

## CELL AVERAGING CHEBYSHEV METHODS FOR HYPERBOLIC PROBLEMS

WEI CAI

Department of Mathematics  
University of North Carolina at Charlotte  
Charlotte, NC 28223, U.S.A.

DAVID GOTTLIEB

Division of Applied Mathematics, Brown University  
Providence, RI 02912, U.S.A.

AMI HARTEN\*

Tel-Aviv University, Tel Aviv, Israel

**Abstract** — This paper describes a cell averaging method for the Chebyshev approximations of first order hyperbolic equations in conservation form.

We present formulas for transforming between pointwise data at the collocation points and cell averaged quantities, and vice-versa. This step, trivial for the finite difference and Fourier methods, is nontrivial for the global polynomials used in Spectral methods.

We then prove that the cell averaging methods presented are stable for linear scalar hyperbolic equations and present numerical simulations of shock-density wave interaction using the new cell averaging Chebyshev methods.

### 1. INTRODUCTION

In this paper we introduce a new spectral technique for the numerical solution of nonlinear hyperbolic equations. This technique, as almost every modern finite difference scheme for shock computations, is based on the cell averaged form of the equations. This is essential for finite difference shock capturing techniques and it is our experience that it plays an essential role in a successful spectral simulation of problems that involve shock waves [1].

Consider the nonlinear hyperbolic equation

$$\begin{cases} U_t + F_x(U) = 0, & x \in [-1, 1], \\ U(x, 0) = U_0(x), \end{cases} \quad (1.1)$$

with appropriate boundary conditions.

The cell averaged form of (1.1) is obtained by integrating (1.1) between any two points  $-1 \leq a < b \leq 1$  to get

$$\frac{\partial}{\partial t} \bar{U} \equiv \frac{\partial}{\partial t} \frac{1}{b-a} \int_a^b U(x) dx = -\frac{1}{b-a} (F(U(b)) - F(U(a))). \quad (1.2)$$

Let  $\bar{u}(x, t)$  be an approximation to  $\bar{U}(x, t)$  at time  $t$ . Following Harten [2] we express the approximation to  $\bar{U}(x, t + \tau)$  by

$$\bar{u}(x, t + \tau) = \mathcal{A} E(t) \mathcal{R}(\cdot; \bar{u}), \quad (1.3)$$

---

\*Computing facility is provided by a research grant from the North Carolina Supercomputer Center. This work has been supported in part by AFOSR Grant 85-0303 and by NSF Grant DMS-8810150 and ASC-9005874. The second author was partially supported by DARPA-URI Contract N00014-86-k0754.

where  $\mathcal{A}$  is the cell averaging operator and  $E(t)$  is the exact time evolution operator corresponding to (1.1). Throughout the paper we will not distinguish between  $\mathcal{R}(\cdot; \bar{u})$  and  $\mathcal{R}\bar{u}$ . The operator  $\mathcal{R}(\cdot; \bar{u})$  is of extreme importance, it represents the way we reconstruct  $u$  from its given cell average values

$$\bar{u}_{j-\frac{1}{2}} = \frac{1}{x_{j-1} - x_j} \int_{x_j}^{x_{j-1}} u(x) dx,$$

where  $\{x_j\}_{j=1}^N$  are the grid points. For finite difference schemes  $u$  is a piecewise polynomial of low degree, so that the reconstruction itself is simple. It becomes complicated only if one imposes also the requirement that the reconstruction should be essentially nonoscillatory. In [1] we have presented an essentially non-oscillatory Fourier method based on the cell averaging formulation (1.2). In that case the transformations between the cell averages and the point values are simple and can be carried out efficiently by the Fast Fourier Transformations (FFT). This can be attributed to the fact that the boundary conditions are periodic and that the cell average of a trigonometric function is proportional to the function itself. However, for Chebyshev methods, the cell averaging operation (denoted by the operator  $\mathcal{A}$ ) is not simple and more so is  $\mathcal{R}(\cdot; \bar{u})$ . As a matter of fact not only the formulation but also the implementation of  $\mathcal{A}$  and  $\mathcal{R}$  is not straightforward.

In this paper we formulate the cell averaging technique for the Chebyshev method. We will discuss its stability for linear problems and show an example of its applicability to nonlinear system of equations by simulating the problem of shock-density wave interaction. The cell averaging formulation is an essential part of the numerical code.

The outline of the paper is as follows: In Section 2 we show how to reconstruct efficiently point values of a polynomial from its cell averages and vice versa. In Section 3 we introduce the new numerical technique and show its stability for linear problems. Section 4 is devoted to numerical results obtained by using the new method.

## 2. CELL AVERAGES AND POINT VALUES

In this section we will discuss the cell averaging operator  $\mathcal{A}$  and the reconstruction operator  $\mathcal{R}$  in the context of the Chebyshev methods. In these methods the approximations are taken from the space of polynomials of degree  $N$ . It is therefore clear that  $\mathcal{A}$  and  $\mathcal{R}$ , when restricted to the polynomial space, can be expressed as matrices  $\mathbf{A}_N$  and  $\mathbf{R}_N$ . We will give the explicit formulas for these matrices. We start by discussing the operator  $\mathcal{A}$ .

Assume that  $f(x)$  is in  $C^r[-1, 1]$ ,  $r \geq 0$ . Let

$$x_j = \cos\left(\frac{j\pi}{N}\right), \quad 0 \leq j \leq N$$

be the Chebyshev-Lobatto points in  $[-1, 1]$ . For later use, define

$$x_{j-1/2} = \cos\left(\frac{(j-1/2)\pi}{N}\right), \quad 1 \leq j \leq N.$$

The cell averaged function  $\bar{f}(x)$  of  $f(x)$  is defined as

$$\bar{f}(x) = \mathcal{A}f =: \frac{1}{h_2(x) - h_1(x)} \int_{h_1(x)}^{h_2(x)} f(x) dx \quad \text{for } -1 < x < 1, \quad (2.1)$$

where

$$\begin{aligned} h_1(x) &= \cos\left(\cos^{-1} x - \frac{\Delta\theta}{2}\right), \\ h_2(x) &= \cos\left(\cos^{-1} x + \frac{\Delta\theta}{2}\right), \\ \Delta\theta &= \frac{\pi}{N}. \end{aligned} \quad (2.2)$$

The reason for the definition in (2.1) is that

$$\bar{f}(x_{j-1/2}) = \frac{1}{x_{j-1} - x_j} \int_{x_j}^{x_{j-1}} f(x) dx, \quad \text{for } 1 \leq j \leq N. \quad (2.3)$$

As stated before, we are interested in  $\mathcal{A}$  operating in a polynomial space. In Lemma 1, we show that the result of  $\mathcal{A}$  on a polynomial is still a polynomial of the same degree.

LEMMA 1. Let  $U_k(x) = 1/(k+1)T'_{k+1}(x)$ ,  $k \geq 0$ , be the second kind of the Chebyshev polynomials. Then

$$\bar{U}_k(x) = \mathcal{A}U_k = \sigma_k U_k(x), \quad (2.4)$$

where

$$\sigma_k = \frac{\sin(k+1)\frac{\Delta\theta}{2}}{(k+1)\sin\frac{\Delta\theta}{2}}. \quad (2.5)$$

PROOF. Substituting  $U_k(x)$  in the right hand side of (2.1) and making the transformation  $x = \cos\theta$ ,  $0 \leq \theta \leq \pi$ , we have

$$\bar{U}_k(x) = \frac{1}{\cos(\cos^{-1}x + \frac{\Delta\theta}{2}) - \cos(\cos^{-1}x - \frac{\Delta\theta}{2})} \int_{\cos(\cos^{-1}x - \Delta\theta/2)}^{\cos(\cos^{-1}x + \Delta\theta/2)} U_k(x) dx.$$

Since

$$T_k(x) = \cos(k\theta), \quad \text{then} \quad U_k(x) = \frac{\sin(k+1)\theta}{(k+1)\sin\theta}$$

and, therefore,

$$\begin{aligned} \bar{U}_k(x) &= \frac{1}{2(k+1)\sin\frac{\Delta\theta}{2}\sin\theta} \int_{\theta - \Delta\theta/2}^{\theta + \Delta\theta/2} \sin(k+1)\theta d\theta \\ &= \frac{1}{2(k+1)\sin\frac{\Delta\theta}{2}\sin\theta} (-1) \frac{\cos(k+1)\theta}{k+1} \Big|_{\theta - \Delta\theta/2}^{\theta + \Delta\theta/2} \\ &= \frac{\sin(k+1)\frac{\Delta\theta}{2}}{(k+1)\sin\frac{\Delta\theta}{2}} \left( \frac{1}{k+1} \frac{\sin(k+1)\theta}{\sin\theta} \right), \end{aligned}$$

i.e.,

$$\bar{U}_k(x) = \sigma_k U_k(x).$$

From Lemma 1, one gets

COROLLARY 1. The cell averaged function of any polynomial of order  $N$  is a polynomial of the same order.

PROOF. Any polynomial of degree  $N$  has the expression

$$f(x) = \sum_{k=0}^N a_k U_k(x), \quad (2.6)$$

where  $a_k$  are constants.

Therefore, by Lemma 1

$$\bar{f}(x) = \sum_{k=0}^N \sigma_k a_k U_k(x). \quad (2.7)$$

Thus, Lemma 1 gives explicitly the eigenvalues  $\sigma_k$  of the matrix  $\mathbf{A}_N$  (the restriction of  $\mathcal{A}$  to polynomials of degree  $N$ ). These eigenvalues are uniformly bounded from above and below, in fact

$$\frac{2}{\pi} \leq \sigma_k \leq \frac{\pi}{2}, \quad 0 \leq k \leq N. \quad (2.8)$$

If  $f(x)$  is a polynomial of degree  $N$ , then it is uniquely determined by its values  $f(x_j)$ ,  $j = 0, \dots, N$ . So theoretically  $\bar{f}(x_{j-1/2})$ ,  $j = 1, \dots, N$  can be determined. Therefore the transformation from  $f(x_j)$ ,  $j = 0, \dots, N$  to  $\bar{f}(x_{j-1/2})$ ,  $j = 1, \dots, N$  is well defined and we only need to address the issue of its efficient implementation.

In general it is known that

$$f(x) = \sum_{k=0}^N a_k T_k(x), \quad (2.9)$$

where

$$\begin{aligned} a_k &= \frac{2}{N \bar{c}_k} \sum_{j=0}^N f(x_j) T_k(x_j), \\ \bar{c}_0 &= \bar{c}_N = 2, \\ \bar{c}_k &= 1 \quad \text{if } k \neq 0 \text{ or } N. \end{aligned} \quad (2.10)$$

Alternatively,

$$f(x) = \sum_{k=0}^N b_k U_k(x), \quad (2.11)$$

$$\begin{aligned} b_k &= \frac{a_k}{2}, & k = N-1 \text{ or } N, \\ b_k &= \frac{1}{2} (\bar{c}_k a_k - a_{k+2}), & 0 \leq k \leq N-2 \end{aligned} \quad (2.12)$$

and therefore, by Corollary 1,

$$\bar{f}(x) = \sum_{k=0}^N \sigma_k b_k U_k(x). \quad (2.13)$$

Now  $\{\bar{f}(x_{j-1/2})\}_{j=1}^N$  are obtained by substituting  $x_{j-1/2}$  in (2.13) (this can be carried out using the FFT).

We note that Equations (2.9)–(2.13) describe how to get the vector  $\bar{f}(x_{j-1/2})$ ,  $j = 1, \dots, N$  from the vector  $f(x_i)$ ,  $i = 0, \dots, N$ . Denote by  $\mathbf{A}_N$  the  $N \times (N+1)$  matrix defined by this transformation. We note that  $\mathbf{A}_N$  can be written explicitly. In fact the polynomial  $f(x)$  has an unique representation as

$$f(x) = \sum_{j=0}^N f(x_j) g_j(x), \quad (2.14)$$

where

$$g_j(x) = \frac{(1-x^2) T'_N(x) (-1)^{j+1}}{\bar{c}_j N^2 (x-x_j)} = \frac{2}{\bar{c}_j N} \sum_{k=0}^N \frac{T_k(x_j) T_k(x)}{\bar{c}_k},$$

with  $\bar{c}_k$  defined in (2.10).

It follows upon substituting (2.14) in (2.1) and using the fact that  $\bar{T}_0(x) = U_0(x)$ ,  $\bar{T}_1(x) = 2\sigma_1 U_1(x)$  and  $\bar{T}_k(x) = 1/2(\sigma_k U_k(x) - \sigma_{k-2} U_{k-2}(x))$  for  $k \geq 2$ ,

$$\bar{f}(x) = \sum_{\ell=0}^N f(x_\ell) \bar{g}_\ell(x), \quad (2.15)$$

where

$$\bar{g}_\ell(x) = \frac{2}{\bar{c}_\ell N} \left[ \frac{1}{2} + 2\sigma_1 T_1(x_\ell) U_1(x) + \sum_{k=2}^N \frac{T_k(x_\ell) (\sigma_k U_k(x) - \sigma_{k-2} U_{k-2}(x))}{2\bar{c}_k} \right], \quad (2.16)$$

with  $\sigma_k$  defined in (2.5).

Setting  $x = x_{j-1/2}$  in (2.16)

$$(\mathbf{A}_N)_{j\ell} = \bar{g}_\ell(x_{j-1/2}), \quad 0 \leq \ell \leq N, \quad 1 \leq j \leq N. \quad (2.17)$$

Thus, we have outlined two procedures to get  $\bar{f}(x_{j-1/2})$  from  $f(x_j)$ , one uses the FFT and another uses matrix vector multiplications.

We are now ready to discuss the reconstruction operator  $\mathcal{R}(\cdot; \bar{f})$ . Note that in the beginning of the solution process (1.3), we only have the values  $\bar{f}_{j-1/2}$ ,  $j = 1, \dots, N$ , thus we need another piece of information in order to define uniquely the  $N$ -th degree polynomial  $f(x)$ . This piece of information is provided by the boundary condition. For simplicity, we assume that the boundary condition is of the form

$$f(1) = f(x_0) = f_0. \quad (2.18)$$

The reconstruction is done in two steps. Define first a  $(N-1)$ -th polynomial  $\bar{f}_{N-1}(x)$  which collocates  $\bar{f}(x)$  at  $\{x_{j-1/2}\}_{j=1}^N$ , i.e.,  $\bar{f}_{N-1}(x_{j-1/2}) = \bar{f}(x_{j-1/2})$ ,  $1 \leq j \leq N$ , it is readily verified that

$$\bar{f}_{N-1}(x) = \sum_{k=0}^{N-1} c_k T_k(x), \quad (2.19)$$

where

$$c_k = \frac{2}{\bar{c}_k N} \sum_{j=1}^N \bar{f}(x_{j-1/2}) T_k(x_{j-1/2}), \quad \bar{c}_k \text{ is same as in (2.10)}. \quad (2.20)$$

Alternatively we have,

$$\bar{f}_{N-1}(x) = \sum_{k=0}^{N-1} b_k U_k(x), \quad (2.21)$$

where

$$\begin{aligned} b_k &= \frac{c_k}{2}, & k &= N-2, N-1, \\ b_k &= \frac{1}{2} (\bar{c}_k c_k - c_{k+2}), & 0 \leq k &\leq N-3. \end{aligned} \quad (2.22)$$

Now, by Corollary 1,

$$f_{N-1}(x) = \sum_{k=0}^{N-1} \frac{b_k}{\sigma_k} U_k(x), \quad (2.23)$$

is a polynomial of degree  $N-1$  such that  $\mathcal{A}f_{N-1} = \bar{f}_{N-1}$ .

Generally,  $f_{N-1}(x)$  does not satisfy the boundary condition (2.18). There are two ways to modify  $f_{N-1}(x)$  so that the boundary condition (2.18) is satisfied. In the first way we can add to  $f_{N-1}(x)$  an  $N$ -th degree polynomial  $Q(x)$  such that

$$\bar{Q}(x_{j-1/2}) = 0, \quad 1 \leq j \leq N, \quad (2.24)$$

and

$$f_{N-1}(x) + Q(1) = f_0. \quad (2.25)$$

It can be verified that in order to satisfy (2.24)

$$Q(x) = c ((1-x^2) T'_N(x))'. \quad (2.26)$$

Let  $f(x)$  be the sum of  $f_{N-1}(x)$  and  $Q(x)$ ,

$$\begin{aligned} f(x) &= f_{N-1}(x) + c \left( (1-x^2) T'_N(x) \right)' \\ &= f_{N-1}(x) - c \left[ x T'_N(x) + N^2 T_N(x) \right]. \end{aligned} \quad (2.27)$$

The last equality follows from the Chebyshev equation, and the constant  $c$  is now determined by the condition  $f(x_0) = f_0$ , i.e.,

$$c = -\frac{1}{2N^2} (f_0 - f_{N-1}(1)). \quad (2.28)$$

Finally, given  $\bar{f}(x_{j-1/2})$ ,  $j = 1, \dots, N$  and  $f_0$  we can get

$$f(x_i) = f_{N-1}(x_i) - c \left[ x_i T'_N(x_i) + N^2 T_N(x_i) \right], \quad i = 0, \dots, N, \quad (2.29)$$

where  $f_{N-1}(x_i)$  can be evaluated from (2.23) by using the FFT. Note that in this procedure we change the values of  $f_{N-1}(x_i)$  at all the grid points.

Denote by  $\mathbf{R}_N$  the  $(N+1) \times (N+1)$  matrix transforming  $f_0$  and  $\bar{f}(x_{j-1/2})$ ,  $1 \leq j \leq N$  to  $f(x_j)$ ,  $0 \leq j \leq N$ , i.e.,

$$(f(x_0), \dots, f(x_N))^T = \mathbf{R}_N \left( f(x_0), \bar{f}(x_{\frac{1}{2}}), \dots, \bar{f}(x_{N-\frac{1}{2}}) \right)^T. \quad (2.30)$$

As before we can write  $\mathbf{R}_N$  explicitly. Equation (2.19) can be rewritten as

$$\bar{f}_{N-1}(x) = \sum_{j=1}^N \bar{f}(x_{j-1/2}) \bar{h}_j(x), \quad (2.31)$$

where  $\bar{h}_j(x_{k-1/2}) = \delta_{jk}$  and  $h_j(x)$  are polynomials of degree  $N-1$ , explicitly

$$\bar{h}_j(x) = (-1)^{j+1} N \sin \left( \left( j - \frac{1}{2} \right) \frac{\pi}{N} \right) \frac{T_N(x)}{x - x_{j-1/2}}. \quad (2.32)$$

It can be shown that  $\bar{h}_j(x)$  is the cell averaged polynomial of

$$h_j(x) = \sum_{k=0}^{N-1} \lambda_k U_k(x), \quad (2.33)$$

with  $\lambda_k$  defined as

$$\begin{aligned} \lambda_k &= \frac{1}{N \sigma_k} T_k(x_{j-1/2}), & \text{if } k = N-2, N-1, \\ \lambda_k &= \frac{1}{N \sigma_k} (T_k(x_{j-1/2}) - T_{k+2}(x_{j-1/2})), & \text{if } 0 \leq k \leq N-3. \end{aligned} \quad (2.34)$$

As a result of (2.31) and (2.33) polynomial  $f_{N-1}(x)$  takes the form

$$f_{N-1}(x) = \sum_{j=1}^{N-1} \bar{f}(x_{j-1/2}) h_j(x), \quad (2.35)$$

and by (2.27)

$$f(x) = f_{N-1}(x) - \frac{1}{2N^2} (f_0 - f_{N-1}(1)) \left( (1-x^2) T'_N(x) \right)'.$$

Using (2.31) we get

$$\begin{aligned} f(x) &= \sum_{j=1}^N \bar{f}(x_{j-1/2}) h_j(x) - \frac{1}{2N^2} \left( f_0 - \sum_{j=1}^N \bar{f}(x_{j-1/2}) h_j(1) \right) ((1-x^2) T'_N(x))' \quad (2.36) \\ &= \sum_{j=1}^N \bar{f}(x_{j-1/2}) \left[ h_j(x) + \frac{h_j(1)}{2N^2} ((1-x^2) T'_N(x))' \right] - \frac{f_0}{2N^2} ((1-x^2) T'_N(x))'. \end{aligned}$$

Substituting  $x = x_i$  into (2.36) gives

$$(\mathbf{R}_N)_{ij} = \begin{cases} \delta_{ij}, & \text{if } i = 0, \\ -\frac{1}{2N^2} ((1-x_i^2) T'_N(x_i))', & \text{if } j = 0, \\ h_j(x_i) + \frac{h_j(1)}{2N^2} ((1-x_i^2) T'_N(x_i))', & 1 \leq i, j \leq N. \end{cases} \quad (2.37)$$

To summarize we state

LEMMA 2. Let  $f(x) = \mathcal{R}(\cdot; \bar{f})$  be the  $N$ -th polynomial defined in (2.27) then

$$(\mathcal{A}f)(x_{j-1/2}) = \bar{f}(x_{j-1/2}) \quad \text{for } 1 \leq j \leq N. \quad (2.38)$$

A different way to modify  $f_{N-1}(x)$  in (2.23), in order to satisfy the boundary condition (2.18), is to add to it a polynomial  $Q_1(x)$  of degree  $N$  such that the point values  $f_{N-1}(x_i)$  remain unchanged and the new polynomial satisfies the boundary condition. Thus instead of (2.27) we define

$$f(x) = f_{N-1}(x) + (f_0 - f_{N-1}(1)) \frac{(1+x) T'_N(x)}{2N^2}. \quad (2.39)$$

Computationally (2.39) is simpler than (2.23) and (2.27). However, note that in this case

$$(\mathcal{A}f)(x_{j-1/2}) \neq \bar{f}(x_{j-1/2}), \quad (2.40)$$

which is in contrast to (2.38). The matrix corresponding to (2.40) can be formed similarly as in (2.37).

### 3. CELL AVERAGING CHEBYSHEV (CAC) METHOD AND LINEAR STABILITY

In this section, we will establish the stability of the Cell Averaging Methods, presented in Section 2, when applied to a first order scalar hyperbolic equation. It is tempting to try to obtain stability in the  $L^1$  norm because of the way the method (1.3) is presented. However we will only give the stability estimate based on a weighted Chebyshev norm.

Consider the initial boundary value problem of the scalar hyperbolic equation

$$\begin{cases} U_t = U_x, & x \in [-1, 1], \\ U(x, 0) = U_0(x), \\ U(1, t) = 0. \end{cases} \quad (3.1)$$

The cell averaged form of (3.1) is

$$\frac{\partial}{\partial t} \bar{U}(x, t) + \frac{1}{h_2(x) - h_1(x)} [U(h_2(x), t) - U(h_1(x), t)] = 0, \quad (3.2)$$

where  $h_1(x)$  and  $h_2(x)$  are defined in (2.2).

We follow the notation (1.3), let the  $N^{\text{th}}$  degree polynomial  $\bar{u}(x, t)$  be the approximation to (3.2). Our aim is to find the error equation for  $\mathcal{R}(\cdot; \bar{u})$  defined either by (2.27) or (2.39). From (1.3)

$$\bar{u}(x, t + \tau) = \mathcal{A} E \mathcal{R} \bar{u}(x, t). \quad (3.3)$$

Applying the reconstruction operator  $\mathcal{R}$  we get

$$\mathcal{R} \bar{u}(x, t + \tau) = \mathcal{R} A E \mathcal{R} \bar{u}(x, t). \quad (3.4)$$

From Equation (3.4) it is clear that  $\mathcal{R} \bar{u}$  satisfies exactly the equation

$$\frac{\partial \mathcal{R} \bar{u}}{\partial t} = \frac{\partial \mathcal{R} \bar{u}}{\partial x} + \tau [(1 - x^2) T'_N]' \quad (3.5)$$

for the reconstruction procedure (2.27) and

$$\frac{\partial \mathcal{R} \bar{u}}{\partial t} = \frac{\partial \mathcal{R} \bar{u}}{\partial x} + \tau_1 (1 + x) T'_N \quad (3.6)$$

for (2.39).  $\tau$  and  $\tau_1$  are quantities depending on time  $t$ .

Note that (3.6) is the error equation for the Chebyshev collocation method. For a scalar linear equation the CAC method corresponding to (3.6) is equivalent to the known collocation method [3]. It remains to investigate the stability of (3.5). For simplicity, in the remaining of this section we denote  $\mathcal{R} \bar{u}$  by  $u$ . From the construction of  $\mathcal{R}$  in Section 2 we know that  $u(x, t)$  satisfies the boundary condition in (3.1), i.e.,  $u(1, t) = 0$ .

It is interesting to note from (3.6) that the CAC method with reconstruction operator (2.27) is equivalent - for constant coefficient case - to the collocation method where the grid points are the zeros of the polynomial

$$Q(x) = [(1 - x^2) T'_N(x)]'. \quad (3.7)$$

Note that by using standard identities

$$Q(x) = -[x T'_N(x) + N^2 T_N(x)]. \quad (3.8)$$

The term  $\tau$  in the right hand side of (3.5) is determined by the boundary condition  $u(1, t) = \mathcal{R} \bar{u}(1, t)$ . In fact substituting  $x = 1$  in the equation (3.5) and noting that  $\frac{\partial u}{\partial t}(1, t) = 0$  we get for  $\tau$

$$\tau = \frac{u_x(1, t)}{2N^2} \quad (3.9)$$

An alternative expression can be obtained by equating the highest coefficient on the both sides of (3.5), thus if  $u$  is expanded as

$$u(x) = \sum_{k=0}^N \hat{u}_k T_k(x) \quad (3.10)$$

then

$$\tau = -\frac{1}{N(N+1)} \frac{\partial \hat{u}_N}{\partial t}. \quad (3.11)$$

Before stating the main stability result of this paper, we state the following lemma.

LEMMA 3. If  $f(x) = \sum_{k=0}^{4N-1} a_k T_k(x)$ , then

$$\frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} f(x_j) = \int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx + \pi a_{2N} \quad (3.12)$$

where  $\bar{c}_j$  is defined in (2.10);  $x_j$   $0 \leq j \leq N$  are the Chebyshev-Lobatto points.

PROOF. See [4].

The main stability result follows from the following lemma.

LEMMA 4. Let  $U(x, t)$  be the solution of (3.2) and  $u(x, t) = \mathcal{R} \bar{u}$  be the approximation to  $U(x, t)$  by the CAC method (1.3) with the reconstruction operator (2.27) or (2.30), then

$$\frac{\partial}{\partial t} \left[ \frac{\pi}{2N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u^2(x_j, t) + \frac{\pi N}{2(N+1)} \hat{u}_N^2(t) \right] \leq 0. \quad (3.13)$$



PROOF. We multiply (3.5) by  $\frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t)$  and sum from the points  $x_N = -1$  to  $x_1$  to get

$$\begin{aligned} \frac{\partial}{\partial t} \frac{\pi}{2N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u^2(x_j, t) &= \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) u_x(x_j, t) \\ &\quad - \tau N^2 \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) T_N(x_j). \end{aligned} \quad (3.14)$$

We treat the two terms in the right hand side of (3.14) separately.

$$\begin{aligned} I &= \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) u_x(x_j, t) \\ &= \frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) u_x(x_j, t) + \frac{\pi}{N} u_x^2(1, t). \end{aligned} \quad (3.15)$$

By noting the exactness of the Gauss Lobatto formula one gets

$$I = \int_{-1}^1 \frac{1+x}{1-x} u(x, t) \frac{\partial}{\partial x} u(x, t) \frac{dx}{\sqrt{1-x^2}} + \frac{\pi}{N} u_x^2(1, t), \quad (3.16)$$

and integration by part yields

$$I = - \int_{-1}^1 \frac{1+\frac{x}{2}}{\sqrt{1-x^2}} \frac{u^2(x, t)}{(1-x)^2} dx + \frac{\pi}{N} u_x^2(1, t). \quad (3.17)$$

We use again the Gauss Lobatto formula to get

$$I = - \frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} \left(1 + \frac{x_j}{2}\right) \frac{u^2(x_j, t)}{(1-x_j)^2} + \frac{\pi}{N} u_x^2(1, t) \quad (3.18)$$

and, therefore,

$$I \leq - \frac{3}{4} \frac{\pi}{N} u_x^2(1, t) + \frac{\pi}{N} u_x^2(1, t) = \frac{\pi}{4N} u_x^2(1, t). \quad (3.19)$$

We turn now to the second term in (3.14)

$$II = -\tau N^2 \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) T_N(x_j) \quad (3.20)$$

or

$$II = -\tau N^2 \frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) T_N(x_j) - \tau N^2 \frac{\pi}{N} u_x(1, t). \quad (3.21)$$

Using (3.9) and (3.11) for  $\tau$  one gets

$$II = \frac{N}{N+1} \frac{\partial \hat{u}_N}{\partial t} \frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) T_N(x_j) - \frac{\pi}{2N} u_x^2(1, t). \quad (3.22)$$

As  $(1+x)/(1-x) u(x, t) T_N(x)$  is a polynomial of degree less than  $4N-1$ , one gets using Lemma 3

$$\frac{\pi}{N} \sum_{j=0}^N \frac{1}{\bar{c}_j} \frac{1+x_j}{1-x_j} u(x_j, t) T_N(x_j) = \int_{-1}^1 \frac{1+x}{1-x} \frac{u T_N}{\sqrt{1-x^2}} dx + \pi a_{2N}, \quad (3.23)$$

where  $a_{2N}$  is the  $2N$  coefficient in the expansion of  $(1+x)/(1-x) u(x, t) T_N(x)$ .

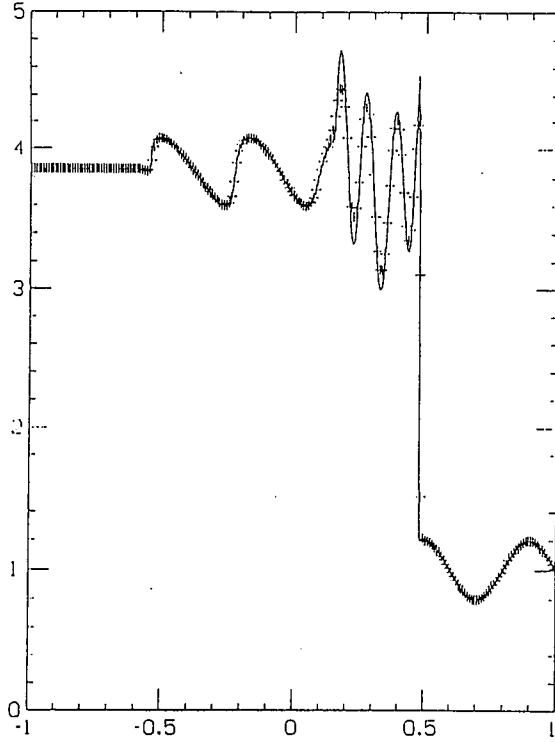


Figure 3. The third order ENO scheme: density,  $N = 220$ , time  $t = 0.3$ . (+) - numerical solutions, solid line - exact solutions.

The reconstruction operator  $\mathbf{R}_N$  yields spectrally accurate point value approximations to the exact solutions if the exact solutions are smooth, thus  $\bar{u}_{j-\frac{1}{2}}$ ,  $j = 1, \dots, N$  can be expected to approximate their cell averages accurately. However, if the exact solutions are discontinuous, the point values  $u_j$ ,  $j = 0, \dots, N$  obtained by  $\mathbf{R}_N$  will be oscillatory as the result of the Gibbs phenomenon. In [1] we proposed a practical way to obtain essentially nonoscillatory spectral reconstruction to a discontinuous function from its oscillatory Fourier approximations. The key idea there is to augment the Fourier space by adding simple discontinuous functions whose locations and magnitudes of discontinuities are approximations to those of the shock waves in the numerical solutions. In our computations of CAC method, we extend this idea along the same line to obtain essentially nonoscillatory reconstructions. The estimates on these reconstructions will be appearing in a separate work. We refer the reader to [1] for further details.

Now consider the Riemann problem for the Euler equations for a polytropic gas

$$\mathbf{U}_t(x, t) + \mathbf{f}_x(\mathbf{U}) = 0, \quad (4.2)$$

where  $\mathbf{U}(x, t)$  and  $\mathbf{f}(\mathbf{U})$  are defined as

$$\begin{aligned} \mathbf{U} &= (p, M, E)^\top, & -1 \leq x \leq 1, \\ \mathbf{f}(\mathbf{U}) &= q\mathbf{U} + (0, P, qP)^\top, \end{aligned} \quad (4.3)$$

where

$$P = (\gamma - 1)(E - \frac{1}{2} \rho q^2), \quad M = \rho q, \quad (4.4)$$

with  $\gamma = 1.4$  and the initial conditions are as follows

$$\begin{aligned} (\rho_L, q_L, P_L) &= (3.857143, 2.629369, 10.33333), & \text{when } x < -0.8, \\ (\rho_R, q_R, P_R) &= (1 + \epsilon \sin 5\pi x, 0, 1), & \text{when } x > -0.8, \end{aligned} \quad (4.5)$$

where  $\epsilon = 0.2$ .

The solutions to (4.1)-(4.4) model the interaction between a moving shock wave and disturbances. Note that in the right state of the density a sinusoidal perturbation of magnitude  $\epsilon = 0.2$  is superposed upon a constant state. From linear analysis it can be shown that the disturbances will interact with the shock wave. A density wave of different frequency and magnitude will emerge behind the shock wave. Also the disturbance in the density field will perturb the velocity and pressure fields behind the shock wave. The numerical solution of this Riemann problem mandates a high order scheme in order to capture the fine structures in the solutions for an correct interpretation of the physical process. We test on this problem with second order MUSCL scheme [5] and third order point value version ENO finite difference scheme [6] and the CAC spectral method proposed in this paper. Our numerical results have shown clearly the advantage of a higher order numerical scheme for problems with complicated structures.

Figures 1-3, show the density profiles obtained by the three methods mentioned above. Figure 1, is the result using CAC spectral method. The solid lines are the solutions taken from [6] using the third order ENO finite difference method with  $N = 800$  which we take as the exact solutions. Figure 2 and Figure 3, are the results with the second order MUSCL scheme [5] and the third order ENO finite difference scheme respectively. In all three cases we use the same amount of mesh points  $N = 220$ . All the results are plotted at the same time  $t = 0.3$ .

#### REFERENCES

1. W. Cai, D. Gottlieb, C.W. Shu, Essentially non-oscillatory spectral Fourier method for shock wave calculations *Math. Comput.*, **52**, (186), 389-410, (April, 1989).
2. A. Harten, On the nonlinearity of modern shock capturing schemes, Icase Report No. 86-69, (October, 1986).
3. D. Gottlieb and E. Turkel, Topics in spectral methods, In *Lecture Notes in Mathematics*, No. 1127, (Edited by A. Dold and B. Eckman), pp. 115-155, (1983).
4. D. Funaro and D. Gottlieb, A new method of imposing boundary conditions in pseudospectral approximation of hyperbolic equations, *Math. Compt.*, **51**, (184), 599-613 (October, 1988).
5. S. Osher, Convergence of generalized MUSCL schemes, *SIAM J. Numer. Anal.*, **22**, 947-961, (1985).
6. C.W. Shu and S. Osher, Efficient implementation of essentially nonoscillatory shock-capturing schemes, II, *J. Compt. Phys.*, **83**, (1), page 32 (July, 1989).