JOURNAL OF COMPUTATIONAL PHYSICS 22, 501–516 (1976)

The Application of the Spectral Method to Nonlinear Wave Propagation

HANS SCHAMEL AND KLAUS ELSÄSSER

Ruhr-Universität Bochum, Theoretische Physik I, 4630 Bochum, West Germany

Received September 29, 1975; revised February 13, 1976

We integrate the Korteweg–de Vries/Burgers equation numerically by using the spectral and pseudospectral method, respectively. Comparing the results with analytic solutions, we show that the aliasing interactions within the pseudospectral method lead to errors increasing in time, while the spectral method gives the correct time evolution. It is shown both analytically and by the numerical solutions that three invariants of the Korteweg–de Vries equation are conserved by both; therefore the number of invariants of any scheme is not decisive for a good approximation of the continuous solutions. Finally, we apply the spectral method to calculate the time evolution of turbulent sound waves in one and two space dimensions.

1. INTRODUCTION

In periodic systems a competitive numerical procedure to finite difference methods is the Fourier method. It represents a special case of Galerkin's method with Fourier modes as the appropriate eigenfunctions. The time evolution of a given system is followed by integrating the corresponding set of equations in Fourier space. Derivatives in real space are well represented, and for a given number of degrees of freedom (i.e., number of gridpoints) the Fourier method is more accurate than finite difference methods which usually show phase errors [1]. Also invariants can be incorporated more easily (see Section 4). The drawback of this method, on the other hand, is that nonlinear terms which are local in $\mathbf{x}$ space are convolution sums in Fourier space, and therfore instead of N^d calculations where d is the number of space dimensions, and N is the number of modes in each direction, $(N^d)^2$ calculations are involved. One circumvents this ineffective procedure by transforming back to real space, performs there the local product, and returns then to k space. Using time saving fast Fourier transform methods, the number of operations is reduced to $rN^d \log_2 N^d$, where r is the number of Fourier

transformations involved. In the last step, however, additional terms appear which are due to the truncation of Fourier expansion. The question is then whether these terms called aliasing or umklapp terms should be removed or not, in order to provide an approximation of the continuous problem. The first case will be named spectral method (SM), the second case pseudospectral method (PSM). Orszag [1] first concluded from studies of the passive scalar convection problem and the Taylor–Green vortex–decay problem that alias-free schemes (SM) are the appropriate ones for simulations of rotational incompressible flows. His decision was, however, based on a programming error for the PSM [2], and a repetition of the calculations gave similar accurate results for both methods; therefore they concluded that "aliasing error" is a misnomer. For future spectral calculations they recommend therfore PSM since it is at least twice as efficient, and is easier to handle. SM has been used by Salu and Knorr [3] to simulate two-dimensional guiding center plasmas. Kreiss and Oliger [4] and Fornberg [5] used a Fourier representation for space derivatives and compared this with finite difference methods. When this method of advancing the unknowns in time in **x** space is translated to **k** space it becomes equivalent to PSM (see Section 2). In this paper we present arguments for the use of alias-free schemes. We solved numerically two well-known equations for which exact analytic solutions are known and have, therefore, additional information to test the importance of aliasing interactions. What we found is that the aliasing interactions alter the nonlinear behavior and therefore the time evolution of the system. The differences of both methods originate in the small scale structure[1], i.e., at scales comparable with the grid distance and extend in the course of time on the whole spectrum. We therefore propose the use of SM to give the best resolution for a given number of *gridpoints* (*modes*). We mention that for both schemes the invariants computed are well preserved so that on this basis a decision in favor of one scheme cannot be made.

The present paper is organized in the following way. In Section 2 we briefly discuss the origin of aliasing interactions and quote a procedure to remove them. In Section 3 the first-order KdV–Burgers' equation and a corresponding second-order equation in time are transformed into infinite Fourier space; two solutions, the soliton and the smoothed sawtooth, are presented. In Section 4 we discuss the truncated system of equations for the KdV equation and show that the first three invariants of this equation are conserved. The numerical results are presented in Section 5, and the application to random acoustic waves follows in the last section.

[1] Similar differences have been reported in the paper of Fox and Orszag [2], but have been accounted for by SM.

2. The Aliasing Interactions

Given a grid in $\mathbf{x}$ (resp. $\mathbf{k}$) space with N^d equally spaced mesh points the transformation from one space to the other is accomplished by the truncated Fourier expansion

$$\begin{aligned} u(\mathbf{x}) &= \sum_{\mathbf{k}\in B_k} \tilde{u}(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}), \qquad \mathbf{x}\in B_X, \\ \tilde{u}(\mathbf{k}) &= (1/N^d) \sum_{\mathbf{x}\in B_X} u(\mathbf{x}) \exp(-i\mathbf{k}\cdot\mathbf{x}), \qquad \mathbf{k}\in B_k, \end{aligned} \tag{1}$$

where the d-dimensional boxes B_X (resp. B_k) are given by

$$\begin{aligned} B_X &= \{\mathbf{x} \mid \mathbf{x} = (1/N)(n_1, n_2, \ldots, n_d);\ 0 \leqslant n_i < N;\ i = 1, 2, 3, \ldots, d\}, \\ B_k &= \{\mathbf{k} \mid \mathbf{k} = 2\pi(m_1, m_2, \ldots, m_d);\ -K \leqslant m_i < K;\ i = 1, 2, 3, \ldots, d\}, \end{aligned} \tag{2}$$

n_i and m_i being integers and $N = 2K$ (usually a power of 2). Later on we will also refer to the box B_K which is B_k without the factor 2π. An important property of this transformation is that

$$\sum_{\mathbf{x}\in B_X} \exp(i\mathbf{k}\cdot\mathbf{x}) = N^d \delta_{\mathbf{k},2\pi\mathbf{e}N}, \tag{3}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_d)$; $e_i = 0, \pm 1, \pm 2, \pm 3, \ldots$; $\delta_{\mathbf{k},\mathbf{q}} = \delta_{k_1,q_1}\delta_{k_2,q_2}\cdots\delta_{k_d,q_d}$; and $\delta_{k,q}$ is the Kronecker symbol. The corresponding orthogonality condition for the infinite Fourier series reads instead

$$\int_0^1 d\mathbf{x} \exp(i\mathbf{k}\cdot\mathbf{x}) = \delta_{\mathbf{k},0}. \tag{3a}$$

The truncation yields to additional terms on the r.h.s. of Eq. (3), i.e., the components of $\mathbf{k}$ are only determined modulo $2\pi N$. The terms with $\mathbf{k} + 2\pi N\mathbf{e}$ are aliases of $\mathbf{k}$ on the discrete grid. Thus if we transform the local product $F(\mathbf{x}) = u(\mathbf{x})\, v(\mathbf{x})$ which stands, e.g., for hydrodynamic nonlinearities we get for the Fourier transforms (denoted by tilde):

$$\tilde{F}(\mathbf{k}) = \sum_{\mathbf{p},\mathbf{q}} \tilde{u}(\mathbf{p})\, \tilde{v}(\mathbf{q})\, \delta_{\mathbf{p}+\mathbf{q}-\mathbf{k},0} + \sum_{\mathbf{p},\mathbf{q},\mathbf{e}} \tilde{u}(\mathbf{p})\, \tilde{v}(\mathbf{q})\, \delta_{\mathbf{p}+\mathbf{q}-\mathbf{k},2\pi\mathbf{e}N}, \tag{4}$$

$\mathbf{p}, \mathbf{q}, \mathbf{k} \in B_k$; $e_i = 0, \pm 1$; $\mathbf{e} \neq (0, 0, \ldots)$. The second sum represents the aliasing interactions. They come into play for large enough arguments. In one dimension, e.g., one can see that there are no aliasing contributions if all three arguments satisfy $0 \leqslant |p|, |q|, |k| < (4\pi/3)K$. Although the involved functions will be, in general, decreasing functions of the wave vector, this, however, does not imply that the aliasing interactions play a negligible role. This will be shown by our numerical examples.

In the examples, two different schemes are tested. In the first scheme, corresponding to SM, all aliasing interactions are eliminated by a procedure shown below. This alias-free scheme represents a continuous periodic system which is transformed to infinite discrete **k** space and then truncated by introducing a cutoff. In second scheme, PSM is obtained first by discretizing the **x** space and then by transforming the equations into **k** space by means of the truncated Fourier expansion (1).

The dealiasing procedure has been described by Patterson and Orszag [6]. It consists of introducing shifted grids in **x** space and relating the dependent variables on it. The aliasing-free $\tilde{F}(\mathbf{k})$ is then obtained by forming the expression

$$\sum_{\hat{\mathbf{e}}} (2N)^{-d} \sum_{\mathbf{x} \in B_X} u(\mathbf{x} + \hat{\mathbf{e}}/2N)\, v(\mathbf{x} + \hat{\mathbf{e}}/2N) \exp[-i\mathbf{k} \cdot (\mathbf{x} + \hat{\mathbf{e}}/2N]). \tag{5}$$

The summation over $\hat{e}$ where $\hat{e}_i = 0, 1$; $i = 1, 2, \ldots, d$. Therefore $2^d - 1$ more Fourier transformations are involved in SM than in PSM at every time step.

Fox and Orszag [2], and also Orszag [7], pointed out that the error made in truncating the spectral series is more uniformly distributed in SM than in PSM; this is achieved by a nonlocal representation of the nonlinear term (see Eq. (5)), a device which has also been successfully used in finite difference methods (see, e.g., [8–10]).

3. Two Nonlinear Equations

The two equations used to test both schemes are

(i) the Korteweg–de Vries–Burgers' equation

$$\varphi_t + (1 + \varphi)\, \varphi_x - (\nu/2)\, \varphi_{xx} + (\lambda_D^2/2)\, \varphi_{xxx} = 0; \tag{6}$$

(ii) the second-order model equation

$$\varphi_{tt} - \nabla^2\varphi - \nabla^2\varphi^2 - \nu\nabla^2\varphi_t - \lambda_D^2(\nabla^2)^2\, \varphi = 0. \tag{7}$$

Both equations describe the propagation of finite amplitude ion sound waves in a lossy plasma (note that the ion sound speed has been chosen as unity). Whereas in Eq. (6) only waves propagating in one direction can be followed, Eq. (7) includes all directions of propagation. As shown earlier [11] this second-order scalar equation can be derived from the full hydrodynamic set of equations if the resonant interactions between three modes are properly taken into account. The modifications due to the changed nonresonant interactions are expected to be a higher-order effect in the small amplitude.

Both equations can be represented by an amplitude equation of the type

$$(\partial/\partial t)\, C_{\mathbf{k}_1}(t) = -i\omega_{k_1}\left(C_{\mathbf{k}_1} + \tfrac{1}{2}\sum_{\mathbf{k}_2,\mathbf{k}_3}{}' v_{-\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3} C_{\mathbf{k}_2} C_{\mathbf{k}_3}\right) + D_{\mathbf{k}_1}. \tag{8}$$

The first equation reduces to this form by applying the (infinite) Fourier transformation

$$\varphi(x, t) = \sum_{k=-\infty}^{+\infty} C_k(t) \exp(ikx), \tag{9}$$

with

$$\begin{aligned} \omega_{k_1} &= k_1(1 - (k_1\lambda_D)^2/2), \\ v_{k_1,k_2,k_3} &= (k_1/\omega_{k_1})\, \delta_{k_1+k_2+k_3,0}, \\ D_{k_1} &= -(\nu/2)\, k_1^2 C_{k_1} \end{aligned} \tag{10}$$

(the reality condition reads $C_{-k} = C_k^*$). To transform the multidimensional Eq. (7) into Eq. (8) we set

$$\begin{aligned} \varphi(\mathbf{x}, t) &= \sum_{\alpha=\pm 1} \sum_{k=-\infty}^{+\infty} C_{\mathbf{k}}^{\alpha}(t) \exp(i\mathbf{k}\cdot\mathbf{x}), \\ \dot{\varphi}(x, t) &= \sum_{\alpha=\pm 1} \sum_{k=-\infty}^{+\infty} (-i\omega_k^{\alpha})\, C_{\mathbf{k}}^{\alpha}(t) \exp(i\mathbf{k}\cdot\mathbf{x}). \end{aligned} \tag{11}$$

In writing (11) we used the quantum mechanic gauge condition which becomes with $C_{\mathbf{k}}^{\alpha}(t) = c_{\mathbf{k}}^{\alpha} \exp(-i\omega_k^{\alpha} t)$,

$$\sum_{\alpha=\pm 1} \dot{c}_{\mathbf{k}}^{\alpha} \exp(-i\omega_k^{\alpha} t) = 0. \tag{12}$$

It can be shown that the amplitude equation satisfies Eq. (12) for all times. For a given $\mathbf{k}$ the index α represents the direction of propagation. The reality condition reads

$$C_{-\mathbf{k}}^{-\alpha} = C_{\mathbf{k}}^{\alpha *} \qquad \omega_{-k}^{-\alpha} = -\omega_k^{\alpha}. \tag{13}$$

It relates $C_{\mathbf{k}}^{-}$ with $C_{\mathbf{k}}^{+}$. With this transformation, the complex function $C_{\mathbf{k}}^{+}(t)$ is completely determined by the two real functions $\varphi(\mathbf{x}, t)$ and $\dot{\varphi}(\mathbf{x}, t)$ through

$$C_{\mathbf{k}}^{+}(t) = \tfrac{1}{2}\left[\int_0^1 d\mathbf{x}\{\varphi(\mathbf{x}, t) + i\dot{\varphi}(\mathbf{x}, t)/|\, \omega_k^{+}\,|\}\right] \exp(-i\mathbf{k}\cdot\mathbf{x}), \tag{14}$$

and vice versa. If we identify $C_{\mathbf{k}}^{+}$ with $C_{\mathbf{k}}$ and extend the sum in (8) to include also the summation over α_2, α_3, $(= \pm 1)$ indicated by prime, then Eq. (7) reduces to Eq. (8) with the expressions

$$\begin{aligned} \omega_{k_1} &\equiv \omega_{k_1}^{+} = + \mid k_1 \mid (1 - (k_1\lambda_D)^2/2), \\ v_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3} &= (k_1^2/\omega_{k_1}^2)\, \delta_{\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3,0}\,, \\ D_{\mathbf{k}_1} &= -(\nu/2)\, k_1^2 (C_{\mathbf{k}_1} - C_{-\mathbf{k}_1}^{*}). \end{aligned} \tag{15}$$

Note that the Debye length λ_D has to be so small that $(k_1\lambda_D)^2$ times a typical wave amplitude can be neglected as must be assumed in deriving (6) and (7). The initial value problems (6) and (7) are therefore transformed to an initial value problem (8) in Fourier space. Since only a finite number of modes can be represented in a computer we have to introduce a cutoff, i.e., a maximal wavenumber (in our case $k_{MAX} = 2\pi K$). This leads us to the problem discussed above.

For the test of both schemes in one dimension we use two known exact solutions of Eqs. (6) and (7).

(a) The solitary wave ($\nu = 0$)

$$\varphi(x, t) = u_0 + \Delta u \operatorname{sech}^2[(\Delta u/6\lambda_D^2)^{1/2}(x - ct)], \tag{16}$$

$$c = 1 + u_0 + \Delta u/3 \qquad \text{for Eq. (6)}, \tag{16a}$$

$$= (1 + 2u_0 + 2\,\Delta u/3)^{1/2} \qquad \text{for Eq. (7)}. \tag{16b}$$

The soliton approximates long wavelength periodic cnoidal waves. The reason for introducing u_0 instead of zero is that we set $C_{k_1=0} = \int_0^1 dx\, \varphi(x, t) = 0$ for all times. This implies

$$u_0 = -2\lambda_D(6\,\Delta u)^{1/2} \tanh[(\Delta u/24)^{1/2}/\lambda_D]. \tag{17}$$

(b) the smoothed sawtooth ($\lambda_D = 0$)

$$\varphi(x, t) = \frac{\Delta u(t)}{2}\left\{-\tanh\left[\frac{\Delta u(t)(x - t)}{2\nu}\right] + 2(x - t)\right\}, \tag{18}$$

$$\Delta u(t) = \frac{\Delta u(0)}{1 + \Delta u(0)\, t} \tag{18a}$$

(Saffman [12]). The first solution is stationary in time whereas the second solution is time-dependent.

In the case of the KdV–equation ($\nu = 0$) there exists an infinite number of invariants, the first of which we have already mentioned,

$$I_1 = \int_0^1 dx\, \varphi(x, t) = C_0\,.$$

The next two are

$$I_2 = \int_0^1 dx\, \varphi^2(x,t) = \sum_{k_1=-\infty}^{+\infty} |C_{k_1}|^2, \tag{19a}$$

$$I_3 = \int_0^1 dx[\varphi^3/3 - \lambda_D{}^2(\varphi_x)^2/2]$$

$$= \tfrac{1}{3} \sum_{k_1,k_2,k_3=-\infty}^{+\infty} C_{k_1}C_{k_2}C_{k_3}\delta_{k_1+k_2+k_3,0} - (\lambda_D{}^2/2) \sum_{k_1=-\infty}^{+\infty} k_1{}^2 |C_{k_1}|^2. \tag{19b}$$

In the next section we show that analogous invariants exist for the truncated set of equations.

4. The Truncated KdV System and Its Invariants

The truncation of the system of equations (8) for the first-order KdV equation will be accomplished to maintain the symmetry in **k** space. For this reason we define a symmetric box in **k** space which is defined by

$$B_K{}^- = \{K_1 \mid -K+1 \leqslant K_1 \leqslant K-1\},$$

and equals B_K, except for the cutoff term ($K_1 = -K$) which is excluded in $B_K{}^-$. The truncated system corresponding to (8) is then given by

$$(\partial/\partial t)\, C_{K_1} = -i\omega_{k_1}\Big(C_{K_1} + \tfrac{1}{2} \sum_{\substack{K_2,K_3 \in B_K \\ e}} v_{-K_1,K_2,K_3} C_{K_2} C_{K_3}\Big), \qquad K_1 \in B_K{}^-, \tag{20a}$$

$$(\partial/\partial t)\, C_{K_1} = C_{K_1} = 0, \qquad K_1 = -K, \tag{20b}$$

where ($k_1 = 2\pi K_1$; $C_{K_1} \equiv C_{k_1}$)

$$v_{K_1,K_2,K_3} = (k_1/\omega_{k_1})\, \delta_{K_1+K_2+K_3,eN}\,, \tag{21}$$

$$\omega_{k_1} = k_1(1 - (k_1\lambda_D)^2/2). \tag{22}$$

For SM, e is zero, whereas $e = 0, \pm 1$ in the case of PSM, including therefore the aliasing interactions. This system has the important property to maintain the reality condition $C_{-K_1} = C_{K_1}^*$, $K_1 \in B_K{}^-$ and $C_{-K} = C_{-K}^*$. With this definition we only make use of $N-1$ degrees of freedom instead of N. Note that for the second-order scalar equation, K_1 in (20a) need not be restricted to $B_K{}^-$ to maintain the reality condition. In the latter case, K_1 may belong to the whole box B_K, and the reality condition is satisfied automatically by relating $C_K{}^-$ to $C_K{}^+$ via (13).

It is now easily seen from (20) that if $C_{K_1=0}$ is zero initially it will remain zero so that I_1 is conserved.

After truncation the second invariant becomes

$$I_2 = \sum_{K_1 \in B_K} | C_{K_1} |^2. \tag{23}$$

We show that the expression

$$\dot{I}_2 = \sum_{K_1 \in B_K} C^*_{K_1} \dot{C}_{K_1} + \text{c.c.}$$

vanishes. Inserting (20) we get

$$\dot{I}_2 = \sum_{K_1 \in B_K} C^*_{K_1} \left[-i\omega_{K_1} C_{K_1} - i\pi K_1 \sum_{\substack{K_2, K_3 \in B_K \\ e}} C_{K_2} C_{K_3} \delta_{K_2+K_3-K_1, eN} \right] + \text{c.c.} \tag{24}$$

Since $C^*_{-K} = 0$, we make no error if (20a) is inserted for $\dot{C}_{-K}$ instead of $\dot{C}_{-K} = 0$. The first term cancels out the corresponding complex conjugate term; therefore

$$\dot{I}_2 = -i\pi \sum_{\substack{K_1, K_2, K_3 \in B_K \\ e}} K_1 (C^*_{K_1} C_{K_2} C_{K_3} - C_{K_1} C^*_{K_2} C^*_{K_3}) \, \delta_{K_2+K_3-K_1, eN} \, .$$

The second term equals the first. To see this we introduce the negative of K_1, K_2, K_3, e in the second term and use the reality condition $C_{-K_1} = C^*_{K_1}$ to get the first term. Replacing K_1 by its negative and symmetrizing it, we get

$$\dot{I}_2 = 2\pi i \sum_{\substack{K_1, K_2, K_3 \in B_K \\ e}} ((K_1 + K_2 + K_3)/3) \, C_{K_1} C_{K_2} C_{K_3} \delta_{K_1+K_2+K_3, eN} \, , \tag{25}$$

which is zero, independent of whether the aliasing terms are removed ($e = 0$; SM) or not ($e = 0, \pm 1$; PSM). In the latter case, the terms with $e = \pm 1$ cancel because of periodicity of C_{K_1} ($= C_{K_1+N}$).

The proof of the invariance of I_3 is strongly simplified by the use of the Hamilton formalism. The third invariant becomes

$$I_3 = \sum_{K_1 \in B_K} (\delta\omega_{k_1}/k_1) \, | C_{K_1} |^2 + \tfrac{1}{3} \sum_{\substack{K_1, K_2, K_3 \in B_K \\ e}} C_{K_1} C_{K_2} C_{K_3} \delta_{K_1+K_2+K_3, eN} \tag{26}$$

for the truncated Fourier variables, where $\delta\omega_{k_1} = -k_1(k_1\lambda_D)^2/2$. This expression is therefore different for both methods, due to the aliasing interactions in the trilinear term. We define a new variable by

$$A_{K_1} = (\omega_{k_1}/k_1)^{1/2} \, C_{K_1} \, , \tag{27}$$

which satisfies

$$\dot{A}_{K_1} = -i\omega_{k_1}\left[A_{K_1} + \tfrac{1}{2}\sum_{\substack{K_2,K_3\in B_K^- \\ e}} (k_1k_2k_3/\omega_{k_1}\omega_{k_2}\omega_{k_3})^{1/2}\, A_{K_2}A_{K_3}\delta_{K_2+K_3-K_1,eN}\right] \qquad K_1 \in B_K^-, \quad (28)$$

and

$$\dot{A}_{-K} = A_{-K} = 0.$$

This expression allows us to find a Hamiltonian

$$H = (1/2!)\sum_{K_1\in B_K^-} |A_{K_1}|^2 + (1/3!)\sum_{\substack{K_1,K_2,K_3\in B_K^- \\ e}} (k_1k_2k_3/\omega_{k_1}\omega_{k_2}\omega_{k_3})^{1/2}\, A_{K_1}A_{K_2}A_{K_3}\delta_{K_1+K_2+K_3,eN}\,, \quad (29)$$

from which (28) follows by means of the Hamiltonian equation of motion

$$\dot{A}_{K_1} = -i\omega_{k_1}(\partial H/\partial A_{-K_1}) \qquad K_1 \in B_K\,. \quad (30)$$

Since H does not depend on A_{-K} the equation for the cutoff term $\dot{A}_{-K} = 0$ is also obtained by (30). H is a constant of motion

$$\dot{H} = \sum_{K_1\in B_K^-} \dot{A}_{K_1}(\partial H/\partial A_{K_1}) = \sum_{K_1\in B_K^-} (-i\omega_{k_1})(\partial H/\partial A_{-K_1})(\partial H/\partial A_{K_1}) = 0.$$

This expression is zero because ω_{k_1} is antisymmetric with respect to K_1 while the remainder is symmetric. This is obviously true for both methods. In terms of the original variable C_{K_1}, the Hamiltonian can be written as

$$H = (1/2!)\sum_{K_1\in B_K} (1 - (k_1\lambda_D)^2/2)\,|C_{K_1}|^2 + (1/3!)\sum_{K_1,K_2,K_3\in B_K} C_{K_1}C_{K_2}C_{K_3}\delta_{K_1+K_2+K_3,eN}\,, \quad (31)$$

where we used (22) and added terms which are zero by using B_K instead of B_K^-. H is a linear combination of I_2, I_3; $H = (I_2 + I_3)/2$, from which follows that I_3 is invariant. Both methods preserve, therefore, the first three invariants, whereas the third invariant is not conserved by the finite difference scheme of Zabusky and Kruskal [8]. We note that conservation is meant in the sense of semiconservation because the error made by time-differencing is disregarded.

5. Numerical Results

The system we followed numerically is (20a) with $K_1 \in B_K$ which is appropriate for the second-order scalar equation. For the first-order KdV–Burgers' equation,

the reality condition is not conserved if Eq. (20a) instead of Eq. (20b) is used at the cutoff $K_1 = -K$, but the numerical solutions showed to be not sensitive to this change.

The main part of our numerical program, the subroutine FALT 2 which computes the convolution sum, had been tested independently. Given two complex (in general, three-dimensional) arrays $\tilde{u}(\mathbf{k}_1)$, $\tilde{v}(\mathbf{k}_1)$ $(\mathbf{k}_1 \in B_k)$. FALT 2 computes

$$\tilde{F}(\mathbf{k}_1) = \sum_{\substack{\mathbf{k}_2, \mathbf{k}_3 \in B_k \\ e}} \tilde{u}(\mathbf{k}_2)\, \tilde{v}(\mathbf{k}_3)\, \delta_{\mathbf{k}_2+\mathbf{k}_3-\mathbf{k}_1, 2\pi \mathbf{e} N} \qquad (k_1 \in B_k)$$

without ($LF = 1$) and with aliasing terms ($LF = 0$). For $LF = 1$ we used the dealiasing procedure of Patterson and Orszag [6] mentioned in Section 2. All parts of this subroutine had been checked by choosing

$$u(k_1) = v(k_1) = \exp[ik_1/N],$$

for which $\tilde{F}(k_1)$ is known.

We followed the truncated system in time starting with $C_K(0)$ which is given by the inverse Fourier transformation of $\varphi(x, 0)$ (and $\dot{\varphi}(x, 0)$ in the second-order case (see Eq. (14)). The time integration had been carried out by a stable fourth-order integration procedure which uses a predictor–corrector method to measure the local truncation error (IBM subroutine HPCG). The error weights have been chosen to be proportional to the inverse input amplitude $|\, C_K(0)|$ so that the relative truncation error δ could be estimated. If δ exceeded the upper bound $\delta_c = 10^{-5}$ the time increment $\Delta T = 0.02$ had been halved. The computer time T was the real time t times $N/4$ so that $T = 1$ represents the period of the mean mode $|\, k\,| = 2\pi N/4$; we usually chose $N = 64$. In the case of the soliton solving KdV equation we computed the two invariants I_2 and I_3 given by (23) and (26) to have a further control of the accuracy of the schemes.

In Fig. 1 the soliton with $\lambda_D = 10^{-2}$, $\Delta u = 0.2$ is shown at $t = 1.25$ corresponding to 1000 time steps.[2] Initially the soliton was centered at $x = -0.5$ (which corresponds to $x = 0.5$ because of periodicity). φ_{num1} (dotted curve) was obtained with the alias-free scheme corresponding to SM; in φ_{num2} (dashed curve) the aliasing interaction was not eliminated (PSM). We see that only the alias-free solution agrees with the analytical solution φ_{analyt} (full curve). The aliasing interactions weaken the nonlinearity. The soliton decays into a smaller one plus a dispersive ion acoustic wave packet which is typical for weaker nonlinearities. For both runs the invariants are found to be conserved within a relative error of 2×10^{-3}. Note that the delay of the smaller soliton results from the weaker nonlinearity and is not due to linear phase errors which are zero for both methods.

[2] Only one-half of the periodicity interval [0, 1) is plotted.

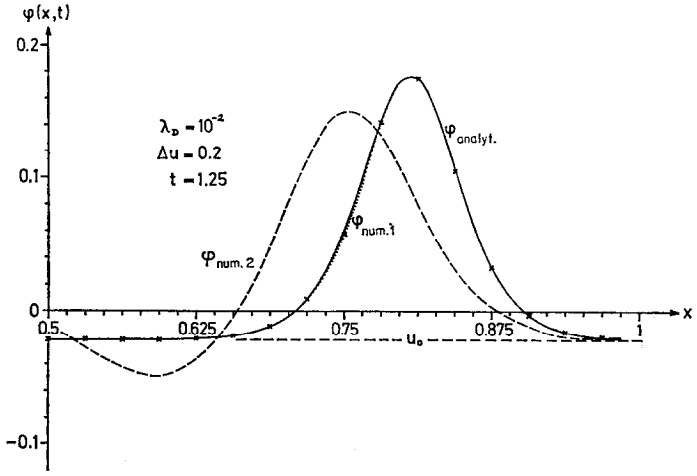

FIG. 1. The KdV soliton at $t = 1.25$. φ_{analyt} corresponds to Eq. (16). φ_{num1} is computed with SM (dotted line $N = 64$, crosses $N = 32$). φ_{num2} is computed with PSM (dashed line $N = 64$).

Figure 2 shows the corresponding spectrum for the alias-free scheme at 10 equidistant time steps ($\Delta t = 0.125$). The spectrum is defined in Eq. (32). Only near the cutoff where the spectrum is 10 orders of magnitude smaller are some changes seen (blocking error). This is a direct computer output showing again the stationary character of the alias-free solution. The corresponding spectrum for PSM is seen in Fig. 3. The change of the spectrum sets in at high wavenumbers and propa-

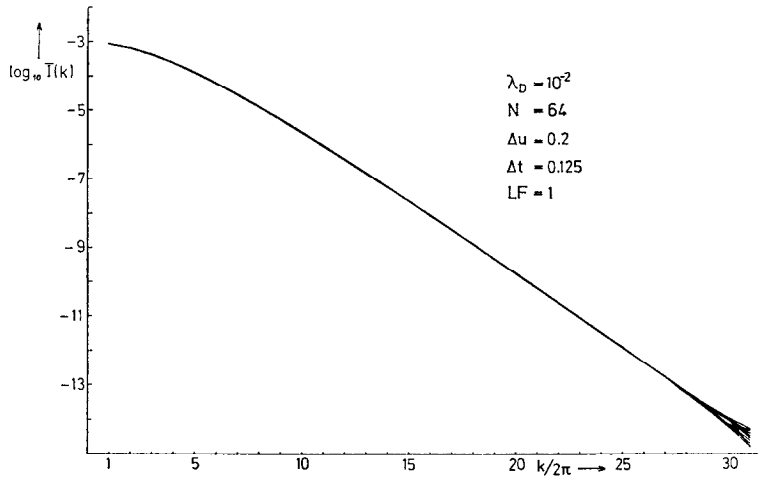

FIG. 2. The spectrum corresponding to φ_{num1} ($N = 64$) at 10 equidistant time-steps.

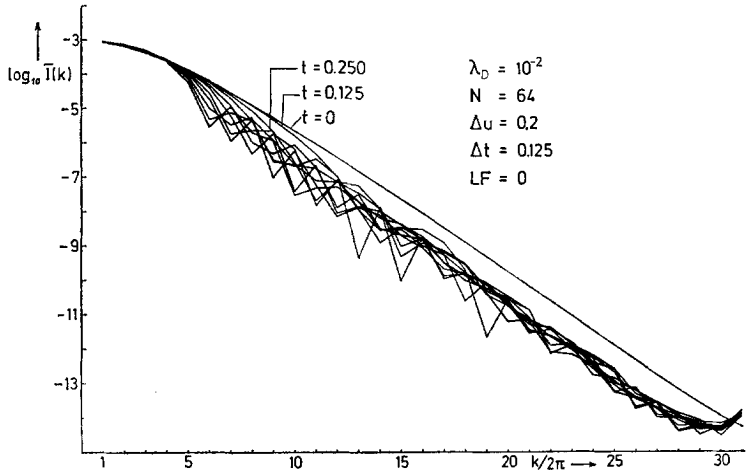

FIG. 3. The spectrum corresponding to φ_{num2} ($N = 64$) at 10 equidistant time-steps.

gates to smaller k. This is not surprising because the aliasing interactions $\sum C_{K_2} C_{K_3} \delta_{K_2+K_3-K_1, eN}$ will affect most strongly amplitudes C_{K_1} with K_1 near the cutoff ($| K_1 | \simeq K$). In this region the aliasing terms are of the same order as the true interaction terms[3]. The system tries to adjust to an asymptotic state which is characterized by a broader soliton whose spectrum decreases faster, superimposed by a dispersive wave train, as seen from the jagged part of the spectrum. We realize that in spite of an energy range of more than 10 orders of magnitude, the aliasing interactions come into play, modifying the whole range of the spectrum. The dealizasing procedure therefore, in long time runs, cannot be avoided by shifting the cutoff to higher k, i.e., by including more modes. This is supported by a run with the alias-free scheme and half the number of modes ($N = 32$). The soliton at $t = 1.25$ (crosses in Fig. 1) coincides with the analytic soliton.[4] The spectrum for this run seen in Fig. 4 confirms the high quality of the spectral method.

Similar results are obtained for the smoothed sawtooth (18). Figure 5 shows

[3] An example: $K_2 = -K/2$, $K_3 = -K/2$, $K_1 = -K$ (true nonlinear term), $K_2 = K/2$, $K_3 = K/2$, $K_1 = -K$ (aliasing term with $e = 1$).

[4] One might assume that there was a programming error in scheme 2, but we have several arguments against this assumption:

(i) Three invariants are conserved.

(ii) Scheme 2 (PSM) is simpler than scheme 1 (SM), which obviously works well.

(iii) A loop in FALT 2 is used only once at each step in scheme 2 but twice for scheme 1. Therefore, if scheme 2 is wrong, then it is probable that scheme 1 is wrong also.

(iv) FALT 2, which could be the only source of error, has been tested independently.

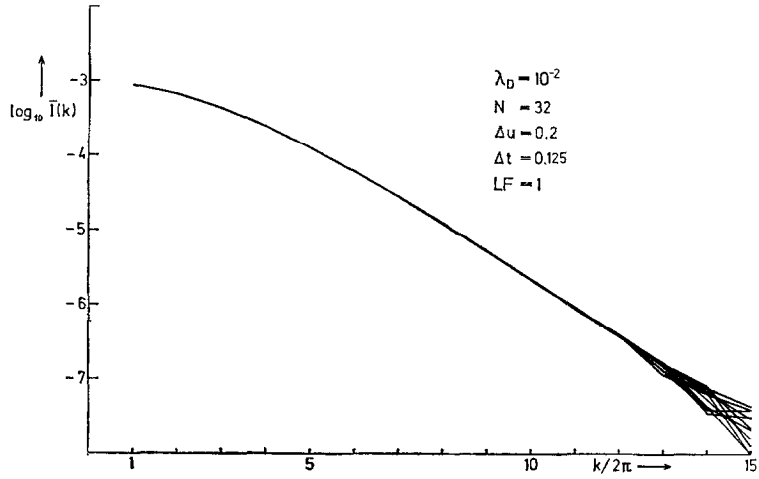

FIG. 4. The spectrum corresponding to φ_{num1} ($N = 32$) at 10 equidistant time-steps.

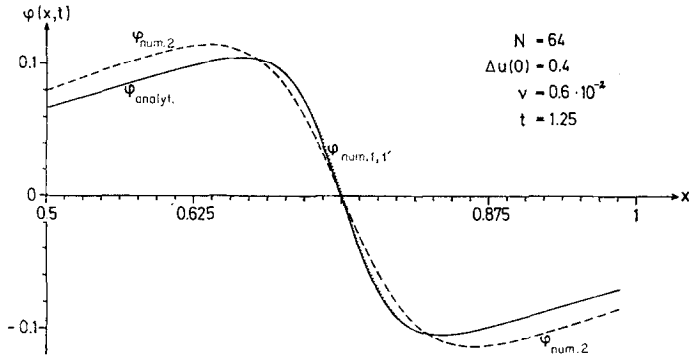

FIG. 5. The smoothed sawtooth at $t = 1.25$; φ_{analyt} corresponds to Eq. (18). φ_{num1} is computed with SM using Burgers' equation, Eq. (6) (dotted line $N = 64$). $\varphi_{\text{num1'}}$ is computed with SM using the second-order equation, Eq. (7) (identical with dotted line). φ_{num2} is computed with PSM using Burgers' equation.

runs for $\Delta u(0) = 0.4$, $\nu = 0.6 \times 10^{-2}$. Except for $\varphi_{\text{num1'}}$, which is calculated using the second-order equation, all solutions are obtained by integrating the truncated form of Eq. (8) corresponding to Burgers' equation. Again the alias-free scheme 1 (1') agrees with the analytical result, whereas scheme 2 fails. As in the previous case, the aliasing interactions weaken the nonlinearity. The steepening process leading to steeper gradients and therefore stronger dissipation is less pronounced. These results therefore suggest the use of alias-free schemes in studying nonlinear wave phenomena.

6. Application and Summary

As an application we followed the evolution of turbulent sound waves ($\lambda_D = 0$; $d = 2$ or $d = 1$) which are governed by the second-order equation, Eq. (7). We prescribed initially $C(0)$ by assuming random phases and a power-law spectrum

$$| C_{\mathbf{K}} |^2 \propto | \mathbf{K} |^{-s}, \qquad s = 6.5.$$

The constant proportionality had been chosen such that

$$\sum_{\mathbf{K}} | C_{\mathbf{K}} |^2 = \epsilon^2,$$

where the small parameter ϵ was given.

In Table I we summarize the main parameters for three runs ($N = 32$). Figure 6

TABLE I

Parameters for Three Runs

	Run number		
Parameter	1	2	3
d	2	2	1
ν	2×10^{-3}	2×10^{-2}	2×10^{-2}
ϵ	0.1	0.1	0.1

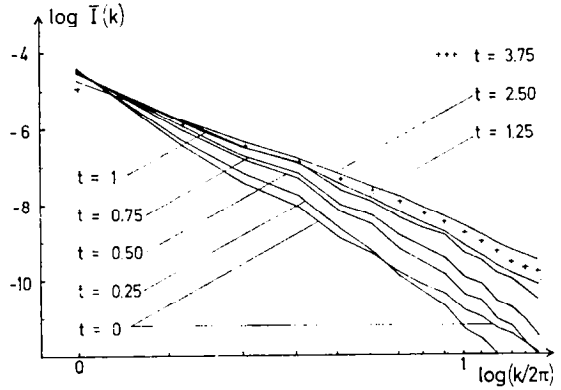

Fig. 6. The spectrum for sound waves of small but finite dissipation (run 1) at seven equidistant time-steps.

shows the time evolution of the spectrum for run number 1. The spectrum is defined by

$$\begin{aligned} \bar{I}(k) &= |C_k|^2 + |C_{-k}|^2 && d = 1, \\ &= \sum_k (|C_k|^2/2\pi\kappa), \qquad \kappa = \left(\sum k^2/n\right)^{1/2}, && d = 2. \end{aligned} \tag{32}$$

(In the last expression the sum is taken over all modes lying in an annulus of thickness 2π; n is the number of modes in it.) We see a flattening of the spectrum indicating the steepening process. The finite damping ν suppresses the accumulation of energy at the cutoff as observed in inviscid runs [11, 13]. At $t \approx 1$ a quasi-steady state is reached where the steepening process is nearly balanced by dissipation. In Fig. 7 the decay of the total wave energy,

$$H = \sum_{\mathbf{k}} |C_k|^2 + (1/3!) \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3}{}' C_{\mathbf{k}_1} C_{\mathbf{k}_2} C_{\mathbf{k}_3} \delta_{\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3,0}\,, \tag{33}$$

is shown for different runs with the same initial conditions. The upper curve shows the relaxation of run 1 into a state with approximately constant energy dissipation rate (constant slope of $H(t)$, while the two other curves (runs 2, 3) show

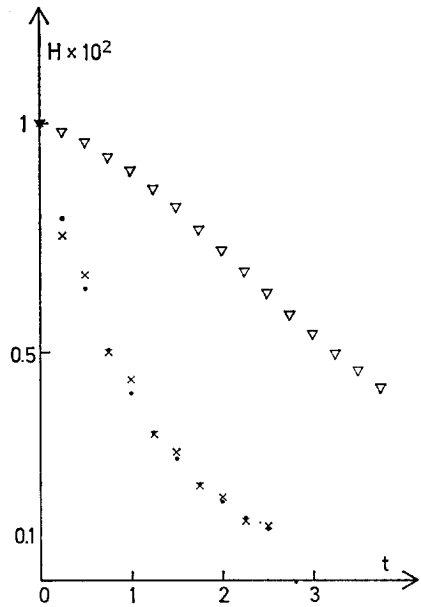

FIG. 7. The total wave energy as a function of time (▽▽▽, run 1; ···, run 2; × × ×, run 3).

the immediate decay of energy. The Reynolds number R, defined for the most energetic mode with wavenumber k_0 by

$$R = |\omega_{k_0}| \epsilon^2/\nu\epsilon k_0^2 = \epsilon/\nu k_0 ,$$

is 18 for run 1, but 10 times smaller for runs 2 and 3. Therefore only for high enough Reynolds numbers can quasi-equilibrium spectra be obtained.

Other examples for application of the spectral method are the nonlinear Schrödinger equation and mode-coupling equations describing parametric processes and involving two different groups ov waves (high-frequency and low-frequency).

In summary, by following two exact solutions of special nonlinear hyperbolic equations we could show that aliasing is indeed an error. The error originates in the small-scale structure and extends in the course of time to larger scales, thereby modifying the whole spectrum. We found that during the evolution the invariants are well preserved, which allows therefore no distinction about the quality of approximation. In addition to the numerical conservation we also have shown analytically that the first three invariants of the KdV equation are conserved by the truncated system. For first-order scalar equations (e.g., two-dimensional incompressible Navier–Stokes), the simplest way to define a truncated system which conserves reality properties and invariants for all time is to set the cutoff term identically to zero. For higher-order scalar equations, the same restriction can be shown to ensure similar invariance properties.

Acknowledgment

We would like to thank Professor G. Knorr for useful conversations.

References

1. S. A. Orszag, *J. Fluid Mech.* **49** (1971), 75.
2. D. G. Fox and S. A. Orszag, *J. Computational Phys.* **11** (1973), 612.
3. Y. Salu and G. Knorr, *J. Computational Phys.* **17** (1975), 68.
4. H. O. Kreiss and J. Oliger, *Tellus* **24** (1972), 199.
5. B. Fornberg, *SIAM J. Numer. Anal.* **12** (1975), 509.
6. G. S. Patterson and S. A. Orszag, *Phys. Fluids* **14** (1971), 2538.
7. S. A. Orszag, *Stud. in Appl. Math.* **51** (1972), 253.
8. N. J. Zabusky and M. D. Kruskal, *Phys. Rev. Lett.* **15** (1965), 240.
9. A. Arakawa, *J. Computational Phys.* **1** (1966), 119.
10. A. C. Vliegenthart, *J. Engrg. Math.* **5** (1971), 137.
11. K. Elsässer and H. Schamel, Report MPI-PAE/Astro 59, 1973.
12. P. G. Saffman, *in* "Topics in Nonlinear Physics" (N. J. Zabusky, Ed.), p. 547, Springer–Verlag, Berlin/Heidelberg/New York, 1968.
13. K. Elsässer and H. Schamel, *Z. Phys. B* **23** (1976), 89.