

# A Fourier Method for Nonsmooth Hyperbolic Problems

Knut S. Eckhoff\*

J. H. Olfsnes\*

## Abstract

Nonsmooth solutions of hyperbolic systems are computed by a modified Fourier-Galerkin method. The described approach is seen to give substantially improved accuracy compared to more traditional methods. Discontinuities are accurately resolved already on coarse grids, and the fine-structure of structured solutions is resolved on relatively coarse grids as well. The accuracy is seen to be of high order, and even for very long term integrations the global error can be kept very small if the grid is sufficiently refined.

**Key words:** hyperbolic systems, discontinuous functions, Fourier series, Bernoulli polynomials.

**AMS subject classifications:** 65M70, 65M20, 35L45, 35L67.

## 1 Introduction

As is well known [4], spectral projections of discontinuous functions exhibit global oscillations which are particularly strong near the discontinuities. This is known as the Gibbs phenomenon, and the reduced accuracy of spectral approximations due to this oscillatory behavior, makes it necessary to modify traditional spectral methods. In particular, for solutions of hyperbolic problems containing shock discontinuities, the literature most often suggests modifications by application of various filtering techniques. It has been shown that the computations may be stabilized by such techniques, and that high order accuracy may be recovered - at least away from the discontinuities.

For initial value problems for linear hyperbolic equations with discontinuous initial data, Majda et al. [16] proposed a modification of the Fourier collocation method

\*Dept. of Mathematics, University of Bergen, Johs. Brunsgr. 12, N-5008 Bergen, Norway.

ICOSAHOM'95: Proceedings of the Third International Conference on Spectral and High Order Methods. ©1996 Houston Journal of Mathematics, University of Houston.

which incorporated smoothing filters applied to the initial data and to the spatial differential operator. For nonlinear hyperbolic equations further difficulties are connected with the fact that the nonlinearities may lead to formation of shock discontinuities even when the initial data are smooth. Tadmor [18], [19] proved that it is possible to maintain stable spectral approximations while retaining spectral accuracy for nonlinear conservation laws, by application of his Spectral Viscosity method with appropriate use of post-processing filters.

In this paper we are concerned with the application of the modified Fourier method presented in [9] and further developed in [10]. The method is designed for studying nonsmooth solutions of well-posed initial value problems for systems of hyperbolic equations, and is a *shock-capturing* method where generalized step-functions are utilized in the spectral scheme as well as in the post-processing of the numerical solutions. The idea of introducing step-functions in the reconstruction of discontinuous functions was initiated by Gottlieb et al. [14], and has been further developed in [1], [2], [3], [7], [8], [9], [10], and [12].

The generalized step-functions utilized in the reconstruction of discontinuous  $2\pi$ -periodic functions from truncated Fourier series expansions in [9], [10] were introduced in [8], and were denoted by  $U_n(\xi)$ ,  $n = 0, 1, 2, \dots$ . On the interval  $0 \leq \xi < 2\pi$ , the family of  $2\pi$ -periodic functions  $U_n(\xi)$  is given by

$$(1) \quad U_n(\xi) = -\frac{(2\pi)^n}{(n+1)!} B_{n+1} \left( \frac{\xi}{2\pi} \right),$$

where  $B_j(x)$ ,  $j = 1, 2, \dots$  are the Bernoulli polynomials [11]. The function  $U_n(\xi)$  is of finite regularity for each  $n = 1, 2, \dots$ , with derivatives  $U_n^{(p)}(\xi) = U_{n-p}(\xi)$  continuous everywhere for  $p = 0, \dots, n-1$ , but with  $U_n^{(n)}(\xi) = U_0(\xi)$  only piecewise continuous with jump-discontinuities of magnitude +1 at  $\xi = 2m\pi$ ,  $m = 0, \pm 1, \pm 2, \dots$ . In fact,  $U_0(\xi)$  is a saw-tooth function, which on the interval  $(-2\pi, 2\pi)$  is given by

$$(2) \quad U_0(\xi) = \begin{cases} \frac{1}{2\pi}(-\pi - \xi) & \text{if } -2\pi < \xi < 0 \\ \frac{1}{2\pi}(\pi - \xi) & \text{if } 0 < \xi < 2\pi. \end{cases}$$

There are no jumps at the singularity locations for the higher order derivatives  $U_n^{(p)}(\xi)$ ,  $p \geq n+1$ . The Fourier coefficients for the functions  $U_n(\xi)$ ,  $n = 0, 1, 2, \dots$  are given by

$$(3) \quad (\widehat{U_n})_0 = 0, \quad (\widehat{U_n})_k = \frac{1}{2\pi(i k)^{n+1}}; \quad k \neq 0.$$

## 2 Linear hyperbolic problems

In this section we shall consider numerical solutions of periodic initial value problems for linear hyperbolic systems of the form

$$(4) \quad \begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + L\mathbf{u} &= \mathbf{0}, \\ L &= A(x, t) \frac{\partial}{\partial x} + B(x, t), \end{aligned}$$

subject to nonsmooth initial data  $\mathbf{u}(x, 0) = \mathbf{u}_0(x)$ . Here,  $\mathbf{u} = \{u_1, \dots, u_m\}^T$  denote the dependent variables, and  $A, B$  are given  $m \times m$  matrices with smooth coefficients which are  $2\pi$ -periodic with respect to  $x$ . We restrict ourselves to studying solutions  $\mathbf{u}(x, t)$  of (4) which for each  $t$  are piecewise smooth on  $[0, 2\pi]$  and  $2\pi$ -periodic with respect to  $x$ .

Following [8],  $\mathbf{u}(x, t)$  may for each  $t \geq 0$ , for an arbitrarily given integer  $Q \geq 0$  and for some finite integer  $R$  be written

$$(5) \quad \mathbf{u}(x, t) = \mathbf{u}^Q(x, t) + \mathbf{V}(x, t),$$

where

$$(6) \quad \mathbf{V}(x, t) = \sum_{n=0}^Q \sum_{j=1}^R \mathbf{a}_j^n(t) U_n(x - x_j(t)).$$

Here,  $x = x_j(t)$ ,  $j = 1, \dots, R$ , denote the characteristic curves associated with (4) across which the solution itself and/or some (or all) of its spatial derivatives suffer jump-discontinuities. Each of these characteristic curves is passing through one of the initial singularity locations at  $t = 0$ . In (5)  $\mathbf{a}_j^n(t)$  denotes the jump in the  $n$ th spatial derivative of  $\mathbf{u}$  across  $x = x_j(t)$  at the time  $t$ , and for each  $t$  the function  $\mathbf{u}^Q(x, t)$  is  $2\pi$ -periodic and at least  $Q$  times continuously differentiable everywhere with respect to  $x$ .

For any given even integer  $N > 0$  and at each instant  $t \geq 0$ , we may, to a piecewise smooth solution  $\mathbf{u}(x, t)$  of (4), associate a truncated Fourier series

$$(7) \quad P_N \mathbf{u}(x, t) = \sum_{k=-N/2+1}^{N/2-1} \hat{\mathbf{u}}_k(t) e^{ikx},$$

where

$$(8) \quad \hat{\mathbf{u}}_k(t) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{u}(x, t) e^{-ikx} dx.$$

Substituting (5) into (8), we get in view of (3) that  $\hat{\mathbf{u}}_0(t) = (\widehat{\mathbf{u}}^Q)_0(t)$  and

$$(9) \quad \hat{\mathbf{u}}_k(t) = (\widehat{\mathbf{u}}^Q)_k(t) + \sum_{n=0}^Q \sum_{j=1}^R \frac{\mathbf{a}_j^n(t) e^{-ikx_j(t)}}{2\pi(i k)^{n+1}}, \quad k \neq 0.$$

For  $N$  sufficiently large, approximate singularity locations and jumps can be computed from  $P_N \mathbf{u}(x, t)$  by the reconstruction algorithm described in [8]. Alternatively [5], [10], these quantities may be determined by integrating respectively, the *characteristic equations* and the *transport equations* associated with (4). When, in addition to  $P_N \mathbf{u}(x, t)$ , the singularity locations  $x_j(t)$  and the corresponding jumps  $\mathbf{a}_j^n(t)$  occurring in (5), (9) are known for each  $t$ ,  $P_N \mathbf{u}^Q(x, t)$  can clearly be determined from (9). Hence  $\mathbf{u}(x, t)$  may be approximately reconstructed on the form (5) with  $\mathbf{u}^Q(x, t)$  replaced by  $P_N \mathbf{u}^Q(x, t)$ . It follows from [8] that for each  $t$  the reconstructed solution is globally  $O(N^{-(Q+1)})$  accurate.

In view of the above considerations, it is reasonable to seek numerical solutions of (4) which accurately approximate  $P_N \mathbf{u}(x, t)$  in Fourier space, even though  $P_N \mathbf{u}(x, t)$  by itself may not provide an accurate approximation for the exact solution in physical space. This is the philosophy set forth in the modified Fourier method described in [9], [10], which can be summarized as follows: In (4) the matrices  $A(x, t)$  and  $B(x, t)$  with *smooth* coefficients are approximated by  $P_N A(x, t)$  and  $P_N B(x, t)$ , i.e.

$$(10) \quad L \rightarrow L_N = P_N A(x, t) \frac{\partial}{\partial x} + P_N B(x, t).$$

The following initial value problem

$$(11) \quad \begin{aligned} \frac{\partial \mathbf{u}_N}{\partial t} + P_N [L_N \mathbf{u}_{2N-2}] &= \mathbf{0} \\ \mathbf{u}_N(x, 0) &= P_N \mathbf{u}_0(x), \end{aligned}$$

then has a trigonometric polynomial solution

$$(12) \quad \mathbf{u}_N(x, t) = \sum_{k=-N/2+1}^{N/2-1} \tilde{\mathbf{u}}_k(t) e^{ikx},$$

which constitutes an accurate numerical approximation to  $P_N \mathbf{u}(x, t)$ . When accurate information about the solution is needed,  $\mathbf{u}_N(x, t)$  is post-processed by reconstruction to the form (5).

In cases where the coefficients  $A, B$  are independent of  $x$ , we have

$$(13) \quad P_N [L_N \mathbf{u}_{2N-2}] = L \mathbf{u}_N.$$

The problem (11) is, therefore, in this case readily seen to be equivalent to the problem resulting from application of the ordinary Fourier-Galerkin method to (4) in  $S_N = \text{span}\{e^{ikx} | -N/2 + 1 \leq k \leq N/2 - 1\}$ . No numerical dispersion or diffusion is introduced through the spatial discretization in this case.

If, on the other hand,  $A$  and  $B$  do depend on  $x$ , the solution  $\mathbf{u}_N$  of (11) is formally seen to depend on the Fourier-Galerkin solution  $\mathbf{u}_{2N-2}(x, t)$  in  $S_{2N-2}$ . To circumvent this difficulty, the modified Fourier method advocated in [9], [10] approximates  $\mathbf{u}_{2N-2}(x, t)$  at each instant  $t$  to the order  $O(N^{-(Q+1)})$  by de-truncating  $\mathbf{u}_N(x, t)$ . This is achieved by putting

$$(14) \tilde{\mathbf{u}}_k(t) = \sum_{n=0}^Q \sum_{j=1}^R \frac{\mathbf{a}_j^n(t) e^{-ikx_j(t)}}{2\pi(ik)^{n+1}}, \quad N/2 \leq |k| \leq N-2.$$

Here the approximate singularity locations  $x_j(t)$  and the corresponding jumps  $\mathbf{a}_j^n(t)$  are calculated either from  $\mathbf{u}_N(x, t)$  by application of the reconstruction algorithm [8], or by the method of characteristics.

The main feature of the modified method (11), (14) is that it reduces to a minimum the numerical dispersion and the numerical diffusion in the spectral approximation  $\mathbf{u}_N(x, t)$  of the nonsmooth solution  $\mathbf{u}(x, t)$ . As pointed out for the case of variable coefficients in [6], [9], the standard Fourier-Galerkin method for (4) introduces dispersive and diffusive effects by the way  $L\mathbf{u}$  is projected onto  $S_N$ . This is expected to represent the major source of error for traditional methods when dealing with nonsmooth solutions of (4).

The de-truncation (14) may also formally be written

$$(15) \quad \mathbf{u}_{2N-2}(x, t) = \mathbf{u}_N^Q(x, t) + P_{2N-2}\tilde{\mathbf{V}}(x, t),$$

where  $\tilde{\mathbf{V}}(x, t)$  is an approximation for  $\mathbf{V}(x, t)$  in (5) and  $\mathbf{u}_N^Q(x, t)$  is an approximation for  $P_N\mathbf{u}^Q(x, t)$  determined by

$$(16) \quad \mathbf{u}_N^Q(x, t) = \mathbf{u}_N(x, t) - P_N\tilde{\mathbf{V}}(x, t)$$

In view of (15), it follows that (11) is formally approximately equivalent to

$$(17) \quad \frac{\partial \mathbf{u}_N}{\partial t} + G_N[L_N \mathbf{u}_N^Q] + P_N[L_N \tilde{\mathbf{V}}] = \mathbf{0},$$

where  $G_N$  is the standard Fourier-Galerkin projection operator onto  $S_N$ . Thus, the sources of numerical dispersion and numerical diffusion stemming from the spatial discretization is, in the modified method, essentially limited to the truncation error associated with  $L_N$  and to the discretization of  $L_N\mathbf{u}^Q$ . As a result, accurate approximations for discontinuous solutions of (1) can be obtained on relatively coarse grids.

### 3 Nonlinear conservation laws

In this section we shall consider numerical solutions of periodic initial value problems for nonlinear conservation laws of the form

$$(18) \quad \frac{\partial \mathbf{u}}{\partial t} + \frac{\partial}{\partial x} \mathbf{f}(\mathbf{u}) = \mathbf{0}.$$

Here, as in the preceding section,  $\mathbf{u} = \{u_1, \dots, u_m\}^T$  denote the dependent variables. The given flux function  $\mathbf{f} = \{f_1, \dots, f_m\}^T$  is assumed to be smooth with respect to  $\mathbf{u}$ . We continue to restrict ourselves to studying solutions  $\mathbf{u}(x, t)$  of (18) which for each  $t$  are piecewise smooth on  $[0, 2\pi]$  and  $2\pi$ -periodic with respect to  $x$ .

We shall in this section consider a modified Fourier method (partially discussed in [9]) for studying solutions of (18) containing shock discontinuities, adapting several of the features discussed in the previous section. The method is geared at handling the propagation of shock-solutions, but is not designed to accurately handle the process where shocks are actually formed. In the short time interval where a shock is formed in an area where the solution earlier was smooth, other methods may, therefore, be more appropriate [18], [19].

Thus, we shall consider the solution of (18) subject to initial data  $\mathbf{u}(x, 0) = \mathbf{u}_0(x)$ , where  $\mathbf{u}_0(x)$  is a  $2\pi$ -periodic discontinuous function which is piecewise smooth on  $[0, 2\pi]$ , and which corresponds to some weak entropy solution of (18). We restrict ourselves to studying solutions  $\mathbf{u}(x, t)$  of (18) in some finite time interval  $t \in [0, T]$  which is such that  $\mathbf{u}(x, t)$  for each  $t$  is piecewise smooth on  $[0, 2\pi]$  and has no other shock discontinuities than those originating from the initial ones. In this time interval,  $\mathbf{u}(x, t)$  then has a representation of the form (5) for some integer  $R$ , and for each  $j = 1, \dots, R$ ,  $x = x_j(t)$  denote the curves across which  $\mathbf{u}(x, t)$  suffer shock discontinuities.

Since  $\mathbf{f}(\mathbf{u})$  is assumed to be smooth with respect to  $\mathbf{u}$ , it is clear that  $\mathbf{f}(\mathbf{u}(x, t))$  is a piecewise smooth function with respect to  $x$  for each  $t$ , if  $\mathbf{u}(x, t)$  is piecewise smooth for each  $t$ . The singularities for  $\mathbf{f}(\mathbf{u}(x, t))$  are clearly located at the same points where  $\mathbf{u}(x, t)$  is singular. Thus, a representation analogous to (5) is valid

$$(19) \quad \mathbf{f}(\mathbf{u}(x, t)) = \mathbf{f}^Q(x, t) + \mathbf{W}(x, t),$$

where, as usual,  $\mathbf{f}^Q(x, t)$  denotes the  $Q$  times continuously differentiable part; and  $\mathbf{W}(x, t)$  denotes the singular part. We clearly have  $\mathbf{W}(x, t) \equiv \mathbf{0}$  if  $\mathbf{V}(x, t) \equiv \mathbf{0}$ , but otherwise the relations between the smooth and the singular parts of  $\mathbf{f}(\mathbf{u}(x, t))$  and  $\mathbf{u}(x, t)$  are relatively complicated, in general. We shall, therefore, confine our discussion here to the important special case where each component of  $\mathbf{f}(\mathbf{u})$  is a quadratic form with respect to  $\mathbf{u}$ , i.e.  $f_j(\mathbf{u}) = \mathbf{u} \cdot A_j \mathbf{u}$ ,

$j = 1, \dots, m$ , where  $A_j$  are given  $m \times m$  matrices. We then for  $j = 1, \dots, m$  have that

$$(20) \quad \begin{aligned} f_j(\mathbf{u}(x, t)) &= \mathbf{u}^Q(x, t) \cdot A_j \mathbf{u}^Q(x, t) \\ &+ \mathbf{u}^Q(x, t) \cdot A_j \mathbf{V}(x, t) + \mathbf{V}(x, t) \cdot A_j \mathbf{u}^Q(x, t) \\ &+ \mathbf{V}(x, t) \cdot A_j \mathbf{V}(x, t). \end{aligned}$$

Clearly, the first term on the right hand side in (20) can only contribute to the smooth part  $\mathbf{f}^Q(x, t)$  of  $\mathbf{f}(\mathbf{u}(x, t))$ . In general, however, the last three terms in (20) all contribute to both the smooth part  $\mathbf{f}^Q(x, t)$  and to the singular part  $\mathbf{W}(x, t)$  of  $\mathbf{f}(\mathbf{u}(x, t))$ .

In order to be able to calculate a numerical solution of (18) which accurately approximates  $P_N \mathbf{u}(x, t)$  in Fourier space, we again note that a trigonometric polynomial  $\mathbf{u}_N(x, t)$  given by (12) and representing a shock-solution, may be accurately reconstructed as

$$(21) \quad \mathbf{v}_N(x, t) = \mathbf{u}_N^Q(x, t) + \tilde{\mathbf{V}}(x, t).$$

Again,  $\tilde{\mathbf{V}}(x, t)$  is an approximation for  $\mathbf{V}(x, t)$  determined by the reconstruction algorithm described in [8], and  $\mathbf{u}_N^Q(x, t)$  is given by (16).

A numerical solution  $\mathbf{u}_N(x, t)$  of (18) which accurately approximates  $P_N \mathbf{u}(x, t)$  in Fourier space may therefore [9] be calculated by solving the following initial value problem

$$(22) \quad \begin{aligned} \frac{\partial \mathbf{u}_N}{\partial t} + \frac{\partial}{\partial x} [P_N \mathbf{f}(\mathbf{v}_N)] &= \mathbf{0} \\ \mathbf{u}_N(x, 0) &= P_N \mathbf{u}_0(x). \end{aligned}$$

The spatial discretization of the nonlinear terms relevant for the semi-discrete formulation (22) is of particular interest here. For quadratic nonlinear terms, we note that the projection of the first term on the right hand side in (20) may be accurately calculated by the standard Fourier-Galerkin projection in  $S_N$  of a product of smooth functions. The projections of the second and the third term on the right hand side in (20) correspond to the approximate projections presented for the linear case (11), where an approximation for the product of a smooth function and a singular function was handled by de-truncation. For the last term in (20), we note that a product of piecewise polynomials is itself a piecewise polynomial. The last term is clearly a linear combination of products of the type  $U_k(x - \gamma)U_l(x - \mu)$ , and if we let  $K = k + l$  and let  $\gamma, \mu$  be any two constants such that  $\gamma \neq \mu$ , then the following relations can be shown to hold:

$$U_k(x - \gamma)U_l(x - \mu) = C_{k,l,\gamma,\mu}$$

$$(23) \quad \begin{aligned} &+ \sum_{j=k}^{K+1} \binom{j}{k} U_{K-j}(\gamma - \mu) U_j(x - \gamma) \\ &+ \sum_{j=l}^{K+1} \binom{j}{l} U_{K-j}(\mu - \gamma) U_j(x - \mu). \end{aligned}$$

Here,  $U_n(\xi)$  is given by (1) also for  $n = -1$ , and the constant  $C_{k,l,\gamma,\mu}$  may readily be determined from (23). Moreover, if  $\gamma = \mu$ , (23) is valid when  $\gamma - \mu$  and  $\mu - \gamma$  are replaced by  $0^+$  and  $0^-$ , respectively. The Fourier coefficients associated with the last term in (20) are therefore readily obtainable in view of (3).

## 4 Linear numerical examples

In this section, the performance of the method described in section 2 is demonstrated by applications to one test problem with constant coefficients and one with variable coefficients. For the cases considered, the Fourier coefficients corresponding to the numerical solution  $\mathbf{u}_N$  determined by (11) are advanced in time by applying a (4)5 order Runge-Kutta method with step size control due to Dormand and Prince [15]. The product terms treated by the de-truncation method are computed by employing  $2N$ -point FFT. By Method 1 we shall refer to solutions with singularity locations and jumps determined by the characteristic equations and the transport equations, respectively, and by Method 2 we shall refer to solutions where the reconstruction algorithm is utilized instead.

The first test case considered is the linear constant coefficient problem

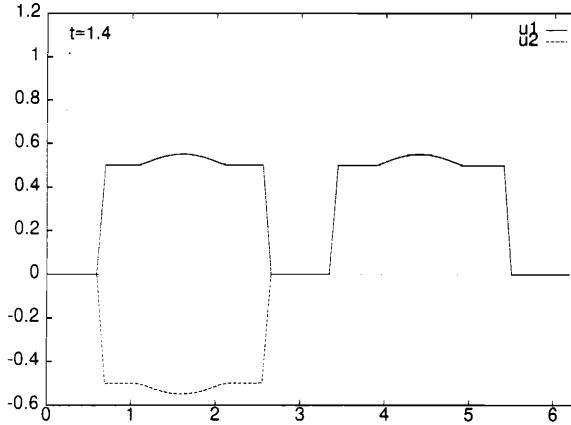
$$(24) \quad \begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \frac{\partial}{\partial x} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \\ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= \begin{pmatrix} g(x) \\ 0 \end{pmatrix} \text{ at } t = 0, \end{aligned}$$

where

$$(25) \quad g(x) = \begin{cases} 0 & 0 \leq x < 2 \\ 1 & 2 < x < 2.5 \\ 1 + 0.1 \sin [\pi(x - 2.5)] & 2.5 \leq x \leq 3.5 \\ 1 & 3.5 < x < 4 \\ 0 & 4 < x < 2\pi, \end{cases}$$

The solution of (24) is readily found to be given by

$$(26) \quad u_1(x, t) = \frac{1}{2}[g(x + t) + g(x - t)],$$

Figure 1: Computed solution of (24),  $N = 64, Q = 1$ .

$$(27) \quad u_2(x, t) = \frac{1}{2}[g(x-t) - g(x+t)].$$

Consequently, the singularities in  $g(x)$  are transported with half their initial strength along the characteristic curves  $x-t = c_1$  and  $x+t = c_2$ . Thus for  $t > 0$  fixed, the solution normally will have 8 singularity locations on  $[0, 2\pi]$ , with jump discontinuities in the derivatives at locations distinct from the points where the solution itself is discontinuous. As pointed out in section 2, the numerical solution  $\mathbf{u}_N$  of (24), (25) obtained from (11) is for this case equivalent to the Fourier-Galerkin solution, and we actually have  $\mathbf{u}_N \equiv P_N \mathbf{u}$ .

The considered example (24), (25) is primarily intended as a demonstration of the accuracy of the reconstruction algorithm [8] for piecewise smooth solutions on the form (5) from knowledge of  $P_N \mathbf{u}$ . The reconstructed solution is plotted in Figure 1 and error results for Method 1 and Method 2 are presented in Figure 2, where straight lines are fitted to the error data by means of least square linear regression.

As pointed out in section 2, we would expect the reconstructed solution to be  $O(N^{-(Q+1)})$  accurate, and since the solution in this case is seen to have no jump discontinuities in its second derivative, we would actually expect to do even better for  $Q = 1$ . For Method 1, these presumptions are clearly confirmed by the results in Figure 2 where the accuracy appears to be at least  $O(N^{-1})$  for  $Q = 0$  and at least  $O(N^{-3})$  for  $Q = 1$ . As for Method 2, we note that no convergence is achieved with  $Q = 0$ , i.e., when the discontinuities are the only singularity locations counted ( $R = 4$ ). In this case the reconstruction algorithm is not accurate due to the presence of the discontinuities in the derivatives at locations distinct from the locations of the discontinuities in the function itself [8]. We note,

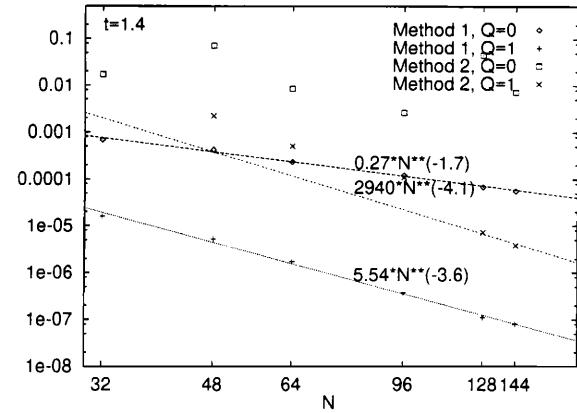


Figure 2: Maximum RMS-error in the two computed solution components of (24).

however, that Method 2 appeared to be  $O(N^{-1})$  accurate for the case  $Q = 0, R = 8$ , i.e. when all the singularity locations were counted as discontinuities.

For the case  $Q = 1$ , we first note that the application of the reconstruction algorithm with 8 singularity locations requires 24 Fourier coefficients with *positive* wave numbers. Method 2, therefore, does not apply for  $N = 32$  in this case. In order to reflect the asymptotic behavior of Method 2, the corresponding straight lines in Figure 2 were fitted to the error data obtained for  $N = 128, 144, 192$  where the last point is not included in the figure. From the obtained results, Method 2 appears to be at least  $O(N^{-3})$  accurate.

As our next example, we consider the problem

$$(28) \quad \frac{\partial u}{\partial t} + a(x) \frac{\partial u}{\partial x} = 0, \quad u(x, 0) = b(x),$$

where

$$(29) \quad b(x) = \begin{cases} 1 + 0.1 \sin^4 \left( \frac{\pi x}{2} \right) & 0 < x < 2 \\ 0 & 2 < x < 2\pi, \end{cases}$$

and where

$$(30) \quad a(x) = \begin{cases} 1 & 0 \leq x \leq 2 \\ 1 + c[(x-2)(x-4)]^2 & 2 < x < 4 \\ 1 & 4 \leq x < 2\pi. \end{cases}$$

Here,  $c$  is a constant for which we have considered two different values:  $c = 0.1$  and  $c = 0.2$ . The solution of (28) is given by

$$(31) \quad u(x, t) = B(\phi(x) - t) \quad \text{where } \phi(x) = \int_0^x \frac{d\tau}{a(\tau)},$$

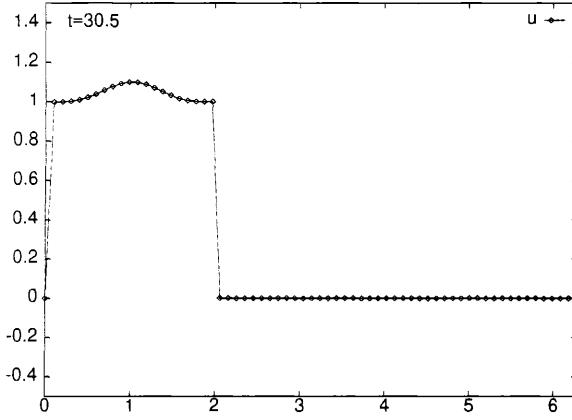


Figure 3: Computed solution of (28),  $N = 64, c = 0.2$  (Method 2)

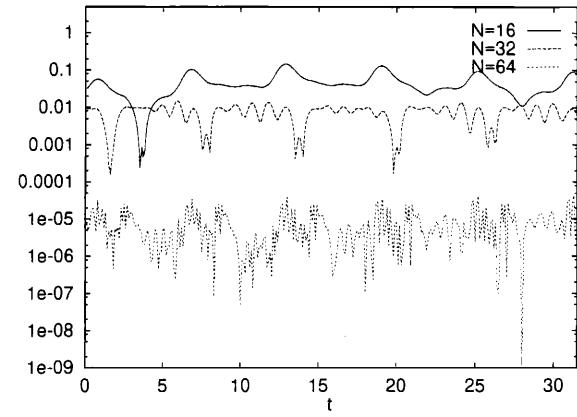


Figure 4: Absolute error in computed discontinuity location,  $c = 0.2$ .

and where  $B(\xi)$  is the  $\phi(2\pi)$ -periodic function which is such that  $B(\phi(x)) \equiv b(x)$ . Since  $a(x)$  is positive everywhere, this constitutes a wave propagating to the right retaining essentially the form of  $b(x)$ . For  $c > 0$ , however, the solution is modified by a stretching in the interval  $2 < x < 4$ , where the amount of stretching is determined by the magnitude of  $c$ . The wave regains its original form after its trailing front has passed  $x = 4$ . When  $a(x)$  is replaced by  $P_N a(x)$ , the exact solution (31) of (28) has for each  $t$  jump discontinuities at two distinct points in each period, while there are no jumps in its first three derivatives. In the computations we therefore let  $Q = 0$ . The time range for the computations corresponds to somewhat more than five time periods for the choice  $c = 0.2$ . A plot of the computed solution after approximately five periods is shown in Figure 3. Figure 4 shows the error in the discontinuity location computed in Method 2, and corresponding to the leading front of the wave.

From (30),  $a(x)$  is seen to be a continuously differentiable function which has jump discontinuities in its second derivative. Hence [4], the approximation by  $P_N a$  is only  $O(N^{-2})$  accurate in  $L^2(0, 2\pi)$ . Consequently we should, a priori, not expect the computed solution to be more accurate than that. In the implementation of Method 1 we have calculated the singularity locations from the characteristics corresponding to the exact  $a(x)$  and not to  $P_N a(x)$ , and the results shown in Figure 5 and Figure 6 for Method 1 indicate the above expected accuracy. We note, however, that the effect of the oscillatory behavior of  $P_N a(x)$  relative to the exact values of  $a(x)$  is not accounted for by the above  $L^2$  argument. We may, therefore, hope for a faster convergence than  $O(N^{-2})$  for Method 2. This is actually confirmed by the results shown in Figures 5 and 6. In view

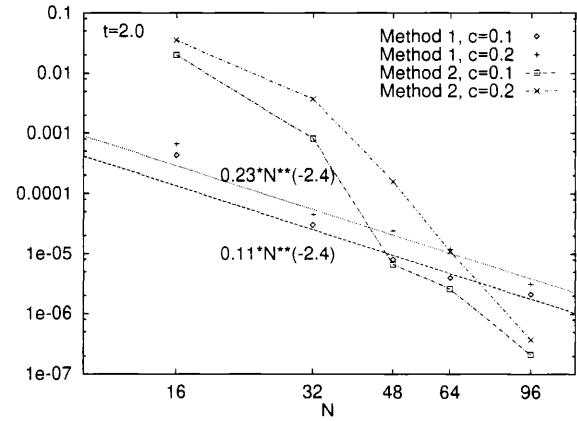


Figure 5: Calculated error for the computed solution of (28).

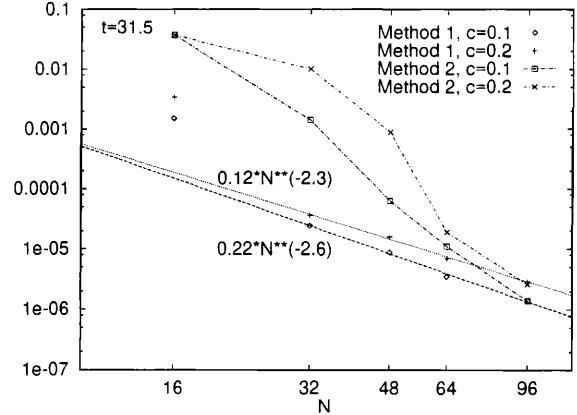


Figure 6: Calculated error for the computed solution of (28).

of these results, we may anticipate a similar improvement in accuracy for Method 1 if it is based on the characteristics corresponding to  $P_{Na}(x)$ .

In order to avoid instabilities for Method 2 within the time range considered we had, in the computations, to disregard a few of the highest order Fourier coefficients in  $u_N$  when we applied the reconstruction algorithm for the de-truncation at each time step. This is done since the highest order modes were polluted with errors which presumably stemmed from the numerical dispersion resulting from the relatively slow convergence of  $P_{Na}$ . An alternative way of stabilizing the computations is, presumably, to utilize some weak filtering.

## 5 Nonlinear numerical examples

As our first nonlinear example to be solved numerically by the method described in section 3, we take the widely used test case which is normally referred to as the inviscid Burgers equation:

$$(32) \quad \begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left( \frac{1}{2} u^2 \right) &= 0 \\ u(x, 0) = u_0(x) &= \sin(x). \end{aligned}$$

For this case the solution  $u(x, t)$  develops one shock in each period, starting at  $t = 1$  with an infinitely steep gradient at the point  $x = \pi$ , and then at later times  $t > 1$  the solution has a jump discontinuity at the same location. In order to prevent nonlinear instabilities from ruining the computations during the transition from a smooth solution to a solution with shocks, the numerical solution  $u_{2N}(x, t)$  of (32) has been computed by the Fourier-Galerkin method augmented with the spectral vanishing viscosity regularization [18] in the time interval up until the shock discontinuity has formed. At later times, the method described in section 3 is applied and is seen to be more accurate - especially for long term integrations on relatively coarse grids.

In order to know when the procedure described in section 3 should be invoked, the question of how one can decide from the knowledge of  $u_{2N}$  when a shock has appeared in the solution must be answered. For this purpose we have implemented a shock-test which is based on the assumption that the highest order modes of the truncated series  $u_N$  contains sufficiently accurate information about the shock shortly after it has formed. As soon as the discontinuity has been detected the computations are carried on by the method described in section 3 on the coarser grid corresponding to  $u_N$ . The discontinuity location and the jump

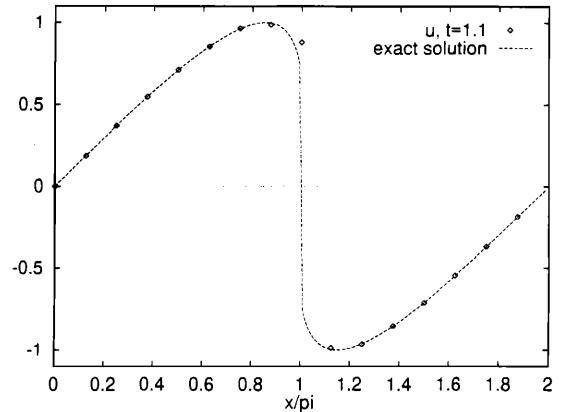


Figure 7: Computed solution of (32).  $N = 16$ .

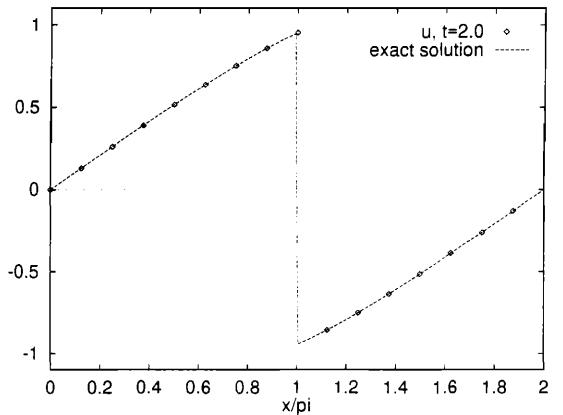


Figure 8: Computed solution of (32),  $N = 16$ .

are computed at each time step applying the reconstruction algorithm [8]. Since the exact solution has no jumps in its derivatives, the method is implemented with  $Q = 0$ . Plots of the computed solution are shown in Figures 7 and 8. Figures 9 and 10 show the error in the computed solution in logarithmic scale. Note that the peaks in the error shown in Figure 9 are due to the insufficient resolution of the curved structure close to the shock at  $t = 1.1$  rather than to remaining Gibbs oscillations. In fact, at  $t = 2.0$  where this curved structure is no longer present, the error is evenly spread throughout the domain (Figure 10). This indicates that the shock location and the shock strength are accurately determined in the calculations.

Our final test case is an initial value problem for the

shallow water equations

$$(33) \quad \begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= -\frac{\partial}{\partial x} \begin{pmatrix} \frac{1}{2}u_1^2 + gu_2 \\ u_1u_2 \end{pmatrix} \\ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ h(x) \end{pmatrix} \text{ at } t = 0. \end{aligned}$$

Here,  $u_1$  corresponds to the velocity and  $u_2$  to the depth, while  $g$  is the constant acceleration of gravity. As our initial condition we choose the symmetric bell-shaped function

$$(34) \quad h(x) = \begin{cases} h_0 & 0 \leq x < \pi/2 \\ h_0 + C[(x - \frac{\pi}{2})(\frac{3\pi}{2} - x)]^4 & \pi/2 \leq x \leq 3\pi/2 \\ h_0 & 3\pi/2 < x < 2\pi, \end{cases}$$

where  $h_0$  is the depth of the undisturbed water, and  $C > 0$  is a constant which determines the size of the initial surface elevation, which is seen to have its maximum at  $x = \pi$ . With the symmetric initial condition (34), the  $2\pi$ -periodic solution of (33) is also seen to give the physical relevant solution of the problem (33) when fixed walls are introduced at  $x = 0$  and  $x = 2\pi$ .

When the system is released, the initial surface elevation first descends and then splits into two symmetric waves traveling in opposite directions, leaving a region of undisturbed water between them. The velocity develops into an N-wave [20] with zero velocity at the symmetry point  $x = \pi$  (Figure 11). Due to the effect of nonlinearity, the waves steepen and eventually break, i.e. shock discontinuities form in the corresponding weak entropy solution of (33).

In the computations we let  $h_0 = 0.5$ , and  $C$  was chosen to correspond to a maximum initial surface elevation of 0.2. The standard Fourier-Galerkin method with dealiasing was used for computing the solution  $\mathbf{u}_N$  up until wave breaking occurred. Spectral viscosity regularization was not added in the pre-shock computations in this test case, since the solution computed by the Fourier method appears to be much less prone to nonlinear instabilities during the transition to a discontinuous solution than for the Burgers equation.

Shortly before the formation of the shocks, however, the highest order modes in  $\mathbf{u}_N$  are polluted and therefore do not contain sufficiently accurate information for the reconstruction algorithm to be applied. An integer parameter  $D$  has thus been introduced, and a shock criterion has been implemented based on the highest order modes in  $\mathbf{u}_{N-2D}$  instead, from which also approximate shock locations were computed. By observing that the corresponding computed

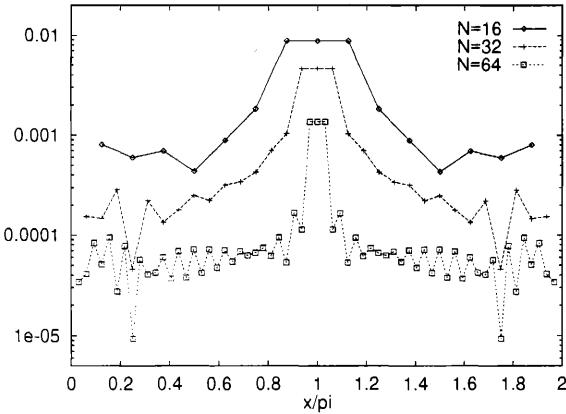


Figure 9: Pointwise error in computed solution,  $t = 1.1$

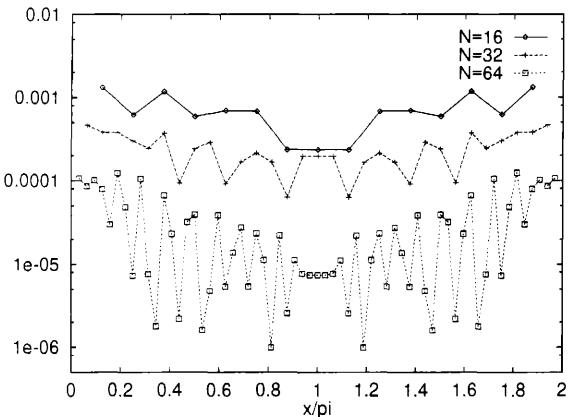
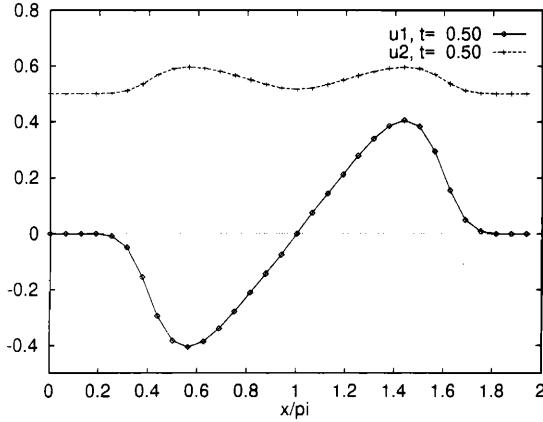
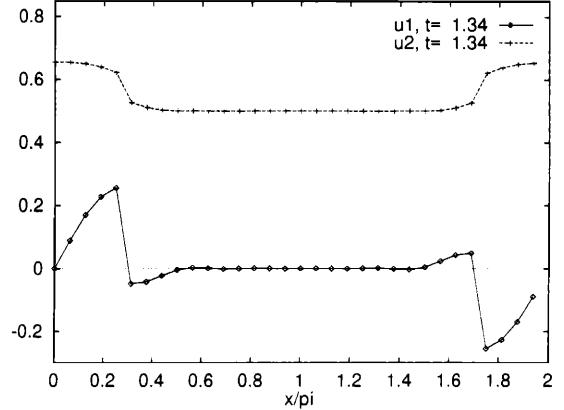


Figure 10: Pointwise error in computed solution,  $t = 2.0$

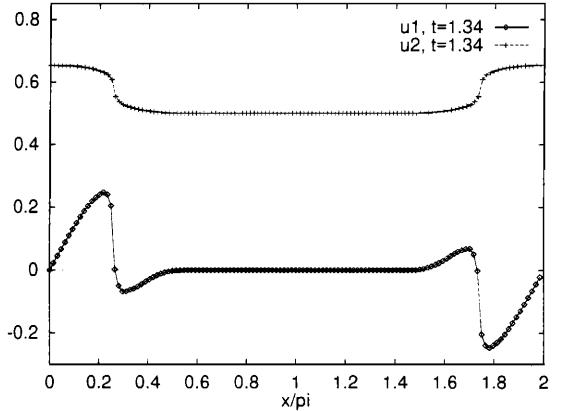
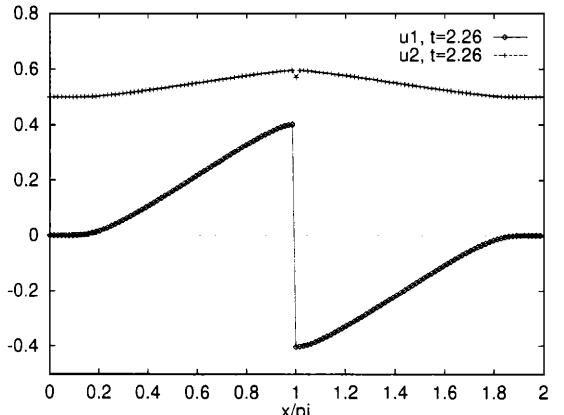
Figure 11: Computed solution of (33),  $N = 32$ .Figure 12: Computed solution of (33),  $N = 32$ .

shock locations for depth and velocity, respectively, approach each other as the shocks are formed, shocks in the computations were considered to have been formed in the solution when those computed location pairs were close enough. In the computations reported here for  $N = 64$  and  $N = 128$ ,  $D$  was set to 10 and 25, respectively. At the instant  $t = t_s$ , where the discontinuities were detected according to this criterion, the reconstruction of  $\mathbf{u}_N$  on the form (21) was invoked. From this time on the computations were continued by the procedure described in section 3 with  $Q = 0$ . We remark here that a weak filtering of the smooth part  $\mathbf{u}_N^Q$  of the solution at  $t = t_s$ , and at subsequent times, was necessary to stabilize the computations; and for that purpose an 8th order exponential cuff-off filter was implemented.

Plots of the computed discontinuous solution are shown in Figures 12-15. Figures 12 and 13 are approximately 0.1 time units after the shocks where detected. Figure 14 is shortly before the shock waves collide, and Figure 15 is after the collision.

## 6 Conclusions

We have, in this paper, presented applications of a modified Fourier method for computing nonsmooth solutions of hyperbolic problems. The numerical solution is sought as an accurate approximation for the truncated Fourier series associated with the exact solution. By utilizing step-functions, the nonsmooth solution is accurately reconstructed from its truncated spectral approximation. The reconstructed solution is used to avoid numerical dispersion and diffusion in connection with the computation for variable coefficients and nonlinear terms. The described approach is seen to give substantially improved accuracy

Figure 13: Computed solution of (33),  $N = 128$ .Figure 14: Computed solution of (33),  $N = 128$ .

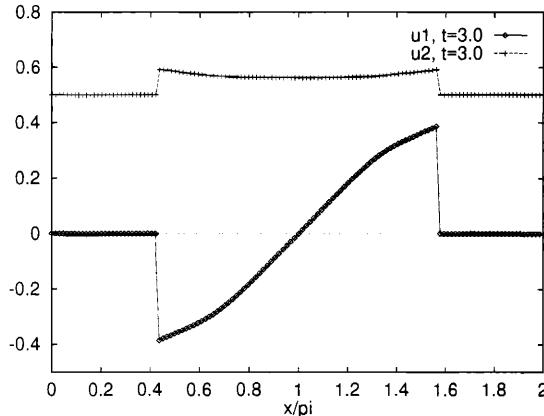


Figure 15: Computed solution of (33),  $N = 128$ .

compared to more traditional methods.

We would, finally, like to give a brief remark on the post-processing of the numerical solutions. Referring to (21), we recall that when the discontinuities are accurately captured, the global error is mainly due to the truncation error associated with  $\mathbf{u}_N^Q$ . In the cases where finite regularity of  $\mathbf{u}^Q$  is the dominant source of this truncation error, it is reasonable to expect that the global error could be further improved by additional post-processing applied to  $\mathbf{u}_N^Q$ . In this connection, the filters described by Vandeven [17] and the Gegenbauer reconstruction method due to Gottlieb et al. [13] seem particularly interesting.

## 7 Acknowledgments

J. H. Rolfsnes thankfully acknowledges the support from the Research Council of Norway.

## References

- [1] S. Abarbanel and D. Gottlieb. Information content in spectral calculations. In E.M. Murman and S.S. Abarbanel, editors, *Progress in Scientific Computing*, Proc. U.S.-Israel Workshop, 1984, volume 6, pages 345–356. Birkhäuser Boston Inc., 1985.
- [2] S. Abarbanel, D. Gottlieb, and E. Tadmor. Spectral methods for discontinuous problems. Technical Report 85-38, NASA-CR-177974, ICASE, 1985. Also in K. W. Morton and M. J. Baines, editors, *Numerical Methods for Fluid dynamics II*, Proc. Conf. Reading, 1985, pages 129-153, Clarendon Press, Oxford, 1986.
- [3] W. Cai, D. Gottlieb, and C.-W. Shu. Essentially nonoscillatory spectral Fourier methods for shock wave calculations. *Math. Comp.*, 52:389–410, 1989.
- [4] C. Canuto, M.Y. Hussaini, A Quarteroni, and T.A. Zang. *Spectral Methods in Fluid Dynamics*. Springer-Verlag, New York, 1988.
- [5] R. Courant and D. Hilbert. *Methods of Mathematical Physics, Volume II: Partial Differential Equations*. Interscience Publishers (John Wiley & Sons), New York, 1962.
- [6] K.S. Eckhoff. On the handling of linear and nonlinear terms in spectral methods. Technical Report STF11 F91053, SINTEF Multiphase Flow Laboratory, Trondheim, 1991.
- [7] K.S. Eckhoff. Accurate and efficient reconstruction of discontinuous functions from truncated series expansions. *Math. Comp.*, 61:745–763, 1993.
- [8] K.S. Eckhoff. Accurate reconstructions of functions of finite regularity from truncated Fourier series expansions. *Math. Comp.*, 64:671–690, 1995.
- [9] K.S. Eckhoff. On discontinuous solutions of hyperbolic equations. *Comput. Methods Appl. Mech. Engrg.*, (116):103–112, 1994.
- [10] K.S. Eckhoff and J.H. Rolfsnes. On nonsmooth solutions of linear hyperbolic systems. Submitted to *J. Comput. Phys.*, 1994.
- [11] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. C. Tricomi. *Higher Transcendental Functions*. McGraw-Hill, New York, 1953.
- [12] D. Gottlieb. Spectral methods for compressible flow problems. In Soubbaramayer and J. P. Boujot, editors, *Proc. 9. Int. Conf. Num. Meth. Fluid. Dynamics*, Saclay, France, 1984, volume 218 of *Lecture Notes in Physics*, pages 48–61. Springer-Verlag, 1985.
- [13] D. Gottlieb, C.-W. Shu, A. Solomonoff, and H. Vandeven. On the Gibbs phenomenon I: recovering exponential accuracy from the Fourier partial sum of a nonperiodic analytic function. *J. Comput. Appl. Math.*, 43:81–98, 1992.
- [14] D. Gottlieb, L. Lustman, and S. A. Orszag. Spectral calculations of one-dimensional inviscid compressible flows. *SIAM J. Sci. Stat. Comput.*, 2:296–310, 1981.

- [15] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems*. Springer-Verlag, Berlin, 1987.
- [16] A. Majda, J. McDonough, and S. Osher. The Fourier method for nonsmooth initial data. *Math. Comp.*, 32:1041–1081, 1978.
- [17] H. Vandeven. Family of spectral filters for discontinuous problems. *J. Sci. Comput.*, 6:159–192, 1991.
- [18] E. Tadmor. Convergence of spectral methods for nonlinear conservation laws. *SIAM J. Numer. Anal.*, 26:30–44, 1989.
- [19] E. Tadmor. Super viscosity and spectral approximations of nonlinear conservation laws. In K. W. Morton and M. J. Baines, editors, *Numerical Methods for Fluid Dynamics IV*, pages 69–82, Oxford, 1993. Clarendon Press.
- [20] G.B. Whitham. *Linear and nonlinear waves*. Wiley-Interscience, New York, 1974.

