

# Hamiltonian-preserving schemes for the Liouville equation of geometrical optics with discontinuous local wave speeds <sup>\*</sup>

Shi Jin <sup>†</sup> and Xin Wen <sup>‡</sup>

## Abstract

In this paper we construct two classes of Hamiltonian-preserving numerical schemes for a Liouville equation with discontinuous local wave speed. This equation arises in phase space description of the geometrical optics, and has been the foundation of the recently developed level set methods for multi-valued solution in geometrical optics. We extend our previous work in [22] for the semiclassical limit of the Schrödinger equation into this system. The designing principle of the Hamiltonian preservation by building in the particle behavior at the interface into the numerical flux is used here, and as a consequence we obtain two classes of schemes that allow a hyperbolic stability condition. When a plane wave hits an interface, the Hamiltonian preservation is equivalent to Snell's law of refraction in the case when the ratio of wave length over the width of the interface goes to zero, when both length scales go to zero. Positivity, and stabilities in both  $l^1$  and  $l^\infty$  norms, are established for both schemes. The approach also provides a selection criterion for a unique weak solution of the underlying linear hyperbolic equations

---

<sup>\*</sup>Research supported in part by NSF grant No. DMS-0305080, Institute for Mathematics and its Applications (IMA) under a New Direction Visiting Professorship, NSFC under the Project 10228101 and the Basic Research Projects of Tsinghua University under the Project JC2002010.

<sup>†</sup>Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA; and Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China. Email address: jin@math.wisc.edu.

<sup>‡</sup>Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P.R. China. Email address: wenx@mail.tsinghua.edu.cn.

with singular coefficients. Numerical experiments are carried out to study the numerical accuracy.

## 1 Introduction

In this paper, we construct and study numerical schemes for the Liouville equation in  $d$ -dimension:

$$f_t + H_{\mathbf{v}} \cdot \nabla_{\mathbf{x}} f - H_{\mathbf{x}} \cdot \nabla_{\mathbf{v}} f = 0, \quad t > 0, \quad \mathbf{x}, \mathbf{v} \in R^d, \quad (1.1)$$

where the Hamiltonian  $H$  possesses the form

$$H(\mathbf{x}, \mathbf{v}, t) = c(\mathbf{x})|\mathbf{v}| = c(\mathbf{x})\sqrt{v_1^2 + v_2^2 + \cdots + v_d^2} \quad (1.2)$$

with  $c(\mathbf{x})$  being the local wave speed.  $f(t, \mathbf{x}, \mathbf{v})$  is the density distribution of particles depending on position  $\mathbf{x}$ , time  $t$  and the slowness vector  $\mathbf{v}$ . In this paper we are interested in the case when  $c(\mathbf{x})$  contains *discontinuities* due to different media. This discontinuity will generate an *interface* at the point of discontinuity of  $c(\mathbf{x})$ , and as a consequence waves crossing this interface will undergo transmission and reflection. The incident and transmitted waves obey Snell's Law of refraction.

The bicharacteristics of this Liouville equation (1.1) satisfies the Hamiltonian system:

$$\frac{d\mathbf{x}}{dt} = c(\mathbf{x})\frac{\mathbf{v}}{|\mathbf{v}|}, \quad \frac{d\mathbf{v}}{dt} = -c_{\mathbf{x}}|\mathbf{v}|. \quad (1.3)$$

In classical mechanics the Hamiltonian (1.2) of a particle remains a constant along particle trajectory, even across an interface.

This Liouville equation arises in phase space description of the geometrical optics. It is the high frequency limit of the wave equation

$$u_{tt} - c(\mathbf{x})^2 \Delta u = 0, \quad t > 0, \quad \mathbf{x} \in R^d. \quad (1.4)$$

Recently several phase space based level set methods are based on this equation, see [14, 20, 26]. Semiclassical limit of wave equations with transmissions and reflections at the interfaces were studied in [1, 25, 33].

The Liouville equation (1.1) is a linear wave equation, with the characteristic speed determined by bicharacteristic (1.3). If  $c(\mathbf{x})$  is smooth, then the standard numerical methods (for example, the upwind scheme and its higher order extensions) for linear wave equations give satisfactory results. However, if  $c(\mathbf{x})$  is discontinuous, the conventional numerical schemes suffer from two problems. Firstly, the characteristic speed  $c_{\mathbf{x}}$  of the Liouville equation is *infinity* at the discontinuous point of wave speed. When numerically approximating  $c_{\mathbf{x}}$  crossing the interface, the numerical derivative of  $c$  is of  $O(1/\Delta x)$ , with  $\Delta x$  the mesh size in the physical space. Thus an explicit scheme needs time step  $\Delta t = O(\Delta x \Delta v)$  with  $\Delta v$  the mesh size in particle

velocity space. This is very expensive. Moreover, a conventional numerical scheme in general does not preserve a *constant Hamiltonian* across the interface, usually leads to poor or incorrect numerical resolutions by ignoring the discontinuities of  $c(\mathbf{x})$ . Theoretically, there is a uniqueness issue for weak solutions to these linear hyperbolic equations with singular wave speed [6, 8, 17, 29, 30]. It is not clear which weak solution a standard numerical discretization that ignores the discontinuity of  $c(\mathbf{x})$  will select.

In this paper, we construct a class of numerical schemes that are suitable for the Liouville equation (1.1) with a discontinuous local wave speed  $c(\mathbf{x})$ . An important feature of our schemes is that they are consistent with the constant Hamiltonian across the interface. This gives a selection criterion for a unique weak solution to the governing equation. As done in [22] for the Liouville equation for the semiclassical limit of the linear Schrödinger equation, we call such schemes **Hamiltonian-preserving schemes**. A key idea of these schemes is to build the behavior of a particle at the interface—either cross over with a changed velocity or be reflected with a negative velocity (or momentum)—into the numerical flux. This idea was formerly used by Perthame and Semioni in their work [28] to construct a well-balanced kinetic scheme for the shallow water equations with a (discontinuous) bottom topography which can capture the steady state solutions—corresponding to a constant energy—of the shallow water equations when the water velocity is zero. As a consequence, these new schemes allow a typical hyperbolic stability condition  $\Delta t = O(\Delta x, \Delta v)$ .

We extend both classes of the Hamiltonian-preserving schemes developed in [22] here. One (called *Scheme I*) is based on a finite difference approach, and involves interpolation in the velocity space. The second (called *Scheme II*) uses a finite volume approach, and numerical quadrature rule in the velocity space is needed. These new schemes allow a typical hyperbolic stability condition  $\Delta t = O(\Delta x, \Delta v)$ . We will also establish the positivity and stability theory for both schemes. It is proved that Scheme I is positive,  $l^\infty$  contracting, and  $l^1$  stable under a hyperbolic stability condition, while Scheme II is positive,  $l^\infty$  stable and  $l^1$  contracting under the same stability condition.

By building in the wave behavior at the interface, we have also provided a selection principle to pick up the weak solution to this linear hyperbolic equation with singular coefficients. For a plane wave hitting a interface, it selects the solution that describes the interface condition in geometrical optics governed by **Snell's Law of refraction** when the wave length is much shorter than the width of the interface while both lengths go to zero.

In geometrical optics applications, one has to solve the Liouville equation like (1.1) with *measure-valued* initial data

$$f(\mathbf{x}, \mathbf{v}, 0) = \rho_0(\mathbf{x})\delta(\mathbf{v} - \mathbf{u}_0(\mathbf{x})), \quad (1.5)$$

see for example [32, 11, 20]. The solution at later time remains measure-valued (with finite or even infinite number of concentrations—corresponding to *multivalued*

solutions in the physical space). Computation of multivalued solutions in geometrical optics and more generally in nonlinear PDEs has been a very active area of research, see [2, 3, 5, 4, 7, 13, 9, 10, 12, 15, 16, 14, 21, 26, 31, 35].

Numerical methods for the Liouville equation with measure-valued initial data (1.5) could easily suffer from poor resolution due to the numerical approximation of the initial data as well as numerical dissipation. The level set method proposed in [19, 20] decomposes  $f$  into  $\phi$  and  $\psi_i (i = 1, \dots, d)$  where  $\phi$  and  $\psi_i$  solve the same Liouville equation with initial data

$$\phi(\mathbf{x}, \mathbf{v}, 0) = \rho_0(\mathbf{x}), \quad \psi_i(\mathbf{x}, \mathbf{v}, 0) = v_i - u_{i0}(\mathbf{x}), \quad (1.6)$$

respectively. (We remark here that the common zeroes of  $\psi_i$  give the multivalued velocity, see [21, 7, 19, 20]). This allows the numerical computations for bounded rather than measure-valued solution of the Liouville equation, which greatly enhances the numerical resolution (see [20]). The moments can be recovered through

$$\rho(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{v}, t) d\mathbf{v} = \int \phi(\mathbf{x}, \mathbf{v}, t) \prod_{i=1}^d \delta(\psi_i) d\mathbf{v}, \quad (1.7)$$

$$\mathbf{u}(\mathbf{x}, t) = \frac{1}{\rho(\mathbf{x}, t)} \int f(\mathbf{x}, \mathbf{v}, t) \mathbf{v} d\mathbf{v} = \int \phi(\mathbf{x}, \mathbf{v}, t) \mathbf{v} \prod_{i=1}^d \delta(\psi_i) d\mathbf{v} / \rho(\mathbf{x}, t). \quad (1.8)$$

Thus one only involves numerically the delta-function at the output time!

Numerical computations of multivalued solution for smooth  $c(\mathbf{x})$  using this technique were given in [20]. In this paper we will also give numerical examples using this technique with a discontinuous  $c(\mathbf{x})$ .

The more general case where the transmission and reflection co-exist will be studied in a forthcoming paper. There again, the Hamiltonian preservation provides an equivalent weak solution obeying Snell's law for a plane wave.

This paper is organized as follows. In Sections 2, we first show that the usual finite difference scheme to solve the Liouville equation with a discontinuous wave speed suffers from the severe stability constraint. We then present the designing principle of our Hamiltonian-preserving scheme by describing the behavior of waves at an interface. We present Scheme I in 1d in Section 3 and study its positivity and stability in both  $l^\infty$  and  $l^1$  norms. Scheme II in 1d is presented and studied in Section 4. We extend these schemes to higher dimension in Section 5 in the simple case of interface aligning with the grids and a plane wave. Numerical examples are given in Section 6 to verify the accuracy of the schemes constructed in this paper. We make some concluding remarks in Section 7.

In the sequel, when we describe our scheme in the 1D case, we employ an uniform mesh with grid points at  $x_{i+\frac{1}{2}}, i = 0, \dots, N$ , in the  $x$ -direction and  $\xi_{j+\frac{1}{2}}, j = 0, \dots, M$  in the  $\xi$ -direction. The cells are centered at  $(x_i, \xi_j), i = 1, \dots, N, j = 1, \dots, M$  with  $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$  and  $\xi_j = \frac{1}{2}(\xi_{j+\frac{1}{2}} + \xi_{j-\frac{1}{2}})$ . The mesh size is denoted by  $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}, \Delta \xi = \xi_{j+\frac{1}{2}} - \xi_{j-\frac{1}{2}}$ . We also assume a uniform time step  $\Delta t$  and the discrete time is given by  $0 = t_0 < t_1 < \dots < t_L = T$ . We introduce mesh ratios

$\lambda_x^t = \frac{\Delta t}{\Delta x}$ ,  $\lambda_\xi^t = \frac{\Delta t}{\Delta \xi}$ , assumed to be fixed. We define the cell average of  $f$  as

$$f_{ij} = \frac{1}{\Delta x \Delta \xi} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} f(x, \xi, t) d\xi dx.$$

Throughout the paper we use  $C$  for a generic positive constant independent of the mesh size and time step.

## 2 The designing principle of the Hamiltonian-preserving scheme

### 2.1 Deficiency of the usual finite difference schemes

Consider the numerical solution of the 1D Liouville equation

$$f_t + c(x)\text{sign}(\xi)f_x - c_x|\xi|f_\xi = 0 \quad (2.1)$$

with a discontinuous wave speed  $c(x)$ .

A typical semi-discrete finite difference method for this equation is

$$\partial_t f_{ij} + c_i \text{sign}(\xi_j) \frac{f_{i+\frac{1}{2},j} - f_{i-\frac{1}{2},j}}{\Delta x} - Dc_i |\xi_j| \frac{f_{i,j+\frac{1}{2}} - f_{i,j-\frac{1}{2}}}{\Delta \xi} = 0, \quad (2.2)$$

where the numerical fluxes  $f_{i+\frac{1}{2},j}$ ,  $f_{i,j+\frac{1}{2}}$  are defined in the upwind manner, and  $Dc_i$  is some numerical approximation of  $c_x$  at  $x = x_i$ .

Such a discretization suffers from two problems:

- If an explicit time discretization is used, the CFL condition for this scheme requires the time step to satisfy

$$\Delta t \max_i \left[ \frac{c_i}{\Delta x} + \frac{|Dc_i| \max_j |\xi_j|}{\Delta \xi} \right] \leq 1. \quad (2.3)$$

Since the wave speed  $c(x)$  is discontinuous at some points,  $\max_i |Dc_i| = O(1/\Delta x)$ , so the CFL condition (2.3) requires  $\Delta t = O(\Delta x \Delta \xi)$ .

- The above discretization in general does not preserve a constant Hamiltonian  $H = c|\xi|$  across the discontinuities of  $c$ , thus may not produce numerical solutions consistent to, for example, Snell's Law of refraction.

## 2.2 Behavior of waves at the interface

When a wave moves with its density distribution governed by the Liouville equation (1.1), *The Hamiltonian  $H = c|\mathbf{v}|$  should be preserved across the interface:*

$$c^+|\mathbf{v}^+| = c^-|\mathbf{v}^-| \quad (2.4)$$

where the superscripts  $\pm$  indicates the right and left limits of the quantity at the interface.

We will discuss the wave behavior in 1D and 2D respectively.

- The 1D case

The 1D case is simpler. Consider the case when, at an interface, the characteristic on the left of the interface is given by  $\xi^- > 0$ . Then the particle definitely crosses the interface and with  $\xi^+ = \frac{c^-}{c^+}\xi^-$ .

- The 2D case, when a plane incident wave hits an interface that aligns with the grid.

In the 2D case,  $\mathbf{x} = (x, y)$ ,  $\mathbf{v} = (\xi, \eta)$ . Consider the case that the interface is a line parallel to the  $y$ -axis, and  $c = c(x)$ . The incident wave has velocity  $(\xi^-, \eta^-)$  to the left side of the interface, with  $\xi^- > 0$ . Since  $c$  is only a function of  $x$ , Clearly, (1.3 implies that  $\eta$  will not change across the interface, while  $\xi$  has three possibilities:

- 1)  $c^- > c^+$ . In this case, the local wave speed decreases, so the wave will cross the interface and increase its  $\xi$  value in order to maintain a constant Hamiltonian. (2.4) implies

$$\xi^+ = \sqrt{\left(\frac{c^-}{c^+}\right)^2 (\xi^-)^2 + \left[\left(\frac{c^-}{c^+}\right)^2 - 1\right] (\eta^-)^2}$$

- 2)  $c^- < c^+$  and  $\left(\frac{c^-}{c^+}\right)^2 (\xi^-)^2 + \left[\left(\frac{c^-}{c^+}\right)^2 - 1\right] (\eta^-)^2 > 0$ . In this case the wave can also cross the interface with a loss in the  $\xi$  value. (2.4) still gives

$$\xi^+ = \sqrt{\left(\frac{c^-}{c^+}\right)^2 (\xi^-)^2 + \left[\left(\frac{c^-}{c^+}\right)^2 - 1\right] (\eta^-)^2}$$

- 3)  $c^- < c^+$  and  $\left(\frac{c^-}{c^+}\right)^2 (\xi^-)^2 + \left[\left(\frac{c^-}{c^+}\right)^2 - 1\right] (\eta^-)^2 < 0$ . In this case, there is no possibility for the wave to cross the interface, so the wave will be reflected with velocity  $(-\xi^-, \eta^-)$ .

If  $\xi^- < 0$ , similar behavior can also be analyzed using the constant Hamiltonian condition (2.4).

*Remark 2.1.* In general, one can not define a unique weak solution to a linear hyperbolic equation with singular (discontinuous or measure-valued) coefficients. By using the wave behavior described above, we give a *selection criterion* on a unique weak solution. This weak solution is the one when the wave length of the incident wave is much smaller than the width of the interface, both of which go to zero. It is equivalent to Snell's Law of refraction:

$$\frac{\sin \theta_i}{c_-} = \frac{\sin \theta_t}{c_+} \quad (2.5)$$

where  $\theta_i$  and  $\theta_t$  stand for angles of incident and transmitted waves. This is to say:

$$\frac{\eta^-}{c_- \sqrt{(\xi^-)^2 + (\eta^-)^2}} = \frac{\eta^+}{c_+ \sqrt{(\xi^+)^2 + (\eta^+)^2}} \quad (2.6)$$

If  $c = c(x)$ , then (1.3) implies that

$$\eta_+ = \eta_- , \quad (2.7)$$

Clearly (2.6) and (2.7) implies (2.4).

Of course this is not the only physically relevant possibility to choose a weak solution. In particular, this principle excludes the more general case when the reflection and transmission waves coexist. It applies to the case when the wave length of the incident wave is much shorter than the width of the interface as both lengths go to zero. The more general case of transmission and reflection will not be treated in this paper, but will be a topic of a forthcoming paper.

The main ingredient in the *well-balanced* kinetic scheme by Perthame and Semioni [28] for the shallow water equations with topography was to build in the Hamiltonian-preserving mechanism into the numerical flux in order to preserve the steady state solution of the shallow water equations when the water velocity is zero. This is achieved using the fact that the density distribution  $f$  remains unchanged along the characteristic, thus

$$f(t, x^+, \xi^+) = f(t, x^-, \xi^-) \quad (2.8)$$

at a discontinuous point  $x$  of  $c(x)$ , where  $\xi^+$  is defined through the constant Hamiltonian condition (2.4).

In this paper, we use this mechanism for the numerical approximation to the Liouville equation (2.1),(5.1) with a discontinuous wave speed. This approximation, by its design, maintains a constant Hamiltonian modulus the numerical approximation error across the interface. In [22] we introduced the Hamiltonian-preserving schemes for the Liouville equation arising from the semiclassical limit of the linear Schrödinger equation by incorporating this particle behavior into the numerical flux.

### 3 Scheme I: a finite difference approach

#### 3.1 A Hamiltonian-preserving numerical flux

We now describe our first finite difference scheme (called *Scheme I*) for the Liouville equation with a discontinuous local wave speed.

Assume that the discontinuous points of wave speed  $c$  are located at the grid points. Let the left and right limits of  $c(x)$  at point  $x_{i+1/2}$  be  $c_{i+1/2}^+$  and  $c_{i+1/2}^-$  respectively. Note that if  $c$  is continuous at  $x_{j+1/2}$ , then  $c_{i+1/2}^+ = c_{i+1/2}^-$ . We approximate  $c$  by a piecewise linear function

$$c(x) \approx c_{j-1/2}^+ + \frac{c_{j+1/2}^- - c_{j-1/2}^+}{\Delta x}(x - x_{j-1/2}).$$

We also define the averaged wave speed as  $c_i = \frac{c_{i-1/2}^+ + c_{i+1/2}^-}{2}$ . We will adopt the flux splitting technique used in [28]. The semidiscrete scheme (with time continuous) reads

$$(f_{ij})_t + \frac{c_i \text{sign}(\xi_j)}{\Delta x} (f_{i+1/2,j}^- - f_{i-1/2,j}^+) - \frac{c_{i+1/2}^- - c_{i-1/2}^+}{\Delta x \Delta \xi} |\xi_j| (f_{i,j+1/2} - f_{i,j-1/2}) = 0, \quad (3.1)$$

where the numerical fluxes  $f_{i,j+1/2}$  is defined using the upwind discretization. Since the characteristics of the Liouville equation maybe different on the two sides of the interface, the corresponding numerical fluxes should also be different. The essential part of our algorithm is to define the splitted numerical fluxes  $f_{i+1/2,j}^-$ ,  $f_{i-1/2,j}^+$  at each cell interface. We will use (2.8) to define these fluxes.

Since  $\xi^+$  determined by the constant Hamiltonian condition (2.4) may not be a grid point, we have to compute it approximately. The first approach is to locate the two cell centers that bound  $\xi^+$ , then use a linear interpolation to evaluate the needed numerical flux at  $\xi^+$ . The detailed algorithm to generate the numerical flux is given below.

**Algorithm I**

- if  $\xi_j > 0$

$$f_{i+1/2,j}^- = f_{ij},$$

$$\xi' = \frac{c_{i+1/2}^+}{c_{i+1/2}^-} \xi_j$$

if  $\xi_k \leq \xi' < \xi_{k+1}$  for some  $k$

$$\text{then } f_{i+1/2,j}^+ = \frac{\xi_{k+1} - \xi'}{\Delta \xi} f_{i,k} + \frac{\xi' - \xi_k}{\Delta \xi} f_{i,k+1}$$

- if  $\xi_j < 0$

$$f_{i+\frac{1}{2},j}^+ = f_{i+1,j},$$

$$\xi' = \frac{c_{i+\frac{1}{2}}^-}{c_{i+\frac{1}{2}}^+} \xi_j$$

if  $\xi_k \leq \xi' < \xi_{k+1}$  for some  $k$

$$\text{then } f_{i+\frac{1}{2},j}^- = \frac{\xi_{k+1} - \xi'}{\Delta\xi} f_{i+1,k} + \frac{\xi' - \xi_k}{\Delta\xi} f_{i+1,k+1}$$

The above algorithm for evaluating numerical fluxes is of first order. One can obtain a second order flux by incorporating the slope limiter, such as van Leer or minmod slope limiter [24], into the above algorithm. This can be achieved by replacing  $f_{ik}$  with  $f_{ik} + \frac{\Delta x}{2} s_{ik}$ , and replacing  $f_{i+1,k}$  with  $f_{i+1,k} - \frac{\Delta x}{2} s_{i+1,k}$  in the above algorithm for all the possible index  $k$ , where  $s_{ik}$  is the slope limiter in the  $x$ -direction.

After the spatial discretization is specified, one can use any time discretization for the time derivative.

### 3.2 Positivity and $l^\infty$ contraction

Since the exact solution of the Liouville equation is positive when the initial profile is, it is important that the numerical solution inherits this property.

We only consider the scheme using the first order numerical flux, and the forward Euler method in time. Without loss of generality, we consider the case  $\xi_j > 0$  and  $c_{i+\frac{1}{2}}^- < c_{i-\frac{1}{2}}^+$  for all  $i$  (the other cases can be treated similarly with the same conclusion). The scheme reads

$$\frac{f_{ij}^{n+1} - f_{ij}^n}{\Delta t} + c_i \frac{f_{ij} - (d_1 f_{i-1,k} + d_2 f_{i-1,k+1})}{\Delta x} - \frac{c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+}{\Delta x} \xi_j \frac{f_{ij} - f_{i,j-1}}{\Delta\xi} = 0,$$

where  $d_1, d_2$  are non-negative and  $d_1 + d_2 = 1$ . We omit the superscript  $n$  of  $f$ . The above scheme can be rewritten as

$$\begin{aligned} f_{ij}^{n+1} &= \left( 1 - c_i \lambda_x^t - \frac{|c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+|}{\Delta x} |\xi_j| \lambda_\xi^t \right) f_{ij} + c_i \lambda_x^t (d_1 f_{i-1,k} + d_2 f_{i-1,k+1}) \\ &+ \frac{|c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+|}{\Delta x} |\xi_j| \lambda_\xi^t f_{i,j-1}. \end{aligned} \quad (3.2)$$

Now we investigate the positivity of scheme (3.2). This is to prove that if  $f_{ij}^n \geq 0$  for all  $(i, j)$ , then this is also true for  $f^{n+1}$ . Clearly one just needs to show that all

the coefficients before  $f^n$  are non-negative. A sufficient condition for this is clearly

$$1 - c_i \lambda_x^t - \frac{|c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+|}{\Delta x} |\xi_j| \lambda_\xi^t \geq 0,$$

or

$$\Delta t \max_{i,j} \left[ \frac{c_i}{\Delta x} + \frac{\frac{|c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+|}{\Delta x} |\xi_j|}{\Delta \xi} \right] \leq 1. \quad (3.3)$$

This CFL condition is similar to the CFL condition (2.3) of the usual finite difference scheme *except* that the quantity  $\frac{|c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+|}{\Delta x}$  now represents the wave speed gradient at its *smooth* point, which has a *finite* upper bound. Thus our scheme allows a time step  $\Delta t = O(\Delta x, \Delta \xi)$ , a significant improvement over a standard discretization.

According to the study in [27], our second order scheme, which incorporates slope limiter into the first order scheme, is positive under the half CFL condition, namely, the constant on the right hand side of (3.3) is 1/2.

The above conclusion are analyzed based on forward Euler time discretization. One can draw the same conclusion for the second order TVD Runge-Kutta time discretization [34].

The  $l^\infty$ -contracting property of this scheme follows easily, because the coefficients in (3.2) are positive and the sum of them is 1.

### 3.3 The $l^1$ -stability of Scheme I

In this section we prove the  $l^1$ -stability of Scheme I (with the first order numerical flux and the forward Euler method in time). For simplicity, we consider the case when the wave speed has only one discontinuity at grid point  $x_{m+\frac{1}{2}}$  with  $c_{m+\frac{1}{2}}^- > c_{m+\frac{1}{2}}^+$ , and  $c'(x) > 0$  at smooth points. The other cases, namely, when  $c'(x) \leq 0$ , or the wave speed having several discontinuity points with increased or decreased jumps, can be discussed similarly. Denote  $\lambda_c \equiv c_{m+\frac{1}{2}}^+ / c_{m+\frac{1}{2}}^- < 1$ .

We consider the general case that  $\xi_1 < 0, \xi_M > 0$ . For this case, the study in [20] suggests that the computational domain should exclude a set  $O_\xi = \{(x, \xi) \in R^2 \mid \xi = 0\}$  which causes singularity in the velocity field. For example, we can exclude the following index set

$$D_o = \left\{ (i, j) \mid |\xi_j| < \frac{\Delta \xi}{2} \right\},$$

from the computational domain.

Since  $c(x)$  has a discontinuity, we also define an index set

$$D_l^4 = \{(i, j) | x_i \leq x_m, \xi_j < \lambda_c \xi_1\}.$$

Due to the slowness change across the wave speed jump at  $x_{m+\frac{1}{2}}$ ,  $D_l^4$  represents the area where waves come from outside of the domain  $[x_1, x_N] \times [\xi_1, \xi_M]$ . In order to implement our scheme conveniently, this index set is also excluded from the computational domain. Thus the computational domain is chosen as

$$E_d = \{(i, j) | i = 1, \dots, N, j = 1, \dots, M\} \setminus \{D_o \cup D_l^4\}. \quad (3.4)$$

A sketch of  $E_d$  and  $D_l^4$  is shown in Figure 4.1 in section 4.2.

As a result of excluding the index set  $D_o$  from the computational domain, the computational domain is split into two independent parts

$$E_d = \{(i, j) \in E_d | \xi_j > 0\} \cup \{(i, j) \in E_d | \xi_j < 0\} \equiv E_d^+ \cup E_d^-.$$

The  $l^1$ -stability study of Scheme I can be carried out in these two domains respectively. In the following we prove the  $l^1$ -stability of Scheme I in the domain  $E_d^-$ . The study in the domain  $E_d^+$  can be made similarly.

We define the  $l^1$ -norm of a numerical solution  $u_{ij}$  in the set  $E_d^-$  to be

$$|f|_1 = \frac{1}{N_d^-} \sum_{(i,j) \in E_d^-} |f_{ij}|$$

with  $N_d^-$  being the number of elements in  $E_d^-$ .

Given the initial data  $f_{ij}^0, (i, j) \in E_d^-$ . Denote the numerical solution at time  $T$  to be  $f_{ij}^L, (i, j) \in E_d^-$ . To prove the  $l^1$ -stability, we need to show that  $|f^L|_1 \leq C|f^0|_1$ .

Due to the linearity of the scheme, the equation for the error between the analytical and the numerical solution is the same as (3.2), so in this section,  $f_{ij}$  will denote the error. We assume there is no error at the boundary, thus  $f_{ij}^n = 0$  at the boundary. If the  $l^1$ -norm of the error introduced at each time step in incoming boundary cells is ensured to be  $o(1)$  part of  $|u^n|_1$ , our following analysis still applies.

Now denote

$$A_i = \frac{1}{\Delta x} \left| c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+ \right|. \quad (3.5)$$

Assume an upper bound for the wave speed slopes is  $A_u$ ,  $A_i < A_u, \forall i$ . These notations will be used below as well as in the stability proof of Scheme II. Assume the wave speed are subject to a lower bound  $C_m, c_i > C_m > 0, \forall i$ .

When  $\xi_j < 0$ , Scheme I is given by

1) if  $i \neq m$ ,

$$f_{ij}^{n+1} = \left(1 - A_i |\xi_j| \lambda_\xi^t - c_i \lambda_x^t\right) f_{ij} + A_i |\xi_j| \lambda_\xi^t f_{i,j+1} + c_i \lambda_x^t f_{i+1,j}, \quad (3.6)$$

2)

$$\begin{aligned} f_{m,j}^{n+1} &= \left(1 - A_m |\xi_j| \lambda_\xi^t - c_m \lambda_x^t\right) f_{m,j} + A_m |\xi_j| \lambda_\xi^t f_{m,j+1} \\ &+ c_m \lambda_x^t (d_{j,k} f_{m+1,k} + d_{j,k+1} f_{m+1,k+1}), \end{aligned} \quad (3.7)$$

where  $0 \leq d_{jk} \leq 1$  and  $d_{jk} + d_{j,k+1} = 1$ . In (3.7)  $k$  is determined by  $\xi_k \leq \frac{\xi_j}{\lambda_c} < \xi_{k+1}$ .

When summing up all absolute values of  $f_{ij}^{n+1}$  in (3.6)-(3.7), one typically gets the following inequality

$$|f^{n+1}|_1 \leq \frac{1}{N_d^-} \sum_{(i,j) \in E_d^-} \alpha_{ij} |f_{ij}^n|, \quad (3.8)$$

where the coefficients  $\alpha_{ij}$  are positive. One can check that, under the CFL condition (3.3),  $\alpha_{ij} \leq 1$  except for possibly  $(i, j) \in D_{m+1}^-$  defined as

$$D_{m+1}^- = \{(i, j) \in E_d^- | i = m + 1\}.$$

We next derive the bound for  $M^-$  defined as

$$M^- = \max_{(m+1,j) \in D_{m+1}^-} \alpha_{m+1,j}.$$

Define the set

$$S_j^{m+1} = \left\{ j' | \xi_{j'} < 0, \left| \frac{\xi_{j'}}{\lambda_c} - \xi_j \right| < \Delta \xi \right\} \quad \text{for } (m+1, j) \in D_{m+1}^-.$$

Let the number of elements in  $S_j^{m+1}$  be  $N_j^{m+1}$ . One can check that  $N_j^{m+1} \leq 2\lambda_c + 1$  because every two elements  $j'_1, j'_2 \in S_j^m$  satisfy  $\left| \frac{\xi_{j'_1}}{\lambda_c} - \frac{\xi_{j'_2}}{\lambda_c} \right| \geq \frac{\Delta \xi}{\lambda_c}$ .

On the other hand, one can easily check from (3.6) and (3.7), for  $(m+1, j) \in D_{m+1}^-$ ,

$$\alpha_{m+1,j} < 1 - c_{m+1} \lambda_x^t + c_m \lambda_x^t (2\lambda_c + 1) = 1 + (c_m + c_{m+1}) \lambda_x^t + O(\Delta x),$$

so for sufficiently small  $\Delta x$ ,  $M^-$  can be bounded by

$$M^- < 1 + 2(c_m + c_{m+1}) \lambda_x^t.$$

Denote  $M' = 2(c_m + c_{m+1}) \lambda_x^t$ . From (3.8),

$$|f^{n+1}|_1 < |f^n|_1 + \frac{M'}{N_d^-} \sum_{(m+1,j) \in D_{m+1}^-} |f_{m+1,j}^n|. \quad (3.9)$$

We now establish the following theorem:

**Theorem 3.1.** *The scheme (3.6), (3.7) is  $l^1$ -stable*

$$|f^L|_1 < C|f^0|_1.$$

*Proof.* From (3.9),

$$|f^L|_1 < |f^0|_1 + \frac{M'}{N_d^-} \sum_{n=0}^{L-1} \left\{ \sum_{(m+1,j) \in D_{m+1}^-} |f_{m+1,j}^n| \right\}. \quad (3.10)$$

It remains to estimate

$$S = \sum_{n=0}^{L-1} \left\{ \sum_{(m+1,j) \in D_{m+1}^-} |f_{m+1,j}^n| \right\}. \quad (3.11)$$

Define the set

$$S_r = \{(i, j) \mid x_i > x_{m+\frac{1}{2}}, (m+1, j) \in D_{m+1}^-\}.$$

$\forall (i, j) \in S_r$ , due to the zero boundary condition and the upwind nature of the scheme, one has

$$f_{ij}^n = \sum_{(p,q) \in S_r, p \geq i} \beta_{pq}^{ijn0} f_{pq}^0, \quad (i, j) \in S_r \quad (3.12)$$

with  $\beta_{pq}^{ijn0} \geq 0$ .

Notice  $D_{m+1}^- \subset S_r$ ,

$$S \leq \sum_{(p,q) \in S_r} \left( \sum_{n=0}^{L-1} \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{m+1,jn0} \right) |f_{pq}^0| \equiv \sum_{(p,q) \in S_r} F(p, q) |f_{pq}^0|, \quad (3.13)$$

where we have defined

$$F(p, q) = \sum_{n=0}^{L-1} \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{m+1,jn0}, \quad (p, q) \in S_r. \quad (3.14)$$

Thus the next step is to estimate these coefficients. Define

$$\beta_{pq}^{ij0} = \sum_{n=0}^{\infty} \beta_{pq}^{ijn0}, \quad (i, j), (p, q) \in S_r, p \geq i,$$

then (3.14) gives

$$F(p, q) = \sum_{(m+1,j) \in D_{m+1}^-} \sum_{n=0}^{L-1} \beta_{pq}^{m+1,jn0} \leq \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{m+1,j0}, \quad (p, q) \in S_r.$$

We first evaluate  $\sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij0}$  when  $i = p$ . Denote  $c_1^{ij} = 1 - A_i |\xi_j| \lambda_\xi^t - c_i \lambda_x^t$ ,  $c_2^{ij} = A_i |\xi_j| \lambda_\xi^t$ ,  $c_3^i = c_i \lambda_x^t$ . Assume the constant in the right hand side of CFL condition (3.3) is  $C_F < 1$ . Then  $c_1^{ij}$ ,  $c_2^{ij}$ ,  $c_3^i$  are all bounded by constant less than 1.  $c_1^{ij} \leq 1 - C_m \lambda_x^t$ ,  $c_2^{ij} < C_F$ ,  $c_3^i \leq C_F$ . From scheme (3.6), it can be directly computed

$$\sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{pj0} < \sum_{n=0}^{\infty} (1 - C_m \lambda_x^t)^n + \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} (C_F)^n = \frac{1}{C_m \lambda_x^t} + \frac{C_F}{(1 - C_F)^2}. \quad (3.15)$$

We now study the relation between  $\sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij0}$  and  $\sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,j0}$  when  $i < p$ . From scheme (3.6),

$$\beta_{pq}^{ij,n+1,0} = c_1^{ij} \beta_{pq}^{ijn0} + c_2^{ij} \beta_{pq}^{i,j+1,n0} + c_3^i \beta_{pq}^{i+1,jn0}. \quad (3.16)$$

Summing up  $j$  in (3.16) gives

$$\begin{aligned} \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij,n+1,0} &= \sum_{(m+1,j) \in D_{m+1}^-} (c_1^{ij} + c_2^{i,j-1}) \beta_{pq}^{ijn0} + c_3^i \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,jn0} \\ &< (1 - c_3^i + A_u \lambda_\xi^t \Delta \xi) \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ijn0} + c_3^i \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,jn0}, \end{aligned} \quad (3.17)$$

then a sum of  $n$  from 0 to  $\infty$  in (3.17) gives

$$(c_3^i - A_u \lambda_\xi^t \Delta \xi) \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij0} < c_3^i \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,j0},$$

so

$$\begin{aligned} \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij0} &< \frac{c_3^i}{c_3^i - A_u \lambda_\xi^t \Delta \xi} \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,j0} \\ &< \left(1 + \frac{A_u}{C_m} \Delta x + o(\Delta x)\right) \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,j0}. \end{aligned}$$

Thus for sufficiently small  $\Delta x$ , one has

$$\sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{ij0} < \left(1 + \frac{2A_u}{C_m} \Delta x\right) \sum_{(m+1,j) \in D_{m+1}^-} \beta_{pq}^{i+1,j0}, \quad i < p. \quad (3.18)$$

We now can evaluate  $F(p, q)$  for  $(p, q) \in S_r$ . From the definition of  $S_r$ , when  $(p, q) \in S_r$ , one has  $p \geq m + 1$ .

$$\begin{aligned}
F(p, q) &\leq \sum_{(m+1, j) \in D_{m+1}^-} \beta_{pq}^{m+1j0} < \left(1 + \frac{2A_u}{C_m} \Delta x\right) \sum_{(m+1, j) \in D_{m+1}^-} \beta_{pq}^{m+2j0} \\
&< \cdots < \left(1 + \frac{2A_u}{C_m} \Delta x\right)^{p-m-1} \sum_{(m+1, j) \in D_{m+1}^-} \beta_{pq}^{pj0} \\
&< \exp\left(\frac{2A_u}{C_m} (x_N - x_1)\right) \sum_{(m+1, j) \in D_{m+1}^-} \beta_{pq}^{pj0} \\
&< \exp\left(\frac{2A_u}{C_m} (x_N - x_1)\right) \left[ \frac{1}{C_m \lambda_x^t} + \frac{C_F}{(1 - C_F)^2} \right] \equiv C_T. \tag{3.19}
\end{aligned}$$

Therefore, from (3.13) one gets

$$S_2 \leq \sum_{(p, q) \in S_r} F(p, q) |f_{pq}^0| < C_T \sum_{(p, q) \in S_r} |f_{pq}^0| \leq C_T \sum_{(p, q) \in E_d} |f_{pq}^0| = C_T N_d |f^0|_1. \tag{3.20}$$

Combing (3.10) and (3.20),

$$\begin{aligned}
|f^L|_1 &< |f^0|_1 + C_T M' |f^0|_1 \\
&= [1 + C_T M'] |f^0|_1 \\
&\equiv C |f^0|_1
\end{aligned}$$

where  $C \equiv 1 + C_T M'$ . Thus Theorem 3.1 is proved.  $\square$

One can prove the similar conclusion for index set  $E_d^+$ .

## 4 Scheme II: a finite volume approach

### 4.1 A Hamiltonian-preserving numerical flux

In this section we derive another flux based on the finite volume approach which results in an  $l^1$ -contracting scheme. We call this scheme as *Scheme II*.

Consider the semidiscrete flux splitting scheme

$$(f_{ij})_t + \frac{\text{sign}(\xi_j)}{\Delta x} (c_{i+\frac{1}{2}}^- f_{i+\frac{1}{2}, j}^- - c_{i-\frac{1}{2}}^+ f_{i-\frac{1}{2}, j}^+) - \frac{c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+}{\Delta x \Delta \xi} |\xi_j| (f_{i, j+\frac{1}{2}} - f_{i, j-\frac{1}{2}}) = 0. \tag{4.1}$$

In the finite volume approach, the numerical fluxes are regarded as integral of solution along the cell interface which depends on the sign of  $\xi_j$  and  $\frac{c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+}{\Delta x}$ . To

illustrate the basic idea, we assume  $\xi_j > 0$ ,  $\frac{c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+}{\Delta x} < 0$ . In this case

$$\begin{aligned} f_{i+\frac{1}{2},j}^- &= \frac{1}{\Delta \xi} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} f \left( x_{i+\frac{1}{2}}^-, \xi, t \right) d\xi, \\ f_{i,j+\frac{1}{2}} &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f \left( x, \xi_{j+\frac{1}{2}}^-, t \right) dx \end{aligned}$$

where  $x_{i+\frac{1}{2}}^-, \xi_{j+\frac{1}{2}}^-$  are the limit from the negative coordinate in the  $x$ -and  $\xi$ -direction, taking into account that  $f(x, \xi, t)$  may be discontinuous at the grid point  $x = x_{i+\frac{1}{2}}$  and  $\xi = \xi_{j+\frac{1}{2}}$ .

By using condition (2.8):

$$f_{i+\frac{1}{2},j}^+ = \frac{1}{\Delta \xi} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} f \left( x_{i+\frac{1}{2}}^+, \xi, t \right) d\xi, = \frac{1}{\Delta \xi} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} \bar{f} \left( x_{i+\frac{1}{2}}^-, \xi, t \right) d\xi, \quad (4.2)$$

where  $\bar{f}$  is defined as

$$\bar{f} \left( x_{i+\frac{1}{2}}^-, \xi, t \right) = f \left( x_{i+\frac{1}{2}}^-, \frac{c_{i+\frac{1}{2}}^+}{c_{i+\frac{1}{2}}^-} \xi, t \right).$$

Using change of variable on (4.2) leads to

$$\begin{aligned} f_{i+\frac{1}{2},j}^+ &= \frac{1}{\Delta \xi} \int_{\xi_{j-\frac{1}{2}}}^{\xi_{j+\frac{1}{2}}} f \left( x_{i+\frac{1}{2}}^-, \frac{c_{i+\frac{1}{2}}^+ \xi}{c_{i+\frac{1}{2}}^-}, t \right) d\xi \\ &= \frac{c_{i+\frac{1}{2}}^-}{c_{i+\frac{1}{2}}^+} \frac{1}{\Delta \xi} \int_{c_{i+\frac{1}{2}}^+ \xi_{j-\frac{1}{2}} / c_{i+\frac{1}{2}}^-}{c_{i+\frac{1}{2}}^+ \xi_{j+\frac{1}{2}} / c_{i+\frac{1}{2}}^-} f \left( x_{i+\frac{1}{2}}^-, \xi, t \right) d\xi. \end{aligned} \quad (4.3)$$

The integral in (4.3) will be approximated by a quadrature rule. Since the end point  $c_{i+\frac{1}{2}}^+ \xi_{j+\frac{1}{2}} / c_{i+\frac{1}{2}}^-$  in (4.3) may not be a grid point in the  $\xi$ -direction, special care needs to be taken at both ends of the interval

$$\left[ c_{i+\frac{1}{2}}^+ \xi_{j-\frac{1}{2}} / c_{i+\frac{1}{2}}^-, c_{i+\frac{1}{2}}^+ \xi_{j+\frac{1}{2}} / c_{i+\frac{1}{2}}^- \right]. \quad (4.4)$$

We propose the following evaluation of the splitted fluxes  $f_{i+\frac{1}{2},j}^\pm$  in (4.1)

### Algorithm II

- if  $\xi_j > 0$

$$f_{i+\frac{1}{2},j}^- = f_{ij},$$

$$\xi'_1 = \frac{c_{i+\frac{1}{2}}^+}{c_{i+\frac{1}{2}}^-} \xi_{j-\frac{1}{2}}, \quad \xi'_2 = \frac{c_{i+\frac{1}{2}}^+}{c_{i+\frac{1}{2}}^-} \xi_{j+\frac{1}{2}}$$

\* if  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi'_2 \leq \xi_{k+\frac{1}{2}}$  for some  $k$

$$f_{i+\frac{1}{2},j}^+ = f_{ik}$$

\* else  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi_{k+\frac{1}{2}} < \dots < \xi_{k+s-\frac{1}{2}} < \xi'_2 \leq \xi_{k+s+\frac{1}{2}}$  for some  $k, s$

$$f_{i+\frac{1}{2},j}^+ = \frac{C_{i+\frac{1}{2}}^-}{C_{i+\frac{1}{2}}^+} \left\{ \frac{\xi_{k+\frac{1}{2}} - \xi'_1}{\Delta\xi} f_{ik} + f_{i,k+1} + \dots + f_{i,k+s-1} + \frac{\xi'_2 - \xi_{k+s-\frac{1}{2}}}{\Delta\xi} f_{i,k+s} \right\}$$

\* end

• if  $\xi_j < 0$

$$f_{i+\frac{1}{2},j}^+ = f_{i+1,j},$$

$$\xi'_1 = \frac{c_{i+\frac{1}{2}}^-}{c_{i+\frac{1}{2}}^+} \xi_{j-\frac{1}{2}}, \quad \xi'_2 = \frac{c_{i+\frac{1}{2}}^-}{c_{i+\frac{1}{2}}^+} \xi_{j+\frac{1}{2}}$$

\* if  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi'_2 \leq \xi_{k+\frac{1}{2}}$  for some  $k$

$$f_{i+\frac{1}{2},j}^- = f_{i+1,k}$$

\* else  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi_{k+\frac{1}{2}} < \dots < \xi_{k+s-\frac{1}{2}} < \xi'_2 \leq \xi_{k+s+\frac{1}{2}}$  for some  $k, s$

$$f_{i+\frac{1}{2},j}^- = \frac{C_{i+\frac{1}{2}}^+}{C_{i+\frac{1}{2}}^-} \left\{ \frac{\xi_{k+\frac{1}{2}} - \xi'_1}{\Delta\xi} f_{i+1,k} + f_{i+1,k+1} + \dots + f_{i+1,k+s-1} + \frac{\xi'_2 - \xi_{k+s-\frac{1}{2}}}{\Delta\xi} f_{i+1,k+s} \right\}$$

\* end

• end

*Remark 4.1.* The above Algorithm uses a first order quadrature rule at the ends of the interval (4.4), thus it is of first order even if the slope limiters in  $x$ -direction are incorporated into the algorithm. One can also use a second order quadrature rule at the ends of intervals (4.4). But the resulting second order scheme is no longer  $l^1$ -contracting, which is the property of Scheme II, as will be proved in the next subsection. One can still prove that this scheme is  $l^1$ -stable, similar to the property of Scheme I. Compared with Scheme I, this scheme is also second order accurate and  $l^1$ -stable, but more complex to implement. We will not present the detail of this numerical scheme in this paper.

## 4.2 The $l^1$ -contraction, $l^\infty$ -stability and positivity of Scheme II

In this subsection we study the  $l^1$  and  $l^\infty$  stability of Scheme II. Its positivity is obvious under the CFL condition (3.3).

**Theorem 4.1.** *If the forward Euler time discretization is used, then the flux given by Algorithm II yields the scheme (4.1) which is  $l^1$ -contracting and  $l^\infty$ -stable.*

*Proof.* In this proof we only discuss the case when the wave speed has one discontinuity at grid point  $x_{m+\frac{1}{2}}$  with  $c_{m+\frac{1}{2}}^- > c_{m+\frac{1}{2}}^+$ , and  $c'(x) < 0$  at smooth points. The other situations can be discussed similarly.

We consider the general case that  $\xi_1 < 0, \xi_M > 0$ . We assume the mesh is such that 0 is a grid point in  $\xi$ -direction. In this case, the index set

$$D_o = \left\{ (i, j) \mid |\xi_j| < \frac{\Delta\xi}{2} \right\}$$

that needs to be excluded from the computational domain is null. In this case, the cell interface  $\{(x, \xi) \mid \xi = 0\}$  is actually the computational domain boundary where appropriate boundary conditions should be imposed [20]. As discussed in section 3.3, the computational domain is chosen as

$$E_d = \{(i, j) \mid i = 1, \dots, N, j = 1, \dots, M\} \setminus D_o^4$$

where

$$D_o^4 = \left\{ (i, j) \mid x_i \leq x_m, \xi_{j-\frac{1}{2}} < \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-} \xi_{\frac{1}{2}} \right\}.$$

Define some subsets of  $E_d$

$$\begin{aligned} D_m^+ &= \left\{ (m, j) \mid \xi_j \geq \frac{\Delta\xi}{2} \right\}, \\ D_{m+1}^+ &= \left\{ (m+1, j) \mid \xi_j \geq \frac{\Delta\xi}{2} \right\}, \\ D_m^- &= \left\{ (m, j) \mid \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-} \xi_{\frac{1}{2}} \leq \xi_{j-\frac{1}{2}} \leq -\Delta\xi \right\}, \\ D_{m+1}^- &= \left\{ (m+1, j) \mid \xi_j \leq -\frac{\Delta\xi}{2} \right\}. \end{aligned}$$

These domains are shown in Figure 4.1.

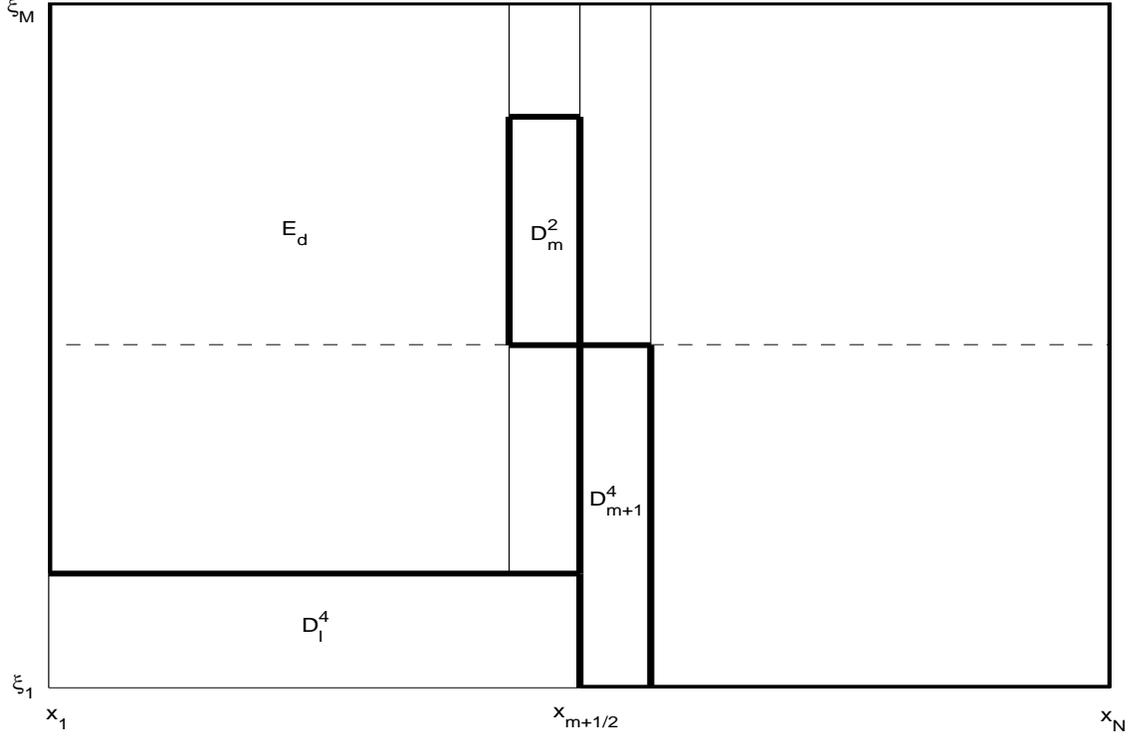


Figure 4.1 Sketch of the index sets  $D_m^+, D_{m+1}^+, D_m^-, D_{m+1}^-, D_l^4$ .

Recall the definition of  $A_i$  in (3.5). Our scheme (4.1) with Algorithm II can be made precise as

1) if  $\xi_j > 0, i \neq m + 1$ ,

$$f_{ij}^{n+1} = \left(1 - A_i \xi_j \lambda_\xi^t - c_{i+\frac{1}{2}}^- \lambda_x^t\right) f_{ij} + A_i \xi_j \lambda_\xi^t f_{i,j-1} + c_{i-\frac{1}{2}}^+ \lambda_x^t f_{i-1,j}, \quad (4.5)$$

2) if  $\xi_j < 0, i \neq m$ ,

$$f_{ij}^{n+1} = \left(1 - A_i |\xi_j| \lambda_\xi^t - c_{i-\frac{1}{2}}^+ \lambda_x^t\right) f_{ij} + A_i |\xi_j| \lambda_\xi^t f_{i,j-1} + c_{i+\frac{1}{2}}^- \lambda_x^t f_{i+1,j}, \quad (4.6)$$

3) if  $\xi_j > 0$ ,

$$f_{m+1,j}^{n+1} = \left(1 - A_{m+1} \xi_j \lambda_\xi^t - c_{m+\frac{3}{2}}^- \lambda_x^t\right) f_{m+1,j} + A_{m+1} \xi_j \lambda_\xi^t f_{m+1,j-1} + c_{m+\frac{1}{2}}^+ \lambda_x^t f_{m+\frac{1}{2},j}^+, \quad (4.7)$$

4) if  $\xi_j < 0$ ,

$$f_{mj}^{n+1} = \left(1 - A_m |\xi_j| \lambda_\xi^t - c_{m-\frac{1}{2}}^+ \lambda_x^t\right) f_{mj} + A_m |\xi_j| \lambda_\xi^t f_{m,j-1} + c_{m+\frac{1}{2}}^- \lambda_x^t f_{m+\frac{1}{2},j}^-, \quad (4.8)$$

where we omit the superscript  $n$  on the right hand side.

By summing up (4.5)-(4.8) for  $(i, j) \in E_d$ , one typically gets the following expression

$$\begin{aligned} \sum_{(i,j) \in E_d} |f_{ij}^{n+1}| &\leq \sum_{(i,j) \in E_d} \alpha_{ij} |f_{ij}| + \sum_{(m+1,j) \in D_{m+1}^+} c_{m+\frac{1}{2}}^+ \lambda_x^t |f_{m+\frac{1}{2},j}^+| + \sum_{(m,j) \in D_m^-} c_{m+\frac{1}{2}}^- \lambda_x^t |f_{m+\frac{1}{2},j}^-| \\ &\equiv I_1 + I_2 + I_3. \end{aligned} \quad (4.9)$$

As in the proof of stability of Scheme I, we assume that  $f$  satisfies the zero boundary condition. In this situation, the coefficients  $\alpha_{ij}$  in (4.9) satisfy

$$\alpha_{ij} \leq 1, \quad (i, j) \in E_d \setminus \{D_m^+ \cup D_{m+1}^-\}, \quad (4.10)$$

$$\alpha_{ij} \leq 1 - c_{m+\frac{1}{2}}^- \lambda_x^t, \quad (i, j) \in D_m^+, \quad (4.11)$$

$$\alpha_{ij} \leq 1 - c_{m+\frac{1}{2}}^+ \lambda_x^t, \quad (i, j) \in D_{m+1}^-. \quad (4.12)$$

We now study the relation between  $I_2$  and  $\sum_{(m,j) \in D_m^+} |c_{m+\frac{1}{2}}^- \lambda_x^t f_{mj}|$ . Let

$$p_{M+1} = \frac{c_{m+\frac{1}{2}}^+ \xi_{M+\frac{1}{2}}}{c_{m+\frac{1}{2}}^-},$$

and assume

$$\xi_{k-\frac{1}{2}} < p_{M+1} \leq \xi_{k+\frac{1}{2}} \leq \xi_{M+\frac{1}{2}}.$$

Assume  $\xi_{J_2-\frac{1}{2}} = 0$  for some  $J_2$ , since

$$\frac{1}{\lambda_x^t c_{m+\frac{1}{2}}^-} I_2 \leq \sum_{j=J_2}^{k-1} |f_{mj}| + \frac{p_{N+1} - \xi_k}{\Delta \xi} |f_{mk}| \leq \sum_{(m,j) \in D_m^+} |f_{mj}|,$$

thus

$$I_2 \leq \sum_{(m,j) \in D_m^+} \left| c_{m+\frac{1}{2}}^- \lambda_x^t f_{mj} \right|. \quad (4.13)$$

Similarly, one gets

$$I_3 \leq \sum_{(m+1,j) \in D_{m+1}^-} \left| c_{m+\frac{1}{2}}^+ \lambda_x^t f_{m+1,j} \right|. \quad (4.14)$$

Combining (4.9), (4.10), (4.11), (4.12), (4.13) and (4.14) gives

$$\sum_{(i,j) \in E_d} |f_{ij}^{n+1}| \leq \sum_{(i,j) \in E_d} |f_{ij}^n|. \quad (4.15)$$

This is the  $l^1$ -contracting property of Scheme II.

Next we prove the  $l^\infty$ -stability. Observing that the coefficients on the right hand side of (4.5)-(4.8) are positive, it remains to check the *sum of these coefficients* (SC).

In (4.5), the SC is

$$\text{SC}_1 = 1 + (c_{i-\frac{1}{2}}^+ - c_{i+\frac{1}{2}}^-)\lambda_x^t < 1 + A_u\Delta t. \quad (4.16)$$

In (4.6), the SC is

$$\text{SC}_2 = 1 + (c_{i+\frac{1}{2}}^- - c_{i-\frac{1}{2}}^+)\lambda_x^t < 1 + A_u\Delta t. \quad (4.17)$$

Now we derive the SC in (4.8). Denote

$$\xi'_1 = \frac{c_{m+\frac{1}{2}}^-}{c_{m+\frac{1}{2}}^+}\xi_{j-\frac{1}{2}}, \quad \xi'_2 = \frac{c_{m+\frac{1}{2}}^-}{c_{m+\frac{1}{2}}^+}\xi_{j+\frac{1}{2}}. \quad (4.18)$$

The condition  $c_{m+\frac{1}{2}}^+ < c_{m+\frac{1}{2}}^-$  gives  $\xi'_2 - \xi'_1 > \Delta\xi$ . Therefore, it is impossible that  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi'_2 \leq \xi_{k+\frac{1}{2}}$  for any  $k$ . Assume  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi_{k+\frac{1}{2}} < \dots < \xi_{k+s-\frac{1}{2}} < \xi'_2 \leq \xi_{k+s+\frac{1}{2}}$  with  $s \geq 1$ . In this case

$$\begin{aligned} f_{m+\frac{1}{2},j}^- &= \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-} \left\{ \frac{\xi_{k+\frac{1}{2}} - \xi'_1}{\Delta\xi} f_{m+1,k} + f_{m+1,k+1} + \dots \right. \\ &\quad \left. + f_{m+1,k+s-1} + \frac{\xi'_2 - \xi_{k+s-\frac{1}{2}}}{\Delta\xi} f_{m+1,k+s} \right\}. \end{aligned} \quad (4.19)$$

Substituting (4.19) into (4.8) yields the evaluation

$$\begin{aligned} \text{SC}_3 &= 1 - c_{m-\frac{1}{2}}^+\lambda_x^t + c_{m+\frac{1}{2}}^-\lambda_x^t \left[ \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-} \left( \frac{\xi_{k+\frac{1}{2}} - \xi'_1}{\Delta\xi} + \frac{\xi_{k+\frac{3}{2}} - \xi_{k+\frac{1}{2}}}{\Delta\xi} + \dots + \frac{\xi'_2 - \xi_{k+s-\frac{1}{2}}}{\Delta\xi} \right) \right] \\ &= 1 - c_{m-\frac{1}{2}}^+\lambda_x^t + c_{m+\frac{1}{2}}^-\lambda_x^t \\ &< 1 + A_u\Delta t \end{aligned} \quad (4.20)$$

Now we consider case (4.7). Denote

$$\xi'_1 = \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-}\xi_{j-\frac{1}{2}}, \quad \xi'_2 = \frac{c_{m+\frac{1}{2}}^+}{c_{m+\frac{1}{2}}^-}\xi_{j+\frac{1}{2}}. \quad (4.21)$$

In this case, we know  $\xi'_2 - \xi'_1 < \Delta\xi$ . So there are two cases  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi'_2 \leq \xi_{k+\frac{1}{2}}$  or  $\xi_{k-\frac{1}{2}} \leq \xi'_1 < \xi_{k+\frac{1}{2}} < \xi'_2 \leq \xi_{k+\frac{3}{2}}$  corresponding respectively to

$$f_{m+\frac{1}{2},j}^+ = f_{mk} \quad (4.22)$$

or

$$f_{m+\frac{1}{2},j}^+ = \frac{c_{m+\frac{1}{2}}^-}{c_{m+\frac{1}{2}}^+} \left\{ \frac{\xi_{k+\frac{1}{2}} - \xi_1'}{\Delta\xi} f_{mk} + \frac{\xi_2' - \xi_{k+\frac{1}{2}}}{\Delta\xi} f_{m,k+1} \right\}. \quad (4.23)$$

Similar to the deduction of (4.20), one can check, for both cases, that

$$\begin{aligned} \text{SC}_4 &= 1 - c_{m+\frac{3}{2}}^- \lambda_x^t + c_{m+\frac{1}{2}}^+ \lambda_x^t \\ &< 1 + A_u \Delta t. \end{aligned} \quad (4.24)$$

Combining (4.16), (4.17), (4.20) and (4.24), one gets

$$|f^{n+1}|_\infty < (1 + A_u \Delta t) |f^n|_\infty,$$

thus

$$|f^L|_\infty < (1 + A_u \Delta t)^L |f^0|_\infty < e^{A_u T} |f^0|_\infty. \quad (4.25)$$

This is the  $l^\infty$ -stability property of Scheme II. □

## 5 The schemes in higher dimensions

Consider the 2D Liouville equation

$$f_t + \frac{c(x,y)\xi}{\sqrt{\xi^2 + \eta^2}} f_x + \frac{c(x,y)\eta}{\sqrt{\xi^2 + \eta^2}} f_y - c_x \sqrt{\xi^2 + \eta^2} f_\xi - c_y \sqrt{\xi^2 + \eta^2} f_\eta = 0. \quad (5.1)$$

We employ an uniform mesh with grid points at  $x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}, \xi_{k+\frac{1}{2}}, \eta_{l+\frac{1}{2}}$  in each direction. The cells are centered at  $(x_i, y_j, \xi_k, \eta_l)$  with  $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$ ,  $y_j = \frac{1}{2}(y_{j+\frac{1}{2}} + y_{j-\frac{1}{2}})$ ,  $\xi_k = \frac{1}{2}(\xi_{k+\frac{1}{2}} + \xi_{k-\frac{1}{2}})$ ,  $\eta_l = \frac{1}{2}(\eta_{l+\frac{1}{2}} + \eta_{l-\frac{1}{2}})$ . The mesh size is denoted by  $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ ,  $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$ ,  $\Delta\xi = \xi_{k+\frac{1}{2}} - \xi_{k-\frac{1}{2}}$ ,  $\Delta\eta = \eta_{l+\frac{1}{2}} - \eta_{l-\frac{1}{2}}$ . We define the cell average of  $f$  as

$$f_{ijkl} = \frac{1}{\Delta x \Delta y \Delta\xi \Delta\eta} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{\xi_{k-\frac{1}{2}}}^{\xi_{k+\frac{1}{2}}} \int_{\eta_{l-\frac{1}{2}}}^{\eta_{l+\frac{1}{2}}} f(x, y, \xi, \eta, t) d\eta d\xi dy dx.$$

Similar to the 1D case, we approximate  $c(x, y)$  by a piecewise bilinear function, and for convenience, we always provide two interface values of  $c$  at each cell interface. When  $c$  is smooth at a cell interface, the two potential interface values are identical. We also define the averaged wave speed in a cell by averaging the four cell interface wave speed values

$$c_{ij} = \frac{c_{i-\frac{1}{2},j}^+ + c_{i+\frac{1}{2},j}^- + c_{i,j-\frac{1}{2}}^+ + c_{i,j+\frac{1}{2}}^-}{4}.$$

The 2D Liouville equation (5.1) can be semi-discretized as

$$\begin{aligned}
(f_{ijkl})_t &+ \frac{c_{ij}\xi_k}{\Delta x \sqrt{\xi_k^2 + \eta_l^2}} \left( f_{i+\frac{1}{2},jkl}^- - f_{i-\frac{1}{2},jkl}^+ \right) \\
&+ \frac{c_{ij}\eta_l}{\Delta y \sqrt{\xi_k^2 + \eta_l^2}} \left( f_{i,j+\frac{1}{2},kl}^- - f_{i,j-\frac{1}{2},kl}^+ \right) \\
&- \frac{c_{i+\frac{1}{2},j}^- - c_{i-\frac{1}{2},j}^+}{\Delta x \Delta \xi} \sqrt{\xi_k^2 + \eta_l^2} \left( f_{ij,k+\frac{1}{2},l} - f_{ij,k-\frac{1}{2},l} \right) \\
&- \frac{c_{i,j+\frac{1}{2}}^- - c_{i,j-\frac{1}{2}}^+}{\Delta y \Delta \eta} \sqrt{\xi_k^2 + \eta_l^2} \left( f_{ijk,l+\frac{1}{2}} - f_{ijk,l-\frac{1}{2}} \right) \\
&= 0,
\end{aligned}$$

where the interface values  $f_{ij,k+\frac{1}{2},l}, f_{ijk,l+\frac{1}{2}}$  are provided by the upwind approximation, and the splitted interface values  $f_{i+\frac{1}{2},jkl}^-, f_{i-\frac{1}{2},jkl}^+, f_{i,j+\frac{1}{2},kl}^-, f_{i,j-\frac{1}{2},kl}^+$  should be obtained using similar but slightly different algorithm for the 1D case. For example, to evaluate  $f_{i+\frac{1}{2},jkl}^\pm$  we can extend Algorithm I as

**Algorithm I in 2D**

- if  $\xi_k > 0$

$$f_{i+\frac{1}{2},jkl}^- = f_{ijkl},$$

$$\star \text{ if } \left( \frac{C_{i+\frac{1}{2},j}^+}{C_{i+\frac{1}{2},j}^-} \right)^2 \xi_k^2 + \left[ \left( \frac{C_{i+\frac{1}{2},j}^+}{C_{i+\frac{1}{2},j}^-} \right)^2 - 1 \right] \eta_l^2 > 0$$

$$\xi' = \sqrt{\left( \frac{C_{i+\frac{1}{2},j}^+}{C_{i+\frac{1}{2},j}^-} \right)^2 \xi_k^2 + \left[ \left( \frac{C_{i+\frac{1}{2},j}^+}{C_{i+\frac{1}{2},j}^-} \right)^2 - 1 \right] \eta_l^2}$$

if  $\xi_{k'} \leq \xi' < \xi_{k'+1}$  for some  $k'$

$$\text{then } f_{i+\frac{1}{2},jkl}^+ = \frac{\xi_{k'+1} - \xi'}{\Delta \xi} f_{ij,k',l} + \frac{\xi' - \xi_{k'}}{\Delta \xi} f_{ij,k'+1,l}$$

$\star$  else

$$f_{i+\frac{1}{2},jkl}^+ = f_{i+1,j,k',l} \text{ where } \xi_{k'} = -\xi_k$$

$\star$  end

- if  $\xi_k < 0$

$$f_{i+\frac{1}{2},jkl}^+ = f_{i+1,jkl},$$

$$\star \text{ if } \left( \frac{C_{i+\frac{1}{2},j}^-}{C_{i+\frac{1}{2},j}^+} \right)^2 \xi_k^2 + \left[ \left( \frac{C_{i+\frac{1}{2},j}^-}{C_{i+\frac{1}{2},j}^+} \right)^2 - 1 \right] \eta_l^2 > 0$$

$$\xi' = -\sqrt{\left(\frac{C_{i+\frac{1}{2},j}^-}{C_{i+\frac{1}{2},j}^+}\right)^2 \xi_k^2 + \left[\left(\frac{C_{i+\frac{1}{2},j}^-}{C_{i+\frac{1}{2},j}^+}\right)^2 - 1\right] \eta_l^2}$$

if  $\xi_{k'} \leq \xi' < \xi_{k'+1}$  for some  $k'$

then  $f_{i+\frac{1}{2},jkl}^- = \frac{\xi_{k'+1} - \xi'}{\Delta\xi} f_{i+1,j,k',l} + \frac{\xi' - \xi_{k'}}{\Delta\xi} f_{i+1,j,k'+1,l}$

☆ else

$$f_{i+\frac{1}{2},jkl}^- = f_{i,j,k',l} \text{ where } \xi_{k'} = -\xi_k$$

☆ end

The flux  $f_{i,j+\frac{1}{2},kl}^\pm$  can be constructed similarly.

The 2d version of Scheme II can be constructed similarly.

As introduced in section 2.2, the essential difference between 1D and 2D split flux definition is that in 2D case, the phenomenon that a wave is reflected at the interface does occur. While in 1D, a wave is transmitted across an interface with a change of slowness.

Since the gradient of the wave speed at its smooth points are bounded by an upper bound, this scheme similar to the 1D scheme, is also subject to a hyperbolic CFL condition under which the scheme is positive, and Hamiltonian preserving.

## 6 Numerical examples

In this section we present numerical examples to demonstrate the validity of the proposed schemes and to study their accuracy. In the numerical computations the second order TVD Runge-Kutta time discretization [34] is used. In Example 6.2 we compare the results of Scheme I and Scheme II. In other Examples, we presents the numerical results using Scheme I.

**Example 6.1.** An 1D problem with exact  $L^\infty$ -solution. Consider the 1D Liouville equation

$$f_t + c(x)\text{sign}(\xi)f_x - c_x|\xi|f_\xi = 0 \quad (6.1)$$

with a discontinuous wave speed given by

$$c(x) = \begin{cases} 0.6 & x < 0 \\ 0.5 & x > 0 \end{cases}.$$

The initial data is given by

$$f(x, \xi, 0) = \begin{cases} 1 & x < 0, \xi > 0, \sqrt{x^2 + \xi^2} < 1, \\ 1 & x > 0, \xi < 0, \sqrt{x^2 + \xi^2} < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (6.2)$$

as shown in Figure 6.1 which depicts the non-zero part of  $f(x, \xi, 0)$ .

The exact solution at  $t = 1$  is given by

$$f(x, \xi, 1) = \begin{cases} 1 & 0 < x < 0.5, \quad 0 < \xi < 1.2\sqrt{1 - (1.2x - 0.6)^2}; \\ 1 & 0 < x < 0.5, \quad -\sqrt{1 - (x + 0.5)^2} < \xi < 0; \\ 1 & -0.4 < x < 0, \quad 0 < \xi < \sqrt{1 - (x - 0.6)^2}; \\ 1 & -0.6 < x < 0, \quad -\frac{1}{1.2}\sqrt{1 - \left(\frac{x}{1.2} + 0.5\right)^2} < \xi < 0; \\ 0 & \text{otherwise,} \end{cases} \quad (6.3)$$

as shown in Figure 6.2.

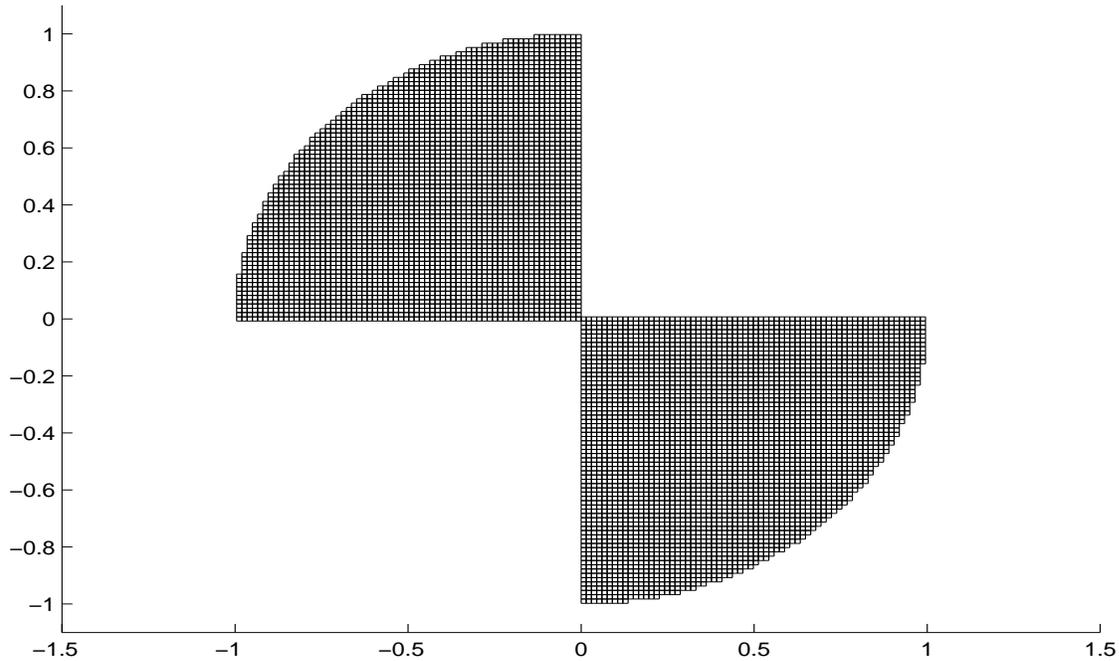


Figure 6.1 Example 6.1, non-zero part of initial data  $f(x, \xi, 0)$  in (6.2). The horizontal axis is position, the vertical axis is slowness quantity.

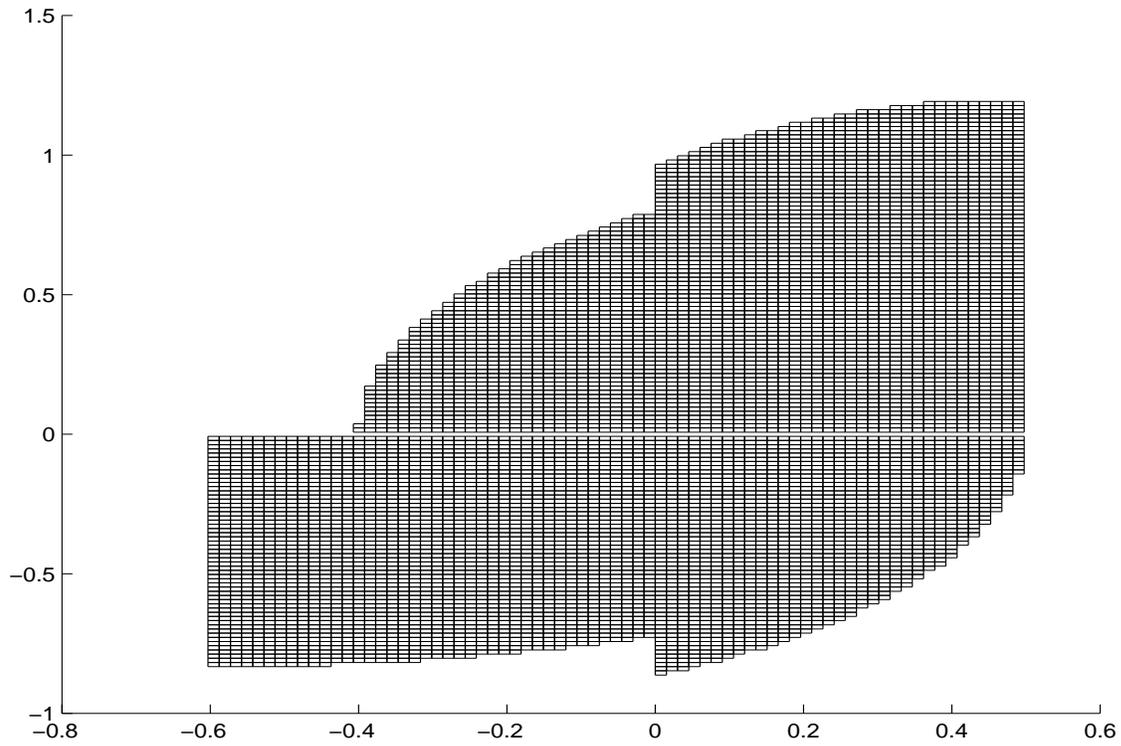


Figure 6.2 Example 6.1, non-zero part of the exact solution  $f(x, \xi, 1)$  depicted on a  $200 \times 201$  cell. The horizontal axis is position, the vertical axis is slowness.

The numerical solution computed with a  $200 \times 201$  cell on the domain  $[-1.5, 1.5] \times [-1.5, 1.5]$  using Scheme I is shown in Figure 6.3. The time step is chosen as  $\Delta t = \frac{1}{2} \Delta \xi$ . It shows a good agreement with the exact solution.

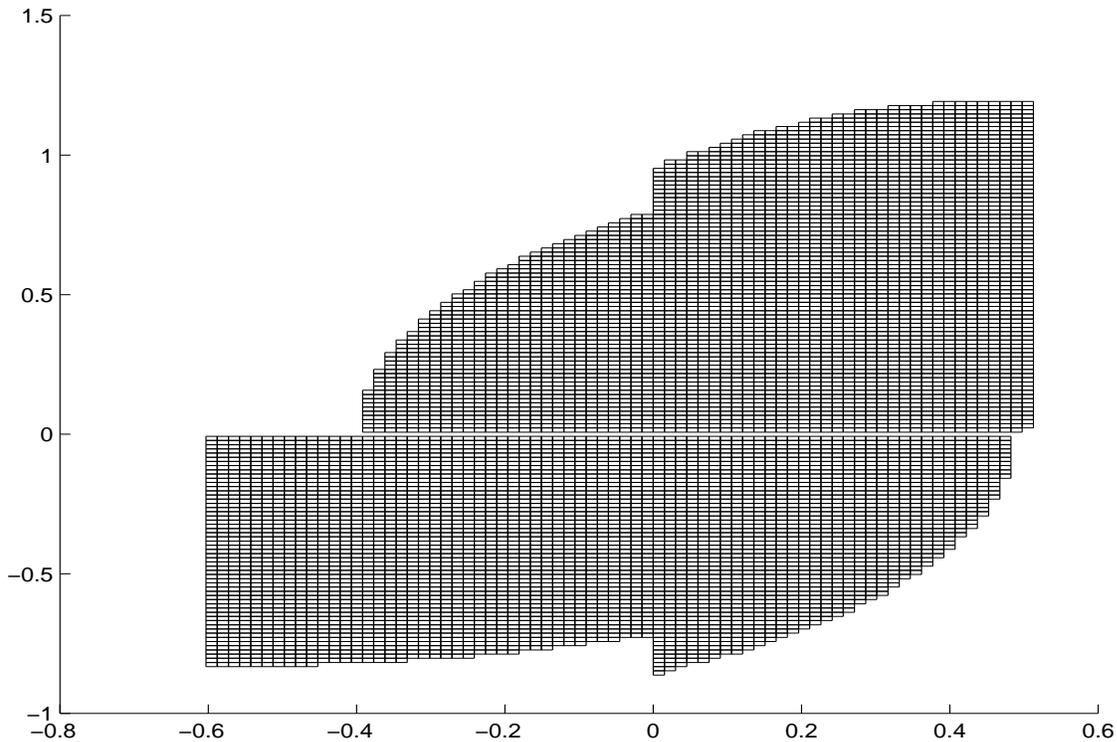


Figure 6.3 Example 6.1, the part of numerical solution of  $f(x, \xi, t)$  at  $t = 1$  where the numerical solution  $> 0.5$ . This numerical solution is computed on a  $200 \times 201$  cell. The horizontal axis is position, the vertical axis is slowness.

Table 1 compares the  $l^1$ -error of the numerical solutions computed by Scheme I using  $50 \times 51$ ,  $100 \times 101$  and  $200 \times 201$  cells respectively. This comparison shows that the convergence rate of the numerical solution in  $l^1$ -norm is about half order. This agrees with the well established theory [23], [36], that the  $l^1$ -error by finite difference scheme for a discontinuous solution of a linear hyperbolic equation is at most half order.

Table 1  $l^1$  error of numerical solutions on different mesh

grid points	$50 \times 51$	$100 \times 101$	$200 \times 201$
	0.269575	0.171837	0.102073

**Example 6.2.** Computing the physical observables of an 1D problem with delta-type solution. Consider the 1D Liouville equation (6.1), where the wave speed

is

$$c(x) = \begin{cases} \frac{1}{e-1}, & x \leq -1; \\ \frac{1}{e-1} + 1 + x, & -1 < x < 0; \\ \frac{1}{e-1} + 0.5 - x, & 0 < x < 1; \\ \frac{1}{e-1} - 0.5, & x \geq 1, \end{cases}$$

and the initial data is given by

$$f(x, \xi, 0) = \delta(\xi - w(x)) \quad (6.4)$$

with

$$w(x) = \begin{cases} 0.8, & x \leq -1.5; \\ 0.8 - \frac{0.8}{(1.5)^2}(x + 1.5)^2, & -1.5 < x \leq 0; \\ -0.8 + \frac{0.8}{(1.5)^2}(x - 1.5)^2, & 0 < x < 1.5; \\ -0.8, & x \geq 1.5. \end{cases} \quad (6.5)$$

Figure 6.4 plots  $w(x)$  in dashed line.

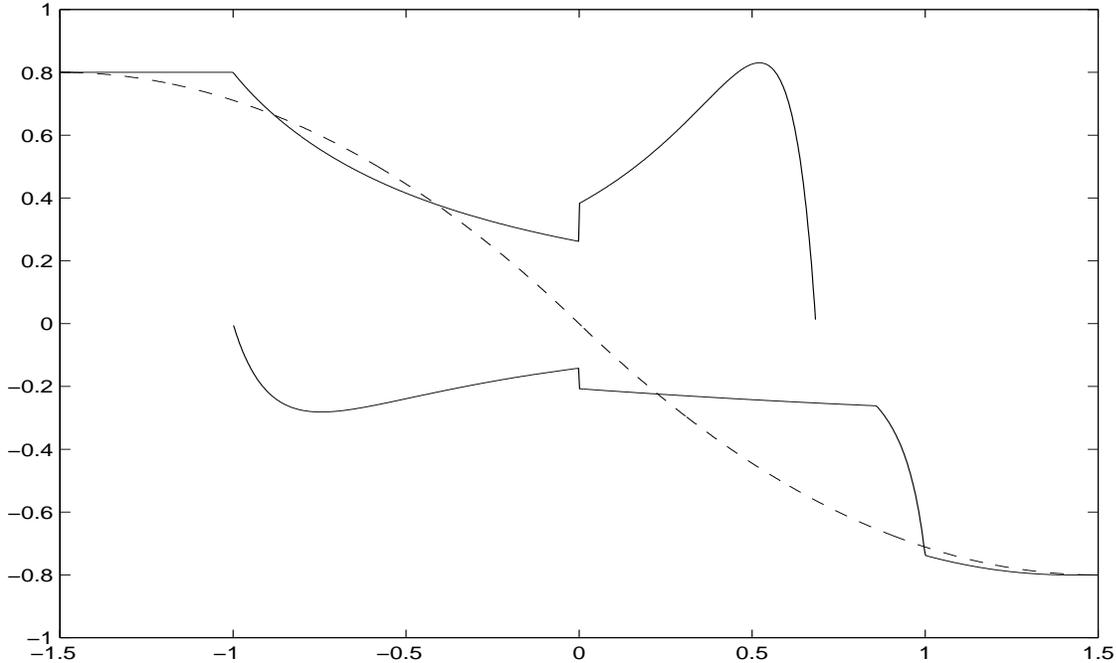


Figure 6.4 Example 6.2, velocity. Dashed line: initial velocity  $w(x)$ ; Solid line: velocity at  $t = 1$ . The horizontal axis is position, the vertical axis is slowness quantity.

In this example we are interested in the approximation of the moments, such as the density

$$\rho(x, t) = \int f(x, \xi, t) d\xi,$$

and the averaged velocity

$$u(x, t) = \frac{\int f(x, \xi, t) \xi d\xi}{\int f(x, \xi, t) d\xi}.$$

These quantities are computed by techniques described in the Introduction. We first solve the level set function  $\psi$  and modified density function  $\phi$  which satisfy the Liouville equation (6.1) with initial data  $\xi - w(x)$  and 1 respectively. Then the desired physical observables  $\rho$  and  $u$  are computed from the numerical singular integrals (1.7), (1.8), which are computed by technique described in [19].

The exact velocity and corresponding density at  $t = 1$  are given in Appendix A. Figure 6.4 shows  $u$  in solid line.

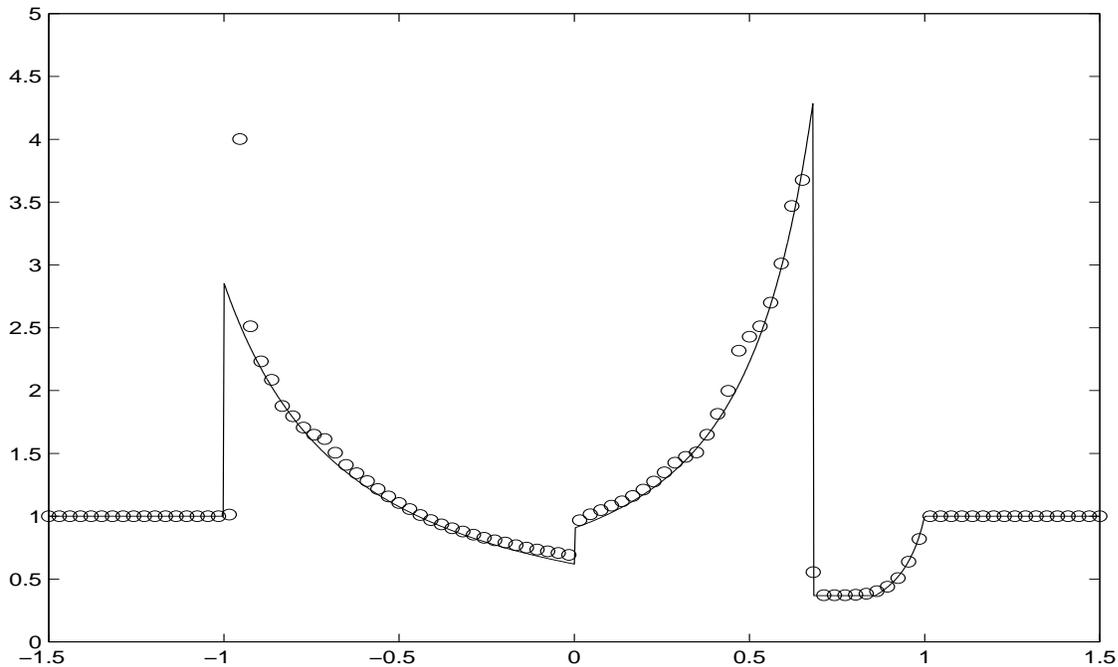


Figure 6.5 Example 6.2, density in physical space at  $t = 1$ . Solid line: exact solution; 'o': numerical solution using a  $100 \times 81$  cell.

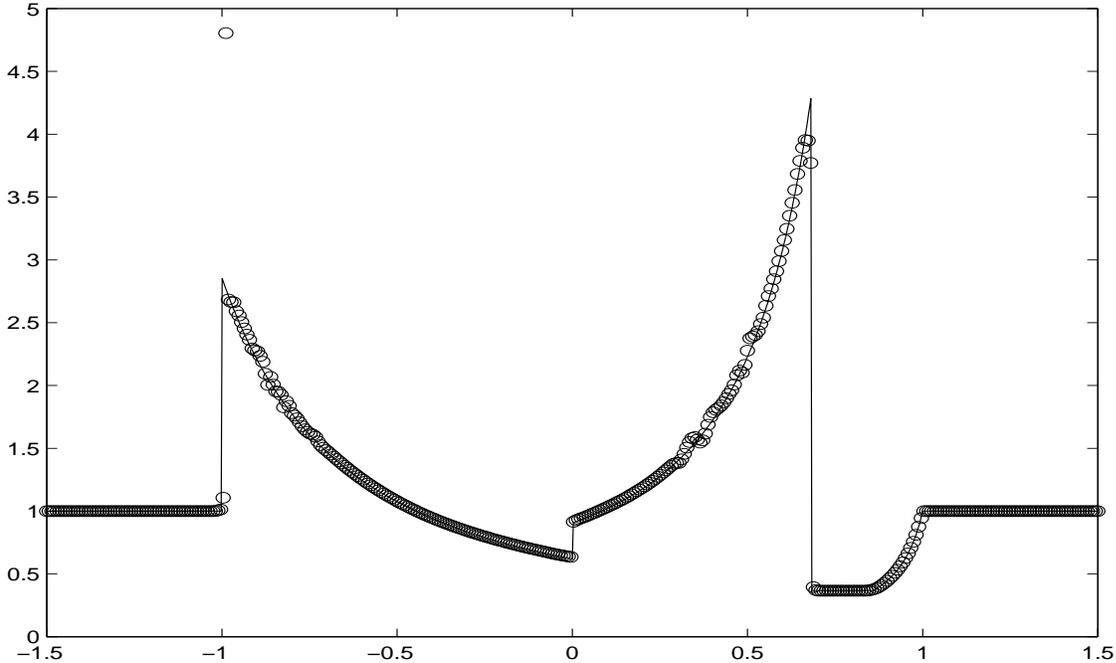


Figure 6.6 Example 6.2, density in physical space at  $t = 1$ . Solid line: exact solution; 'o': numerical solution using a  $400 \times 321$  cell.

In the computation of this example, the time step is chosen as  $\Delta t = \frac{1}{3}\Delta\xi$ . Figures 6.5-6.6 show the numerical solutions of density with different meshes using Scheme I together with the exact density. Figure 6.7-6.8 show the numerical solutions of averaged velocity in different meshes using Scheme I together with the exact averaged velocity.

Table 2 presents the  $l^1$ -error of numerical densities computed with several different meshes on the domain  $[-1.5, 1.5] \times [-1.2, 1.2]$ . Table 3 presents the  $l^1$ -error of numerical averaged velocity. It can be observed that the  $l^1$ -convergence rate of the numerical solutions is about first order. The first order accuracy here is contributed by the numerical evaluation of delta-type integral (1.7), (1.8) used to recover the moments. It should be remarked here the  $l^1$ -convergence rate of the numerical solutions reported in [22] is only halfth order due to the discontinuities in the level set function  $\psi$  arising from the reflection of particles by a potential barrier. In the Liouville equation of geometrical optics, there is no reflection phenomenon in 1D case and thus the accuracy degeneration in moment evaluation does not occur here.

In comparison, Scheme II generally has a slightly larger numerical errors than Scheme I.

Table 2  $l^1$  error of numerical densities by different meshes

grid points	$100 \times 81$	$200 \times 161$	$400 \times 321$
Scheme I	0.301738	0.117921	0.060890
Scheme II	0.301454	0.120410	0.062534

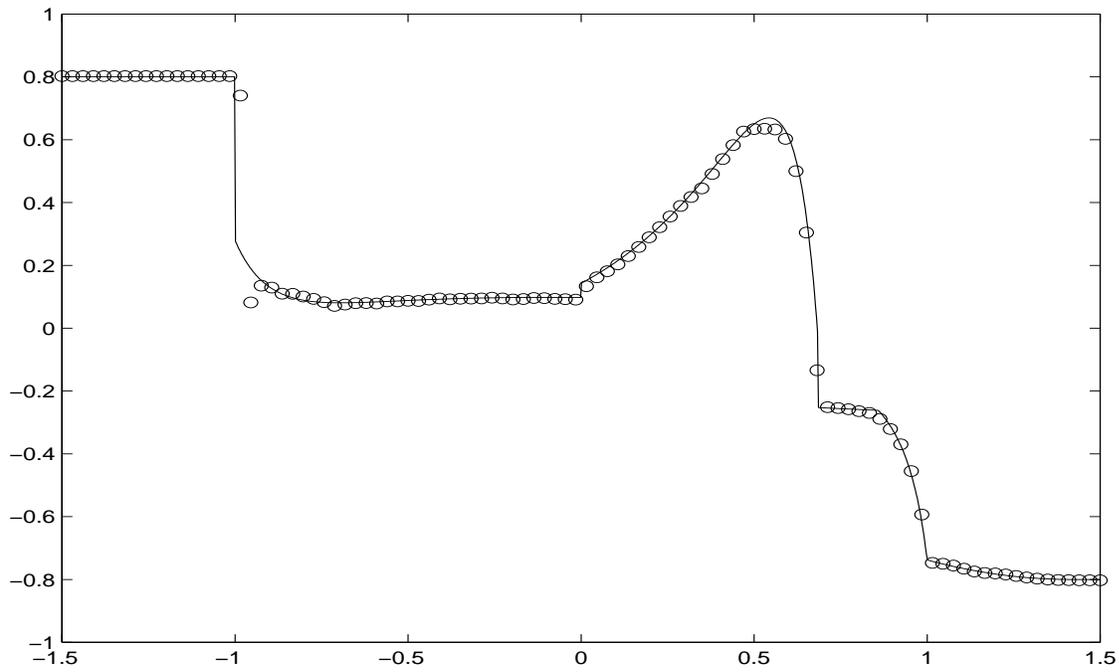


Figure 6.7 Example 6.2, averaged velocity in physical space at  $t = 1$ . Solid line: exact solution; 'o': numerical solution using a  $100 \times 81$  cell.

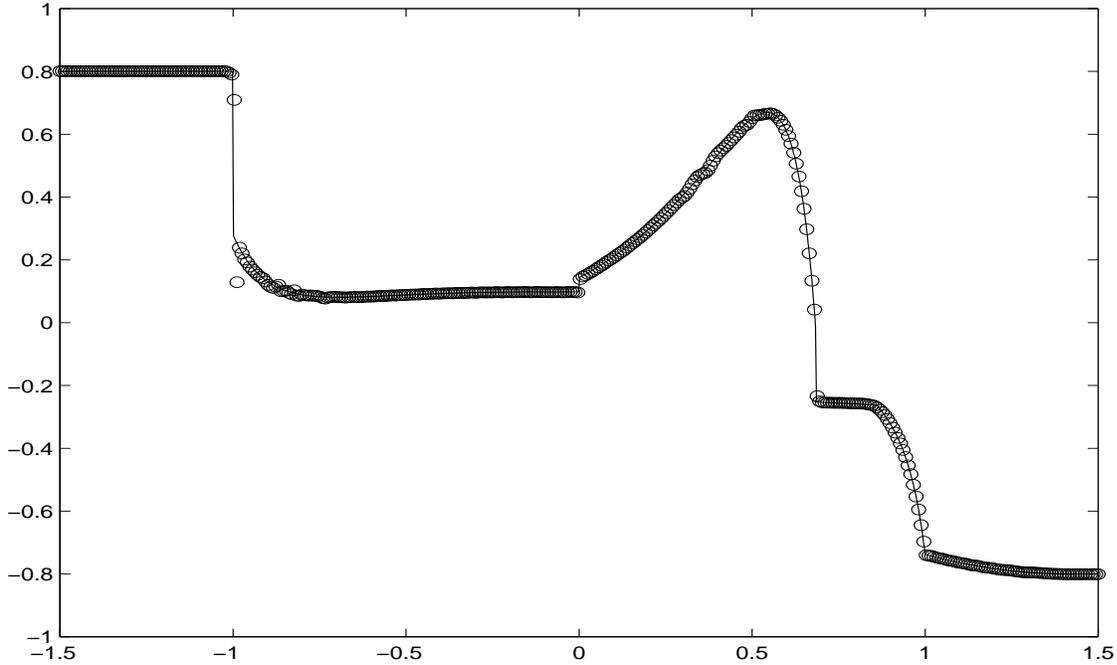


Figure 6.8 Example 6.2, averaged velocity in physical space at  $t = 1$ . Solid line: exact solution; 'o': numerical solution using a  $400 \times 321$  cell.

Table 3  $l^1$  error of numerical averaged velocity by different mesh

grid points	$100 \times 81$	$200 \times 161$	$400 \times 321$
Scheme I	0.041166	0.021229	0.008956
Scheme II	0.041249	0.022450	0.009877

**Example 6.3.** Computing the physical observables of a 2D problem with a delta-type solution. Consider the 2D Liouville equation (5.1) with a discontinuous wave speed given by

$$c(x, y) = \begin{cases} \sqrt{0.6}, & x > 0, y > 0, \\ \sqrt{0.8}, & \text{else} \end{cases}$$

and with a delta-type initial data

$$f(x, y, \xi, \eta, 0) = \rho(x, y, 0)\delta(\xi - p(x, y))\delta(\eta - q(x, y)),$$

where

$$\rho(x, y, 0) = \begin{cases} 0 & x > -0.1, y > -0.1 \\ 1 & \text{else} \end{cases},$$

$$p(x, y) \equiv q(x, y) = 0.6.$$

In this example we aim at computing the numerical density which is the first moment of this delta-type solution

$$\rho(x, y, t) = \int \int f(x, y, \xi, \eta, t) d\xi d\eta.$$

The computational domain is chosen to be  $[x, y, \xi, \eta] \in [-0.2, 0.2] \times [-0.2, 0.2] \times [0.3, 0.9] \times [0.3, 0.9]$ .

Set  $D_1 = \frac{0.4\sqrt{4}}{\sqrt{15}} - \frac{0.2\sqrt{2}}{3}$ ,  $D_2 = \sqrt{2}$ ,  $D_3 = \sqrt{\frac{9}{8}}$ , the exact density at  $t = 0.4$  is

$$\rho(x, y, 0.4) = \begin{cases} 1 & x < 0 \text{ or } y < 0 \\ D_3 & 0 \leq x \leq D_1, y \geq D_2 x \\ D_3 & 0 \leq y \leq D_1, y \leq \frac{x}{D_2} \\ 0 & \text{otherwise} \end{cases},$$

as shown in Figure 6.9 plotted on  $50^2$  space mesh.

In the computation of this example, the time step is chosen as  $\Delta t = \frac{1}{2}\Delta x$ . Figures 6.10-6.11 show respectively the numerical solutions of density with  $26^4$  and  $50^4$  phase space meshes using Scheme I.

Table 4 presents the  $l^1$  errors on  $[0, 0.2] \times [0, 0.2]$  of numerical densities computed by Scheme I with several different meshes in phase space. The convergence order is about  $1/2$ . In this example, since the initial density is discontinuous, the modified density function  $\phi$  is also discontinuous in the zero level set in phase space, which contribute to the halfth order accuracy in  $l^1$ -convergence rate of density evaluated by formula (1.7).

Table 4  $l^1$  error of numerical densities  
on  $[0, 0.2] \times [0, 0.2]$  using different meshes

grid points	$14^4$	$26^4$	$50^4$
	0.012411	0.010044	0.007741

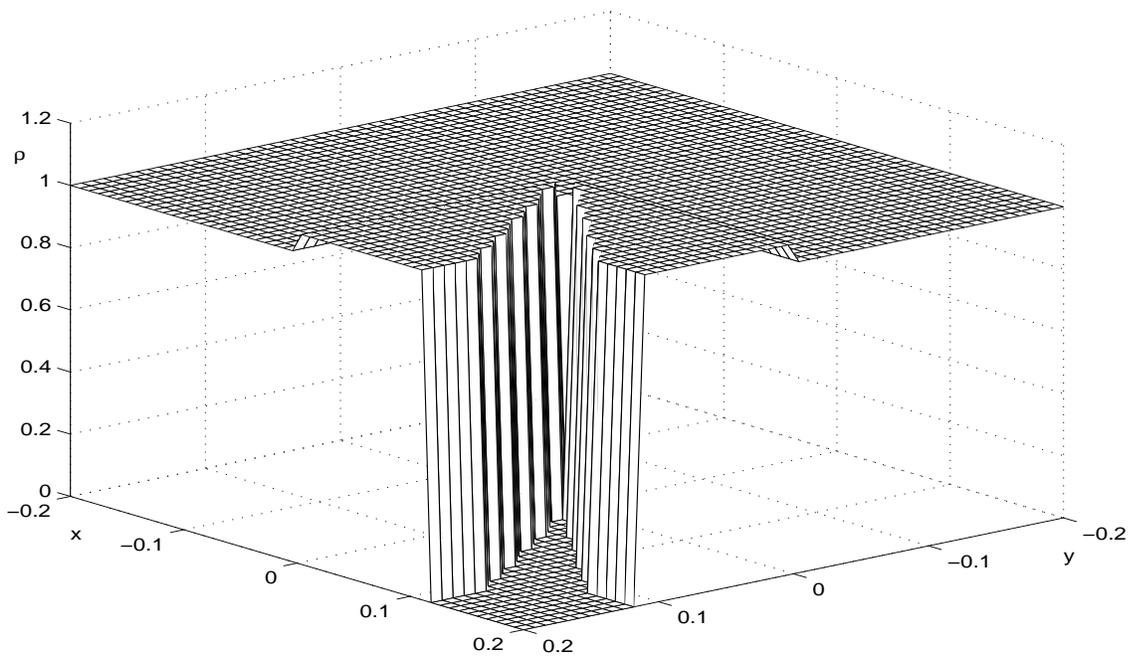


Figure 6.9 Example 6.3, exact density at  $t = 0.4$  on  $50^2$  space mesh.

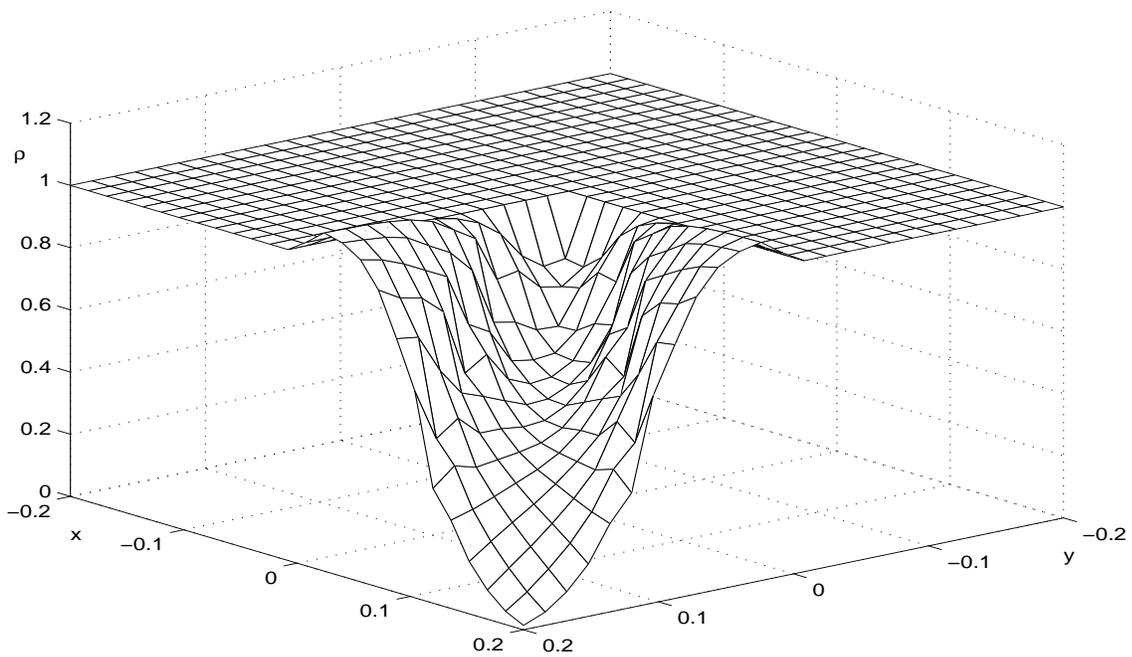


Figure 6.10 Example 6.3, numerical density at  $t = 0.4$  using  $26^4$  phase space mesh.

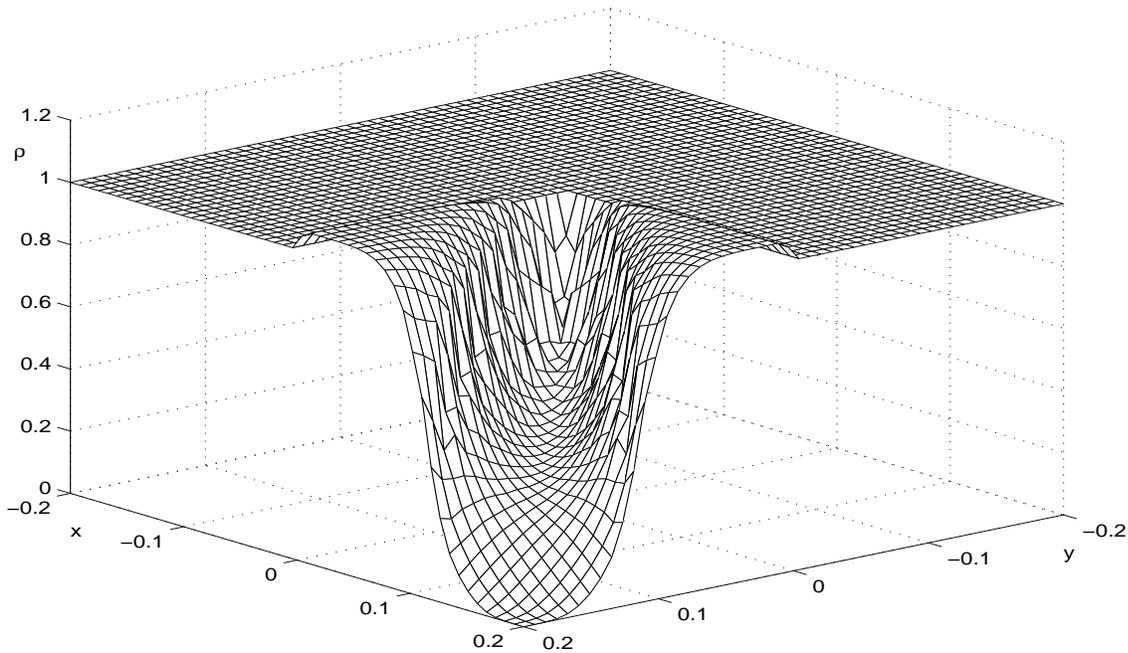


Figure 6.11 Example 6.3, numerical density at  $t = 0.4$  using  $50^4$  phase space mesh.

## Appendix A

This Appendix gives the exact multivalued velocity profile and density at  $t = 1$  for the problem in example 6.2. Note the multivalued velocity  $\omega$  is the common zeroes of  $\psi_i$  defined in the introduction, while the averaged velocity  $u$  is given by (1.8).

- In the domain  $-1.5 < x < -1$ ,  $\omega(x)$  is single phased given by  $\omega(x) = 0.8$  and the corresponding density is the constant 1.
- In the domain  $-1 < x < 0$ ,  $\omega$  has two phases. Set

$$x_1 = \frac{\ln((e-1)x + e) - 1}{e-1} - 1,$$

$$x_2 = \left(\frac{1}{2} + \frac{1}{e-1}\right) \left(1 - \frac{1}{(e-1)\left(\frac{e}{e-1} + x\right)}\right),$$

then

$$\omega_i(x) = \frac{w(x_i)c(x_i)}{c(x)}, \quad i = 1, 2.$$

The densities are given by

$$\rho_1(x) = \frac{1}{(e-1)x + e},$$

$$\rho_2(x) = \frac{\frac{1}{2} + \frac{1}{e-1}}{(e-1)\left(\frac{e}{e-1} + x\right)^2}.$$

- In the domain  $0 < x < \frac{e+1}{2e}$ ,  $\omega$  has two phases. Set

$$\begin{aligned} x_1 &= \frac{1}{(e-1)\left(1 - \frac{2x(e-1)}{e+1}\right)} - \frac{e}{e-1}, \\ x_2 &= \frac{e+1}{2e} + \frac{x}{e}, \end{aligned}$$

then

$$\omega_i(x) = \frac{w(x_i)c(x_i)}{c(x)}, \quad i = 1, 2.$$

The densities are given by

$$\begin{aligned} \rho_1(x) &= \frac{2}{(e+1)\left(1 - \frac{2x(e-1)}{e+1}\right)^2}, \\ \rho_2(x) &= \frac{1}{e}. \end{aligned}$$

- In the domain  $\frac{e+1}{2e} < x < 1$ ,  $\omega$  is single phased. Set

$$x_1(x) = \begin{cases} \frac{e+1}{2e} + \frac{x}{e} & x < \frac{1}{2}(e-1) \\ 1 + \frac{3-e}{2(e-1)}\left(1 - \ln\left(\frac{2(e-1)}{3-e}\left(\frac{e+1}{2(e-1)} - x\right)\right)\right) & x \geq \frac{1}{2}(e-1) \end{cases},$$

then

$$\omega(x) = \frac{w(x_1)c(x_1)}{c(x)}$$

with the corresponding density

$$\rho(x) = \begin{cases} \frac{1}{e} & x < \frac{1}{2}(e-1) \\ \frac{3-e}{2(e-1)\left(\frac{e+1}{2(e-1)} - x\right)} & x \geq \frac{1}{2}(e-1) \end{cases}.$$

- In the domain  $1 < x < 1.5$ ,  $\omega$  is single phased given by

$$\omega(x) = \begin{cases} -0.8 + \frac{0.8}{(1.5)^2}\left(x - 2 + \frac{1}{e-1}\right)^2 & 1 < x < 2 - \frac{1}{e-1} \\ -0.8 & 2 - \frac{1}{e-1} < x < 1.5 \end{cases},$$

the corresponding density is the constant 1.

## 7 Conclusion

In this paper, we construct and study two classes of Hamiltonian-preserving schemes for the Liouville equation arising in the phase space description of geometrical optics. These schemes are effective when the local wave speed is discontinuous, corresponding to different media. These schemes have a hyperbolic CFL condition, which is a significant improvement over a conventional discretization. The main idea is to build in the wave behavior at the interface—which conserves the Hamiltonian—into the numerical flux, as was previously done in [28, 22]. This gives a selection criterion on the choice of unique weak solution to this linear hyperbolic equation with singular coefficients. It allows the wave to be transmitted obeying Snell’s law of refraction, or be reflected. We established stability theory of these discretizations, and conducted numerical experiments to study the numerical accuracy.

In multidimension, we have presented the scheme only in the simple case when a plane wave hits the interface that aligns with the grids, and when the reflection and transmission of waves do not occur simultaneously. For the more general cases of both reflection and transmissions, curved interface, etc. the principle of Hamiltonian-preserving can still be used, however, a different construction of numerical flux at the interface is needed. In addition, the same idea can also be extended to problems with external fields, such as the electrical or electromagnetic fields. There Vlasov-Poisson or Vlasov-Maxwell systems arise. Currently we are exploring the Hamiltonian-preserving schemes in these more general applications.

## References

- [1] G. Bal, J.B. Keller, G. Papanicolaou and L. Ryzhik, Transport theory for acoustic waves with reflection and transmission at interfaces, *Wave Motion*, 30, 303-327 (1999).
- [2] J.-D. Benamou, Big Ray Tracing: Multivalued Travel Time Field Computation Using Viscosity Solutions of the Eikonal Equation, *J. Comput. Phys.* 128, 463-474 (1996).
- [3] J.-D. Benamou, Direct computation of multivalued phase space solutions for Hamilton-Jacobi equations, *Comm. Pure Appl. Math.* 52(11), 1443-1475 (1999).
- [4] J.-D. Benamou, An introduction to Eulerian geometrical optics(1992-2002), *J. Sci. Comp.* 19(1-3), 63-93, Dec. (2001).
- [5] J.-D. Benamou and I. Sollic, An Eulerian Method for Capturing Caustics, *J. Comput. Phys.* 162(1), 132-163 (2000).

- [6] F. Bouchut and F. James, One-dimensional transport equations with discontinuous coefficients, *Nonlinear Analysis, Theory, Methods and Applications* 32, 891-933, 1998.
- [7] L.-T. Cheng, H.-L. Liu and S. Osher, Computational high-frequency wave propagation using the Level Set method, with applications to the semi-classical limit of Schrödinger equations, *Comm. Math. Sci.* 1, 593-621, 2003.
- [8] I. Capuzzo Dolcetta and B. Perthame, On some analogy between different approaches to first order PDEs with non-smooth coefficients, *Adv. Math. Sci. Appl.*, (1996).
- [9] B. Engquist and O. Runborg, Multi-phase computations in geometrical optics, *J. Comput. Appl. Math.* 74, 175-192 (1996).
- [10] B. Engquist and O. Runborg, Multiphase computations in geometrical optics, In M. Fey and R. Jeltsch, editors, *Hyperbolic Problems: Theory, Numerics, Applications*, volume 129 of *Internat. Ser. Numer. Math.*, Zrich, Switzerland, 1998. ETH Zentrum.
- [11] B. Engquist and O. Runborg, Computational high frequency wave propagation, *Acta Numerica* 12, 181-266 (2003).
- [12] B. Engquist, O. Runborg and A.-K. Tornberg, High-Frequency Wave Propagation by the Segment Projection Method, *J. Comput. Phys.* 178(2), 373-390 (2002).
- [13] E. Fatemi, B. Engquist and S. Osher, Numerical solution of the high frequency asymptotic expansion for the scalar wave equation, *J. Comput. Phys.* 120, 145-155 (1995).
- [14] S. Fomel and J.A. Sethian, Fast Phase Space Computation of Multiple Arrivals, *Proc. Natl. Acad. Sci. USA* 99(11), 7329-7334 (2002).
- [15] L. Gosse, Using K-branch entropy solutions for multivalued geometric optics computations, *J. Comp. Phys.* 180, 155-182, 2002.
- [16] L. Gosse, Multiphase semiclassical approximation of an electron in a one-dimensional crystalline lattice II. Impurities, confinement and Bloch oscillations, *J. Comp. Phys.* 201, 344-375, 2004.
- [17] L. Gosse and F. James, Numerical approximations of one-dimensional linear conservation equations with discontinuous coefficients, *Math. Comp.* 987-1015, 2000.
- [18] L. Gosse, S. Jin and X.T. Li, On Two Moment Systems for Computing Multiphase Semiclassical Limits of the Schrodinger Equation *Math. Model Methods Appl. Sci.* 13, 1689-1723, 2003.

- [19] S. Jin, H.L. Liu, S. Osher and R. Tsai, Computing multivalued physical observables for the semiclassical limit of the Schrodinger equation, *J. Comp. Phys.*, in press.
- [20] S. Jin, H.L. Liu, S. Osher and R. Tsai, Computing multi-valued physical observables for high frequency limit of symmetric hyperbolic systems, *J. Comp. Phys.*, submitted.
- [21] S. Jin and S. Osher, A level set method for the computation of multi-valued solutions to quasi-linear hyperbolic PDE's and Hamilton-Jacobi equations, *Comm. Math. Sci.* 1(3), 575-591 (2003).
- [22] S. Jin and X. Wen, Hamiltonian-preserving schemes for the Liouville equation with discontinuous potentials, submitted to *Communications in Mathematical Sciences*.
- [23] N.N. Kuznetsov, On stable methods for solving non-linear first order partial differential equations in the class of discontinuous functions, *Topics in Numerical Analysis III (Proc. Roy. Irish Acad. Conf.)*(J.J.H.Miller, ad.), Academic Press, London, 183-197 (1977).
- [24] R.J. LeVeque, *Numerical Methods for Conservation Laws*, Birkhauser-Verlag, Basel, 1990.
- [25] L. Miller, Refraction of high frequency waves density by sharp interfaces and semiclassical measures at the boundary, *J. Math. Pures Appl.* IX 79, 227-269, (2000).
- [26] S. Osher, L.-T. Cheng, M. Kang, H. Shim and Y.-H. Tsai, Geometric optics in a phase-space-based level set and Eulerian framework, *J. Comput. Phys.* 179(2), 622-648 (2002).
- [27] B. Perthame and C.W. Shu, On positivity preserving finite volume schemes for Euler equations, *Numer. Math.* 73, 119-130 (1996).
- [28] B. Perthame and C. Simeoni, A kinetic scheme for the Saint-Venant system with a source term, *CALCOLO* 38(4), 201-231 (2001).
- [29] G. Petrova and B. Popov, Linear transport equations with discontinuous coefficients, *J. Math. Anal. Appl.* 260, 307-324, (2001).
- [30] F. Poupand and M. Rascle, Measure solutions to the linear multidimensional transport equation with non-smooth coefficients, *Comm. PDEs* 22, 337-358, 1997.
- [31] O. Runborg, Some new results in multiphase geometrical optics, *M2AN Math. Model. Numer. Anal.* 34, 1203-1231 (2000).

- [32] L. Ryzhik, G. Papanicolaou and J. Keller, Transport equations for elastic and other waves in random media, *Wave Motion* 24, 327-370 DEC (1996).
- [33] L. Ryzhik, G. Papanicolaou and J. Keller, Transport equations for waves in a half space, *Comm. PDE's*, 22, 1869-1910 (1997).
- [34] C.W. Shu and S. Osher, Efficient implementation of essentially non-oscillatory shock capturing scheme, *J. Comput. Phys.* 77, 439-471 (1988).
- [35] W. Symes and J. Qian, A slowness matching Eulerian method for multivalued solutions of Eikonal equations, *J. Sci. Comp.* 19(1-3), 501-526, (2003). Special issue in honor of the sixtieth birthday of Stanley Osher.
- [36] T. Tang and Z.H. Teng, The sharpness of Kuznetsov's  $O(\sqrt{\Delta x})$   $L^1$ -error estimate for monotone difference schemes, *Math. Comp.* 64, 581-589 (1995).