Projection Dynamics in Godunov-Type Schemes I: to the Physical Understanding of Post-Shock Oscillations

Kun Xu* and Jishan Hu†
Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
(submitted to Journal of Computational Physics)

Abstract

There are two stages in the 1st-order Godunov-type schemes to update flow variables, the gas evolution stage for the numerical fluxes across cell interface and the projection stage for the reconstruction of constant state inside each cell. Ideally, the evolution stage should be based on the exact Euler solutions, the so-called Riemann solver. In this paper, we will show that some anomalous phenomena, such as post-shock oscillations, are caused from the underlying unsteady dissipative mechanism from the projection stage. Based on physical model, we are going to analyze and evaluate quantitatively the projection dynamics and compare our theoretical analysis with numerical observations, such as the relation between oscillation amplitude and the shock speed. The conclusion is that any flow solvers based on the exact Euler solutions are not adequate to compensate the projection errors. In order to get a correct representation of flow motion in the discretized space and time, the consistent dissipative terms must be added in the flux functions and solve the Navier-Stokes-like equations directly.

1 Introduction

In the past two decades there have been tremendous progress in the development of numerical methods for solving the the Euler equations. The most promising method is the Godunov scheme where the Riemann solution is used for the flux evaluation, followed by the projection stage to construct constant states inside each cell. While the Godunov-type schemes are often held up to be models of robustness and accuracy, they can on occasions fail quite spectacularly. Even the 1st-order Godunov method may give anomalous results, i.e. the post-shock oscillation and pressure wiggles in the multicomponent interface [9, 4, 3]. In an important paper, Quirk catalogued a number of situations in which anomalous behaviour is known to occur [8]. Since then, a few papers have been proposed to explain this kinds of phenomenon. For example, for the post-shock oscillations, Arora and Roe interpreted that the

*Email:makxu@uxmail.ust.hk
†Email:majhu@uxmail.ust.hk

intermediate states inside the shock layer do not lie on the Hugoniot curve [1]. Hence, the Riemann solver generates a whole fan of waves and induces the post-shock oscillations. They also emphasized the potential importance to study the gas-kinetic BGK scheme[12]. Jin and Liu explained this in the context of traveling wave solution and suggested that the unsteadiness in the momentum spike is the cause for the oscillations [3]. Karni and Canic noticed that upwind scheme and Lax Friedrichs scheme behave differently here. They concluded that the small viscosity in Roe's scheme at the shock region contributes to the oscillations, and derived the parabolic modified equation plus a small perturbation in the shock layer[4].

The current work benefits from all previous experience and is focusing on the study of dynamical effects on the projection stage. Instead of guessing the underlying wave equations for the "numerical fluids", we construct a physical model for the dynamical averaging and derive the governing equations. As we know, the 1st-order Godunov-method always gives smeared numerical solutions, and it is usually interpreted as artificial viscosity effects. If the numerical fluxes are based on the exact Euler solutions, the dissipation must be added somehow in the projection stage. Our focal point in this paper is to figure out quantitatively the time-dependent dissipative mechanism in the projection stage for the moving shock case. Based on this model, we are going to explain numerical observations, such as the constant momentum peak value and the relation between the oscillation amplitude and shock speed.

We do not intend to compare different numerical schemes in this paper. We believe that certain anomalous phenomena are not caused originally by the flux functions, but by the dynamical effects on the projection stage. These anomalous behaviour are universal and intrinsic to any finite volume average taking shock capturing schemes. However, other anomalous phenomena, such as negative temperature, expansion shocks, carbuncle phenomena, and odd-even decoupling, are caused by the specific flux function. Different from the errors generated in the projection stage, these errors are not intrinsic and can be fixed case by case.

The main conclusion in this paper is that solving the Euler equations, even for the inviscid flow, is not adequate enough to correctly describe flow motion in the discretized space. We have to modify the governing equations by including appropriate viscous terms in the gas evolution stage in order to compensate the projection errors.

2 Post-Shock Oscillations

We consider the shock capturing schemes for the one dimensional Euler equations:

$$\begin{cases} \rho_t + (\rho U)_x = 0, \\ (\rho U)_t + (\rho U^2 + P)_x = 0, \\ E_t + (EU + PU)_x = 0, \end{cases} \tag{2.1}$$

where ρ is the density, U the velocity, ρU the momentum, $E = \frac{1}{2}\rho U^2 + \rho e$ the total energy, e the internal energy, and P the pressure. We assume that the gas is a γ-law gas, i.e. $P = (\gamma - 1)\rho e$, $1 \leq \gamma \leq 3$. In the following, we are going to construct a physical model for the shock capturing schemes in solving the above equations. At the same time, the observed artifacts, such as the post-shock oscillations, are well explained.

2.1 Physical Modeling

In order to fully understand the dynamic effects in the projection stage, let's construct a physical model. This model is general and is not limited to the shocks. Suppose we have a flow discontinuity inside a numerical cell j. The left and right states in cell j are (ρ_1, U_1, E_1) and (ρ_2, U_2, E_2) in regions $[x_{j-1/2}, x_{j-1/2} + \alpha\Delta x]$ and $[x_{j-1/2} + \alpha\Delta x, x_{j+1/2}]$ respectively, where Δx is the cell size, see Fig.(1). Since we are focusing on the physical behaviour, in the following we first assume $\alpha = \frac{1}{2}$. The projection averaging is based on the conservation of mass, momentum and energy,

$$\left\{ \begin{array}{l} \frac{1}{2}\rho_1 + \frac{1}{2}\rho_2 = \bar{\rho_j}, \\ \frac{1}{2}\rho_1 U_1 + \frac{1}{2}\rho_2 U_2 = \bar{\rho_j}\bar{U_j}, \\ \frac{1}{2}E_1 + \frac{1}{2}E_2 = \bar{E_j}, \end{array} \right. \tag{2.2}$$

where $\bar{\rho_j}, \bar{U_j}, \bar{E_j}$ are average density, velocity and energy in cell j. From above equations, we have

$$\bar{U_j} = \frac{\rho_1 U_1 + \rho_2 U_2}{\rho_1 + \rho_2}. \tag{2.3}$$

After projection, the kinetic energy in cell j becomes

$$\bar{E_k} = \frac{1}{2}(\rho_1 + \rho_2)\bar{U_j}^2. \tag{2.4}$$

However, the original kinetic energy before the projection is

$$E_k = \frac{1}{2}\rho_1 U_1^2 + \frac{1}{2}\rho_2 U_2^2. \tag{2.5}$$

From Eq.(2.4) and Eq.(2.5), we know $E_k \geq \bar{E_k}$ and the lost kinetic energy due to averaging is

$$\Delta E_k = E_k - \bar{E_k} = \frac{1}{2}\frac{\rho_1\rho_2}{\rho_1 + \rho_2}(U_2 - U_1)^2. \tag{2.6}$$

Since the total energy is conserved in the projection stage, the decreasing in kinetic energy corresponds to the increasing in thermal energy. So, the projection is actually a dynamical process which transfers the kinetic energy into thermal energy, and this process is finished through the so-called artificial dissipation. From thermodynamics, we can also prove that the entropy is increased in the above averaging process. At the same time, the transition from initial density distributions ρ_1 and ρ_2 to the final uniform density $\frac{1}{2}(\rho_1 + \rho_2)$ is finished through an equivalent mass diffusion process.

The flow variables in each cell are updated through the numerical fluxes in the gas evolution stage and the dissipative mechanism in the projection stage. If it is assumed that the Euler equations are exactly solved in the gas evolution stage, such as the Godunov method, the flow update can be described by the following "Navier-Stokes" equations (the detailed derivation is given in Appendix A),

$$\left\{ \begin{array}{l} \rho_t + (\rho U)_x = \epsilon\rho_{xx}, \\ (\rho U)_t + (\rho U^2 + P)_x = \epsilon(\rho U_x)_x, \\ E_t + (EU + PU)_x = \frac{1}{4}\epsilon(\rho U U_x)_x, \end{array} \right. \tag{2.7}$$

where the flux terms $(\rho U, \rho U^2 + P, EU + PU)$ are obtained from the Euler solutions in the gas evolution stage and the dissipative terms $(\epsilon\rho_{xx}, \epsilon(\rho U_x)_x, \frac{1}{4}\epsilon(\rho U U_x)_x)$ are from the projection stage. The viscous coefficient ϵ is

$$\epsilon \sim \frac{1}{2}\alpha(1 - \alpha)\frac{\Delta^2 x}{\Delta t}, \tag{2.8}$$

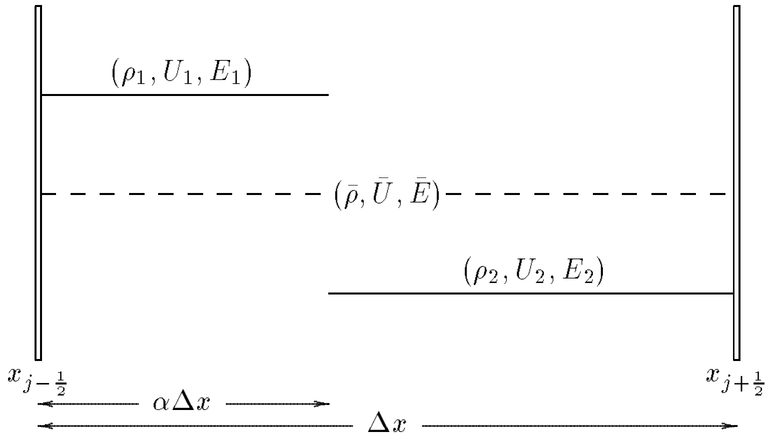

Figure 1: Subcell Mixing in the Projection Stage

which depends on the location of discontinuity as shown in Fig(1). The dependence of ϵ on Δx and Δt is due to the fact that the actual averaging process is finished dynamically through a whole time step inside each cell although it is often interpreted as an instantaneous process taken place at the end of each time step. Notice the differences between diffusion and dissipative terms in Eq.(2.7).

From the above physical model, the mass diffusion equation is naturally obtained, and this mass diffusion has nothing to do with the specific Euler fluxes across cell interface. More over, the viscosity terms in the momentum and energy equations (2.7) are different from the standard Navier-Stokes equations. Consequently, the existence and stability of the shock structure obtained from Eq.(2.7) should be in some ways different from Gilbarg and Paolucci's results for the standard Navier-Stokes equations [2]. In the following, we are going to present the shock structure for Eq.(2.7) and analyze its numerical consequences.

2.2 Stationary Shock Solutions for the "Navier-Stokes" Equations

For the steady state shock solutions with constant ϵ, Eq.(2.7) reduces to

$$\left\{\begin{array}{l} (\rho U)_x = \epsilon \rho_{xx}, \\ (\rho U^2 + P)_x = \epsilon(\rho U_x)_x, \\ (\frac{1}{2}\rho U^3 + \dfrac{\gamma}{\gamma - 1} PU)_x = \dfrac{\epsilon}{4}(\rho U U_x)_x, \end{array}\right. \tag{2.9}$$

and the boundary conditions are assumed to be

$$\left\{\begin{array}{ll} \rho(+\infty) = 1, & \rho_x(+\infty) = 0, \\ U(+\infty) = 1, & U_x(+\infty) = 0, \\ P(+\infty) = 1. & \end{array}\right. \tag{2.10}$$

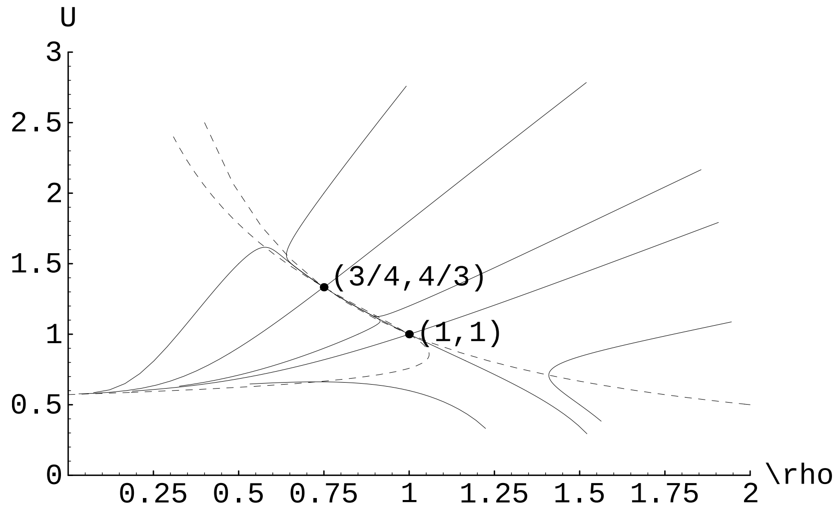

Figure 2: The computed phase plane for Eq.(2.11) with $\epsilon = 1$ and initial conditions(2.10). The dotted points are the two critical points; dash lines are two curves determined by Eq.(2.12).

Integrating Eq.(2.9) once, we obtain

$$\begin{cases} \rho U = \epsilon \rho_x + 1, \\ \rho U^2 + P = \epsilon \rho U_x + 2, \\ \dfrac{1}{2}\rho U^3 + \dfrac{\gamma}{\gamma - 1} P U = \dfrac{\epsilon}{4} \rho U U_x + \left(\dfrac{\gamma}{\gamma-1} + \dfrac{1}{2}\right). \end{cases}$$

After eliminating P from the last two equations in the above system, we have an autonomous system for ρ and U:

$$\begin{cases} \epsilon \rho_x = \rho U - 1, \\ \epsilon U_x = \left(\dfrac{\gamma}{\gamma-1} - \dfrac{1}{4}\right)^{-1} \left[\left(\dfrac{\gamma}{\gamma-1} + \dfrac{1}{2}\right)\rho^{-1}U^{-1} - \dfrac{2\gamma}{\gamma-1}\rho^{-1} + \left(\dfrac{\gamma}{\gamma-1} - \dfrac{1}{2}\right)U^2\right]. \end{cases} \tag{2.11}$$

The above autonomous system has two critical points at

$$(1,1) \quad \text{and} \quad \left(\frac{\gamma+1}{3\gamma-1}, \frac{3\gamma-1}{\gamma+1}\right).$$

The Jacobi matrix of the autonomous system is

$$\mathrm{J} = \epsilon^{-1}\left(\frac{\gamma}{\gamma-1} - \frac{1}{4}\right)^{-1} \begin{pmatrix} \left(\frac{\gamma}{\gamma-1} - \frac{1}{4}\right) U & \left(\frac{\gamma}{\gamma-1} - \frac{1}{4}\right) \rho \\ -\left(\frac{\gamma}{\gamma-1} + \frac{1}{2}\right)\rho^{-2}U^{-1} + \frac{2\gamma}{\gamma-1}\rho^{-2} & -\left(\frac{\gamma}{\gamma-1} + \frac{1}{2}\right)\rho^{-1}U^{-2} + 2\left(\frac{\gamma}{\gamma-1} - \frac{1}{2}\right) U \end{pmatrix}.$$

At the critical point $(1,1)$, the Jacobi matrix has eigenvalues

$$\lambda_{1,2} = \epsilon^{-1} \cdot \frac{7 + \gamma \pm \sqrt{33 - 18\gamma + 49\gamma^2}}{2 + 6\gamma}$$

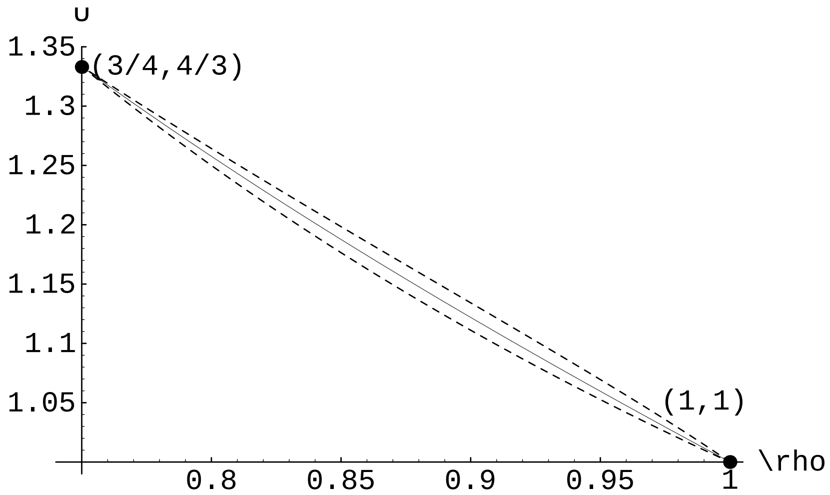

Figure 3: The detailed connection between two node points on the phase plane

Since $\gamma > 1$, these eigenvalues have opposite sign. Hence, the critical point $(1,1)$ is a saddle point. At the critical point $\left(\frac{\gamma+1}{3\gamma-1}, \frac{3\gamma-1}{\gamma+1}\right)$, the Jacobi matrix has eigenvalues

$$\lambda_{1,2} = \frac{\epsilon^{-1}}{2(1+\gamma)(1+3\gamma)}\left[-7+4\gamma+19\gamma^2 \pm \sqrt{33-56\gamma-90\gamma^2+152\gamma^3+217\gamma^4}\right]$$

For $\gamma > 1$, these eigenvalues are positive. Hence, the critical point $\left(\frac{\gamma+1}{3\gamma-1}, \frac{3\gamma-1}{\gamma+1}\right)$ is an un-stable node.

As an example with $\gamma = 1.4$ and $\epsilon = 1$, the calculated phase plane (ρ, U) is given in Fig.(2). From this figure, we can see that the two critical points are located at $(\frac{3}{4}, \frac{4}{3})$ and $(1,1)$. The heteroclinic orbit connects these two points in the direction from $(\frac{3}{4}, \frac{4}{3})$ to $(1,1)$. The blowup view of Fig(2) is shown in Fig.(3), where the heteroclinic orbit is shown clearly and is located within the region enclosed by the curves

$$\rho U - 1 = 0 \quad \text{and} \quad \left(\frac{\gamma}{\gamma-1}+\frac{1}{2}\right)\rho^{-1}U^{-1} - \frac{2\gamma}{\gamma-1}\rho^{-1} + \left(\frac{\gamma}{\gamma-1}-\frac{1}{2}\right)U^2 = 0, \tag{2.12}$$

where $\frac{\gamma+1}{3\gamma-1} \leq \rho \leq 1$. In Fig.(4)-(6), we also present the calculated density, velocity and momentum distributions across stationary shock layer for different $\epsilon = 0.1, 0.05, 0.033$. From these solutions, we know that

(1) ρ increases monotonically while U decreases, as x varies from $-\infty$ to $+\infty$;
(2) the momentum $\rho U \geq 1$ for all $x \in (-\infty, +\infty)$;
(3) the maximum value of momentum ρU is independent of ϵ.

From above analysis, we see that there is a stationary shock structure for the “Navier-Stokes” equations for each constant ϵ.

The property of constant momentum peak value can be explained from the following observations.

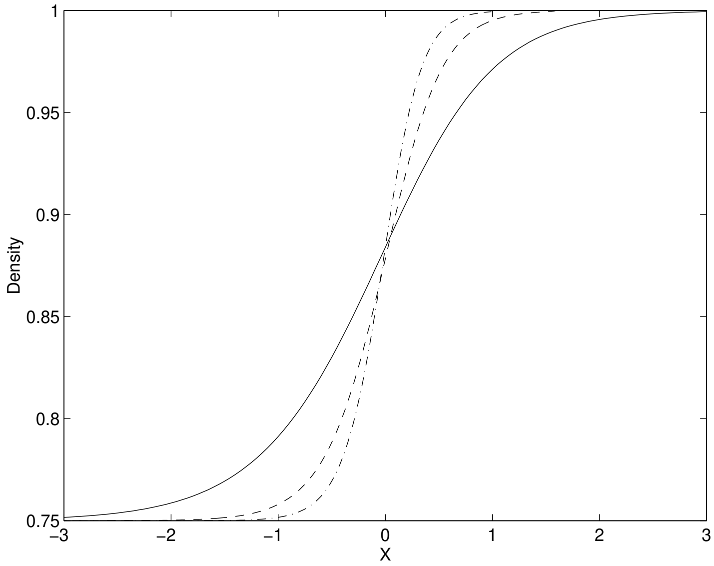

Figure 4: Density distribution inside shock layer for different ϵ, solid line $\epsilon = 0.1$, dash line $\epsilon = 0.05$, dash-dotted line $\epsilon = 0.03$

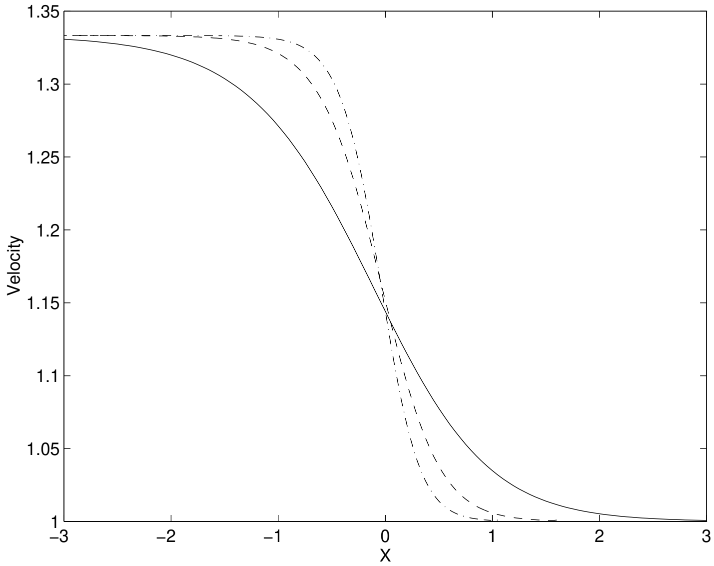

Figure 5: Velocity distribution inside shock layer for different ϵ, solid line $\epsilon = 0.1$, dash line $\epsilon = 0.05$, dash-dotted line $\epsilon = 0.03$

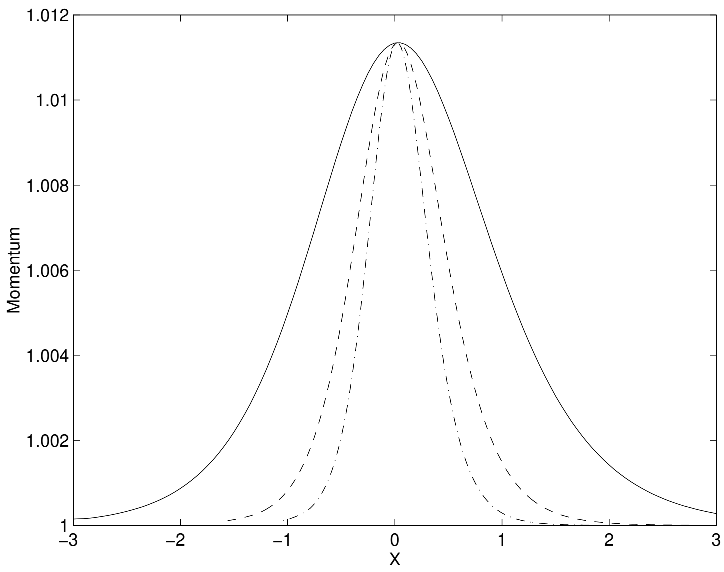

Figure 6: Momentum distribution inside shock layer for different ϵ, solid line $\epsilon = 0.1$, dash line $\epsilon = 0.05$, dash-dotted line $\epsilon = 0.03$

From the continuity equation, we have

$$(\rho U)_{\max} = \epsilon(\rho_x)_{\max} + 1.$$

Since the density variation on the leading order across shock layer is proportional to

$$(\rho_x)_{\max} \sim \frac{\rho_2 - \rho_1}{\delta x} \sim \frac{(\rho_2 - \rho_1)^2}{l\rho_1} \sim \frac{C_1(\rho_2 - \rho_1)^2}{\epsilon\rho_1},$$

where δx is the shock thickness, l the mean free path, (ρ_1, ρ_2) the upstream and downstream densities, and C_1 the upstream sound speed. So, $(\rho U)_{\max}$ can be approximated as

$$(\rho U)_{\max} = \epsilon(\rho_x)_{max} + 1 \sim \frac{C_1(\rho_2 - \rho_1)^2}{\rho_1} + 1,$$

which is independent of ϵ. Theoretically, the maximum peak value $(\rho U)_{\max}$ in the momentum distribution is a constant for all shock capturing schemes if only projection dissipation is involved.

2.3 Numerical Evidence and Physical Explanation

2.3.1 Oscillation Period and Wavelength

From mathematical analysis in the last section, it is shown that there is a steady state "Navier-Stokes" shock structure for each constant ϵ. If ϵ is really constant, such as in the stationary shock case, there should have a steady shock structure and there are no oscillations generated. However, almost all shock-capturing schemes for the moving shocks give post-shock oscillations. In order to explain this, we have to go back to our original physical model. Generally, the discontinuity could be located at any place inside each cell $(x_{j-1/2} + \alpha\Delta x)$ and its location is changing with time. So, α depends on the flow motion. As a consequence, ϵ in the "Navier-Stokes" Equations (2.7) will also depend on time

t. For single shock propagation, $\alpha(t)$ is a periodic function, and the only period existing is the time interval for the shock acrossing a numerical cell,

$$T = \frac{\Delta x}{U_s},$$

where U_s is the shock speed. So, in reality we are actually solving the modified "Navier-Stokes" equations(2.7) with unsteady viscosity coefficients in the moving shock cases. For different viscosity coefficients, there correspond to different shock structure as shown in Fig(4)-(6). The pulsating shock structure generates post-shock oscillations with the same period T. This conclusion is verified by many numerical experiments [3, 1].

The continuing variation of shock structure looks like a periodic forcing term acting on one end of a string. The oscillations are generated and propagated along the string. At the same time, the oscillating amplitude is decaying due to the dissipations. In the downstream flow region, the signal propagating speed depends on the specific waves. For example, the acoustic and entropy waves have speeds $U + C$ and U repectively. So, the wavelength for the post-shock oscillations on each characteristic wave is

$$l_{\text{oscillation}} = \begin{cases} (U + C)\dfrac{\Delta x}{U_s} & \text{acoustic wave,} \\ U\dfrac{\Delta x}{U_s} & \text{entropy wave,} \end{cases} \tag{2.13}$$

where C is the sound speed. There is no $U - C$ characteristic wave carrying oscillations, because it moves back into the shock front. These analysis are also numerically confirmed in [9, 1].

2.3.2 Constant Momentum Peak Value

The "Navier-Stokes" equations (2.7) give a constant momentum peak value. This conclusion is proved numerically in [3], where RELAX, ENO1-Roe, ENO2-Roe, ENO3-Roe, ENO1-LLF, ENO2-LLF and ENO3-LLF give almost the same momentum peak values in different test cases. The constant $(\rho U)_{\max}$ relies on the regularity of the "Navier-Stokes" equations which could support a consistent shock structure. The projection averaging plus Euler fluxes in (2.7) just provide us such equations. However, in the real gas evolution stage, we are NOT exactly solving the Euler equations. Different flux functions in the approximate Riemann solvers give different additional artificial dissipation terms. If we combine the dissipations from both the gas evolution stage and the projection stage, Eq.(2.7) goes to

$$\begin{cases} \rho_t + (\rho U)_x = \epsilon \rho_{xx} + \epsilon_1, \\ (\rho U)_t + (\rho U^2 + P)_x = \epsilon(\rho U_x)_x + \epsilon_2, \\ E_t + (EU + PU)_x = \frac{1}{4}\epsilon(\rho U U_x)_x + \epsilon_3, \end{cases} \tag{2.14}$$

where the so-called artificial dissipations $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the flux functions depend on the specific Riemann solver, and the ability of the above equations to support a regular travelling shock structure depends on the compatibility between $(\epsilon \rho_{xx}, \epsilon(\rho U_x)_x, \frac{1}{4}\epsilon(\rho U U_x)_x)$ and $(\epsilon_1, \epsilon_2, \epsilon_3)$. For the Godunov method, $(\epsilon_1 = 0, \epsilon_2 = 0, \epsilon_3 = 0)$ are satisfied and Eq.(2.14) will automatically generate the oscillations due to the unsteadiness in the viscous coefficient ϵ, but the momentum peak value is a constant which is independent of ϵ. However, if $(\epsilon_1 \neq 0, \epsilon_2 \neq 0, \epsilon_3 \neq 0)$, the final momentum distribution will depend on the specific formats of $(\epsilon_1, \epsilon_2, \epsilon_3)$. For example, in Lax Friedrich scheme the momentum peak value is contaminated by its large dissipative character in the gas evolution stage. The slight variations of $(\rho U)_{\max}$ for different schemes in the moving shock case [3] are also caused by the inconsistency among the dissipative terms in (2.14). As we will show later, the BGK scheme gives no variation of $(\rho U)_{\max}$ at all, which means that its intrinsic flux dissipations can be uniformly combined with the projection dissipations to form a steady travelling shock structure.

Due to the unsteadiness of viscosity coefficient, the time-dependent shock structure generates post-shock oscillations. This behaviour is only related to the momentum spike superficially. Actually, even without mass diffusion terms in the continuity equation (2.7) and with monotonic momentum distribution, the post-shock oscillations will still be generated due to the unsteady nature in the momentum and energy equations. This is also the main reason why the momentum spike can be eliminated through a simple transformation $\rho u \to \rho u - \epsilon \partial_x \rho$, but the post-shock oscillations will remain the same[3]. The use of mass flux instead of cell averaged momentum (updated by Eq.(2.14)) to get a monotonic momentum distribution is basically due to the fact that we are solving the Euler equations (2.1) for the numerical fluxes in the gas evolution stage, where there is no mass diffusion term at all[1]. The loose connection between momentum spike and oscillations can be also illustrated by the numerical results from the BGK scheme in this paper, where there is momentum spike but no post-shock oscillations. As another example, if applying the projection model to the moving contact discontinuity wave, we can get the same mass diffusion equation and consequently the unsteady momentum spike. However, we will not find any numerical oscillations there because of two reasons. The first reason is that there is no time-dependent nonlinear dissipative terms in the momentum and energy equations in Eq.(2.7) due to the equal velocities $U_1 = U_2$ on both sides of the discontinuity, see Eq.(2.6). The second reason is that at the contact discontinuity, although there is time-dependent mass diffusion terms to generate unsteady momentum distribution, the flux functions from the gas evolution stage will not try to steep the flow distribution and compete with the projection dissipations. So, the contact wave is just smearing along with time. So, it is the nonlinear effects of energy redistribution, and the competition between nonlinear steepening and projection dissipations form and support post-shock oscillations.

2.3.3 Oscillation Amplitude vs. Shock Speed

For very slow and very fast moving shocks, the amplitude in the oscillations are usually smaller than that in the intermediate shock speeds. There are detailed observational data about the relation between the oscillation amplitude and the shock speed[1, 5]. In this section, we are going to explain these observational curves.

[1] Usually we write numerical fluxes as

$$F_{j+1/2} = \frac{1}{2}(F_j + F_{j+1}) - Q_{j+1/2}\Delta_{j+1/2}U,$$

in the shock capturing schemes and regard $Q\Delta U$ as artificial viscous terms. This statement is not appropriate. Actually, both $F_{j+1/2}$ and $\frac{1}{2}(F_j + F_{j+1})$ could represent inviscid Euler fluxes, but for different initial conditions. The initial condition for $F_{j+1/2}$ is two constant states at the left and right cells, the other for $\frac{1}{2}(F_j + F_{j+1})$ is the linearly and continuously distributed flow variables across cell interface by connecting the left and right states.

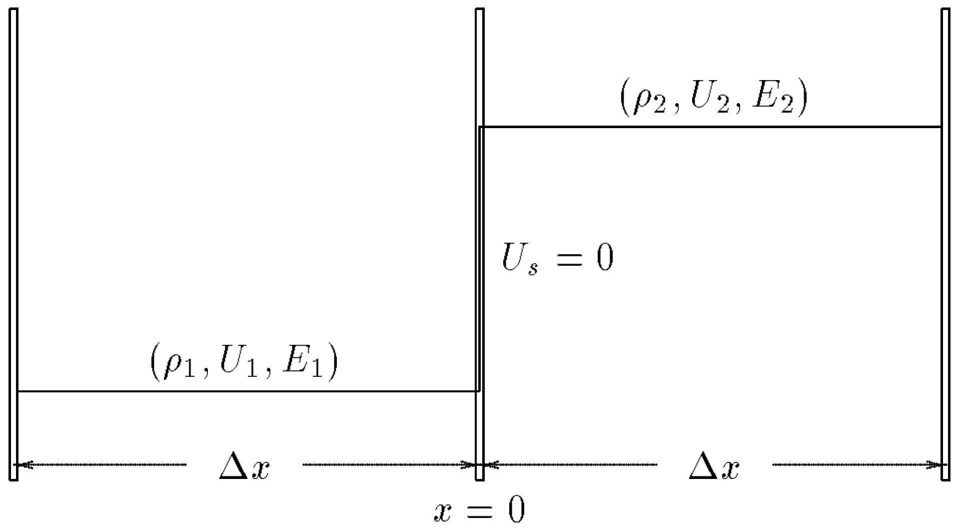

Figure 7: Stationary shock located at $x = 0$

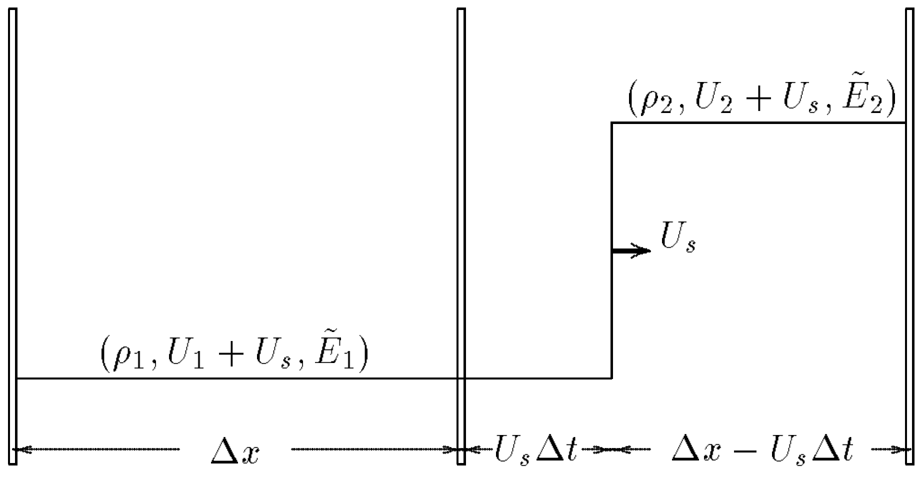

Figure 8: Moving shock with speed U_s at time Δt

In order to quantitatively evaluate the relation between the oscillation amplitude and shock speed, let's consider the following physical model. Initially a stationary shock is located at cell interface $x = 0$ with distributions (ρ_1, U_1, E_1) and (ρ_2, U_2, E_2) on the left and right sides, as shown in Fig(7). The upstream and downstream flow conditions are

$$\begin{cases} \rho_1 = 1, & \\ U_1 = 1, & x \leq 0, \\ P_1 = \frac{1}{\gamma M^2}. & \end{cases} \tag{2.15}$$

and

$$\begin{cases} \rho_2 = \dfrac{(\gamma+1)M^2}{2+(\gamma-1)M^2}\rho_1, & \\ U_2 = (\dfrac{\gamma-1}{\gamma+1} + \dfrac{2}{(\gamma+1)M^2})U_1, & x \geq 0, \\ P_2 = (\dfrac{2\gamma}{\gamma+1}M^2 - \dfrac{\gamma-1}{\gamma+1})P_1, & \end{cases} \tag{2.16}$$

where M is the upstream Mach number and is a free parameter here. From above flow conditions, we can get the sound speeds C_1 and C_2 on both sides,

$$C_1 = \sqrt{\frac{\gamma P_1}{\rho_1}} \qquad \text{and} \qquad C_2 = \sqrt{\frac{\gamma P_2}{\rho_2}}.$$

For moving shocks, the above flow conditions will be changed to $U_1 \to U_1 + U_s$ and $U_2 \to U_2 + U_s$, where U_s is the shock speed, and the shock location after time Δt will be changed to $U_s \Delta t$, see Fig(8). In the cell with shock, the total kinetic energy transferred into the thermal energy due to the projection is

$$\Delta E_k = \frac{1}{2}\frac{\rho_1 \Delta t U_s \rho_2 (\Delta x - \Delta t U_s)}{\rho_1 \Delta t U_s + \rho_2(\Delta x - \Delta t U_s)}(U_1 - U_2)^2, \qquad \text{for} \quad U_s > 0. \tag{2.17}$$

From the CFL condition (CFL number=1),

$$\Delta t = \frac{\Delta x}{|U_s| + Max(|U_1| + C_1, |U_2| + C_2)},$$

Eq.(2.17) goes to

$$\Delta E_k = \frac{1}{2}\frac{\rho_1 \rho_2 U_* (1 - U_*)}{\rho_1 U_* + \rho_2 (1 - U_*)}(U_1 - U_2)^2, \qquad \text{for} \quad U_s > 0, \tag{2.18}$$

where

$$U_* = \frac{|U_s|}{|U_s| + \mathrm{Max}(|U_1| + C_1, |U_2| + C_2)}.$$

The energy fluctuation (2.18) depends mainly on the shock speed and relative velocity, which is independent of numerical cell size. So, the post-shock oscillations can be never eliminated by refining mesh. Similarly, for $U_s < 0$, we have

$$\Delta E_k = \frac{1}{2}\frac{\rho_1 \rho_2 (1 - U_*) U_*}{\rho_1 (1 - U_*) + \rho_2 U_*}(U_1 - U_2)^2, \qquad \text{for} \quad U_s < 0. \tag{2.19}$$

Since all signals can be only propagated downstream after the shock, the continuing energy variations in the shock layer will perturbe the downstream flow motion. The relative energy fluctuation ΔE_k to the total downstream energy density $\tilde{E}_2$ is

$$\frac{\Delta E_k}{\tilde{E}_2} = \frac{\frac{1}{2}\frac{\rho_1 \rho_2 U_*(1-U_*)}{\rho_1 U_* + \rho_2(1-U_*)}(U_1-U_2)^2}{\frac{1}{2}\rho_2(U_2+U_s)^2 + \frac{1}{\gamma-1}P_2}, \qquad \text{for} \quad U_s > 0, \tag{2.20}$$

similarly,

$$\frac{\Delta E_k}{\tilde{E}_2} = \frac{\frac{1}{2}\frac{\rho_1 \rho_2 (1-U_*)U_*}{\rho_1(1-U_*) + \rho_2 U_*}(U_1-U_2)^2}{\frac{1}{2}\rho_2(U_2+U_s)^2 + \frac{1}{\gamma-1}P_2}, \qquad \text{for} \quad U_s < 0. \tag{2.21}$$

Fig.(9) is the plot of relative energy fluctuation $\frac{\Delta E_k}{\tilde{E}_2}$ vs. the shock speed U_s for different Mach number. As we can see, the relative kinetic energy is smaller at both lower and higher shock speeds. Since the energy perturbation acting on the downstream flow looks like a continuing energy source to disturb the main flow motion, it will make the downstream flow oscillating. By the definition of total energy density $E = \frac{1}{2}\rho U^2 + \frac{1}{\gamma-1}P$, we have the energy change

$$\Delta E = \rho U \Delta U + \frac{1}{2}U^2 \Delta\rho + \frac{1}{\gamma-1}\Delta P.$$

Since $C^2 \sim \frac{\gamma \Delta P}{\Delta \rho}$ and from the Riemann invariant $\Delta U \sim \frac{\Delta P}{\rho C}$, we have

$$\Delta E \sim \Delta\rho(\frac{|U|C}{\gamma} + \frac{1}{2}U^2 + \frac{C^2}{\gamma(\gamma-1)}),$$

from which we can get the relative density change in the downstream flow motion due to the energy perturbation ΔE_k,

$$\frac{\Delta\rho}{\rho_2} = \frac{1}{\rho_2}\frac{\Delta E_k}{\frac{|(U_2+U_s)|C_2}{\gamma} + \frac{1}{2}(U_2+U_s)^2 + \frac{C_2^2}{\gamma(\gamma-1)}}.$$

Fig.(10) is the plot of relative density changes for different Mach number. If we reverse the shock direction by reversing velocities U_1 and U_2, and position of ρ_1 and ρ_2 gases in Eq.(2.15) and (2.16), a similar figure can be obtained for the density variations which are shown in Fig.(11). Comparing with the numerical results presented in [5, 1], we can see that the theoretical analysis provide similar distributions. In the real flow calculations, the above theoretical amplitude would be modified and reduced by the dissipative terms in Eq.(2.14), but the overall shape will be kept the same. For example, the fast moving shock creats high frequency modes which are decaying much faster than that of low frequency mode due to the flux dissipation. As a result, the observational amplitude is decaying faster than our theoretical analysis, especially in the high shock speed regions. Also, the shock layer may be smeared by several mesh points, the intermediate states in the shock layer may be different from the upstream and downstream flow conditions. The final observation is a statistical averaging over all possible states in the shock layer. For example, the kinetic energy fluctuation should be modified to

$$\overline{\Delta E_k} = \frac{1}{\beta-1}\int_1^\beta \Delta E_k(\beta')d\beta',$$

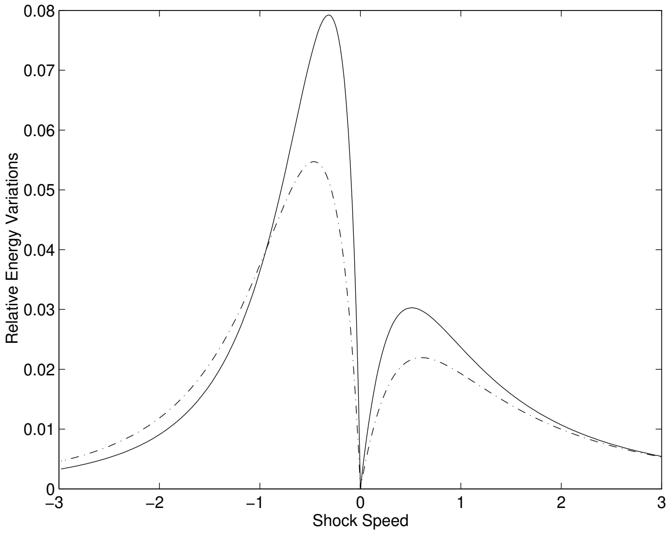

Figure 9: Relative energy variations $\frac{\Delta E_k}{\tilde{E}_2}$ vs. shock speed U_s for different Mach number M; dash-dotted line $M = 3.0$, solid line $M = 20.0$

which is an averaging over all possible density jumps,

$$\beta' = \frac{\rho_2}{\rho_1},$$

and

$$\beta = \frac{(\gamma+1)M^2}{2+(\gamma-1)M^2}.$$

The distribution $\overline{\Delta E_k}$ has the similar shape as the original distributions in (2.18) and (2.19) because the averaging is taken on the similar functions. From above analysis, it is proved once again that the oscillations are generated and maintained nonlinearly by the unsteadiness of dissipative mechanism in the projection stage.

2.4 BGK Scheme for Moving Shocks

In the last section, we have analyzed the mechanism for the generation of post-shock oscillations. The possible way to reduce the oscillations is to have compatible dissipative terms from gas evolution stage to compensate the unsteady projection dissipations in Eq.(2.14). In this section, we are going to present the numerical results from the BGK scheme and show the possible inspiration we can get from it. The BGK scheme is tested for both the slowly moving and stationary shocks with the mesh refinement. In both cases, $\gamma = 1.4$ is used. The detailed numerical flux functions for the 1st-order BGK scheme are given in the Appendix B.

Example 1: We take the following initial data [8] that gives a Mach-3 shock moving to the right with a speed $s = 0.1096$,

$$(2.22)\quad W_1 = \begin{pmatrix} 3.86 \\ -3.1266 \\ 27.0913 \end{pmatrix} \quad \text{if} \quad 0 \leq 0.5; \qquad W_2 = \begin{pmatrix} 1.0 \\ -3.44 \\ 8.4168 \end{pmatrix} \quad \text{if} \quad 0.5 \leq x \leq 1,$$

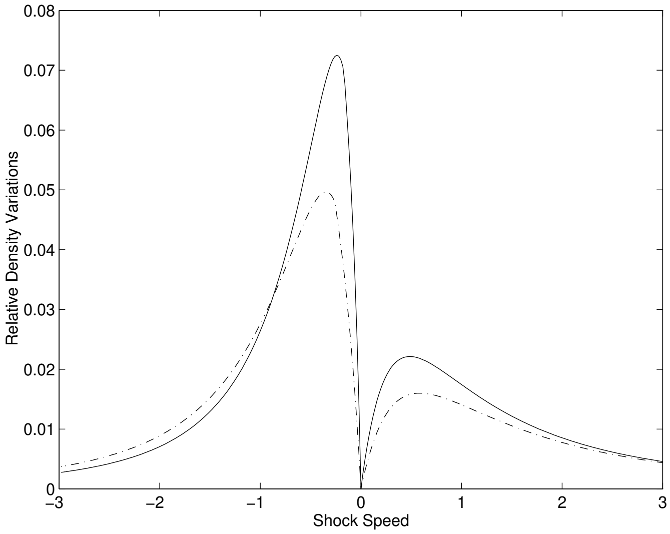

Figure 10: Relative density variations $\frac{\Delta\rho}{\rho_2}$ vs. shock speed U_s for different Mach number M; dash-dotted line $M = 3.0$, solid line $M = 20.0$

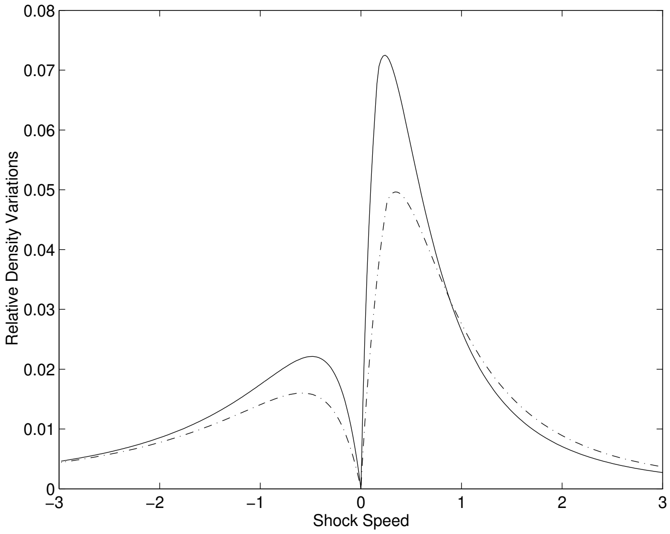

Figure 11: Relative density variations $\frac{\Delta\rho}{\rho_2}$ vs. shock speed U_s for different Mach number M in a reversed shock; dash-dotted line $M = 3.0$, solid line $M = 20.0$.

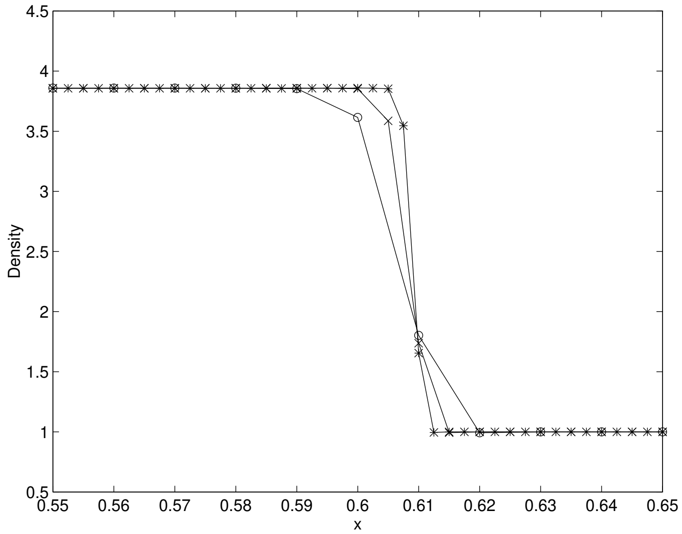

Figure 12: Density distribution for Example 1 using BGK scheme,___o___$\Delta x = 0.01$, ___x___$\Delta x = 0.005$ and ___$*$___$\Delta x = 0.0025$

where W are the mass, momentum and energy densities. We have used $100, 200$ and 400 mesh points in the calculations with CFL number$= 0.65$. The output time is at $t = 0.95$. The density and momentum distributions around shock fronts are shown in Fig(12) and Fig(13). For different viscosity coefficient ϵ, see Eq.(2.8), we observe that the momentum peak value is independent of ϵ, but the numerical shock thickness is reducing along with the decreasing of ϵ. There is no oscillation generated even with two or three mesh points in the shock layer. So, the intrinsic dissipative mechanism in the BGK fluxes are compatible with the projection dissipations to form a numerical shock structure.

Example 2: We take the following initial data [3] that gives a stationary shock.

$$(2.23) \qquad W_1 = \begin{pmatrix} 2/3 \\ 1/\sqrt{2} \\ 8/14 \end{pmatrix} \quad \text{if} \quad 0 \leq 0.5; \qquad W_2 = \begin{pmatrix} 2 \\ 1/\sqrt{2} \\ 23/14 \end{pmatrix} \quad \text{if} \quad 0.5 \leq x \leq 1.$$

We also use $100, 200$ and 400 mesh points in this calculation. The output time is at $t = 2$. The density and momentum distributions around shock front are shown in Fig(14) and Fig(15). Once again, for different ϵ the momentum peak value keeps a constant. Note that the shock location is slightly off the center of the correct location because a small amount of mass has been carried away initially when the shock is settling down to its final numerical structure.

2.5 The Requirement of Consistent Dissipative Mechanism

The BGK scheme gives sharp and oscillation free shock transition. This can be explained by noticing that the BGK scheme is not a scheme for the Euler equations and its flux function has intrinsic dissipations $(\epsilon_1, \epsilon_2, \epsilon_3)$ due to the Non-Equilibrium property in the gas distribution function f at the cell interface. This non-equilibrium mimics the physical mechanism in the construction of numerical shock layer. The non-equilibrium is represented by the fact that the gas distribution deviates

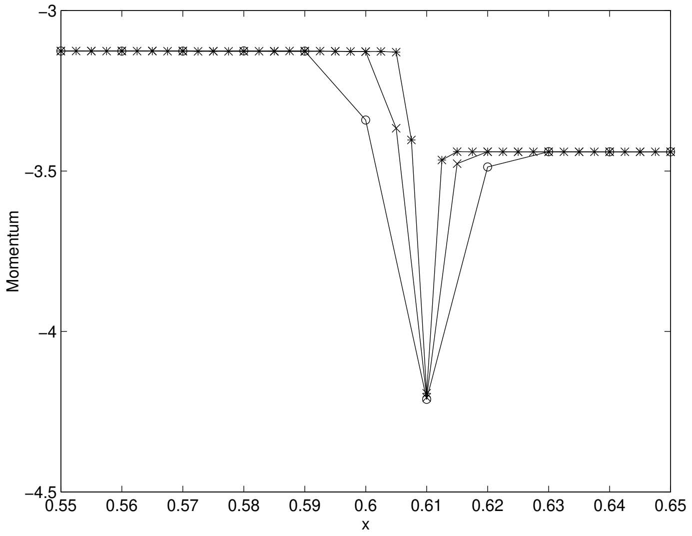

Figure 13: Momentum distribution for Example 1 using BGK scheme,___o___$\Delta x = 0.01$, ___x___$\Delta x = 0.005$ and ___$*$___$\Delta x = 0.0025$

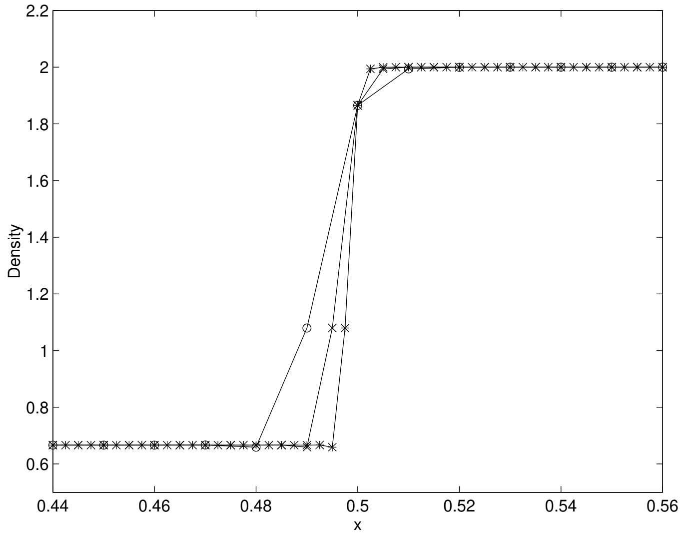

Figure 14: Density distribution for Example 2 using BGK scheme, ___o___$\Delta x = 0.01$, ___x___$\Delta x = 0.005$ and ___$*$___$\Delta x = 0.0025$

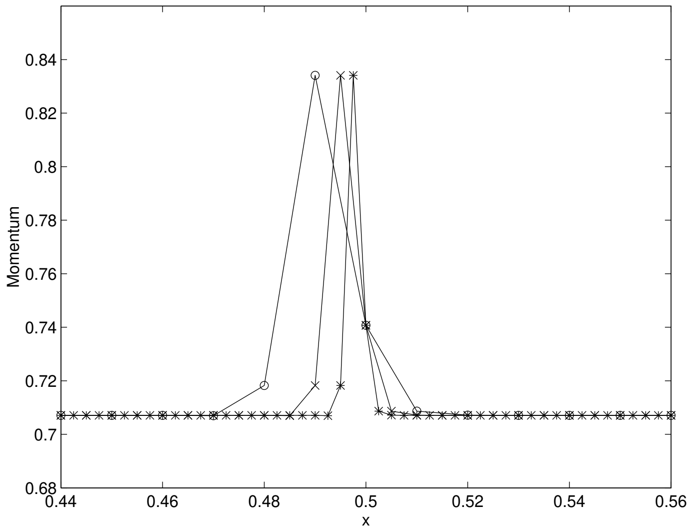

Figure 15: Momentum distribution for Example 2 using BGK scheme,__o__$\Delta x = 0.01$, __x__$\Delta x = 0.005$ and __$*$__$\Delta x = 0.0025$

from a real Maxwellian. The importance of non-Maxwellian is well-recognized in the study of strong shock structure[7]. In the Godunov method, the Euler equations require the flux functions with $(\epsilon_1 = 0, \epsilon_2 = 0, \epsilon_3 = 0)$ in Eq.(2.14), which could not provide the physical dissipations needed together with projection dissipations to form a steady travelling shock structure. For these schemes with implicit artificial dissipations added, such as most TVD and ENO schemes, the shock transitions are sharply resolved. However, due to the inconsistency between the projection and evolution dissipations, the post-shock oscillations are still generated. In other words, the projection stage starts the oscillations through unsteady dissipative mechanism, the Godunov-type schemes could hardly dissipate all oscillating modes consistently in a physical way. To have a consistent dissipative mechanism is critical for any high resolution schemes; otherwise gigantic amount of viscosity is needed to smear shock layer in order to get an oscillation free solution, such as the Lax Friedrich scheme which has $(\epsilon_1 \gg \epsilon\rho_{xx}, \epsilon_2 \gg \epsilon(\rho U_x)_x, \epsilon_3 \gg \frac{1}{4}\epsilon(\rho U U_x)_x)$ in Eq.(2.14). Without having a consistent Navier-Stokes-like governing equations in the gas evolution stage, we have to carefully tune the viscosity terms $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the hope to get a smooth shock transition. For example, as observed in [5, 4], in order to suppress the oscillations in Roe's scheme, the justible parameters should in some ways depend on the time step Δt and specific flow situations; and there is no any uniform criteria found yet. This is not surprising because we are searching something to be compatible with the time-dependent projection dissipations. In fact, many misconducts in the Riemann solvers have been controlled also by the use of artificial viscosity. It is worth point out that Colella and Woodward's PPM scheme[10], which has proved itself to be robust does in fact employ an elaborate artificial dissipation mechanism to supplement the dissipation provided via upwinding. However, without a good physical model, the added artificial dissipations seem not as good as that provided in the BGK model, at least in the moving shock case. In the chemical reactive flow calculations, the requirement of the scheme which has physical Navier-Stokes limit is also important, otherwise the Euler solutions could not converge with mesh refinement[6].

Another purpose to put additional dissipation terms $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the flux functions is to reduce the projection errors by stretching the viscous shock layer. Therefore, the better the Euler equations are solved in the gas evolution stage, the worse the post-shock oscillations are maintained. It is reasonable for Karni and Canic to add viscosity in Roe's Riemann solver to reduce the amplitude in the post-shock oscillations [4]. However, without a good physical model for the modifications, it is hard to tell how much dissipation is enough if we can not be trusted to control it in a physically consistent way. Also, the post-shock oscillations are not formed originally by the numerical flux functions and the use of 3rd order modified equation to explain them is not necessary [4]. Arora and Roe concluded that the oscillation is due to fact that the intermediate states in the shock regions are not located on the Hugoniot curve. This conclusion is based only on the Euler equations. For the "Navier-Stokes" solutions, the states inside the shock layer indeed would not stay on the Hugoniot curve, but they will not generate oscillations either if the dissipation terms in (2.14) are consistent, such as the BGK scheme.

The real behavior of numerical schemes are much more complicated than the physical model outlined in this paper, especially for high-order schemes. However, our model in some ways points out the critical dynamical influence from both the projection and evolution stages on the numerical fluids. In most other analysis [1, 3, 4], the focal point is on the flux function in the gas evolution stage which could hardly explain satisfactorily all observed numerical artifacts. However, how to define the consistency between projection and evolution dissipations, how to explicitly figure out $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the BGK scheme, and how to understand the physical pictures in Eq.(2.14) to form a steady discretized travelling shock structure, are still unknowns. But we believe that the BGK scheme provides something special here and we need pay attention to that.

3 Re-Understanding of Numerical Schemes

There are two process governing the flow update, gas evolution around cell interface and dynamical projection inside each cell. The main dynamical influence in the projection stage is that it provides time-dependent dissipative terms. The modeling and analysis of projection in the last section are very general. Any flow structures with high gradients, where the flow variables change substantially on the scale of cell size, will be effected by the diffusion and dissipation in the averaging process. In order to compensate and reduce the projection errors, we have to in some ways modify the flux function consistently in the gas evaluation stage. Instead of solving the Euler equations, we have to consider the Navier-Stokes-like equations[2]. The post-shock oscillations can be only eliminated by consistent combination of all dissipative terms on the right hand side of Eq.(2.14). This seems contradicting to our general idea where we are focusing on reducing the dissipation $(\epsilon_1, \epsilon_2, \epsilon_3)$ as much as possible in order to get sharp discontinuity representation and correct Euler solutions in the smooth flow regions. In order to get both smooth and discontinuous flow correct, the use of adaptive parameter in controlling the viscous terms $(\epsilon_1, \epsilon_2, \epsilon_3)$ seem necessary. The search for ultimate schemes which are free of any justiable parameters and are robust to describe all kinds of flow situations looks impossible and the goal can be hardly achieved by considering the Euler equations. Instead of blaming the justiable parameters, the important thing is to give a good physical reason for that. To find this reason in some ways depends on a correct understanding of numerical fluids on the discretized space. The dynamical effects in the projection stage for the nonlinear system give us something special phenomena, such as the kinetic energy being transferred into thermal energy. This special feature in the nonlinear system breaks our old believing that any good numerical techniques for solving the linear wave equation ($U_t + aU_x = 0$) could be extended by a simple mechanism into equally good numerical techniques for solving system of conservation laws $\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = 0$. It is rather meaningless to keep on searching good schemes for the scalar equation or linear system in the hope to get good methods for the nonlinear ones. Actually, for the nonlinear system the projection dissipations prevent the numerical "Navier-Stokes" solutions from converging to the Euler solutions even with the viscosity coefficient $\epsilon \rightarrow 0$ in Eq.(2.7). For example, the momentum peak value is independent of ϵ in the numerical shock layer.

In order to reduce the post-shock oscillations, we need to solve correct viscous equations in the gas evolution stage. The traditional extension from the Euler to the Navier Stokes equations by simply splitting the inviscid and viscous terms would probably give us the post-shock oscillations again due to the splitting errors. This involves another hot topic in the CFD community about how to treat the source term in the Godunov-type schemes and how much the errors would be generated if we separate source term from other terms. We believe that the study of numerical shock structure for the nonlinear system will provide us a good base to understand and develop new algorithms. The consistency of the artificial viscosity in the flux functions can be also tested from the shock width and oscillatory pattern, and any claims about the order of accuracy should be tested here rather than in the smooth flow regions. Actually, the numerical shock layer indeed provides us a smooth flow region if we believe that we have to solve the "Navier-Stokes" equations (2.14). The special feature here is that all terms, convection and dissipation, physical and numerical, will behave themselves equally in the numerical shock formation and the final shock transition is finished by their common efforts.

4 Conclusion

In this paper, we are focusing on the projection and evolution dynamics in the Godunov-type schemes. The projection averaging is a purely numerical aspect due to the limited cell size and time step, however it provides dynamical influence on the flow motion. So, numerics and physics are mixed in any schemes, and how to distinguish them and unify them consistently is the main problem we need to pay attention to. Contrarily to the conventional opinion in the shock capturing community that the schemes based closely on the exact Euler solutions will yield robust and accuracy schemes, we find that the projection errors can be hardly eliminated by considering the Euler equations. The

[2] Sometimes the transition from the upstream to downstream in the numerical shock layer is considered to be connected by several small shocks. From physical point of view, it is impossible to reach the same final state by compressing the gas through several shock waves as is reached by compressing with a single shock wave. For example, a strong shock wave propagating through a monatomic gas will yield a density ratio of 4, while two successive strong shock waves could result in a density ratio of 16. So, the transition cells in a numerical shock layer have to be regarded as the shock structure. In other words, the intermediate states can not stay on the Hugoniot curve connecting the unstream and downstream for the nonlinear system.

independence of numerical momentum peak value to the viscosity coefficient also gives us an example that the numerical "Navier-Stokes" solutions will not go to the Euler solutions even in the limit $\epsilon \to 0$ in Eq.(2.7). We also believe that ultimate schemes to correctly describe both smooth and discontinuity flows can be hardly obtained from the Euler equations. Unless the fluxes in the Riemann solver is augmented by consistent dissipations, any Godunov-type schemes will be flawed. Here we advocate the BGK-type scheme again, and the study of its nonlinear dissipative mechanism in order to get more understanding about nonlinear system.

Looking to the future, it is to be hoped that certain criteria can be obtained in the algorithm development for the nonlinear system, and the reliable criteria should largely be based on physical models. It seems that we can not be so optimistic as people expressed in the past years for the upwinding schemes. The more we are pursuing on it, the fewer we feel to get the ultimate truth.

Acknowledgments

Kun Xu wishes to thank Prof. W.H. Hui for helpful discussion about differences between Lagrangian and Eulerian schemes which initiates our work. The current research is supported by the Research Grant Council of Hong Kong through DAG Grant 96/97 for both authors.

Appendix A: The Derivation of the "Navier-Stokes" Equations

As shown in Fig(1), initial discontinuous distributions are evolved into constant equilibrium states in a time scale Δt. If only the projection dynamics being considered, the averaging can be regarded as taking place in an isolated system. In the following, we are going to derive the approximate diffusion and viscous equations in such a system. At the end of this section, the convective fluxes from the gas evolution stage will be added and the full "Navier-Stokes" equations (2.7) are obtained.

1. Continuity Equation: Looking at the subcell $x\epsilon[x_{j-1/2}, x_{j-1/2}+\alpha\Delta x]$, where the density is changed from the initial ρ_1 to the final $\bar{\rho} = \alpha\rho_1 + (1-\alpha)\rho_2$ through the diffusive fluxes at $x = x_{j-1/2} + \alpha\Delta x$. If we assume that the mass diffusive flux is $\eta_1\rho_x$, we have

$$\begin{aligned} \frac{\bar{\rho}-\rho_1}{\Delta t} &= \frac{\eta_1\rho_x}{\alpha\Delta x} \\ &= \frac{\eta_1(\rho_2-\rho_1)}{\alpha\Delta x\frac{1}{2}\Delta x}, \end{aligned}$$

from which the diffusive coefficient η_1 can be obtained $\eta_1 = \frac{1}{2}\alpha(1-\alpha)\frac{\Delta^2 x}{\Delta t^2}$.

2. Momentum Equation: Looking at the subcell $x\epsilon[x_{j-1/2}, x_{j-1/2}+\alpha\Delta x]$ again, where the initial momentum $\rho_1 U_1$ is changed to $\bar{\rho}\bar{U}$ in a time step Δt. With the assumption of viscous flux $\eta_2 U_x$, we have

$$\begin{aligned} \frac{\bar{\rho}\bar{U}-\rho_1 U_1}{\Delta t} &= \frac{\alpha\rho_1 U_1 + (1-\alpha)\rho_2 U_2 - \rho_1 U_1}{\Delta t} \\ &= \frac{\eta_2 U_x}{\alpha\Delta x} \\ &= \frac{\eta_2(U_2-U_1)}{\alpha\Delta x\frac{1}{2}\Delta x}. \end{aligned}$$

With the assumption

$$\rho_2 U_2 - \rho_1 U_1 \sim \rho(U_2 - U_1),$$

we have

$$\eta_2 = \frac{1}{2}\alpha(1-\alpha)\frac{\Delta^2 x}{\Delta t^2}\rho.$$

3. Energy Equation: The dissipative term in the energy equation is to transfer the kinetic energy into thermal energy. At this time, we are going to look at the whole numerical cell Δx, and write the energy dissipation as,

$$\frac{\partial E_k}{\partial t} = (\eta_3 U U_x)_x.$$

From the above equation and Eq.(2.6), we have

$$\begin{aligned}
\frac{\Delta E_k}{\Delta t} &= \frac{\frac{1}{2}\frac{\alpha\rho_1(1-\alpha)\rho_2}{\alpha\rho_1 + (1-\alpha)\rho_2}(U_2 - U_1)^2}{\Delta t} \\
&= \eta_3 \frac{(U_2 - U_1)U_x}{\frac{1}{2}\Delta x} \\
&= \eta_3 \frac{(U_2 - U_1)^2}{\frac{1}{2}\Delta x \frac{1}{2}\Delta x}.
\end{aligned}$$

So, with the assumption

$$\frac{\rho_1\rho_2}{\alpha_1\rho_1 + (1-\alpha)\rho_2} \sim \rho,$$

we get

$$\eta_3 = \frac{1}{8}\alpha(1-\alpha)\frac{\Delta^2 x}{\Delta t^2}\rho.$$

The gas evolution stage and the projection stage are taking place at the same time. While the projection averaging is smearing the subcell structure, the evolution stage provides the dynamical influence from the externals through the numerical fluxes across cell interface. If the Euler equations are precisely solved in the gas evolution stage, the mass, mometum and energy transports through a cell interface are $(\rho U, \rho U^2 + P, EU + PU)$. After combining with the diffusion and dissipative terms in the projection stage, we have the following "Navier-Stokes" Equations for the update of flow variables,

$$\begin{cases}
\rho_t + (\rho U)_x = \epsilon \rho_{xx}, \\
(\rho U)_t + (\rho U^2 + P)_x = \epsilon(\rho U_x)_x, \\
E_t + (EU + PU)_x = \frac{1}{4}\epsilon(\rho U U_x)_x,
\end{cases}$$

where

$$\epsilon = \frac{1}{2}\alpha(1-\alpha)\frac{\Delta^2 x}{\Delta t^2}.$$

Appendix B: 1st-Order BGK Scheme

The numerical fluxes for the 1st-order BGK scheme is based on the combination of two fluxes: the nonequilibrium fluxes from initial gas distribution function f_0 and the equilibrium fluxes from the Maxwellian distribution function g_0. The final distribution function f at $x_{j+1/2}$ is

$$f_{j+1/2} = \frac{1}{2}g_0 + \frac{1}{2}f_0.$$

Note that the above distribution function is different from a single Maxwellian. We shall present the basic formulas in the following, and refer to [11, 12] for more details.
1. Given the initial data $\{\rho_j^n, U_j^n, E_j^n\}$ in each cell j, compute $\{\lambda_j^n\}$ using

$$\lambda = \frac{K+1}{4}\frac{\rho}{E-\frac{1}{2}\rho U^2}$$

where $K=4$ for $\gamma=1.4$ gas.
2. Compute the numerical fluxes from f_0, which are denoted as $\{F^0_{\rho,j+1/2}, F^0_{\rho U,j+1/2}, F^0_{E,j+1/2}\}$.

$$\begin{pmatrix} F^0_{\rho,j+1/2} \\ F^0_{\rho U,j+1/2} \\ F^0_{E,j+1/2} \end{pmatrix} = \rho_j \begin{pmatrix} \frac{U_j}{2}\mathrm{erfc}(-\sqrt{\lambda_j}U_j) + \frac{1}{2}\frac{e^{-\lambda_j U_j^2}}{\sqrt{\pi\lambda_j}} \\ \left(\frac{U_j^2}{2}+\frac{1}{4\lambda_j}\right)\mathrm{erfc}(-\sqrt{\lambda_j}U_j) + \frac{U_j}{2}\frac{e^{-\lambda_j U_j^2}}{\sqrt{\pi\lambda_j}} \\ \left(\frac{U_j^3}{4}+\frac{K+3}{8\lambda_j}U_j\right)\mathrm{erfc}(-\sqrt{\lambda_j}U_j) + \left(\frac{U_j^2}{4}+\frac{K+2}{8\lambda_j}\right)\frac{e^{-\lambda_j U_j^2}}{\sqrt{\pi\lambda_j}} \end{pmatrix}$$

$$+\rho_{j+1} \begin{pmatrix} \frac{U_{j+1}}{2}\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) - \frac{1}{2}\frac{e^{-\lambda_{j+1} U_{j+1}^2}}{\sqrt{\pi\lambda_{j+1}}} \\ \left(\frac{U_{j+1}^2}{2}+\frac{1}{4\lambda_{j+1}}\right)\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) - \frac{U_{j+1}}{2}\frac{e^{-\lambda_{j+1} U_{j+1}^2}}{\sqrt{\pi\lambda_{j+1}}} \\ \left(\frac{U_{j+1}^3}{4}+\frac{K+3}{8\lambda_{j+1}}U_{j+1}\right)\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) - \left(\frac{U_{j+1}^2}{4}+\frac{K+2}{8\lambda_{j+1}}\right)\frac{e^{-\lambda_{j+1} U_{j+1}^2}}{\sqrt{\pi\lambda_{j+1}}} \end{pmatrix}.$$

3. Compute the numerical fluxes $\{F^1_{\rho,j+1/2}, F^1_{\rho U,j+1/2}, F^1_{E,j+1/2}\}$ for the equilibrium states g_0,

$$\begin{pmatrix} F^1_{\rho,j+1/2} \\ F^1_{\rho U,j+1/2} \\ F^1_{E,j+1/2} \end{pmatrix} = \bar{\rho}_{j+1/2} \begin{pmatrix} \bar{U}_{j+1/2} \\ \bar{U}^2_{j+1/2} + \frac{1}{2\lambda_{j+1/2}} \\ \frac{1}{2}\bar{U}^3_{j+1/2} + \frac{K+3}{4\lambda_{j+1/2}}\bar{U}_{j+1/2} \end{pmatrix},$$

where

$$\begin{pmatrix} \bar{\rho}_{j+1/2} \\ \bar{\rho}_{j+1/2}\bar{U}_{j+1/2} \\ \frac{1}{2}\bar{\rho}_{j+1/2}\left(\bar{U}^2_{j+1/2}+\frac{K+1}{2\lambda_{j+1/2}}\right) \end{pmatrix} = \rho_j \begin{pmatrix} \frac{1}{2}\mathrm{erfc}(-\sqrt{\lambda_j}U_j) \\ \frac{1}{2}U_j\mathrm{erfc}(-\sqrt{\lambda_j}U_j) + \frac{1}{2}\frac{e^{-\lambda_j U_j^2}}{\sqrt{\pi\lambda_j}} \\ \frac{1}{2}\left(\frac{U_j^2}{2}+\frac{K+1}{4\lambda_j}\right)\mathrm{erfc}(-\sqrt{\lambda_j}U_j) + \frac{U_j}{4}\frac{e^{-\lambda_j U_j^2}}{\sqrt{\pi\lambda_j}} \end{pmatrix}$$

$$+\rho_{j+1} \begin{pmatrix} \frac{1}{2}\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) \\ \frac{1}{2}U_{j+1}\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) - \frac{1}{2}\frac{e^{-\lambda_{j+1} U_{j+1}^2}}{\sqrt{\pi\lambda_{j+1}}} \\ \frac{1}{2}\left(\frac{U_{j+1}^2}{2}+\frac{K+1}{4\lambda_{j+1}}\right)\mathrm{erfc}(\sqrt{\lambda_{j+1}}U_{j+1}) - \frac{U_{j+1}}{4}\frac{e^{-\lambda_{j+1} U_{j+1}^2}}{\sqrt{\pi\lambda_{j+1}}} \end{pmatrix}.$$

4. Update $\{\rho_j^n, (\rho U)_j^n, E_j^n\}$ to $\{\rho_j^{n+1}, (\rho U)_j^{n+1}, E_j^{n+1}\}$ in a time step Δt through

$$\begin{pmatrix} \rho_j^{n+1} \\ (\rho U)_j^{n+1} \\ E_j^{n+1} \end{pmatrix} = \begin{pmatrix} \rho_j^{n} \\ (\rho U)_j^{n} \\ E_j^{n} \end{pmatrix} + \sigma \begin{pmatrix} F^0_{\rho,j-1/2} - F^0_{\rho,j+1/2} \\ F^0_{\rho U,j-1/2} - F^0_{\rho U,j+1/2} \\ F^0_{E,j-1/2} - F^0_{E,j+1/2} \end{pmatrix} + \sigma \begin{pmatrix} F^1_{\rho,j-1/2} - F^1_{\rho,j+1/2} \\ F^1_{\rho U,j-1/2} - F^1_{\rho U,j+1/2} \\ F^1_{E,j-1/2} - F^1_{E,j+1/2} \end{pmatrix},$$

where $\sigma = 0.325\frac{\Delta t}{\Delta x}$ for CFL number= 0.65.

References

[1] M. Arora and P. Roe, *On Post-Shock Oscillations Due to Shock Capturing Schemes in Unsteady Flows*, to appear in J. Comput. Phys., 1996.

[2] D. Gilbarg and D. Paolucci, *The Structure of Shock Waves in the Continuum Theory of Fluids*, J. Rational Mech. Anal., Vol.2, 617 (1953).

[3] S. Jin and J. Liu, *The Effects of Numerical Viscosity I: Slowly Moving Shocks*, J. Comput. Phys., Vol. 126, 373-389 (1996).

[4] S. Karni and S. Canic, *Computations of Slowly Moving Shocks*, submitted to JCP, 1996.

[5] H.C. Lin, *Dissipation Additions to Flux-Difference Splitting*, J. Comput. Phys., Vol.117, 20-27 (1995).

[6] D. Lindstrom, *Numerical Computation of Viscous Detonation Waves in Two Space Dimensions*, Report No. 178/1996, Department of Computing, Uppsala University (1996).

[7] H.M. Mott-Smith, Phys. Rev. Vol.82, 885 (1951).

[8] J. Quirk, *A Contribution to the Great Riemann Solver Debate*, Int. J.for Num. Methods in Fluids, Vol. 18, 555-574 (1994).

[9] T.W. Roberts, *The Behavior of Flux Difference Splitting Near Slowly Moving Shock Waves*, J. Comput. Phys., Vol.90, 141-160 (1990).

[10] P. Woodward and P. Colella, *The numerical simulation of two-dimensional fluid flow with strong shocks*, J. Comput. Phys., Vol.54, 115-173 (1984).

[11] K. Xu, *Numerical Hydrodynamics from Gas-Kinetic Theory*, Ph.D. thesis, Columbia University, 1993.

[12] K. Xu, L. Martinelli,and A. Jameson, *Gas-Kinetic Finite Volume Methods, Flux-Vector Splitting and Artificial Diffusion*, J. Comput. Phys., **120**, 48-65, 1995.