), *DN*6 (squares), *DN*8 (crosses), *DN*16 (triangles); $N = 32, 64, 128, 256$.

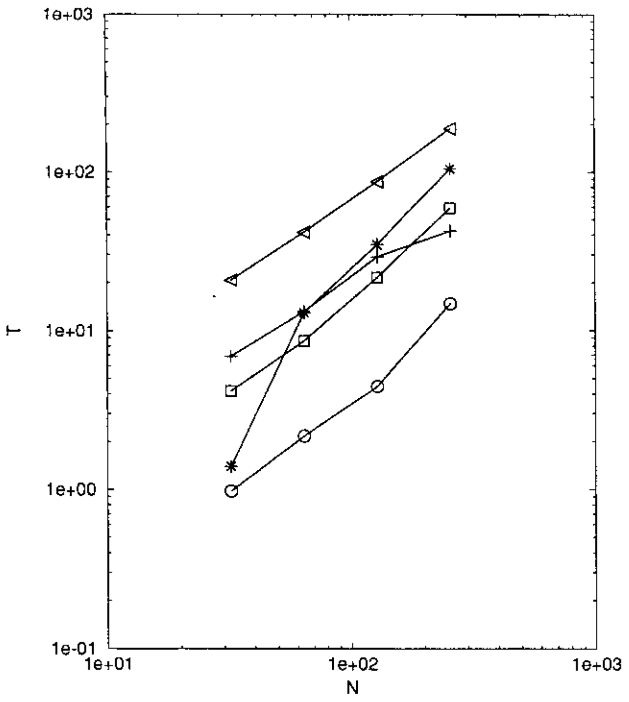

FIG. 8. BQS time comparison of the three methods: FD (circles), FS (stars), *DN*6 (squares), *DN*8 (crosses), *DN*16 (triangles); N = 32, 64, 128, 256.

that for a small number in N the FS is superior to WG, but as the problem gets larger, the increasing WG becomes more cost effective. Figure 8 is a plot of the relation between the wall-clock time T and the resolution N. Disregarding the quality of the solution the graphs show that the FD is the most cost effective method. For small problems the low-genus *DN* is favored over high *DN*, but for larger problems the large *DN* should prove more cost effective. The same can be said of the FS compared with the WG method for any order.

To put the above conclusion in perspective we need to examine the computational cost as a function of the quality of the solution. Figure 9 presents such a relation for the BQS problem. We computed the solution of the BQS using WG *DN*16 with $N = 1024$ and $\Delta t = 10^{-4}$ for the test problem, Eq. (11). We took the norms for this solution as a benchmark since they were the lowest ones computed by any of the methods and any of the discretizations chosen. We chose as a measure of the error of a particular solution the absolute difference in the l_1 norm between the solution and the benchmark at $t = 0.5$. It will be assumed that a measure of the quality of the solution is given by the inverse of the error.

Figure 9 shows that the viability of a particular method depends on the size of error that we are willing to deem acceptable. The graph should be interpreted as being only qualitatively accurate; for very small error values the curves slope up because the error calculation has been effected using the norm of a highly refined discretization which has an accuracy of 10^{-6}. Furthermore, there may be some uncertainty in the cost calculation. The experiment was conducted on a work station which was monitored for activity. However, the work station was connected to a large network and we experienced discrepancies on the wall-clock values over duplicate runs of at most 10%. On the other hand, the computed norms over duplicate runs were replicated to machine precision. The discrepancy in the wall-clock time component of the cost may account for the unexpected dip in the *DN*16 curve.

As evidenced in Fig. 9, for large error values, the FD and FS methods are most cost effective. For a decrease of an order of magnitude in the error, however, the FD cost increases by at least an order of magnitude. Additionally, the graph suggests that for high accuracy the FD and FS are comparable in cost. For small errors, this experiment shows that the WG has a clear advantage over the other methods. The lower genus wavelet solutions were most cost effective.

5. CONCLUDING DISCUSSION

The wavelet–Galerkin solution was qualitatively compared with the solution of finite difference and Fourier pseudo-spectral implementations of the wave equation, the shallow water wave equation, and the Boussinesq system. Time-stability was assured for all three problems and all three methods by repeated selection of a variety of time steps. In this selection process we were guided by the results in [15, 12] for the WG case and in [11] for the Fourier case. Our comparisons were based on

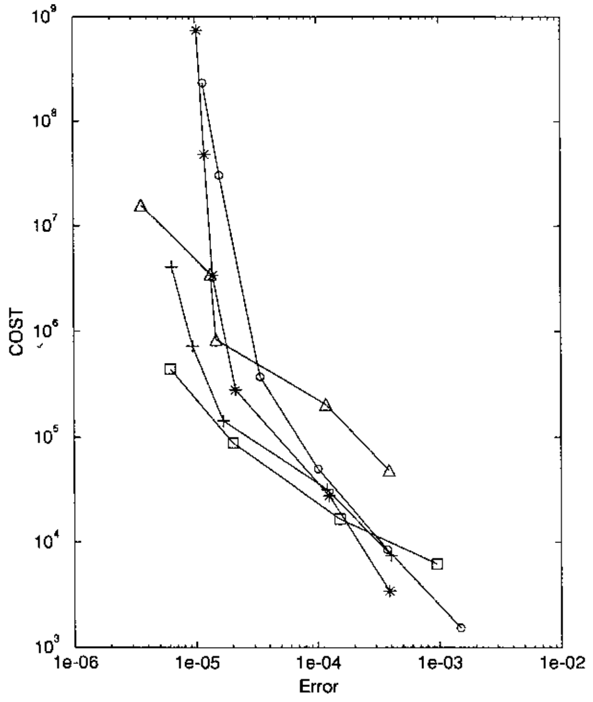

FIG. 9. BQS cost comparison of the three methods as a function of the quality of the solution: FD (circles), FS (stars), *DN*6 (squares), *DN*8 (crosses), *DN*16 (triangles).

the use of the computational efficiency C_{eff} as the merit criterion, which is the reciprocal product of the wall-clock time and the storage requirement.

For the wave equation, based on this criterion, it was found that the FS was the most efficient. The WG was found to be comparable in efficiency to the FD method, requiring less storage but more time than the FD.

Unlike the wave equation problem, in the shallow water wave equation the nature of the solutions may differ considerably from that of the initial conditions. Phase and shape preservation are important issues, and much work has been done creating FD and FS implementations that perform far better in these respects than the particular implementations presented in this study. Nevertheless, these particular implementations are adequate to compare the three methods. Since our merit value C_{eff} does not take into account the regularity of the initial data, our results regarding the computational efficiency cannot be taken to represent the general case. With regards to the qualitative characteristics of the solution for the three methods, we found that for small initial data all methods perform very similarly. However, for large amplitude solutions, particularly when shock-like solutions are involved, the FS develops ever-increasing small-scale oscillations which will eventually spread to the whole domain, but it holds reasonably well to the large-scale features of the solution. The second-order FD solution has the same phase as the FS and very similar large-scale features. At the shock front the FD solution overshoots but the oscillation is confined to the neighborhood of the shock. The WG solution leads in phase, and its shape is similar to its FD counterpart, but the overshoot is spread further away from the shock. Three-point averaging of the solution is found to be effective in improving the shape of the FS and the WG outcomes. For high C_{eff} the WG solution, averaged twice, was best in phase and shape accuracy, while for modest values of C_{eff} and FS solution is best in shape and phase accuracy.

In the BQS problem the challenges are conveying properly the effect of the regularizing operator **L** and efficiently effecting its inversion. Based on our merit criteria the WG method has a distinct advantage over the other two methods. The WG method, which is a particular variant of a finite-element procedure, may be a viable alternative to more traditional counterparts, such as finite-difference and Fourier-pseudospectral methods for problems exemplified by the BQS problem. The FS was the least efficient owing to the fact that the inversion of **L** is an $O(N^2)$ operation as compared to $O(N)$ for the FD and WG implementations. However, it should be remembered that an FFT-based inversion of the operator **L** could have been applied to any of the methods examined here, having a significant effect in bringing down the operation count of the FS method.

Our results have been restricted to one spatial dimension; as far as operational counts are concerned, Weiss and Qian [5] have two-dimensional estimates which when compared to our results indicate that the C_{eff} may not scale according to dimension in a simple way. However, since the genus DN and the resolution $N = 2^p$ both enter in the computational efficiency of the wavelet–Galerkin method, these may be exploited in certain instances to achieve marginal to significant resource economy in the use of the WG approximation in very high resolution studies, regardless of the number of spatial dimensions.

APPENDIX A: THE CONNECTION COEFFICIENTS

The formulation of solutions will require integrations of the form

$$\Omega^{d_1,d_2,\ldots,d_n}_{k_1,k_2,\ldots,k_n} = \langle \varphi^{(d_0)}_{j,k_0}\varphi^{(d_1)}_{j,k_1}\cdots\varphi^{(d_n)}_{j,k_n}\rangle = \int_0^1 \varphi^{(d_0)}_{j,k_0}\varphi^{(d_1)}_{j,k_1}\cdots\varphi^{(d_n)}_{j,k_n}\,dx,$$

where $\varphi^{(d)} = d^d\varphi/dx^d$. This expression is an n-term connection coefficient, or n-tuple. Since φ cannot be represented in closed form for $DN > 2$ and, by construction, has limited regularity, analytic calculation of the integral is infeasible, and the numerical quadrature is often inaccurate as a result of the wildly oscillating nature of the resulting kernels. An alternative approach developed by Latto, Resnikoff, and Tenenbaum [1], which we adopt in this study, recasts the quadrature problem into a linear algebra calculation. The procedure may be used with other wavelet basis provided a moment condition is available and that attention is given to possible wrap-around when the support of the basis functions extends beyond the periodic interval. The following is a brief outline of the calculation of the inner products required in the Galerkin procedure, using the 2-tuple as an example. Complete details appear in [1, 13]. A different and equally viable approach that the reader may wish to consider was proposed by Beylkin in [16].

Integration by parts is performed repeatedly on the 2-tuple integral to obtain

$$\Omega^{d_1,d_2}_{k_1,k_2} = (-1)^{d_1}\Omega^{0,d_2+d_1}_{k_1,k_2},$$

where the periodicity of the wavelets has been invoked. By changing variables, we further reduce the equation to

$$\Omega^{0,d}_{k_1,k_2} = \Omega^{0,d}_{0,k_2-k_1} \equiv \Lambda^{d}_{k_2-k_1},$$

where $d = d_1 + d_2$. From these relations it is clear that any 2-tuple can be represented by a Λ^d_k.

To construct the eigenvector problem, fix d, then solve for $\{\Lambda^d_k\}_{0\le k<2^j}$ by creating a system of 2^j homogeneous relations in Λ^d_k and enough inhomogeneous equations to reduce the dimension of the associated eigenspace to 1. Although we are using the connection-coefficient method for the nonperiodized case, we are computing them for the periodic case (by equivalence), which is where the bounds on k come into play.

First, to form homogeneous relations, we fix $d, j \in \mathbb{N}$, such that $\varphi_j^{(d)}$ is well defined. To simplify notation, denote $\varphi_{j,k}^{(d)} \equiv \Phi_k^d$. Since for every $0 \leq k < 2^j$,

$$\begin{aligned}\Lambda_k^d &= \int \Phi_0(x)\Phi_k^d(x)\,dx \\ &= \int \left(\sum_{m=0}^{N-1} h_m\Phi_m(2x)\right)\left(\sum_{l=0}^{N-1} h_l\Phi_{l+2k}^d(2x)\right) 2^d d(2x) \\ &= 2^d \sum_m \sum_l h_m h_l \int \Phi_m(2x)\Phi_{l+2k}^d(2x)d(2x) \\ &= 2^d \sum_m \sum_l h_m h_l \int \Phi_0(\zeta)\Phi_{l+2k-m}^d(\zeta)\,d\zeta;\end{aligned}$$

thus,

$$\Lambda_k^d = 2^d \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} h_m h_l \Lambda_{l+2k-m}^d$$

where $h_m = c_m/\sqrt{2}$.

This linear homogeneous system can be represented as

$$A\lambda^d = 2^{-d}\lambda^d,$$

where $\lambda^d = \{\Lambda_k^d\}_{0\leq k<2^j}$.

Second, we generate enough inhomogeneous relations to bring down the dimension of the null space to 1. For the 2-tuple case, this means that a single inhomogeneous relation is needed. To this end we avail ourselves of the "moment" property of Daubechies wavelets (cf. [12]), which states that

$$\int \varphi(x)x^k\,dx = 0, \quad k = 0, \ldots, M-1.$$

To generate the inhomogeneous equations, we must first assume $d \leq M - 1$. The moment condition then guarantees that

$$x^d = \sum_{l\in\mathbb{Z}} \tilde{M}_l^d,$$

where $\tilde{M}_l^d = \langle x^d, \varphi_{0,l}\rangle$. Setting $x = 2^j\zeta$ and defining $M_l^d = \langle x^d, \Phi_l\rangle$, we have

$$\tilde{M}_l^d = 2^{dj}2^{j/2}\langle \zeta^d, \Phi_l\rangle = 2^{dj}2^{j/2}M_l^k.$$

This gives the relation

$$\zeta^d = \sum_{l\in\mathbb{Z}} M_l^d\Phi_l(\zeta),$$

which, when differentiated d times, yields

$$d! = \sum_{l\in\mathbb{Z}} M_l^d\Phi_l^d(\zeta).$$

Multiplying by Φ_0^0 and integrating, we obtain

$$\sum_{l\in\mathbb{Z}} M_l^d \int \Phi_0^0(\zeta)\Phi_l^d(\zeta)\,d\zeta = d! \int \varphi_{j,0}(\zeta)\,d\zeta = d!2^{-j/2}.$$

Thus $\sum_l M_l^d\Lambda_l^d = d!2^{-j/2}$. The sum over l is actually over $|l| \leq N - 2$ since the φ's are compactly supported. Thus, by changing the indices of summation by $m = l + 1 + (N - 2)$, the inhomogeneous equations are

$$\sum_{m=1}^{2N-3} \Lambda_m^d M_{m-1-(N-2)}^d,$$

with

$$M_l^d = 2^{-j(2d+1)/2}\tilde{M}_l^d.$$

The linear system formed by the 2^j homogeneous equations and the above inhomogeneous equations has eigenspace dimension 1. Thus, all that remains to specify the system is to calculate $\tilde{M}_l^d$:

$$\begin{aligned}\tilde{M}_l^k = \int x^d\varphi(x-l)\,dx &= \int (y+n)^k\varphi(y)\,dy \\ &= \int \sum_{j=0}^{k} \binom{k}{j} y^j n^{k-j}\varphi(y)\,dy \\ &= \sum_{j=0}^{k} \binom{k}{j} n^{k-j}\tilde{M}_0^j.\end{aligned}$$

Since $\tilde{M}_l^0 = 1$ the above relation is used to evaluate recursively $\tilde{M}_l^d$ for all l. The linear system is now complete and fixes the values of Λ_k^d.

APPENDIX B: TECHNICAL DATA

The codes were executed on a Sparc 10/51 running SunOS 4.1.3U1. The Fortran Sun compiler used was Fortran Version 1.4 with optimization flags turned off. All runs were performed in double-precision arithmetic. Wall-clock times reported apply only to the time integration. Times should be interpreted comparatively, since the code contains many diagnostic operations. All linear algebra operations were performed with general solvers from LAPACK and the FFT's were performed with Paul Swarztrauber's FFTPACK, version 1989.

ACKNOWLEDGMENTS

We thank John Weiss and the referees for their helpful comments and suggestions. This research was supported in part by an appointment to the Distinguished Postdoctoral Research Program sponsored by the U.S. Department of Energy, Office of University and Science Education Programs, and administered by the Oak Ridge Institute for Science and Education. Further support was provided by the Office of Scientific Computing, U.S. Department of Energy under Contract W-31-109-Eng-38.

REFERENCES

1. A. Latto, H. L. Resnikoff, and E. Tenenbaum, *The Evaluation of Connection Coefficients of Compactly Supported Wavelets, Proceedings, French–USA Workshop on Wavelets and Turbulence, Princeton* (Springer-Verlag, New York, 1991).
2. A. Latto and E. Tenenbaum, *C. R. Acad. Sci. Paris* (1990).
3. J. S. Xu and W. C. Shann, *Numer. Math.* **63,** 123 (1992).
4. Z. Qian and J. Weiss, *Appl. Math. Lett.* **6,** 47 (1993).
5. J. Weiss, preprint, 1993.
6. J. Boussinesq, *J. Math. Pures Appl.* **2,** 55 (1872).
7. J. D. Smith, *Numerical Solution of Partial Differential Equations, Finite Difference Methods* (Clarendon Press, Oxford, 1985).
8. G. J. Haltiner and R. T. Williams, *Numerical Prediction and Dynamic Metereology* (Wiley, New York, 1980).
9. J. J. O'Brien, *Advanced Physical Oceanographic Modelling* (Reidel, Dordrecht, 1986).
10. K. Bryan, *J. Comput. Phys.* **4,** 347 (1969).
11. C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics*, Computational Physics Series (Springer-Verlag, New York, 1988).
12. I. Daubechies, *Ten Lectures on Wavelets*, Regional Conference Series in Applied Mathematics, Vol. 61 (SIAM, Philadelphia, 1992).
13. J. M. Restrepo, G. K. Leaf, and G. Schlossnagle, MCS Preprint MCS-P423-0394, (Argonne National Laboratory, Washington, DC, 1994).
14. S. Qian and J. Weiss, *J. Comput. Phys.* **106,** 155 (1993).
15. K. McCormick and R. O. Wells, Rice University Technical Report CML TR91-02, 1992 (unpublished).
16. G. Beylkin, *SIAM J. Numer. Anal.* **6,** 1716 (1992).