UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Uniformly Accurate Schemes for Hyperbolic Systems with Relaxation

Russel E. Caflisch
Shi Jin
Giovanni Russo

May 1994

CAM Report 94-16

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

Uniformly Accurate Schemes for Hyperbolic Systems with Relaxation

Russel E. Caflisch*
Shi Jin†
Giovanni Russo‡

May 9, 1994

Abstract

We develop high resolution shock capturing numerical schemes for hyperbolic systems with relaxation. In such systems the relaxation time may vary from order one to much less than unity. When the relaxation time is small, the relaxation term becomes very strong and highly stiff. Usually one can not decouple the problem into separate regimes and handle different regimes with different methods. Thus it is important to have a scheme that works uniformly with respect to the relaxation time. Using the Broadwell model of the nonlinear Boltzmann equation we analyze some existing schemes and develop a second order scheme that works effectively, with a fixed spatial and temporal discretization, for all range of mean free path. Formal uniform convergence proof for a first order scheme, and numerical convergence proof for the second order method are also presented. This study is motivated by the reentry problem in hypersonic computations.

*Department of Mathematics, University of California, Los Angeles, CA 90024, USA. Research supported in part by the Army Research Office under grant number DAAL03-91-G-0162.

†School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA. Research was supported in part by AFOSR grant F49620-92-J0098.

‡Dipartimento di Matematica, Università dell'Aquila, 67010 L'Aquila, Italy.

1 Introduction

Hyperbolic systems with relaxation are used to describe many physical problems that involve both convection and nonlinear interaction. In the Boltzmann equation from the kinetic theory of rarefied gas dynamics, the collision (relaxation) terms describe the interaction of particles. In viscoelasticity, memory effects are modeled as relaxation. Relaxation occurs in water waves when the gravitational force balances the frictional force of the riverbed. For gas in thermo-nonequilibrium the internal state variable satisfies a rate equation that measures a departure of the system from the local equilibrium. Relaxations also occur in other problems ranging from magnetohydrodynamics to traffic flow.

In such systems, when the nonlinear interactions are strong, the relaxation rate is large. In kinetic theory, for example, this occurs when the mean free path between collisions is small (i.e. the Knudsen number is small). Within this regime, which is referred to as the *fluid dynamic limit*, the gas flow is well-described by the Euler or Navier-Stokes equations of fluid mechanics, except in shock layers and boundary layers. The characteristic length scale of the kinetic description of the gas is the microscopic, collision distance; in the fluid dynamic limit it is replaced by the macroscopic length scale of fluid dynamics. By analogy with the kinetic theory, we shall refer to the limit of large relaxation rate (or small relaxation time) for a general hyperbolic system with relaxation as the fluid dynamic limit.

The fluid dynamic limit is challenging for numerical methods, because in this regime the relaxation terms become stiff. In particular, a standard numerical scheme might fail to give physically correct solutions once the (microscopic) relaxation distance is smaller than the spatial discretization. Although a full simulation of the relaxation process would require a very fine (and expensive) discretization, it may be possible to accurately compute the solution on a coarser fluid dynamic length scale. The goal of this paper is to present a class of numerical methods using implicit finite difference equations that work with uniform accuracy from the rarefied regime to the fluid dynamic limit for the Broadwell model of kinetic theory.

Numerical methods for hyperbolic systems with relaxation terms have attracted a lot of attention in recent years [7], [15], [16], [20], [21]. Studying the numerical behavior for these problems is important not only for the physical applications, but also for the development of new numerical methods for conservation laws, such as kinetic schemes ([14], [8], [22]) and relaxation schemes [17]. Most kinetic or relaxation schemes can be described as fractional step methods, in which the collision step is just a projection of the system into a sort of discrete "local Maxwellian" or local equilibrium. Although the goal of a kinetic scheme or relaxation scheme is different from ours, nevertheless we use them as guidelines for the study of the properties of a numerical scheme near the fluid dynamic regime.

In earlier works on system of equations with relaxation, the goal was to develop robust numerical schemes that handle effectively the stiffness of the problem. In regions where the relaxation time is no longer small and the problem becomes nonstiff, however, these schemes usually may not have high order accuracy uniformly with respect to the wide range of the relaxation time.

Our motivation differs from these earlier approach in that we seek to develop robust numerical schemes that work uniformly for a wide range of relaxation rate. We consider a simpler model of the Boltzmann equation, and we derive a numerical scheme which is of second order uniformly in the mean free path. This is motivated by hypersonic computations of reentry problems. We develop robust schemes that are able to handle all different regimes, from the rarefied gas to the fluid limit (the stiff regime), with a fixed spatial and temporal grids that are independent of the mean free path. Although we develop our method based on the Broadwell model, this scheme can be applied to a much wider class of hyperbolic systems with relaxation terms and to other discrete velocity kinetic models. In particular, it applies to a class of hyperbolic systems with relaxation characterized by Liu [19].

Probably the paper that is closest in spirit to our work is the one by Coron and Perthame [7]. In that paper the authors derive a numerical scheme for solving the BGK model of the Boltzmann equation under a wide range of mean free path. They discretize velocity space and use a splitting scheme. The collision step is treated by a semi-implicit method that guarantees positivity and entropy condition for the time-discrete model. The scheme is first order accurate in space and time.

Development of numerical methods for the problems considered here is considerably aided by knowledge of the equations for the fluid dynamic limit. In other problems, such as stiff detonation waves and oscillatory fluid flows, the corresponding limit may be less well understood. The goal of "numerical homogenization" [9, 10] for such problems is to formulate numerical methods that automatically give the correct macroscopic or homogenized solution.

2 The Broadwell Model

A simple discrete velocity kinetic model for a gas was introduced by Broadwell [2]. It describes a 2-D gas as composed of particles of only four speeds with a binary collision law and spatial variation in only one direction. The gas defined by a density function in phase space satisfies the equation

$$\begin{aligned} \partial_t f + \partial_x f &= \frac{1}{\varepsilon}(h^2 - fg)\,, \\ \partial_t h &= -\frac{1}{\varepsilon}(h^2 - fg)\,, \end{aligned} \tag{2.1}$$

$$\partial_t g - \partial_x g = \frac{1}{\varepsilon}(h^2 - fg),$$

where ε is the mean free path, f, h and g denote the mass densities of gas particles with speed $1, 0$ and -1 respectively in space x and time t. The fluid dynamic moment variables are density ρ, momentum m and velocity u defined by

$$\rho = f + 2h + g, \quad m = f - g, \quad u = \frac{m}{\rho}. \tag{2.2}$$

In addition, define

$$z = f + g. \tag{2.3}$$

Then the Broadwell equations can be rewritten as

$$\partial_t \rho + \partial_x m = 0, \tag{2.4}$$

$$\partial_t m + \partial_x z = 0, \tag{2.5}$$

$$\partial_t z + \partial_x m = \frac{1}{2\varepsilon}(\rho^2 + m^2 - 2\rho z). \tag{2.6}$$

A *local Maxwellian* is a density function that satisfies

$$Q(f, h, g) = h^2 - fg = \rho^2 + m^2 - 2\rho z = 0, \tag{2.7}$$

i.e.

$$z = z_E(\rho, m) \equiv \frac{1}{2\rho}(\rho^2 + m^2) = \frac{1}{2}(\rho + \rho u^2) \tag{2.8}$$

As $\varepsilon \to 0$ Eq. (2.1) or (2.6) gives the local Maxwellian distribution (2.8). Applying (2.8) in (2.5), one gets the fluid dynamic limit described by the following model Euler equations

$$\begin{aligned} \partial_t \rho + \partial_x(\rho u) &= 0, \\ \partial_t(\rho u) + \partial_x\left(\frac{1}{2}(\rho + \rho u^2)\right) &= 0. \end{aligned} \tag{2.9}$$

To the next order, a model Navier-Stokes equation can be derived via the Chapman-Enskog expansion [5]. For a description of the Broadwell model and its fluid dynamic limit see, for example, references [3] and [13].

The Broadwell equations is a prototypical example for more general hyperbolic systems with relaxations in the sense of Whitham [25] and Liu [19]. These problems can be described mathematically by the system of evolutional equations

$$\partial_t U + \partial_x F(U) = -\frac{1}{\varepsilon} R(U), \qquad U \in \mathcal{R}^N. \tag{2.10}$$

We will call this system the *relaxation system*. Here we use the term relaxation in the sense of Whitham [25] and Liu [19]. The relaxation term is endowed with an $n \times N$ constant matrix Q with rank $n < N$ such that

$$QR(U) = 0, \qquad \text{for all} \quad U. \tag{2.11}$$

This yields n independent conserved quantities $v = QU$. In addition we assume that each such v uniquely determines a local equilibrium value $U = \mathcal{E}(v)$ satisfying $R(\mathcal{E}(v)) = 0$ and such that

$$Q\mathcal{E}(v) = v\,, \qquad \text{for all} \quad v\,. \tag{2.12}$$

The image of $\mathcal{E}$ then constitutes the manifold of local equilibria of R.

Associated with Q are n local conservation laws satisfied by every solution of (2.10) and that take the form

$$\partial_t(QU) + \partial_x(QF(U)) = 0\,. \tag{2.13}$$

These can be closed as a reduced system for $v = QU$ if we take the local relaxation approximation

$$U = \mathcal{E}(v)\,, \tag{2.14}$$

$$\partial_t v + \partial_x e(v) = 0\,, \tag{2.15}$$

where the reduced flux f is defined by

$$e(v) \equiv QF(\mathcal{E}(v))\,. \tag{2.16}$$

A system of conservation laws with relaxation is stiff when ε is small compared to the time scale determined by the characteristic speeds of the system and some appropriate length scales. While we mainly concentrate on the Broadwell equation, the analysis as well as the numerical schemes can certainly be applied to this class of relaxation problems. In fact from time to time we will use the general equation (2.10) to simplify the notation.

3 A Brief Review of Previous Methods

We introduce the spatial grid points $x_{j+\frac{1}{2}}$, $j = \cdots, -1, 0, 1, \cdots$ with uniform mesh spacing $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$ for all j. The time level $t_0, t_1, \cdots$ are also spaced uniformly with space step $\Delta t = t^{n+1} - t^n$ for $n = 0, 1, 2, \cdots$. Here the assumption of a uniform grid is only for simplicity. We use U_j^n to denote the cell average of U in the cell $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ at time t^n,

$$U_j^n = \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} U(t^n, x)\,dx\,. \tag{3.1}$$

The most natural way to solve the Broadwell equations is to split an implicit collision step from an explicit convection step. In the collision step, the backward Euler method can be used to achieve numerical stability independent of ε. In

this step it is very convenient to use the fluid variables. In the convection step, the characteristic method can be applied for the free stream. The overall Courant-Friedrichs-Lewy (CFL) number then will be solely determined by the convection step, and it is 1. This simple splitting scheme is as follows:

$$\tilde{\rho}_j = \rho_j^n\,, \quad \tilde{m}_j = m_j^n\,, \quad \tilde{z}_j = z_j^n + \frac{\Delta t}{2\varepsilon}((\rho_j^n)^2 + (m_j^n)^2 - 2\rho_j^n \tilde{z}_j)\,; \tag{3.2}$$

$$f_j^{n+1} = \tilde{f}_{j-1}^n\,, \quad h_j^{n+1} = \tilde{h}_j^n\,, \quad g_j^{n+1} = \tilde{g}_{j+1}\,. \tag{3.3}$$

Between the collision and convection steps the relation (2.2) and (2.3) will be used. This method is going to have the correct fluid limit when $\varepsilon \to 0$ since the first step always yield the correct local Maxwellian $z = (\rho^2 + m^2)/(2\rho)$. Applying it to the next step, one gets a first order approximation to the model Euler equations. However, the overall accuracy is first order, uniformly in ε (see Section 8).

Previously the numerical solution of the Broadwell equations has been considered by several authors [11, 12, 1]. These authors did not consider numerical schemes that work also in the fluid dynamic regime.

In paper [11] a fractional step method, consisting of a convection and a collision step, was analyzed. The author proved that the numerical solution to the scheme converges to the solution of the Broadwell equations. He also showed that in the limit $\varepsilon \to 0$ the scheme is equivalent to Lax-Friedrichs scheme applied to the model Euler equations, provided the initial condition is a local Maxwellian.

While it is a step in the right direction, the method of [11], suffers from several deficiencies that make it inapplicable in practice. For small ε, the scheme does not provide the proper relaxation to a local Maxwellian unless the small mean free path is well resolved temporally, and therefore the scheme is not able to handle initial layers with underresolved temporal discretizations.

In the papers [12] and [1], numerical solutions of the Broadwell model have been obtained. In these papers the aim is to study the behavior of the Broadwell model and to formulate conjectures about the asymptotic behavior of the solutions.

In [12] the authors considered the 1D Broadwell model that originates from the six-velocity model. Their aim is to study the long time behavior of the 1D Broadwell equations with reflecting boundary conditions. For this purpose they use a space-time discrete Broadwell model, which is basically a consistent finite difference scheme for the Broadwell model, with $\Delta x = \Delta t$. The scheme is the following:

$$\begin{aligned} f_i^{n+1} &= f_{i-1}^n + \Delta t Q_i^n \\ h_i^{n+1} &= h_i^n - \Delta t Q_i^n \\ g_i^{n+1} &= g_{i+1}^n + \Delta t Q_i^n \end{aligned} \tag{3.4}$$

with $Q_i^n \equiv (h_i^n)^2 - f_{i-1}^n g_{i+1}^n$. Note that the discrete collision term is different from $(h_i^n)^2 - f_i^n g_i^n$, corresponding to the naive explicit scheme, which is unstable. For this scheme they proved that, provided the total initial mass is less than ε, the scheme preserves positivity and total mass, and the solution is bounded. Moreover, the solution to the discrete scheme converges to the solution of the true Broadwell model, at least when the solution of the latter has been proved to exist. The numerical results suggested that the solution converged toward a stationary uniform state. [1]

In another paper [1] the authors make use of a splitting scheme for the 2D Broadwell equations, in which the collision step is treated by a fully implicit scheme, and the convection step is treated by exact free flow. Their goal is to study the decay of solutions for boundary value problem with specular boundary conditions. They found numerical evidence that the solution decays to a uniform state.

Numerical methods for more general hyperbolic systems with stiff relaxation terms (2.10) have been considered in [20, 21, 16, 15]. There the goal is to develop underresolved numerical schemes ($\Delta x \gg \varepsilon, \Delta t \gg \varepsilon$) that work in the stiff regime. The zero relaxation limit guides the development of these methods. We will use some of the ideas there to study the behavior in the underresolved regime, however here our goal is to develop a scheme that works with a uniformly second order accuracy in all different regimes.

4 The Convection Steps–Upwind Schemes

We use the method of lines – in which the time discretization and spatial discretization are taken separately – for the Broadwell equations. In this section we shall discuss the spatial discretization, which concerns the linear convection terms. Note that the linear convection in the Broadwell equation is of hyperbolic type. Thus it is natural to use upwind schemes.

Consider the convection part of the Broadwell equations in fluid variables

$$\begin{aligned} \partial_t \rho + \partial_x m &= 0, \\ \partial_t m + \partial_x z &= 0, \\ \partial_t z + \partial_x m &= 0, \end{aligned} \tag{4.1}$$

A conservative spatial discretization of (4.1) is

$$\partial_t \rho_j + \frac{m_{j+\frac{1}{2}} - m_{j-\frac{1}{2}}}{\Delta x} = 0,$$

[1] Recently, Illner [23] proved that the solution to the 1D Broadwell model with specular reflection at the boundary converges to a uniform solution.

$$\begin{aligned} \partial_t m_j + \frac{z_{j+\frac{1}{2}} - z_{j-\frac{1}{2}}}{\Delta x} &= 0\,, \\ \partial_t z_j + \frac{m_{j+\frac{1}{2}} - m_{j-\frac{1}{2}}}{\Delta x} &= 0\,. \end{aligned} \tag{4.2}$$

Eq. (4.1) can be diagonalized into the Riemann invariant form

$$\begin{aligned} \partial_t f + \partial_x f &= 0\,, \\ \partial_t h &= 0\,, \\ \partial_t g - \partial_x g &= 0\,, \end{aligned} \tag{4.3}$$

where f, h and g are exactly the original density functions for the Broadwell equation. The connection between (4.1) and (4.3) are established through the definition of the fluid moment variables (2.2), (2.3). Since f and g travel along the constant characteristics with speeds 1 and -1 respectively, upwind schemes can be easily applied to them.

4.1 The Upwind Scheme

The upwind scheme applied to f and g gives

$$f_{j+\frac{1}{2}} = f_j\,, \quad g_{j+\frac{1}{2}} = g_{j+1}\,, \tag{4.4}$$

while h_j is constant. This implies

$$(z+m)_{j+\frac{1}{2}} = (z+m)_j\,, \quad (z-m)_{j+\frac{1}{2}} = (z-m)_{j+1}\,, \tag{4.5}$$

or equivalently in fluid moment variables

$$\begin{aligned} m_{j+\frac{1}{2}} &= \frac{m_{j+1}+m_j}{2} - \frac{z_{j+1}-z_j}{2} \\ z_{j+\frac{1}{2}} &= \frac{z_{j+1}+z_j}{2} - \frac{m_{j+1}-m_j}{2}. \end{aligned} \tag{4.6}$$

Applying (4.6) in (4.2) gives the semi-discrete upwind scheme for the convection step:

$$\begin{aligned} \partial_t \rho_j + \frac{m_{j+1}-m_{j-1}}{2\Delta x} - \frac{z_{j+1}-2z_j+z_{j-1}}{2\Delta x} &= 0\,, \\ \partial_t m_j + \frac{z_{j+1}-z_{j-1}}{2\Delta x} - \frac{m_{j+1}-2m_j+m_{j-1}}{2\Delta x} &= 0\,, \\ \partial_t z_j + \frac{m_{j+1}-m_{j-1}}{2\Delta x} - \frac{z_{j+1}-2z_j+z_{j-1}}{2\Delta x} &= 0\,. \end{aligned} \tag{4.7}$$

4.2 van Leer's MUSCL Scheme

A second order extension of the upwind scheme is van Leer's MUSCL scheme [24]. While the upwind scheme uses piecewise constant interpolation (4.4), the MUSCL uses piecewise linear interpolation, along with slope limiters to eliminate numerical oscillations at discontinuities. Applying MUSCL to the Riemann Invariants f and g, respectively, we obtain

$$f_{j+\frac{1}{2}} = f_j + \frac{1}{2}\Delta x\, \sigma_j^+\,, \quad g_{j+\frac{1}{2}} = g_{j+1} - \frac{1}{2}\Delta x\, \sigma_{j+1}^-. \tag{4.8}$$

Here σ_j^+ and σ_j^- are the slope of f and g on the jth cell respectively. For $w^+ = f, w^- = g$, $\sigma^\pm$ are defined by [18]

$$\sigma_j^\pm = \frac{1}{\Delta x}(w_{j+1}^\pm - w_{j-1}^\pm)\phi(\theta^\pm)\,, \quad \theta^\pm = \frac{w_j^\pm - w_{j-1}^\pm}{w_{j+1}^\pm - w_j^\pm} \tag{4.9}$$

and the slope limit function $\phi(\theta)$ by van Leer is

$$\phi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|}. \tag{4.10}$$

With this limiter the MUSCL scheme is total-variation-diminishing (TVD). Rewriting Eq. (4.8) in the fluid moment variables gives

$$m_{j+\frac{1}{2}} = \frac{m_{j+1} + m_j}{2} - \frac{z_{j+1} - z_j}{2} + \frac{\Delta x}{4}(\sigma_j^+ + \sigma_{j+1}^-)\,, \tag{4.11}$$

$$z_{j+\frac{1}{2}} = \frac{z_{j+1} + z_j}{2} - \frac{m_{j+1} - m_j}{2} + \frac{\Delta x}{4}(\sigma_j^+ - \sigma_{j+1}^-)\,. \tag{4.12}$$

Applying (4.11), (4.12) in (4.2) finally gives

$$\begin{aligned}
\partial_t \rho_j &+ \frac{m_{j+1} - m_{j-1}}{2\Delta x} - \frac{z_{j+1} - 2z_j + z_{j-1}}{2\Delta x} \\
&+ \frac{\Delta x}{4}(\sigma_j^+ + \sigma_{j+1}^- - \sigma_{j-1}^+ - \sigma_j^-) = 0\,, \\
\partial_t m_j &+ \frac{z_{j+1} - z_{j-1}}{2\Delta x} - \frac{m_{j+1} - 2m_j + m_{j-1}}{2\Delta x} \\
&+ \frac{\Delta x}{4}(\sigma_j^+ - \sigma_{j+1}^- - \sigma_{j-1}^+ + \sigma_j^-) = 0\,, \\
\partial_t z_j &+ \frac{m_{j+1} - m_{j-1}}{2\Delta x} - \frac{z_{j+1} - 2z_j + z_{j-1}}{2\Delta x} \\
&+ \frac{\Delta x}{4}(\sigma_j^+ + \sigma_{j+1}^- - \sigma_{j-1}^+ - \sigma_j^-) = 0.
\end{aligned} \tag{4.13}$$

5 The Collision Steps – A Uniformly Second Order Time Discretization

Since our goal is to seek a robust scheme that works uniformly for all range of ε, it is essential that the time discretization is stable for every ε. This is especially important near the fluid regime where the mean free path is small and the problem becomes stiff.

Uniformly numerical stability can be achieved with implicit temporal integrators. Since stiffness appears only through the relaxation term, it is natural to keep the convection terms explicit, while the collision step has to be treated implicitly.

Numerical experience for such relaxation problems shows that this problem is not merely a numerical stability problem. Stable numerical discretization may still produce spurious solutions [20], [21], [15]. For example, the Crank-Nicholson scheme

$$z_j^{n+1} = z_j^n + \frac{\Delta t}{2\varepsilon}[(\rho_j^n)^2 + (m_j^n)^2 - \rho_j^n(z_j^n + z_j^{n+1})]\,, \tag{5.1}$$

coupled with the free streaming convection with $\Delta x = \Delta t$ gives a scheme that is stable independent of ε, but does not have the correct fluid limit. If the initial data are not local Maxwellians, then the departure of the relaxation process from the local Maxwellian persists for all later time, creating an $O(1)$ numerical error (see Figure 4).

Previous results [15] demonstrate that an effective condition for the correct numerical behavior near the fluid regime is that the numerical scheme should possess the correct fluid limit, in the sense that a discrete analogue of the Chapman-Enskog expansion for the continuous equations remains valid for the numerical discretization, and the resulting numerical fluid limit should be a good discretization for the model Euler limit. A sufficient condition to achieve this is that the collision step always projects the nonequilibrium data into a local Maxwellian at every time step.

We now show how to construct a uniformly second order scheme. The basic idea is to combine a high order convection step with an implicit collision step according to the following guidelines:

i) Truncation error analysis is used to obtain second order accuracy in the rarefied regime ($\varepsilon = O(1)$).

ii) The collision step is well-posed $\forall \varepsilon > 0$.

iii) The limit scheme obtained as $\varepsilon \to 0$ is a good discretization of the model Euler equations when $\varepsilon \ll 1$.

It is possible to show that if condition iii) is not satisfied, then the scheme may give the wrong behavior in the fluid dynamic limit (this is the case, for example, of the scheme in [11]).

A generalization of scheme (3.3) applied to system (2.10) can be written in the form:

$$
\begin{aligned}
U_1 &= U^n - \alpha \frac{\Delta t}{\varepsilon} R(U_1)\,, && (5.2)\\
U_2 &= U_1 - \tilde{\alpha} \Delta t \mathcal{D} F(U_1)\,, && (5.3)\\
U_3 &= U_2 - \beta \frac{\Delta t}{\varepsilon} R(U_3) - \gamma \frac{\Delta t}{\varepsilon} R(U_1)\,, && (5.4)\\
U_4 &= U_3 - \tilde{\beta} \mathcal{D} F(U_3)\,, && (5.5)\\
U_5 &= \xi U^n + \eta U_4\,, && (5.6)\\
U^{n+1} &= U_5 - \mu \frac{\Delta t}{\varepsilon} R(U^{n+1})\,. && (5.7)
\end{aligned}
$$

The scheme is a fractional step combination of convection and collision steps. The parameters $\alpha, \beta, \gamma, \mu, \tilde{\alpha}, \tilde{\beta}, \xi, \eta$ will be determined by conditions i)-iii).

Roughly speaking, this splitting scheme mimics the asymptotic process that leads from the Broadwell equations to the model Euler equations. At $t = t^{(1)}$, the stiff source solver gives $R(U_1) \approx 0$, which is the local Maxwellian. Applying it to the next convection step $t = t^{(2)}$ should give a numerical flux that approximates the flux for the model Euler equation. Similar behavior occurs at $t = t^{(3)}$ and $t^{(4)}$. The last step gives $R(U^{n+1}) \approx 0$, guaranteeing the correct local Maxwellian at $t = t^{n+1}$. The above steps are combined in a second order way, aiming at a scheme with uniform second order accuracy.

In order to satisfy condition i) we apply the scheme to the linear system

$$\partial_t U + AU + BU = 0\,, \tag{5.8}$$

where A and B are constant matrices. The exact solution of (5.8) at time $t = \Delta t$ is given by

$$U(\Delta t) = e^{-(A+B)\Delta t} U(0).$$

Applying scheme (5.2–5.7) to (5.8), and write the difference equation in the compact form

$$U^{n+1} = \mathcal{C} U^n\,. \tag{5.9}$$

To achieve a second order accuracy we impose that

$$(\mathcal{C}(\Delta t) - e^{-(A+B)\Delta t}) U(0) = O(\Delta t^3)\,,$$

then the following restrictions on the parameters should be satisfied:

$$
\begin{aligned}
&\xi + \eta = 1\,,\\
&\eta(\tilde{\alpha} + \tilde{\beta}) = 1\,,
\end{aligned}
$$

$$
\begin{aligned}
&\eta(\alpha+\beta+\gamma)+\mu(\xi+\eta)=1\,,\\
&2\eta\tilde{\alpha}\tilde{\beta}=1\,,\\
&2\eta(\alpha\tilde{\alpha}+\alpha\tilde{\beta}+\tilde{\beta}\gamma+\tilde{\beta}\beta)=1\,,\\
&2\eta(\tilde{\alpha}\beta+\mu\tilde{\alpha}+\mu\tilde{\beta})=1\,,\\
&2\eta(\alpha^2+\alpha\gamma+\alpha\beta+\beta\gamma+\beta^2)+2\mu\eta(\alpha+\beta+\gamma)+2(\xi+\eta)\mu^2=1\,.
\end{aligned}
\tag{5.10}
$$

This is a system of seven algebraic equations with eight unknowns. The system can be explicitly solved by expressing all the parameters as function of μ. The solution is given by:

$$
\alpha=0,\quad \beta=\frac{2\mu-1}{2(\mu-1)},\quad \gamma=-\frac{2\mu^2-2\mu+1}{2\mu(\mu-1)},
$$

$$
\tilde{\alpha}=\frac{1}{2\mu},\quad \tilde{\beta}=-\frac{1}{2(\mu-1)},\quad \xi=2\mu^2-2\mu+1,\quad \eta=-2\mu(\mu-1).
$$

We shall use this freedom to satisfy the second condition ii). When applying the scheme to the Broadwell model, the collision step (5.2) becomes

$$
z_1=\frac{z^n+\alpha(\Delta t/2\varepsilon)(\rho_1^2+m_1^2)}{1+\alpha(\Delta t/\varepsilon)\rho_1}.
$$

Therefore for positive data, both numerator and denominator are positive provided $\alpha\geq 0$. The same property holds for step (5.7), provided $\mu\geq 0$.

Step (5.4) becomes:

$$
z_3=\frac{z_2+\beta(\Delta t/2\varepsilon)(\rho_2^2+m_2^2)+\gamma(\Delta t/2\varepsilon)(\rho_1^2+m_1^2-2\rho_1 z_1)}{1+\beta(\Delta t/\varepsilon)\rho_2}.
$$

The denominator is positive for positive data, provided $\beta\geq 0$. We therefore look for a solution of equations (5.10) that satisfies also the restrictions:

$$
\alpha\geq 0,\quad \beta>0,\quad \mu>0\,. \tag{5.11}
$$

A set of parameters satisfying equations (5.10) and the conditions (5.11) is given by:

$$
\xi=5/9,\ \eta=4/9,\ \alpha=0,\ \mu=1/3,\ \beta=1/4,\ \gamma=5/4,\ \tilde{\alpha}=3/2\ \tilde{\beta}=3/4. \tag{5.12}
$$

A- and L-stability analysis for stiff ODE's can be performed on (5.2)-(5.7) by letting $F=0$ and $R(U)=aU$, with $\Re(a)>0$. Set $q=a\Delta t/\varepsilon$. Applying (5.2)-(5.7) in this case, after combining steps, one has

$$
U^{n+1}=M(q)U^n \tag{5.13}
$$

where the amplification factor is given by:

$$
M=\frac{\xi+\eta+(\xi\beta-\eta\gamma)q}{(1+\mu q)(1+\beta q)}=\frac{1+(\mu-1/(2-2\mu))q}{(1+\mu q)(1+q/(2-2\mu))} \tag{5.14}
$$

A simple analysis shows that for $0 < \mu < 1/2$ the region of complex q plane for which $|M(q)| < 1$ includes the half plane $\Re(q) > 0$. Furthermore, since the amplification factor $\to 0$ as $q \to +\infty$, this scheme, as a stiff ODE solver, is L-stable and any oscillations generated by the transient behavior will rapidly decay.

The construction of the time discretization that combines the convection and the relaxation term is similar in spirit to the time discretizations used in [10] and [15]. Both time discretizations of [10] and [15] have negative parameters in the implicit terms, which may cause numerical breakdown in the intermediate regime $\Delta t = O(\varepsilon)$.

A good feature of the time discretization in [15] is that it has the correct initial layer behavior, since in the first step it projects into the local Maxwellian, and in the fluid limit it becomes the second order TVD Runge-Kutta method for the fluid equation. Our scheme here does not have such a mechanism, because as a result of truncation analysis, $\alpha = 0$ in Eq. (5.2), thus the *de facto* first step (5.4) is not a projection into the local Maxwellian. This will introduce an initial disturbance if the initial layer is not well resolved. Although at later time the scheme has the mechanism to project the data into the local Maxwellian, this initial error will remain at all later times, causing nonconservative numerical solutions, thus degrading the quality of the numerical results. This has been demonstrated in [15]. Hence an extra care must be taken to properly handle the initial layer, which will be discussed in the next section.

6 An Initial Layer Fix

One possibility to overcome this burden is simply to resolve the initial layer, by using a time step $\Delta t \ll \varepsilon$ in the first few steps, and then switch to a larger time step. This procedure is very expensive if the problem also contains nonstiff regions where ε is not small and one does not need to resolve the initial layer.

The initial layer analysis performed in [3] indicates that the initial layer projects the initial data into the local Maxwellian. In order to have the correct behavior, the numerical scheme should have the same projection in the first time step. Note that in the first order splitting scheme (3.3), the first step, due to its fully implicit collision term, always projects the initial data into the local Maxwellian, thus can be used for the first time step along with our new splitting scheme. The first order accuracy of this initial step will be overcome by a Richardson type extrapolation on (3.3).

Let us write the exact solution to (2.10) as

$$U(t) = \mathcal{S}(t)U(0)\,, \tag{6.1}$$

where $S(t)$ is the evolutional operator and $U(0)$ is the initial datum. Suppose

the difference operator is denoted by $\mathcal{T}(t)$. Then the numerical solution $u(t)$ is given by

$$u(t) = \mathcal{T}(t)U(0)\,. \tag{6.2}$$

A first iteration of the splitting scheme (3.3) can be written as

$$u_1(\Delta t) = \mathcal{T}(\Delta t)U(0)\,, \tag{6.3}$$

while a two half step iterations give

$$u_2(\Delta t) = \mathcal{T}(\frac{\Delta t}{2})^2 U(0)\,. \tag{6.4}$$

We define our first time step as

$$u(\Delta t) = 2u_2(\Delta t) - u_1(\Delta t)\,. \tag{6.5}$$

This Richardson extrapolation will give a second order approximation, as will be demonstrated now.

Since the splitting scheme (3.3) is globally first order (locally second order), one has

$$u_1(\Delta t) - U(\Delta t) = (\mathcal{T}(\Delta t) - \mathcal{S}(\Delta t))U(0) = C\Delta t^2 + O(\Delta t^3)\,, \tag{6.6}$$

where C depends on $\mathcal{T}, \mathcal{S}$ and $U(0)$. Of course this implies

$$u_1(\frac{\Delta t}{2}) - U(\frac{\Delta t}{2}) = (\mathcal{T}(\frac{\Delta t}{2}) - \mathcal{S}(\frac{\Delta t}{2}))U(0) = \frac{1}{4}C\Delta t^2 + O(\Delta t^3)\,. \tag{6.7}$$

Also,

$$\mathcal{T}(\Delta t) = I + O(\Delta t) \tag{6.8}$$

where I is the identity operator. Now,

$$\begin{aligned} u(\Delta t) &= 2u_2(\Delta t) - u_1(\Delta t) \\ &= 2\mathcal{T}(\frac{\Delta t}{2})^2 U(0) - \mathcal{T}(\Delta t)U(0) \\ &= 2\mathcal{T}(\frac{\Delta t}{2})[\mathcal{S}(\frac{\Delta t}{2}) + \mathcal{T}(\frac{\Delta t}{2}) - \mathcal{S}(\frac{\Delta t}{2})]U(0) \\ &\quad -[\mathcal{S}(\Delta t) + \mathcal{T}(\Delta t) - \mathcal{S}(\Delta t)]U(0) \end{aligned}$$

Making use of (6.6) one has:

$$\begin{aligned} u(\Delta t) &= 2\mathcal{T}(\frac{\Delta t}{2})U(\frac{\Delta t}{2}) + 2\mathcal{T}(\frac{\Delta t}{2})[\frac{1}{4}C\Delta t^2 + O(\Delta t^3)] \\ &\quad -U(\Delta t) - C\Delta t^2 + O(\Delta t^3) \\ &= 2[\mathcal{S}(\frac{\Delta t}{2}) + \mathcal{T}(\frac{\Delta t}{2}) - \mathcal{S}(\frac{\Delta t}{2})]U(\frac{\Delta t}{2}) - U(\Delta t) - \frac{1}{2}C\Delta t^2 + O(\Delta t^3) \\ &= 2U(\Delta t) + 2[\frac{1}{4}C\Delta t^2 + O(\Delta t)^3] - U(\Delta t) - \frac{1}{2}C\Delta t^2 + O(\Delta t^3) \\ &= U(\Delta t) + O(\Delta t^3)\,. \end{aligned}$$

Thus we have a locally third, or globally second order approximation in this first step. In fact this second order scheme alone can be applied to system (2.10). However such scheme would be slightly less efficient that the one we use (5.2–5.7), since one has to integrate the convection part three times in each time step, while (5.2–5.7) only needs two. This Richardson extrapolation (6.5) may also lose positivity of the numerical solution.

7 The Numerical Fluid Dynamic Limits

To show that the scheme has the correct fluid dynamic limit, we assume that the solution is smooth in the sense that all of the spatial derivatives are of $O(1)$. We also assume that $\varepsilon/\Delta t \ll 1, \varepsilon/\Delta x \ll 1$. The asymptotic analysis will be performed with fixed Δt and Δx, while letting $\varepsilon \to 0$.

7.1 The Spatial Discretizations

We first analyze the limiting behavior of the two spatial discretizations discussed in Section 4.

Applying the upwind scheme (4.4) to the Broadwell equation, one has

$$\begin{aligned}
\partial_t \rho_j &+ \frac{m_{j+1} - m_{j-1}}{2\Delta x} - \frac{z_{j+1} - 2z_j + z_{j-1}}{2\Delta x} = 0\,, \\
\partial_t m_j &+ \frac{z_{j+1} - z_{j-1}}{2\Delta x} - \frac{m_{j+1} - 2m_j + m_{j-1}}{2\Delta x} = 0\,, \\
\partial_t z_j &+ \frac{m_{j+1} - m_{j-1}}{2\Delta x} - \frac{z_{j+1} - 2z_j + z_{j-1}}{2\Delta x} = \frac{1}{2\varepsilon}(\rho_j^2 + m_j^2 - 2\rho_j z_j)\,.
\end{aligned} \tag{7.1}$$

If $\varepsilon \to 0$, then the third equation gives

$$z = \frac{1}{2\rho}(\rho^2 + m^2) + O(\varepsilon)\,. \tag{7.2}$$

Applying this to the first two equations implies (after ignoring the $O(\varepsilon)$ term)

$$\begin{aligned}
\partial_t \rho_j &+ \frac{m_{j+1} - m_{j-1}}{2\Delta x} - \frac{\frac{1}{2}(\rho + \rho u^2)_{j+1} - (\rho + \rho u^2)_j + \frac{1}{2}(\rho + \rho u^2)_{j-1}}{2\Delta x} = 0\,, \\
\partial_t m_j &+ \frac{\frac{1}{2}(\rho + \rho u^2)_{j+1} - \frac{1}{2}(\rho + \rho u^2)_{j-1}}{2\Delta x} - \frac{m_{j+1} - 2m_j + m_{j-1}}{2\Delta x} = 0\,.
\end{aligned} \tag{7.3}$$

This gives the limiting scheme of the upwind scheme. It can be easily seen that this is a first order conservative discretization of the model Euler equation (2.9). Thus the upwind scheme has the correct asymptotic limit.

Applying the MUSCL scheme (12) to the Broadwell equations gives

$$\begin{aligned}
\partial_t \rho_j &+ \frac{m_{j+1}-m_{j-1}}{2\Delta x} - \frac{z_{j+1}-2z_j+z_{j-1}}{2\Delta x} \qquad (7.4)\\
&+ \frac{\Delta x}{4}(\sigma_j^+ + \sigma_{j+1}^- - \sigma_{j-1}^+ - \sigma_j^-) = 0\,,\\
\partial_t m_j &+ \frac{z_{j+1}-z_{j-1}}{2\Delta x} - \frac{m_{j+1}-2m_j+m_{j-1}}{2\Delta x}\\
&+ \frac{\Delta x}{4}(\sigma_j^+ - \sigma_{j+1}^- - \sigma_{j-1}^+ + \sigma_j^-) = 0\,, \qquad (7.5)\\
\partial_t z_j &+ \frac{m_{j+1}-m_{j-1}}{2\Delta x} - \frac{z_{j+1}-2z_j+z_{j-1}}{2\Delta x} \qquad (7.6)\\
&+ \frac{\Delta x}{4}(\sigma_j^+ + \sigma_{j+1}^- - \sigma_{j-1}^+ - \sigma_j^-) = \frac{1}{2\varepsilon}(\rho_j^2 + m^2 - 2\rho z).
\end{aligned}$$

As $\varepsilon \to 0$, the last equation again gives (7.2), which can be applied to (7.5) to yield the limiting scheme of the MUSCL:

$$\begin{aligned}
\partial_t \rho_j &+ \frac{m_{j+1}-m_{j-1}}{2\Delta x} - \frac{\frac{1}{2}(\rho+\rho u^2)_{j+1} - (\rho+\rho u^2)_j + \frac{1}{2}(\rho+\rho u^2)_{j-1}}{2\Delta x}\\
&+ \frac{\Delta x}{4}(\sigma_j^{E+} + \sigma_{j+1}^{E-} - \sigma_{j-1}^{E+} - \sigma_j^{E-}) = 0\,, \qquad (7.7)\\
\partial_t m_j &+ \frac{\frac{1}{2}(\rho+\rho u^2)_{j+1} - \frac{1}{2}(\rho+\rho u^2)_{j-1}}{2\Delta x} - \frac{m_{j+1}-2m_j+m_{j-1}}{2\Delta x}\\
&+ \frac{\Delta x}{4}(\sigma_j^{E+} - \sigma_{j+1}^{E-} - \sigma_{j-1}^{E+} + \sigma_j^{E-}) = 0\,, \qquad (7.8)
\end{aligned}$$

where

$$\sigma_j^{E\pm} = \sigma_j^{\pm}|_{z=\frac{1}{2\rho}(\rho^2+m^2)}\,. \qquad (7.9)$$

Note that $\sigma_j^{\pm}$ is defined by ρ, m and z, and the right hand side of (7.9) simply means applying the local Maxwellian $z = (\rho^2+m^2)/(2\rho)$ into $\sigma_j^{\pm}$. One can see that this is a second order conservative discretization of the model equations (2.9). It is close to the second order relaxed scheme of Jin and Xin [17]. Thus the MUSCL also has the correct fluid limit.

7.2 The Temporal Discretizations

We now demonstrate that the new temporal splitting scheme (5.2)-(5.7) has the correct fluid limit in the sense to be specified below. We always assume that $R(U)$ is Lipschitz continuous in U, and $I + \beta\frac{\Delta t}{\varepsilon}R'$ is invertible for all ε and Δt where R' is the Jacobian matrix of R. This invertibility is true for general hyperbolic systems with relaxations classified by Liu [19]. In particular, it is easy to check that the collision term of the Broadwell equations satisfies this condition.

By the initial layer fix, we can assume that

$$R(U^n) = O(\varepsilon)\,. \tag{7.10}$$

Now, we want to show that this condition will imply that

$$R(U_1) \approx 0\,, \quad R(U_3) \approx 0\,, \quad R(U^{n+1}) \approx 0\,, \tag{7.11}$$

up to some error terms depending on ε and Δt. First, since $\alpha = 0$, thus $U_1 = U^n$, and

$$R(U_1) = R(U^n) = O(\varepsilon) \tag{7.12}$$

Using (5.3),

$$U_2 - U_1 = O(\Delta t) \tag{7.13}$$

By (5.4),

$$\begin{aligned} U_3 - U_2 &= -\beta\frac{\Delta t}{\varepsilon}(R(U_3) - R(U_2)) - \beta\frac{\Delta t}{\varepsilon}R(U_2) + O(\Delta t) \\ &= -\beta\frac{\Delta t}{\varepsilon}R'(U^*)(U_3 - U_2) - \beta\frac{\Delta t}{\varepsilon}(R(U_1) + O(\Delta t)) + O(\Delta t) \\ &= -\beta\frac{\Delta t}{\varepsilon}R'(U^*)(U_3 - U_2) + O(\Delta t + \frac{\Delta t^2}{\varepsilon}) \end{aligned} \tag{7.14}$$

where U^* lies between U_2 and U_3. This implies

$$U_3 - U_2 = (I + \beta\frac{\Delta t}{\varepsilon}R'(U^*))^{-1}O(\Delta t + \frac{\Delta t^2}{\varepsilon}) \tag{7.15}$$

$$= O(\frac{\varepsilon}{\Delta t})O(\Delta t + \frac{\Delta t^2}{\varepsilon}) = O(\varepsilon + \Delta t) = O(\Delta t)\,. \tag{7.16}$$

Applying this back to (5.4) then implies

$$R(U_3) = O(\varepsilon)\,. \tag{7.17}$$

Similar arguments also imply

$$R(U_{n+1}) = O(\varepsilon)\,. \tag{7.18}$$

By this argument, for any initial data, with the initial layer fixing, one always has

$$R(U_3) = O(\varepsilon)\,, \qquad R(U^n) = O(\varepsilon)\,, \quad \textit{for} \quad \textit{all} \quad n \geq 1\,. \tag{7.19}$$

Thus this scheme projects the numerical data into the local Maxwellian at every time step. Moreover, applying these into (5.3), (5.5) respectively, after reordering the indices,

$$U_1 = U^n - \tilde{\alpha}\Delta t\mathcal{D}F(U^n)|_{R(U^n)=0}\,, \tag{7.20}$$

$$U_2 = U_1 - \tilde{\beta}\Delta t\mathcal{D}F(U_1)|_{R(U_1)=0}\,, \tag{7.21}$$

$$U^{n+1} = \xi U^n + \eta U_2\,, \tag{7.22}$$

after ignoring the $O(\varepsilon)$ terms. In (7.20), (7.21) $F(U)|_{R(U)=0}$ are the relaxed flux given in (4.6) for the upwind scheme or the second order relaxed flux (4.8) of the MUSCL. This shows that the new splitting scheme indeed has the correct fluid limit in the coarse regime.

7.3 The Intermediate Regime

The intermediate regime is defined as $\varepsilon \ll 1, \Delta t \ll 1$ but $\varepsilon/\Delta t = O(1)$. We now show that the splitting scheme (5.2)-(5.7) also has the correct fluid limit in this intermediate regime. The key is to demonstrate that (7.11) are valid also in this regime.

First, since $\alpha = 0$, (7.12) is valid. Thus (5.4) implies

$$\begin{aligned} U_3 - U_2 &= -\beta\frac{\Delta t}{\varepsilon}R(U_3) + O(\varepsilon) \\ &= -\beta\frac{\Delta t}{\varepsilon}(R(U_3) - R(U_2)) - \beta\frac{\Delta t}{\varepsilon}R(U_2) + O(\varepsilon) \\ &= -\beta\frac{\Delta t}{\varepsilon}R'(U^*)(U_3 - U_2) - \beta\frac{\Delta t}{\varepsilon}(R(U_1) + O(\Delta t)) + O(\varepsilon) \\ &= -\beta\frac{\Delta t}{\varepsilon}R'(U^*)(U_3 - U_2) + O(\Delta t) \end{aligned}$$

Therefore

$$U_3 - U_2 = (1 + \beta\frac{\Delta t}{\varepsilon}R'(U^*))^{-1}O(\Delta t) = O(\Delta t)\,.$$

Applying this to (5.4) gives

$$R(U_3) = O(\Delta t)O(\frac{\varepsilon}{\Delta t}) = O(\varepsilon)\,.$$

Similarly one can also show

$$R(U^{n+1}) = O(\varepsilon)\,.$$

Thus, one always has

$$R(U_3) = O(\varepsilon)\,, \qquad R(U^n) = O(\varepsilon)\,, \quad \text{for all } n \geq 1\,.$$

independent of the initial data. So the solution is always a local Maxwellian. Moreover, as $\varepsilon \to 0$ ($\Delta t \to 0$ as well) one gets (4.7) or (4.8) depending on the order of the numerical fluxes. Thus in the intermediate regime this scheme also has the correct fluid limit.

8 Theoretical Results

In this section we derive several properties of the simple fractional step method (3.3), including positivity, the entropy inequality and convergence in certain regimes.

8.1 Positivity

The velocity densities (f, g, h) are positive if and only if

$$\rho > z > |m|\,,$$

by (2.2) and (2.3). In the collision step $\tilde{\rho} = \rho^n$ and $\tilde{m} = m^n$ are unchanged, while

$$\tilde{z} = (1 + 2\kappa\rho^n)^{-1}(z^n + \kappa((m^n)^2 + (\rho^n)^2))$$

in which $\kappa = \Delta t/(2\varepsilon)$. Since $\tilde{\rho} = \rho^n > z^n > |m^n| = |\tilde{m}|$, it follows easily that

$$\tilde{\rho} > \tilde{z} > |\tilde{m}|.$$

In fact, it is

$$\tilde{z} - |\tilde{m}| = \tilde{z} - |m| = \frac{z - |m| + \kappa(\rho - |m|)^2}{1 + 2\kappa\rho} > 0$$

and

$$\tilde{\rho} - \tilde{z} = \rho - \tilde{z} = \frac{\rho - z + \kappa(\rho^2 - m^2)}{1 + 2\kappa\rho} > 0.$$

So $(\tilde{f}, \tilde{g}, \tilde{h})$ is positive. Positivity of $(\tilde{f}, \tilde{g}, \tilde{h})$ clearly implies positivity of f^{n+1}, g^{n+1} and h^{n+1}, the result of the convective step in (3.3).

8.2 Entropy Inequality

We prove next that the H function, given by

$$H = f \log f + 2h \log h + g \log g,$$

decreases in the collision step at each grid point j. This result is valid both for periodic boundary conditions and for unbounded domain.

Let (ρ, m, z) be the state after the collision step, and $(\tilde{\rho}, \tilde{m}, \tilde{z}) = (\rho, m, z+2\Delta)$ be the state before the collision step, in which

$$\Delta = 2\kappa(h^2 - fg).$$

Since $f = (z+m)/2$, $g = (z-m)/2$ and $h = (\rho - z)/2$, then

$$\tilde{f} = f - \Delta\,, \quad \tilde{g} = g - \Delta\,, \quad \tilde{h} = h + \Delta\,. \tag{8.1}$$

Moreover since (f, g, h) and $(\tilde{f}, \tilde{g}, \tilde{h})$ are all positive, then

$$1 > \max\left(\frac{\Delta}{f}, \frac{\Delta}{g}, -\frac{\Delta}{h}\right).$$

Before the collision the H function is

$$\tilde{H} = \tilde{f}\log\tilde{f} + 2\tilde{h}\log\tilde{h} + \tilde{g}\log\tilde{g}.$$

First rewrite the $\tilde{f}$ term in H as

$$\begin{aligned}
\tilde{f}\log\tilde{f} &= (f-\Delta)\log(f-\Delta) \\
&= f\left(1-\frac{\Delta}{f}\right)\log\left(1-\frac{\Delta}{f}\right) + (f-\Delta)\log f \\
&\geq -\Delta + (f-\Delta)\log f.
\end{aligned}$$

The last bound uses the elementary inequality

$$(1-a)\log(1-a) \geq -a$$

for $1 > a$, in which $a = \Delta/f$. Similarly

$$\tilde{g}\log\tilde{g} \geq -\Delta + (g-\Delta)\log g\,, \quad \tilde{h}\log\tilde{h} \geq \Delta + (h+\Delta)\log h\,. \tag{8.2}$$

Add these together to obtain

$$\begin{aligned}
\tilde{H} &\geq (f-\Delta)\log f + 2(h+\Delta)\log h + (g-\Delta)\log g \\
&= H + \Delta\log(h^2/fg) \\
&= H + 2\kappa(h^2 - fg)\log(h^2/fg) \\
&\geq H.
\end{aligned}$$

Finally note that the total "entropy"

$$\mathcal{H}^n = \Delta x \sum_j H_j^n$$

is preserved by the convection step of the fractional step method (3.3), since the values of f, g and h do not change but only move between j points.

The results of these two subsections are summarized by the following proposition.

Proposition 8.1. *Let (f_j^n, g_j^n, h_j^n) be the solution of the fractional step method (3.3). Assume that $f_j^0 > 0$, $g_j^0 > 0$ and $h_j^0 > 0$ for all j. It follows that:*

(i) $f_j^n > 0, g_j^n > 0, h_j^n > 0$ *for all* $n > 0$ *and all* j.
(ii) Define

$$\mathcal{H}^n = \sum_j (f_j^n \log f_j^n + 2h_j^n \log h_j^n + g_j^n \log_j^n)$$

and suppose that $\mathcal{H}^0 < \infty$. *Then for all* $n \geq 0$

$$\mathcal{H}^{n+1} \leq \mathcal{H}^n.$$

8.3 Formal Analysis of Convergence: Uniform Bounds on Consistency Error

Finally we present a formal analysis of the splitting method which shows that it is first order accurate uniformly in ε, if the underlying Broadwell solution is smooth. This analysis does not apply for the solution in a boundary layer, initial layer or shock, in which numerical results indicate that a half order of accuracy may be lost. The analysis here is only of the consistency error, which is shown to be of size $O(\Delta t)$ uniformly in ε. Analysis of the stability error has not yet been successful.

This estimate will be performed by writing the Broadwell equations and the discretized Broadwell equations for the fluid dynamic variables ρ and m, and for the difference from equilibrium w, defined by

$$w = z - \frac{\rho^2 + m^2}{2\rho} \tag{8.3}$$

Define also

$$z_E(\rho, m) = \frac{\rho^2 + m^2}{2\rho}. \tag{8.4}$$

Whereas the equation for z involves a forcing term of size ε^{-1}, in the w equation the factor ε^{-1} appears only in a decay term.

The Broadwell equations (2.1) can be written in terms of ρ, m and w as

$$\rho_t = -m_x \tag{8.5}$$
$$m_t = -(w + z_E(\rho, m))_x \tag{8.6}$$
$$w_t = -\varepsilon^{-1}\rho w - m_x - z_E(\rho, m)_t \tag{8.7}$$

The equation for w can be written in integral form, for each x, as

$$w(t) = A(t)w^0 + \int_0^t \frac{A(t)}{A(s)} g(s) ds \tag{8.8}$$

in which

$$\begin{aligned} A(t) &= \exp(-\varepsilon^{-1}\int_0^t \rho(t',x)dt') \\ g(t) &= -(m_x + z_E(\rho,m)_t). \end{aligned}$$

The discrete Broadwell equations (with $\Delta t = \Delta x$) can be written in terms of the discretized variables $\tilde{\rho}, \tilde{m}$ and $\tilde{w}$, with the convection and collision steps combined, for each spatial value j, as

$$\tilde{\rho}^{n+1} - \tilde{\rho}^n = -\frac{1}{2}D_0\tilde{m}^n + \frac{1}{2}D^2(\tilde{w}^n + \tilde{z}_E^n) \tag{8.9}$$

$$\tilde{m}^{n+1} - \tilde{m}^n = -\frac{1}{2}D_0(\tilde{w}^n + \tilde{z}_E^n) + \frac{1}{2}D^2\tilde{m}^n \tag{8.10}$$

$$\tilde{w}^{n+1} = (1 + \frac{\Delta t}{\varepsilon}\tilde{\rho}^{n+1})^{-1}(\tilde{w}^n + \Delta t\tilde{g}^{n+1}) \tag{8.11}$$

in which

$$\tilde{g}^{n+1} = -\Delta t^{-1}(\tilde{z}_E^{n+1} - \tilde{z}_E^n + \frac{1}{2}D_0\tilde{m}^{n+1} - \frac{1}{2}D^2(\tilde{w}^{n+1} + \tilde{z}_E^{n+1})).$$

The difference operators D_0 and D^2 are defined as

$$(D_0 f)_j = f_{j+1} - f_{j-1}\,, \quad (D^2 f)_j = f_{j+1} - 2f_j + f_{j-1}\,. \tag{8.12}$$

The equation for $\tilde{w}^{n+1}$ can be considered a linear difference equation in $\tilde{w}^n$. The solution can be written as

$$\tilde{w}^n = \tilde{B}^n w^0 + \Delta t\sum_{k=1}^{n}\frac{\tilde{B}^n}{\tilde{B}^{k-1}}\tilde{g}^k \tag{8.13}$$

in which

$$\tilde{B}^n = \prod_{j=1}^{n}(1 + \frac{\Delta t}{\varepsilon}\tilde{\rho}^j)^{-1}\,, \quad \tilde{B}^0 = 1\,. \tag{8.14}$$

Assume that ρ, m and w are smooth and bounded uniformly in ε (this excludes shocks, boundary layers and initial layers).

The consistency error is defined is the error formed by substituting the continuous solution $(\rho^n, m^n, w^n) = (\rho(n\Delta t), m(n\Delta t), w(n\Delta t))$ into the discrete equations. Define

$$E_1 = (\rho^{n+1} - \rho^n)/\Delta t - \left\{-\frac{1}{2}D_0 m^n + \frac{1}{2}D^2(w^n + z_E^n)\right\}/\Delta t \tag{8.15}$$

$$E_2 = (m^{n+1} - m^n)/\Delta t - \left\{-\frac{1}{2}D_0(w^n + z_E^n) + \frac{1}{2}D^2 m^n\right\}/\Delta t \tag{8.16}$$

$$E_3 = w^n - \left\{B^n w^0 + \Delta t\sum_{k=1}^{n}\frac{B^n}{B^{k-1}}g^k\right\} \tag{8.17}$$

in which B^n and g^n are defined as above in terms of the continuous solution (ρ^n, m^n, w^n). Note that the consistency error for the w equations is defined here in terms of the "integrated solution" rather than the finite difference equation. In fact the error due to the implicit collision operator is better behaved over many time steps than over a single step. This is also the reason that the factor $1/\Delta t$ does not appear in the definition of E_3.

The consistency error serves as a forcing term in a finite difference equation for the total error. Define the difference between the numerical solution $(\tilde{\rho}^n, \tilde{m}^n, \tilde{w}^n)$ and the continuous solution $(\rho^n, m^n, w^n) = (\rho, m, w)(n\Delta t)$ as

$$a^n = \tilde{\rho}^n - \rho^n\,, \quad b^n = \tilde{m}^n - m^n\,, \quad c^n = \tilde{w}^n - w^n\,. \tag{8.18}$$

Denote $A(\rho^n, m^n, w^n)$ and $B(\rho^n, m^n, w^n)$ to be the bracketed convection terms in the ρ and m equations (8.15) and (8.16). Also denote $C[\rho, m, w]$ to be the bracketed term in the w equation (8.17). Then the error quantities (a^n, b^n, c^n) satisfy

$$\begin{aligned} a^{n+1} - a^n &= (A(\rho^n + a^n, m^n + b^n, w^n + c^n) - A(\rho^n, m^n, w^n)) + \Delta t E_1 \\ b^{n+1} - b^n &= (B(\rho^n + a^n, m^n + b^n, w^n + c^n) - B(\rho^n, m^n, w^n)) + \Delta t E_2 \\ c^{n+1} - c^n &= (C[\rho + a, m + b, w + c] - C[\rho, m, w]) + E_3 \end{aligned} \tag{8.19}$$

The first term on the right hand sides of these equations is the stability error, and is approximately a linear operator in (a, b, c). If the difference scheme is uniformly stable, which we have not proved, then the error will be of the size of the consistency error E_1, E_2, E_3.

Under the assumption that the continuous solution ρ, m, w is smooth and that the density ρ is uniformly bounded above and below, i.e.

$$\bar{\rho} < \rho < \bar{c}\bar{\rho} \tag{8.20}$$

we prove the following uniform bound on the consistency error.

Proposition 8.2. *Suppose that ρ, m, w is smooth and that the density ρ satisfies (8.20) for some constants $\bar{\rho}$ and $\bar{c}$. Then the consistency error E_1, E_2, E_3, defined by (8.15), (8.16) and (8.17), for the discretized Broadwell equations (8.9), (8.10) and (8.13) with $\Delta t = \Delta x$ satisfies*

$$|E_1| + |E_2| + |E_3| \leq c\Delta t \tag{8.21}$$

for some constant c that is independent of ε. In other words, the consistency error is uniformly first order.

The estimates on E_1 and E_2 are straightforward; i.e. for some constant c,

$$|E_1| + |E_2| \leq c\Delta t. \tag{8.22}$$

In order to compare the integral (8.8) to the sum (8.17), we first replace the integrand by a piecewise-constant function. Define

$$\bar{A}(s) = A((m-1)\Delta t)\,, \quad \bar{g}(s) = g((m-1)\Delta t)\,. \tag{8.23}$$

for $(m-1)\Delta t \le s < m\Delta t$. For $t = n\Delta t$ define

$$\begin{aligned} \bar{w}(t) &= A(t)w^0 + \int_0^t \frac{A(t)}{\bar{A}(s)}\bar{g}(s)ds \\ &= A(t)w^0 + \Delta t \sum_{k=1}^{n} \frac{A(n\Delta t)}{A((k-1)\Delta t)} g(k\Delta t) \end{aligned} \tag{8.24}$$

Now estimate, for $(m-1)\Delta t \le s < m\Delta t$, $m \le n$,

$$|g(s) - \bar{g}(s)| < c\Delta t \tag{8.25}$$

$$\begin{aligned} \left|\frac{A(t)}{A(s)} - \frac{A(t)}{\bar{A}(s)}\right| &= \left|\frac{A(t)}{A(s)}\right| \cdot \left|1 - \frac{A(s)}{\bar{A}(s)}\right| \\ &\le \exp\left(-\int_s^t \rho(s')ds'/\varepsilon\right) \\ &\quad \left|1 - \exp\left(-\int_{(m-1)\Delta t}^{s} \rho(s')ds'/\varepsilon\right)\right| \\ &\le \min(1, \bar{c}\Delta t\bar{\rho}/\varepsilon)\exp(-(n-m)\Delta t\bar{\rho}/\varepsilon). \end{aligned} \tag{8.26}$$

since $|e^{-\alpha} - 1| \le \min(1, \alpha)$ for any α.

Next estimate the difference between $A(t)$ and B^n. First make a simple general estimate (for $t = n\Delta t$)

$$A(t) \le e^{-t\bar{\rho}/\varepsilon} \le (1 + \frac{\Delta t}{\varepsilon}\bar{\rho})^{-n}, \tag{8.27}$$

$$B^n \le \left(1 + \frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-n}, \tag{8.28}$$

so that

$$|A(t) - B^n| \le 2\left(1 + \frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-n}. \tag{8.29}$$

If $\Delta t\bar{\rho}/\varepsilon$ is small, a more refined estimate is needed. In this case

$$\log\left(1 + \frac{\Delta t}{\varepsilon}\rho^k\right) = \frac{\Delta t}{\varepsilon}\rho^k + O\left(\frac{\Delta t}{\varepsilon}\bar{\rho}\right)^2 \tag{8.30}$$

Then

$$\begin{aligned} A(t) - B^n &= A(t)\left\{1 - \exp\left(\int_0^t \rho(s)ds/\varepsilon - \sum_{k=1}^{n} \log\left(1 + \frac{\Delta t}{\varepsilon}\rho^k\right)\right)\right\} \\ &= A(t)\left\{1 - \exp\left(nO\left(\frac{\Delta t}{\varepsilon}\bar{\rho}\right)^2\right)\right\} \end{aligned} \tag{8.31}$$

in which we ignore a term of size $n\,O(\Delta t^2\bar{\rho}/\varepsilon)$ in the exponential.

Now $n(\Delta t\bar{\rho}/\varepsilon)^2 = (t\bar{\rho}/\varepsilon)(\Delta t\bar{\rho}/\varepsilon)$. If this is larger than 2, we use the estimate (8.29); while if it is less than 2, then

$$|A(t) - B^n| \leq A(t) O\left(\frac{t\bar{\rho}}{\varepsilon}\left(\frac{\Delta t}{\varepsilon}\bar{\rho}\right)\right). \tag{8.32}$$

Combining these estimates, we obtain the bound

$$|A(t) - B^n| \leq \min\left(2, \frac{t\Delta t}{\varepsilon^2}\bar{\rho}^2\right)\left(1 + \frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-n}. \tag{8.33}$$

Finally the estimate on $A(t)/A((k-1)\Delta t) - B^n/B^{k-1}$ is found in a similar way to be

$$\begin{aligned}\left|\frac{A(t)}{A((k-1)\Delta t)} - \frac{B^n}{B^{k-1}}\right| \;\leq\; & \min\left(2, \varepsilon^{-2}(n-k+1)\Delta t^2\bar{\rho}^2\right) \\ & \left(1 + \frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-(n-k)}.\end{aligned} \tag{8.34}$$

Now these basic estimates are combined to estimate E_3. First

$$\begin{aligned} w(t) - \bar{w}(t) &= \int_0^t \left(\frac{A(t)}{A(s)}g(s) - \frac{A(t)}{\bar{A}(s)}\bar{g}(s)\right) ds \\ &\leq \int_0^t \left(\left|\frac{A(t)}{A(s)} - \frac{A(t)}{\bar{A}(s)}\right| |\bar{g}(s)| + \left|\frac{A(t)}{A(s)}\right| |g(s) - \bar{g}(s)|\right) ds \\ &\leq c\Delta t \sum_{m=1}^{n} \min(1, \tilde{c}\Delta t\bar{\rho}\varepsilon^{-1}) \exp(-(n-m)\Delta t\bar{\rho}/\varepsilon) \\ &\quad + c\Delta t \int_0^t e^{-\bar{\rho}(t-s)/\varepsilon} ds \\ &\leq c\Delta t \min(1, \Delta t\bar{\rho}\varepsilon^{-1})(1 - e^{-\Delta t\bar{\rho}/\varepsilon})^{-1} + c\varepsilon\Delta t/\bar{\rho} \\ &\leq \tilde{c}\Delta t \end{aligned} \tag{8.35}$$

since

$$\max_{y>0}\left(\frac{\min(1,y)}{1-\exp(-y)}\right) = \frac{e}{e-1}.$$

Next, making use of (8.33),

$$\begin{aligned} |E_3 + \bar{w}(t) - w^n| &= \Big|(A(t) - B^n)w^0 \\ &+ \Delta t \sum_{k=1}^{n}\left(\frac{A(t)}{A((k-1)\Delta t)}\bar{g}(k\Delta t) - \frac{B^n}{B^{k-1}}g^k\right)\Big| \end{aligned}$$

$$
\begin{aligned}
&\leq |w^0| \min\left(2, \frac{t\Delta t}{\varepsilon^2}\bar{\rho}^2\right)\left(1+\frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-n} \\
&+ c\Delta t \sum_{k=1}^{n}\left(1+\frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-(n-k+1)}\Delta t \\
&+ c\Delta t \sum_{k=1}^{n}\min\left(1, \frac{\Delta t^2\bar{\rho}^2}{\varepsilon^2}(n-k+1)\right)\left(1+\frac{\bar{\rho}\Delta t}{\varepsilon}\right)^{-n+k-1} \qquad (8.36)
\end{aligned}
$$

Denote the three terms on the right side of (8.36) as D_1, D_2, D_3 respectively.

The first term D_1 is bounded by

$$
c\min\left(2, \frac{n\Delta t}{\varepsilon}\bar{\rho}\frac{\Delta t\bar{\rho}}{\varepsilon}\right)\left(1+\frac{\Delta t}{\varepsilon}\bar{\rho}\right)^{-n} \leq 2c\Delta t\,. \qquad (8.37)
$$

In fact, if $\Delta t\bar{\rho}/\varepsilon > 1/2$, then

$$
D_1 < 2c(3/2)^{-n} < 2c\Delta t \qquad (8.38)
$$

while if $\Delta t\bar{\rho}/\varepsilon < 1/2$ then

$$
D_1 < c(1/n)(n\Delta t\bar{\rho}/\varepsilon)^2(1+\Delta t\bar{\rho}/\varepsilon)^{-n} < c/n < c\Delta t\,, \qquad (8.39)
$$

since $x^2(1+x/n)^{-n}$ is bounded by 1 uniformly in $n > 2$ and $x \geq 0$.

Now we consider the second term D_2. Let $q = (1+\bar{\rho}\Delta t/\varepsilon)^{-1}$. Then

$$
\begin{aligned}
D_2 &= c\Delta t^2 \sum_{k=1}^{\infty} q^{n-k+1} = c\Delta t^2 q\frac{1-q^n}{1-q} \\
&\leq c\Delta t^2\frac{q}{1-q} = \frac{c\Delta t\varepsilon}{\bar{\rho}} \qquad (8.40)
\end{aligned}
$$

Finally consider the third term D_3.

If $\Delta t\bar{\rho}/\varepsilon > 2$, then

$$
\begin{aligned}
D_3 &\leq c\Delta t\sum_{k=1}^{n} q^{n-k+1} \\
&= c\Delta t q\frac{1-q^n}{1-q} \leq c\Delta t\frac{q}{1-q} \\
&= c\Delta t\frac{\varepsilon}{\bar{\rho}\Delta t} \leq \frac{c\Delta t}{2}
\end{aligned}
$$

If $\Delta t\bar{\rho}/\varepsilon < 2$, then

$$
\begin{aligned}
D_3 &\leq c\Delta t\left(\frac{\Delta t\bar{\rho}}{\varepsilon}\right)^2\sum_{0}^{\infty} mq^m \\
&= c\Delta t\left(\frac{\Delta t\bar{\rho}}{\varepsilon}\right)^2\frac{q}{(1-q)^2} \\
&= c\Delta t(1+\Delta t\bar{\rho}/\varepsilon) \leq 3c\Delta t \qquad (8.41)
\end{aligned}
$$

since

$$\sum_{0}^{\infty} m x^{m-1} = (1-x)^{-2}.$$

This shows that

$$|E_3 + \bar{w}(t) - w^n| \leq c\Delta t. \tag{8.42}$$

Combine this with (8.35) to obtain

$$|E_3| \leq c\Delta t. \tag{8.43}$$

Together with (8.22), this implies that

$$|E_1| + |E_2| + |E_3| \leq c\Delta t. \tag{8.44}$$

This concludes the formal demonstration that consistency error in the discrete method is first order accurate, uniformly in ε.

9 Numerical Results

In this section we present some numerical examples in order to compare some existing schemes and the new splitting scheme developed in this article. Temporal discretizations to be compared with are the first order splitting scheme (3.3), the new splitting scheme (5.2)-(5.7) and a second order Strang splitting combined with the Crank-Nicholson source term (5.1). Spatial discretizations that we take are the upwind scheme (4.6) or the MUSCL scheme (4.12). Note that when the CFL number = 1 the upwind scheme becomes the characteristic method which is exact for the stream operator. These two spatial discretizations combined with the three different temporal discretizations give six different numerical schemes, which will be abbreviated as SP1 (the basic splitting scheme (3.3)), SCN (the Strang splitting combined with the Crank-Nicholson source term (5.1)), SP1vL (the splitting scheme (3.3) with the stream step replaced by the MUSCL scheme (4.12), SCNvL (5.1) with the stream step replaced by the MUSCL scheme), NSP (the new splitting scheme (5.2)-(5.7) combined with the MUSCL scheme without the initial layer fix), and NSPIF (the new splitting scheme (5.2)-(5.7) combined with the MUSCL scheme with the initial layer fix) respectively.

First we solve the Broadwell equation with the following initial data

$$\rho = 2\,, \quad m = 1\,, \quad z = 1\,, \qquad \text{for} \qquad x < 0\,, \tag{9.1}$$

$$\rho = 1\,, \quad m = 0.13962\,, \quad z = 1\,, \qquad \text{for} \qquad x > 0\,, \tag{9.2}$$

We integrate over domain $[-1, 1]$ with reflecting boundary conditions. We take $\Delta x = 0.01$ and $\Delta t = O(\Delta x)$ (for precise Δt see the Figure captions). We

test six different schemes. The exact solution is obtained using fine grids with $\Delta x = 0.0005$.

First we take $\varepsilon = 1$. This is in the rarefied regime. The numerical solutions of ρ, m and z are depicted with the "exact" solution in Figure 1. In this regime, SP1 yields the best resolution, SP1vL is very diffusive, SP1vL and SCNvL is slightly oscillatory near the discontinuities. NSP and NSPIF are comparable and are slightly more diffusive than SP1.

Next we take $\varepsilon = 10^{-8}$ and compare the behavior of NSP and NSPIF. This is the regime when the mean free path is very small and the limiting Euler equation has a shock wave moving right with a speed $s = 0.86038$ determined by the Rankine-Hugoniot jump condition. The initial data for z is not a local Maxwellian, which yields an initial layer. The results are displayed in Figure 2. The NSP and NSPIF are comparable with respect to the shock, however, without the initial layer fix the NSP create a bump near the initial discontinuity $x = 0$, solely caused by the kinetic effect.

Finally we integrate over $[-1, 1]$ with $\varepsilon = 0.02$. We also take $\Delta x = 0.02$. This is in the intermediate regime where $\varepsilon, \Delta x$ and Δt are of the same order. The results are depicted in Figure 3. It seems that NSP and NSPIF give the best results, especially in the viscous shock region.

We then choose another initial data

$$\rho = 1\,, \quad m = 0\,, \quad z = 1\,, \qquad \text{for} \qquad x < 0.5\,, \tag{9.3}$$

$$\rho = 0.2\,, \quad m = 0\,, \quad z = 1\,, \qquad \text{for} \qquad x > 0.5\,, \tag{9.4}$$

We integrate over domain $[0, 1]$ with reflecting boundary conditions. We take $\Delta x = 0.01$ and $\varepsilon = 10^{-8}$ so the solution is close to that of the model Euler equation. By solving the model Euler equation one obtains a left moving rarefaction wave and a right moving shock wave. The initial data are not in the local Maxwellian. The numerical solutions are plotted in Figure 4. In this fluid limit we observe that the SP1 and SCN are both very diffusive, the SP1vL gives sharp resolution but is oscillatory. With the Crank-Nicholson source, the solution does not yield the correct local Maxwellian for z. Both NSP and NSPIF give good numerical resolution of shock and rarefaction waves and are monotone.

Next we perform the numerical convergence study. We consider an initial value problem with periodic boundary conditions, such that the solution is smooth in a time interval $[0, T]$ for any value of the parameter ε. We compute the error at time T by differencing, i.e. by comparing the result obtained with a given grid $(\Delta x, \Delta t)$ with the one obtained with the grid $(\Delta x/2, \Delta t/2)$.

In Sec. 8 we proved that the consistency error for SP1 is uniformly first order. By truncation analysis we know that the scheme NSP and NSPIF are second order both in space and time, if $\Delta t \ll \varepsilon$, and they are second order also in the fluid regime (for smooth solutions).

The goal of the test is to perform a numerical study of the convergence rate for a wide range of ε, and check whether the convergence is uniform in ε also

in the intermediate regime. The test problem is given by equations (2.1) with periodic boundary conditions: $s(x+L,t) = s(x,t)$ with $s = f, g, h$. The initial data is given by:

$$\rho(x,0) = 1 + a_\rho \sin\frac{2\pi x}{L}, \quad u(x,0) = \frac{1}{2} + a_u \sin\frac{2\pi x}{L},$$

$$m(x,0) = \rho(x,0), m(x,0), \quad z(x,0) = z_E(\rho(x,0), m(x,0)) * \theta_M,$$

where θ_M is a real parameter. If $\theta_M = 1$ then the initial condition is a local Maxwellian, otherwise it is not. If $\theta_M \neq 1$, $\varepsilon \ll 1$, there is an initial layer. The system is integrated for $t \in [0,T]$. The values of the parameters used in the computations are:

$$L = 20,\ T = 30,\ a_\rho = 0.3,\ a_u = 0.1$$

The values of Δx used in the computations are:

$$\Delta x = 0.4, 0.2, 0.1, 0.05, 0.025$$

for the first order scheme SP1 and

$$\Delta x = 1, 0.5, 0.25, 0.125, 0.0625$$

for the second order schemes. The time step is chosen in such a way that CFL condition is satisfied: $\Delta t = \Delta x$ for scheme SP1 and $\Delta t = \Delta x/2$ for the second order schemes. The convergence rate is computed from the error according to the formula:

$$\text{convergence rate}_i = \frac{\log(\text{error}_i/\text{error}_{i+1})}{\log(\Delta x_i/\Delta x_{i+1})}$$

where error_i is obtained by comparing the solution obtained with Δx_i to that obtained with Δx_{i+1}. The errors and convergence rate are computed and plotted as function of ε. For each value of ε, five runs have been done for five different values of Δx, resulting in four error curves and three curves of convergence rate.

Several measures of the error have been used, namely, L_1, L_2, and L_∞ relative norm of the error. The different norms give essentially the same results, therefore we shall show only the L_1 norm.

First we consider the simple splitting scheme SP1. In Figure 5 we show the relative discrete norm of the error in ρ and in m as function of ε (left column) and the corresponding convergence rate (right column).

The initial state is a local Maxwellian in the first two cases and it is not in the last. The convergence rate increases when the mesh becomes finer, and seems to confirm that the scheme is first order, uniformly in ε, for a fine enough mesh.

Next we consider second order schemes. In Figure 6a-b the result of scheme NSP is shown. The initial condition is a local Maxwellian.

As it is evident from the figures, the scheme is second order accurate for small and large values of ε, and there is a slight deterioration of the accuracy in the intermediate regime.

Figure 6c shows the effect of the initial layer if scheme NSP is used, without using Richardson extrapolation for the first step. The accuracy of the scheme of course degrades due to the initial layer. In Figure 6d the convergence rate is shown.

The problem of the initial layer can be overcome by using Richardson extrapolation for the first step (scheme NSPIF, Figure 6e-f). A similar result is obtained by using a scheme entirely based on Richardson extrapolation, in which every step is of the form (6.5).

Acknowledgment

S. Jin and G. Russo would like to thank the Mathematics Department at UCLA for its hospitality during their visits there. S. Jin is grateful for Prof. George Papanicolaou for his constant encouragement and support for this work.

References

[1] A.V.Bobylev, E.Gabetta, L.Pareschi, "On a boundary value problem for the plane Broadwell model. Exact solutions and numerical simulation", to appear on Mathematical Models & Methods in Applied Sciences.

[2] J. E. Broadwell, "Shock structure in a simple discrete velocity gas", Phys. Fluids, vol.7, 1013–1037, 1964.

[3] R. E. Caflisch and G. Papanicolaou, "The fluid-dynamical limit of a nonlinear model Boltzmann equation", Comm. Pure. Appl. Math. 22, 1979, pp. 586-616.

[4] C. Cercignani, "The Boltzmann Equation and its Applications" Springer-Verlag, New York (1988).

[5] S. Chapman and T. G. Cowling, "The Mathematical Theory of Nonuniform Gases", 3rd Ed., Cambridge Univ. Press, Cambridge (1970).

[6] G. Q. Chen, C. D. Levermore and T. P. Liu, "Hyperbolic conservation laws with stiff relaxation terms and entropy", Comm. Pure. Appl. Math., 1993, to appear.

[7] F. Coron and B. Perthame, "Numerical passage from numerical to fluid equations", SIAM J. Numer. Anal. Vol. 28, N.1, pp.26-42.

[8] S. Deshpande, "A second order accurate, kinetic-theory based method for inviscid compressible flow", NASA Langley Tech. paper No. 2613, 1986.

[9] B. Engquist, "Computation of oscillatory solutions to hyperbolic differential equations", Lecture Notes in Mathematics, vol.1270, 10-22, Springer-Verlag, 1987.

[10] B. Engquist and B. Sjogreen, "Robust Difference Approximations of Stiff Inviscid Detonation Waves", SIAM J. Sci. Stat. Comp., to appear.

[11] H. Emamirad, "Fractional step method for Broadwell model", C. R. Acad. Sc. Paris, t.304, Séries II, 487-490, 1987.

[12] Gabetta, E.; Pareschi, L. "Approximating the Broadwell model in a strip." Mathematical Models & Methods in Applied Sciences, March 1992, vol.2, (no.1):1-19.

[13] R. Gatignol, "Théorie cinétique des gas à répartition discrète de vitesses", Lecture Notes in Physics, vol. 36, Springer-Verlag, Heidelberg (1975).

[14] A. Harten, P. D. Lax and B. van Leer, "On upwind differencing and Godunov-type schemes for hyperbolic conservation laws", SIAM Rev. 25, 1983, pp. 35–61.

[15] S. Jin, "Implicit numerical schemes for hyperbolic systems with stiff relaxation terms", J. Comp. Phys., submitted.

[16] S. Jin and C. D. Levermore, "Numerical schemes for hyperbolic systems with stiff relaxation terms", J. Comp. Phys., submitted.

[17] S. Jin and Z. P. Xin, "The relaxation schemes for systems of conservation laws in arbitrary space dimensions", Comm. Pure. Appl. Math., to appear.

[18] R.J. LeVeque, "Numerical Methods for Conservation Laws", Birkhauser Verlag, Basel (1992).

[19] T. P. Liu, "Hyperbolic conservation laws with relaxation", Comm. Math. Phys. 108, 1987, pp. 153–175.

[20] R. B. Pember, "Numerical methods for hyperbolic conservation laws with stiff relaxation, I. Spurious solutions," *SIAM J. Appl. Math.*, 535, 1293–1330 (1993).

[21] R. B. Pember, "Numerical methods for hyperbolic conservation laws with stiff relaxation, II. High Order Godunov methods", SIAM J. Sci. Stat. Comp., 1993, to appear.

[22] B. Perthame, "Second order Boltzmann schemes for compressible Euler equations in one and two space dimensions" *SIAM J. Numer. Anal.*, 291, 1–19 (1992).

[23] T. Platkowski and R. Illner, "Discrete velocity models of the Boltzmann equations: a survey on the mathematical aspects on the theory", SIAM Review, vol.30, 1988, 213–255.

[24] B. val Leer, "Towards the ultimate conservative difference scheme V. A second order sequel to Godunov's method", J. Comp. Phys. 32, 1979, pp. 101–136.

[25] G.B. Whitham, "Linear and Nonlinear Waves", Wiley, New York (1974).

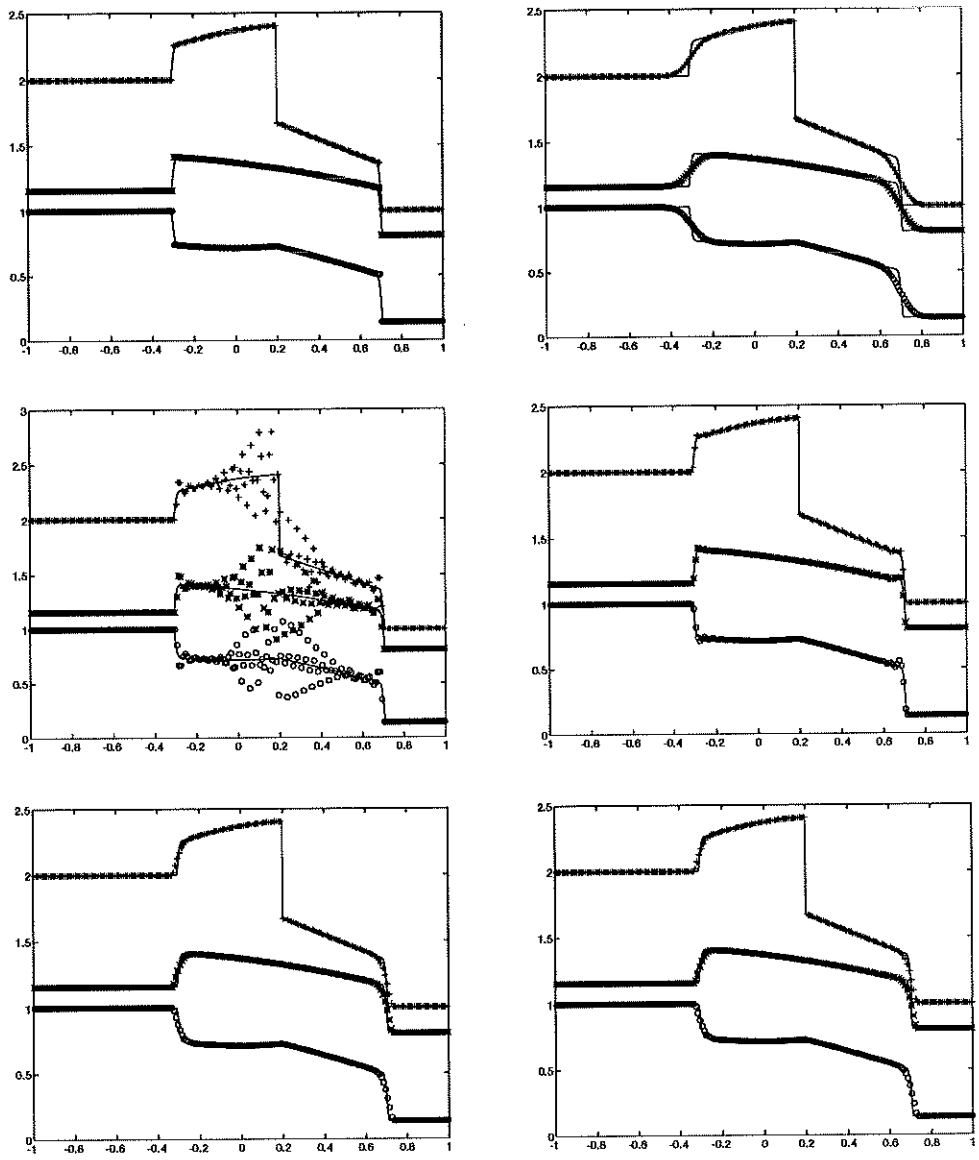

Figure 1. The numerical solutions of ρ ('+'), m ('o') and z ('*') at $t = 0.5$ in $x \in [-1, 1]$ for initial data (9.2) by (from left to right, then top to bottom) SP1, SCN, SP1vL, SCNvL, NSP and NSPIF. $\varepsilon = 1$, $\Delta x = 0.01$. $CFL = 1$ for SP1 and SCN, $CFL = 0.5$ for SCNvL, NSP and NSPIF, and $CFL = 0.25$ for SP1vL.

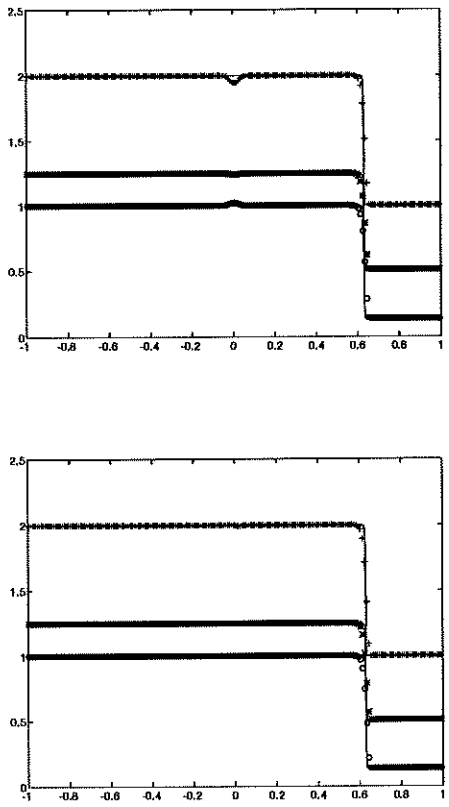

Figure 2. The numerical solutions of ρ ('+'), m ('o') and z ('*') at $t = 0.5$ in $x \in [-1, 1]$ for initial data (9.2) by NSP and NSPIF. $\varepsilon = 10^{-8}$, $\Delta x = 0.01$ and $CFL = 0.5$.

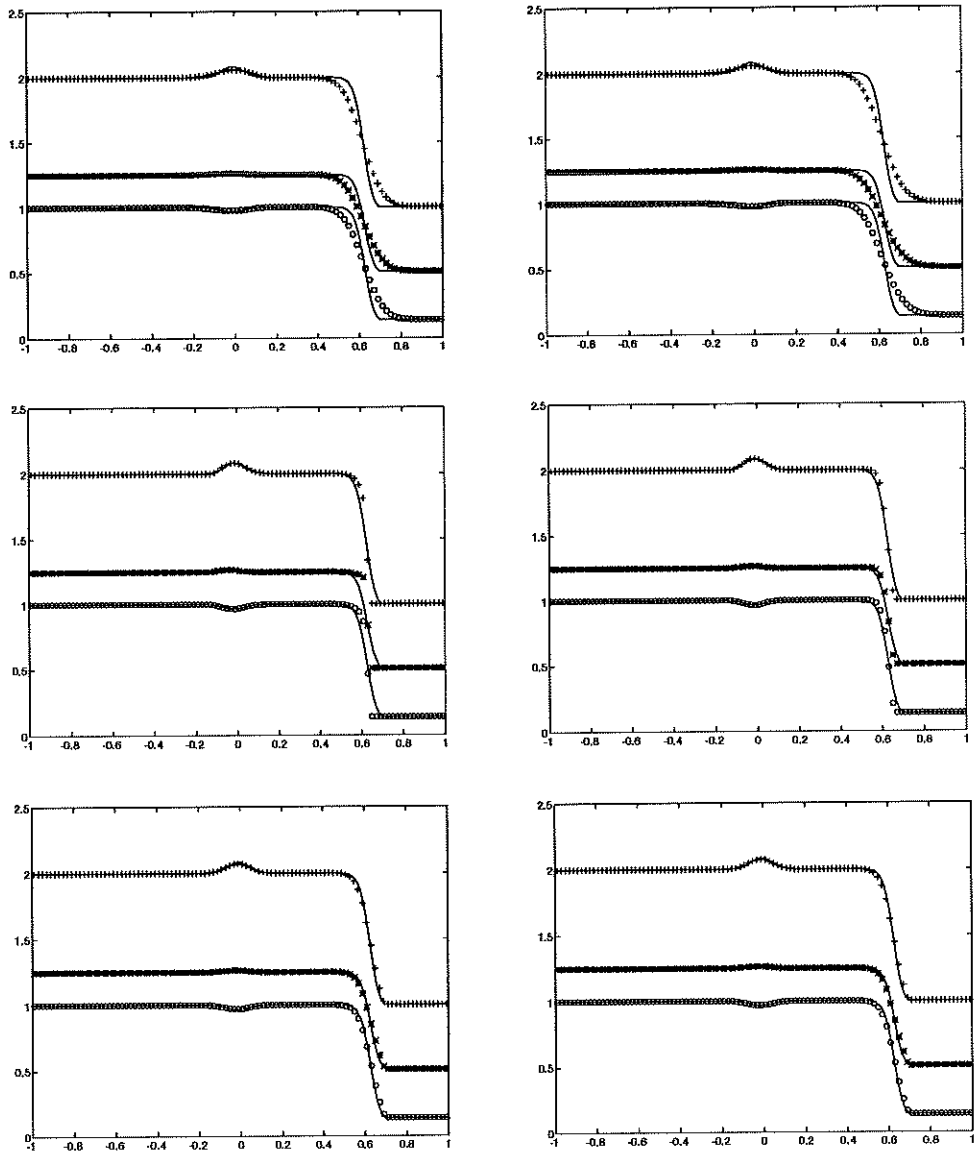

Figure 3. The numerical solutions of ρ ('+'), m ('o') and z ('*') at $t = 0.5$ in $x \in [-1, 1]$ for initial data (9.2) by (from left to right, then top to bottom) SP1, SCN, SP1vL, SCNvL, NSP and NSPIF. $\varepsilon = 0.02$, $\Delta x = 0.02$. $CFL = 1$ for SP1 and SCN , $CFL = 0.5$ for others.

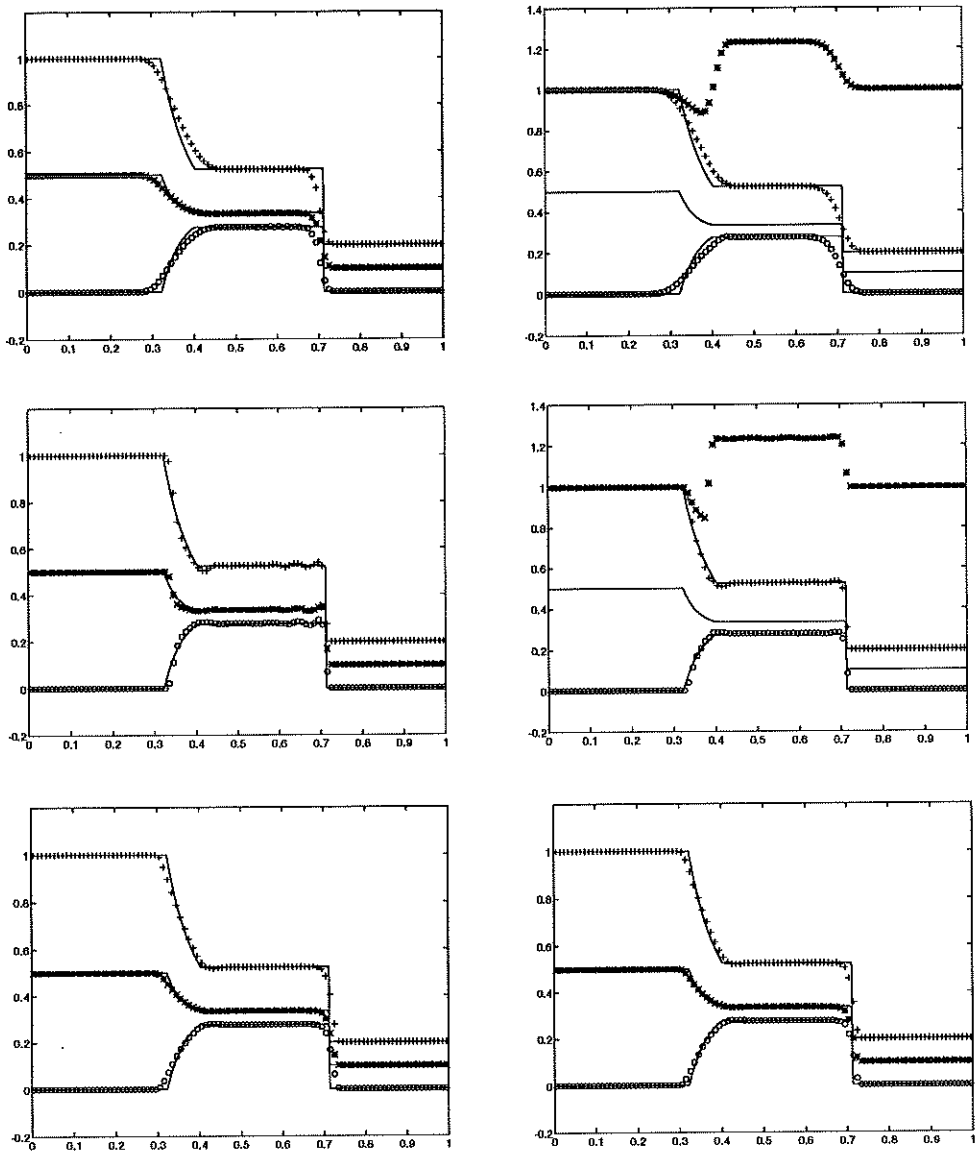

Figure 4. The numerical solutions of ρ ('+'), m ('o') and z ('*') at $t = 0.5$ in $x \in [0, 1]$ for initial data (9.4) by (from left to right, then top to bottom) SP1, SCN, SP1vL, SCNvL, NSP and NSPIF. $\varepsilon = 10^{-8}$, $\Delta x = 0.01$. $CFL = 1$ for SP1 and SCN , $CFL = 0.5$ for others.

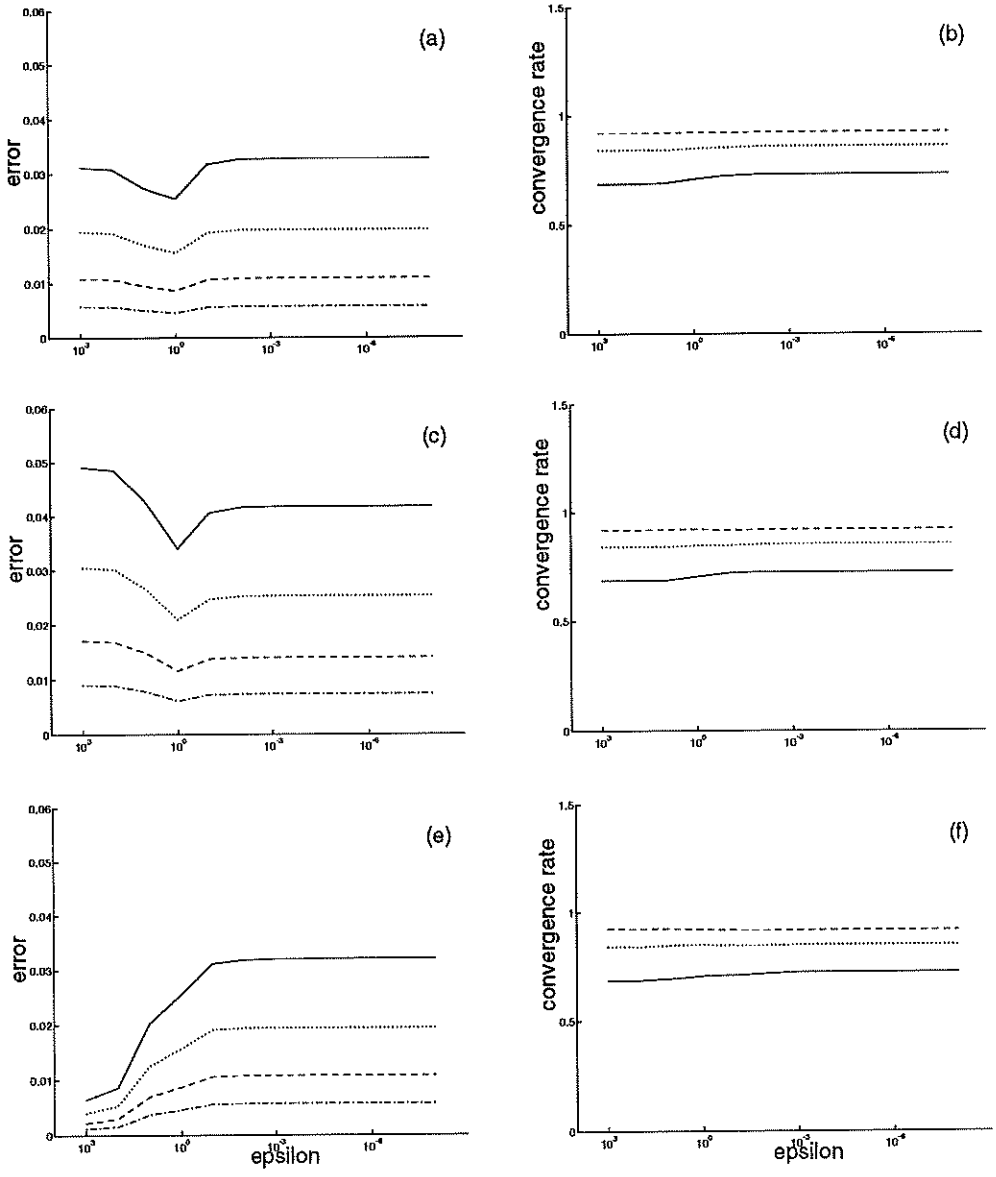

Figure 5. Uniform convergence of the simple splitting scheme SP1. Relative error (left) and convergence rate (right) vs ε for various values of the grid step size Δx. Continuous line: coarsest mesh; dashed line: finest mesh. (a-b): L_1 error in ρ. Initial condition: local Maxwellian. (c-d): L_1 error in m. Initial condition: local Maxwellian. (e-f): L_1 error in ρ. Initial condition: not local Maxwellian.

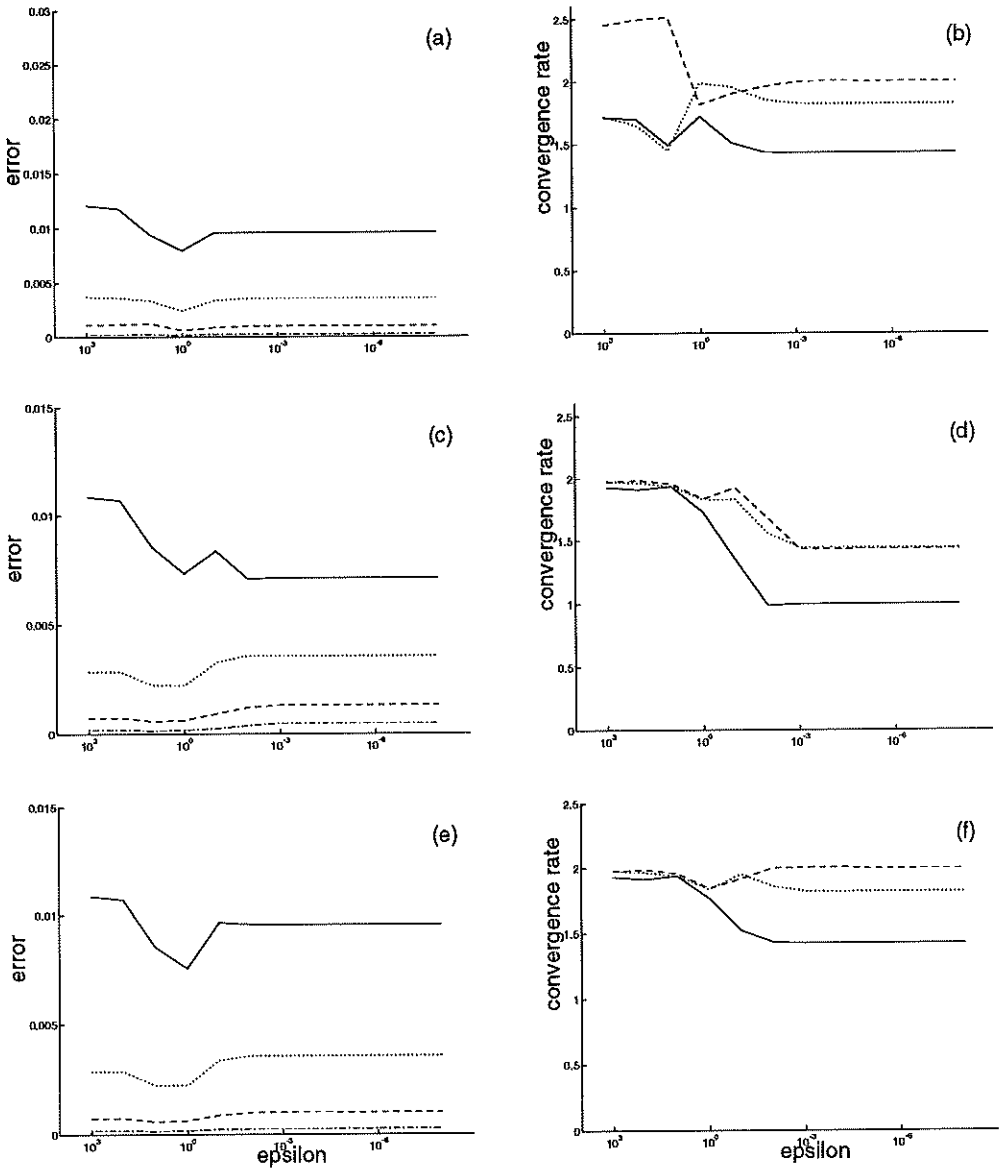

Figure 6. Relative L^1 norm of the error in density and convergence rate vs ε for various values of the grid step size Δx. (a) and (b): Scheme NSP. Initial state: local Maxwellian. (c) and (d): Scheme NSP. The initial state is not a local Maxwellian. (e) and (f): Scheme NSPIF. Same initial state of (c-d).