UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Multi-Phase Computations in Geometrical Optics

Bjorn Engquist
Olof Runborg

November 1995

CAM Report 95-45

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

Multi-Phase Computations in Geometrical Optics

Björn Engquist* Olof Runborg†

1 Introduction

In the direct calculation of wave propagation, the computational effort is larger at higher frequencies. With constant accuracy the work grows algebraically with frequency. For sufficiently high frequencies or short wavelengths it is unrealistic to compute the wave field directly. Fortunately, this is often the regime for which high frequency asymptotic approximations are quite accurate.

Generically, phase and amplitude vary on a much slower scale than the dependent variables in the original wave equations and are thus in principle easier to compute. The geometrical optics type asymptotic expansions are used in many applications, for example in electromagnetic, elastic and acoustic wave propagation.

Traditionally, ray tracing has been the computational method of choice. Recently, however, the geometrical optics approximations are also being solved by partial differential equation (PDE) techniques. This is e.g. done in [5] and within the framework of seismology in [8], [10] and [11]. The PDEs give only one unique phase at each point in space. In this chapter we shall derive equations which allow for multiple phases or crossing

*Department of Mathematics, UCLA Los Angeles and NADA, KTH Stockholm.
Research supported by the grants ARPA/ONR [N00014-92-J-1890], ONR, [N00014-91-J-1034], NSF, [DMS-91-03104].
†NADA, KTH Stockholm

rays. The equations are based on the closure assumption of a finite number of crossing rays for the kinetic formulation of geometrical optics.

High Frequency Asymptotics

When high frequency waves are treated, the computations can be simplified by considering the asymptotic behavior of the solution as the frequency tends to infinity. There are two strongly related ways to formulate this approximation: the PDEs of geometrical optics and ray tracing. Typical wave phenomena, such as diffraction and interference, are lost in the leading terms of the high frequency approximation.

Classical geometrical optics is based on the scalar wave equation,

$$u_{tt} + c\nabla^2 u = 0. \tag{1}$$

Here $c = c(\boldsymbol{x})$ is the local wave velocity of the medium. We also define the *index of refraction* as $\eta = c_0/c$ with the reference velocity c_0 (e.g. the speed of light in vacuum). Geometrical optics considers the case when the solution to (1) can be written as a series expansion of the form:

$$u = e^{i\omega\phi(\boldsymbol{x},t)} \sum_{k=0}^{\infty} w_k(\boldsymbol{x},t)(i\omega)^{-k}. \tag{2}$$

Entering this expression into (1) and summing terms of the same order in ω, to zero, we obtain separate equations for the unknown variables in (2). The phase function ϕ will satisfy the *eikonal equation*,

$$\phi_t + c|\nabla\phi| = 0, \tag{3}$$

and the amplitude coefficients w_k solve the *transport equations*,

$$(w_0)_t + c\frac{\nabla\phi\cdot\nabla w_0}{|\nabla\phi|} + \frac{c^2\nabla^2\phi - \phi_{tt}}{2c|\nabla\phi|}w_0 = 0, \tag{4}$$

$$(w_{k+1})_t + c\frac{\nabla\phi\cdot\nabla w_{k+1}}{|\nabla\phi|} + \frac{c^2\nabla^2\phi - \phi_{tt}}{2c|\nabla\phi|}w_{k+1} + \frac{c^2\nabla^2 w_k - (w_k)_{tt}}{2c|\nabla\phi|} = 0. \tag{5}$$

For large ω only the first term in the expansion (2) is significant, and the problem is reduced to computing the phase ϕ and the first amplitude

term w_0. Note that once ϕ is known, the transport equations are linear equations with variable coefficients. Solving (3) and (4) can be done by finite difference methods.

The problem with the geometrical optics approach is that the class of solutions which justify an expansion of the type (2), is limited. In particular, it does not include solutions with multiple phases, corresponding to crossing waves. In fact, even in the case of a single phase solution, the series does not necessarily converge, for instance when the geometric boundaries create diffraction effects. We shall concentrate on the multiple phase problem and assume the geometrical optics approximations of (3) and (4).

The eikonal equation is a nonlinear PDE which requires extra conditions to have a unique solution. This solution is known as the *viscosity solution* [3]. Of course, it does not have to agree with the correct physical solution in all cases. At points where the correct solution should have a multivalued phase, the viscosity solution picks out the phase corresponding to the first arriving wave.

The eikonal equation's inability to capture multi-phase solutions is related to its nonlinear character. In the case of the linear wave equation, that it approximates, a linear combination of solutions is also a solution. For the nonlinear eikonal equation, this *superposition principle* does not hold. An example is shown in Figure 1.

Solving the eikonal equation numerically as a PDE instead of using ray tracing has recently been used in seismology. This technique is demonstrated in [8], [10] and [11]. For these applications it is of direct interest to determine the first arrival.

A second phase, corresponding to crossing rays was calculated in [5] using two separate eikonal equations. Boundary conditions for the second phase was given at the discontinuity of the first phase or at a geometric reflecting boundary. This boundary could be difficult to determine.

Another way to treat high frequency waves computationally is through ray tracing, which is based on a kinetic formulation. The waves are postulated to be particles (photons) whose trajectories are rays. The *ray vector*, $\boldsymbol{p}$, is defined as the index of refraction multiplied by the unit vector, $\hat{\boldsymbol{s}}$, in the direction of the ray, i.e. $\boldsymbol{p} = \eta\hat{\boldsymbol{s}}$. For simplicity we

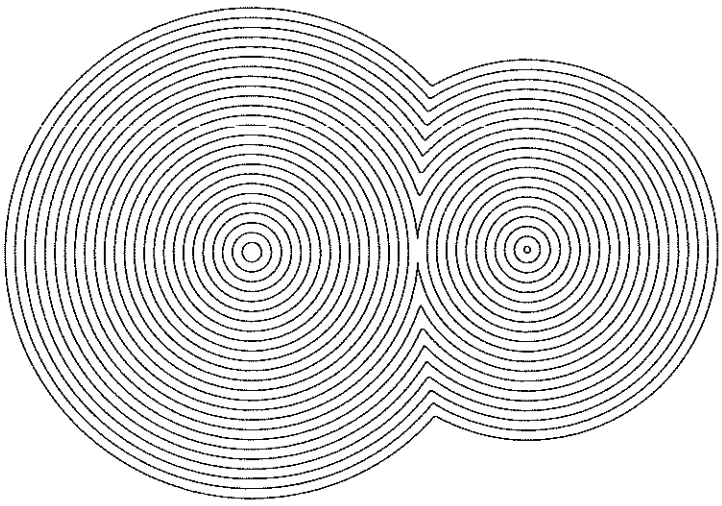

Figure 1: Level curves of ϕ in the solution of the eikonal equation (3) for two interacting waves. Note how the superposition principle does not hold. Instead, the first arriving wave takes precedence over the second at each point.

will henceforth let $c_0 = 1$, so that the velocity vector $\boldsymbol{v} = c\hat{\boldsymbol{s}} = c^2\boldsymbol{p}$. A transport equation for particles in the space $(\boldsymbol{x}, \boldsymbol{p}, t)$ can then be derived. Denoting the density of particles by $f(\boldsymbol{x}, \boldsymbol{p}, t)$ the evolution of f is described by the Vlasov type equation

$$f_t + \boldsymbol{v} \cdot \boldsymbol{\nabla}_x f + c\boldsymbol{\nabla}_x \eta \cdot \boldsymbol{\nabla}_p f = 0. \tag{6}$$

Tracing the particle trajectories of (6) corresponds to ray tracing and also to the method of characteristics for (3) and (6). Since (6) is linear the superposition principle is valid.

Because of the large number of independent variables (six in 3D) it is very hard numerically to solve the full equation (6). If the equation

is solved using ray tracing it is difficult to cover the full domain with rays, [10]. There will often be shadow zones where the field cannot be resolved. It is also hard to determine the derivative of ϕ, which is needed when computing the amplitude.

Moment Formulation

In this chapter we propose a middle way between geometrical optics and the kinetic model. It is a high frequency approximation through which the whole field can be solved. Moreover, the superposition principle holds up to a point; the maximum allowed number of intersecting waves can be chosen arbitrarily, but a higher number means that a larger system of PDEs must be solved. The technique we use to capture multivalued solutions is based on a closure assumption for a system of equations representing the moments (see [1]).

The starting point for this approach is the transport equation (6). Instead of solving the full equation in phase space, we observe that when f is of a simple form in $\boldsymbol{p}$, we can transform (6) to a finite system of moment equations in the reduced space $(\boldsymbol{x}, t)$, analogously to the classical approach of the hydrodynamic limit from a kinetic formulation. In particular we are interested in cases where, for given $\boldsymbol{x}$ and t, the density function f is non-zero only for a finite number of $\boldsymbol{p}$. This corresponds to a finite number of rays in different directions at each point.

This chapter is organized as follows: In Section 2 the moment equations are derived from the kinetic model for high frequency waves. They are equivalent to the equations of geometrical optics. We also explore some theoretical issues and find that the resulting hyperbolic equations are not well-posed in the strong sense. Existence of solutions of unbounded variation is indicated. Next, in Section 3, we describe the numerical approximations we have used to solve these equations for one and two phases. The standard Lax-Friedrich method gives satisfactory results. More elaborate, and less viscous, methods like the Godunov method and the second order TVD Nessyahu-Tadmor scheme, although converging well in L_1 suffer from problems locally and converge poorly in L_∞. For the two phase system, the sensitivity of the equations is

more pronounced and consequently it is harder to find stable numerical methods. After proper initialization, the equations can however be solved with the Lax-Friedrich method. We present computational results in Section 4.

2 Derivation of the Moment Equations

In this section we will derive the system of PDEs that follows from the kinetic model and the assumption that a maximum of N rays pass through any given point in space. The analysis is carried out in two-dimensional space.

The derivation of the moment equations is based on the transport equation (6). This equation comes from the Hamiltonian system

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{\nabla}_p H(\boldsymbol{x}, \boldsymbol{p}, t), \qquad \frac{d\boldsymbol{p}}{dt} = -\boldsymbol{\nabla}_x H(\boldsymbol{x}, \boldsymbol{p}, t), \tag{7}$$

where the corresponding density function $f(\boldsymbol{x}, \boldsymbol{p}, t)$ solves

$$f_t + \boldsymbol{\nabla}_x \cdot (f \boldsymbol{\nabla}_p H) - \boldsymbol{\nabla}_p \cdot (f \boldsymbol{\nabla}_x H) = 0. \tag{8}$$

The equation (6) follows from (8) when

$$H = \frac{|\boldsymbol{p}|^2}{2\eta^2}. \tag{9}$$

The Moment Equations

We start by defining the moments m_{ij}. With $\boldsymbol{p} = (p_1, p_2)$, let

$$m_{ij} = \int_{\mathbf{R}^2} p_1^i p_2^j f \, d\boldsymbol{p}. \tag{10}$$

Next, we multiply (6) by $p_1^i p_2^j$ and integrate over $\mathbf{R}^2$ with respect to $\boldsymbol{p}$. Using the definition (10) we get the following moment equation:

$$\eta^2 (m_{ij})_t + (m_{i+1,j})_x + (m_{i,j+1})_y = i\eta\eta_x m_{i-1,j} + j\eta\eta_y m_{i,j-1}, \tag{11}$$

where we have used the fact that f has compact support in $\boldsymbol{p}$. Since this equation is valid for all $i, j \geq 0$, we have an infinite system of moment equations. For uniformity in notation we have defined $m_{i,-1} = m_{-1,i} = 0$, $\forall i$.

The system (11) is not closed. If truncated at finite i and j, there are more unknown than equations. To close (11) we use the assumption that for fixed values of $\boldsymbol{x}$ and t, the particle density f is non-zero at a maximum of N points, and only when $|\boldsymbol{p}| = \eta(\boldsymbol{x})$. Thus f can be written

$$f(\boldsymbol{x}, \boldsymbol{p}, t) = \sum_{k=1}^{N} g_k \cdot \delta(|\boldsymbol{p}| - \eta, \arg \boldsymbol{p} - \theta_k). \tag{12}$$

The new variables that we have introduced here are $g_k = g_k(\boldsymbol{x}, t)$, which corresponds to the strength (particle density) of ray k, and $\theta_k = \theta_k(\boldsymbol{x}, t)$ which is the direction of the same ray. Inserting (12) into (10) yields

$$m_{ij} = \sum_{k=1}^{N} \eta^{i+j} g_k \cos^i \theta_k \sin^j \theta_k, \tag{13}$$

which is the expression for the moments that we will use.

A system describing N phases, needs $2N$ equations, corresponding to the N ray strengths g_k and their directions θ_k. It is not immediately clear which equations to select among the candidates in (11). Given the equations for a set of $2N$ moments, it must be possible to write the remaining moments of these equations in terms of the leading ones. This is not always true. For instance, with the choice of m_{20} and m_{02}, for $N = 1$, the quadrant of the angle θ cannot be recovered, and therefore in general not the sign of the moments.

We choose here the equations for $m_{k,0}$ and $m_{0,k}$ with $k = 1, \ldots, N$. This system can be closed for $N = 1, 2$. After scaling the moments, $\eta^{i+j} \tilde{m}_{ij} = m_{ij}$, those equations take the following form:

$$\begin{aligned} &\eta^2 (\tilde{m}_{k,0})_{\bar{t}} + (\eta \tilde{m}_{k+1,0})_x + (\eta \tilde{m}_{k,1})_y \\ &\qquad = k(\eta_x \tilde{m}_{k-1,0} - \eta_x \tilde{m}_{k+1,0} - \eta_y \tilde{m}_{k,1}), \end{aligned} \tag{14}$$

$$\begin{aligned} &\eta^2 (\tilde{m}_{0,k})_{\bar{t}} + (\eta \tilde{m}_{1,k})_x + (\eta \tilde{m}_{0,k+1})_y \\ &\qquad = k(\eta_y \tilde{m}_{0,k-1} - \eta_x \tilde{m}_{1,k} + \eta_y \tilde{m}_{0,k+1}). \end{aligned} \tag{15}$$

To simplify notation, we will henceforth write m_{kl} for $\tilde{m}_{kl}$.

We introduce new variables,

$$\begin{aligned}\boldsymbol{u} &= (u_1, u_2, u_3, u_4 \ldots, u_{2N-1}, u_{2N})^T \\ &:= (g_1 \cos\theta_1, g_1 \sin\theta_1, g_2 \cos\theta_2, g_2 \sin\theta_2, \ldots, g_N \cos\theta_N, g_N \sin\theta_N)^T. \quad (16)\end{aligned}$$

These variables have a physical interpretation; the vector (u_{2k-1}, u_{2k}) shows the direction and strength of ray k.

To write out the equations for $\boldsymbol{u}$ in a concise form we need a few definitions. Let $\boldsymbol{I}$ be the 2×2 identity matrix, and

$$\boldsymbol{D}_k = \begin{pmatrix} \cos\theta_k & 0 \\ 0 & \sin\theta_k \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{I} & \cdots & \boldsymbol{I} \\ \boldsymbol{D}_1 & \boldsymbol{D}_2 & \cdots & \boldsymbol{D}_N \\ \boldsymbol{D}_1^2 & \boldsymbol{D}_2^2 & \cdots & \boldsymbol{D}_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{D}_1^{N-1} & \boldsymbol{D}_2^{N-1} & \cdots & \boldsymbol{D}_N^{N-1} \end{pmatrix}. \quad (17)$$

Moreover,

$$\begin{aligned}\boldsymbol{D} &= \mathrm{diag}(\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_N), && (18)\\ \boldsymbol{C} &= \mathrm{diag}(\cos\theta_1 \boldsymbol{I}, \cos\theta_2 \boldsymbol{I}, \ldots, \cos\theta_N \boldsymbol{I}), && (19)\\ \boldsymbol{S} &= \mathrm{diag}(\sin\theta_1 \boldsymbol{I}, \sin\theta_2 \boldsymbol{I}, \ldots, \sin\theta_N \boldsymbol{I}), && (20)\\ \boldsymbol{T} &= \mathrm{diag}(\boldsymbol{I}, 2\boldsymbol{I}, \ldots, N\boldsymbol{I}), && (21)\\ \boldsymbol{R} &= \mathrm{diag}(\underbrace{\eta_x, \eta_y, \ldots, \eta_x, \eta_y}_{2N \text{ elements}}). && (22)\end{aligned}$$

These definitions let us write our PDEs as a system of nonlinear conservation laws with source terms,

$$\eta^2(\boldsymbol{A}\boldsymbol{u})_t + (\eta \boldsymbol{A}\boldsymbol{C}\boldsymbol{u})_x + (\eta \boldsymbol{A}\boldsymbol{S}\boldsymbol{u})_y = \boldsymbol{T}\boldsymbol{A}(\boldsymbol{R}\boldsymbol{D}^{-1} - \eta_x \boldsymbol{C} - \eta_y \boldsymbol{S})\boldsymbol{u}. \quad (23)$$

Note that $\boldsymbol{R}$ and $\boldsymbol{A}$ commute, $\boldsymbol{R}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{R}$.

A Comparison with Geometrical Optics

To see how the moment equations (23) are equivalent formulations of the equations of geometrical optics, we present the following derivation for smooth solutions.

The additional definitions

$$\boldsymbol{\Theta} = \operatorname{diag}(\theta_1 \boldsymbol{I}, \theta_2 \boldsymbol{I}, \ldots, \theta_N \boldsymbol{I}), \tag{24}$$

$$\boldsymbol{g} = (\underbrace{g_1, g_1, g_2, g_2, \ldots, g_N, g_N}_{2N \text{ elements}})^T, \tag{25}$$

will help us write a separated form of (23). Since each element of the matrices only depend on one variable (one of the θ_ks), we can let the prime sign, $'$, denote elementwise differentiation of a matrix. Using the identities $(\boldsymbol{AD})' = \boldsymbol{TAD}'$ and $\boldsymbol{R} = (\eta_x \boldsymbol{D} + \eta_y \boldsymbol{D}')\boldsymbol{C} + (\eta_y \boldsymbol{D} - \eta_x \boldsymbol{D}')\boldsymbol{S}$, the equations (23) can be written as

$$\begin{aligned} &\boldsymbol{AD}(\eta^2 \boldsymbol{g}_t + (\eta \boldsymbol{C}\boldsymbol{g})_x + (\eta \boldsymbol{S}\boldsymbol{g})_y) \\ &\qquad + (\boldsymbol{AD})'(\eta^2 \boldsymbol{\Theta}_t + (\eta \boldsymbol{S})_x - (\eta \boldsymbol{C})_y)\boldsymbol{g} = 0. \end{aligned} \tag{26}$$

Noting that $\boldsymbol{C}$, $\boldsymbol{S}$ and $\boldsymbol{\Theta}$ are diagonal and that they all, together with $\boldsymbol{g}$, have their elements ordered pairwise, a solution is given by solving the N separated systems

$$\begin{aligned} \eta^2(\theta_k)_t + (\eta \sin\theta_k)_x - (\eta \cos\theta_k)_y &= 0, \\ \eta^2(g_k)_t + (\eta g_k \cos\theta_k)_x + (\eta g_k \sin\theta_k)_y &= 0. \end{aligned} \qquad k = 1, \ldots, N. \tag{27}$$

On the other hand, after some algebraic manipulations of (3) and (4) we get

$$\begin{aligned} \eta^2 \tfrac{\partial}{\partial t} \arctan\left(\tfrac{\phi_y}{\phi_x}\right) + \tfrac{\partial}{\partial x}\left(\eta \tfrac{\phi_y}{|\boldsymbol{\nabla}\phi|}\right) - \tfrac{\partial}{\partial y}\left(\eta \tfrac{\phi_x}{|\boldsymbol{\nabla}\phi|}\right) &= 0, \\ \eta^2 \tfrac{\partial}{\partial t} w_0^2 + \tfrac{\partial}{\partial x}\left(\eta w_0^2 \tfrac{\phi_x}{|\boldsymbol{\nabla}\phi|}\right) + \tfrac{\partial}{\partial y}\left(\eta w_0^2 \tfrac{\phi_y}{|\boldsymbol{\nabla}\phi|}\right) &= 0. \end{aligned} \tag{28}$$

If we identify the variables of (23) as

$$g_k = w_{0,k}^2, \qquad \cos\theta_k = \frac{(\phi_k)_x}{|\boldsymbol{\nabla}\phi|}, \qquad \sin\theta_k = \frac{(\phi_k)_y}{|\boldsymbol{\nabla}\phi|}, \qquad k = 1, \ldots, N, \tag{29}$$

they will solve (27) and hence (23).

Analysis of the Conservation Laws

For simplicity we will mainly deal with the single phase, $N = 1$, one-dimensional case where the medium is vacuum, $\eta = \text{const} = 1$. This

system reads:

$$\begin{aligned} (g\cos\theta)_t + (g\cos^2\theta)_x &= 0, \\ (g\sin\theta)_t + (g\cos\theta\sin\theta)_x &= 0. \end{aligned} \tag{30}$$

With the $\boldsymbol{u}$ variables defined in (16) as conserved quantities, the system can be written on the standard form of a conservation law,

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0, \qquad \boldsymbol{f}(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1^2}{\sqrt{u_1^2+u_2^2}} \\ \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} \end{pmatrix}. \tag{31}$$

The Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{u}$ has the following form:

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}} = \boldsymbol{G}\begin{pmatrix} \cos\theta & -\sin\theta \\ 0 & \cos\theta \end{pmatrix}\boldsymbol{G}^{-1}, \qquad \boldsymbol{G} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{32}$$

Thus, the linearized problem has a double real eigenvalue, $\cos\theta$, and an incomplete set of eigenvectors; the system (31) is only *weakly hyperbolic*. In general this means that (31) is not well-posed in the strongly hyperbolic sense. The system is likely to be much more sensitive than regular hyperbolic systems. The solution of the linearized problem with frozen coefficients loses one derivative. The L_2 norm of the solution at time $t > t_0$ can be estimated in terms of the H_1 norm of the initial data at time $t = t_0$. The sensitivity of the equations is reflected in difficulties in finding stable numerical methods to solve them (see Section 3). If the solution has a shock, the double eigenvalue means that there are always two characteristics incident to the shock at each side. Shocks are thus overcompressive.

The existence of solutions to (31) is also an open question. It appears that solutions cannot be expected to be of bounded variation. In fact, analytic and numerical evidence suggest that (31) can have measure valued solutions, i.e. of delta function type (cf. Figure 5). An extended solution concept is needed to accommodate measure valued solutions. This problem was addressed in [2] and [4], where also existence of such solutions for certain conservation laws was proved. Entropy conditions and uniqueness of solutions to (31) are even more uncertain.

The appearance of a delta function is closely linked to when the physically correct solution passes outside the class of solutions that the system (23) describe. If initial data dictates a physical solution with N phases for $t > T$, the system (23) with $M < N$ phases will have a measure valued solution for $t > T$. In the case of (31), a delta function will appear in the solution when multiple phases are present.

The statements above are supported by our numerical simulations. We will consider a one-dimensional example. In vacuum, $\eta = 1$, the separated system (27) can be rewritten as:

$$\begin{array}{rcll} (\theta_k)_t + (\sin\theta_k)_x & = & 0, & \\ (g_k)_t + \cos\theta_k (g_k)_x & = & g_k (\sin\theta_k)_x, & k = 1, \ldots, N. \end{array} \qquad (33)$$

The equation for θ_k is known to develop shocks in finite time. The angle θ_k will be constant along characteristics, which are straight lines corresponding to rays. The shock develops where characteristics cross, i.e. where two wave fields meet. The equation for g_k is an ordinary transport equation with a source term involving the derivative of $\sin\theta_k$. Along characteristics, which are the same for both equations, the source term is zero, except at a shock where it becomes a delta function. The resulting solution for g_k is a delta function where the phase "should" have split into two new phases.

It is interesting again to compare the moment equations with the eikonal and transport equations, (3) and (4). The latter also form a weakly hyperbolic system with the same eigenvalue as (31). As was mentioned before, the viscosity solution picks out the phase corresponding to first arrival where the physically correct phase is multivalued. When wave fields meet, there will therefore in general be a jump in ϕ. At these points the first amplitude coefficient w_0 has a measure. Hence, the two different formulations are similar also in this respect.

For the two-dimensional case, another function, $\boldsymbol{g}$, is added to (31),

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x + \boldsymbol{g}(\boldsymbol{u})_y = 0, \qquad \boldsymbol{g}(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} \\ \frac{u_2^2}{\sqrt{u_1^2+u_2^2}} \end{pmatrix}. \qquad (34)$$

Taking a linear combination of the Jacobians for $\boldsymbol{f}$ and $\boldsymbol{g}$ we get

$$\begin{aligned}\boldsymbol{J}(\theta,\alpha_1,\alpha_2) &:= \alpha_1\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}} + \alpha_2\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}} \\ &= \boldsymbol{G}\left[\alpha_1\begin{pmatrix}\cos\theta & -\sin\theta\\ 0 & \cos\theta\end{pmatrix} + \alpha_2\begin{pmatrix}\sin\theta & \cos\theta\\ 0 & \sin\theta\end{pmatrix}\right]\boldsymbol{G}^{-1}, \quad (35)\end{aligned}$$

with the same rotation $\boldsymbol{G}$ as in (32). Regardless of the choice of (α_1,α_2), we still only have one eigenvalue and an incomplete set of eigenvectors.

In the general case with N phases the governing equations (23) can be written

$$\boldsymbol{F}_0(\boldsymbol{u})_t + \boldsymbol{F}_1(\boldsymbol{u})_x + \boldsymbol{F}_2(\boldsymbol{u})_y = 0, \quad (36)$$

where $\boldsymbol{F}_k$ are rather complicated functions. (In Section 3 the functions are given explicitly for the case $N = 2$.) Denoting the Jacobians of $\boldsymbol{F}_k$ with $\boldsymbol{J}_k$, the following relationship can be derived:

$$\alpha_1\boldsymbol{J}_1 + \alpha_2\boldsymbol{J}_2 = \boldsymbol{J}_0 \cdot \mathrm{diag}(\boldsymbol{J}(\theta_1,\alpha_1,\alpha_2),\ldots,\boldsymbol{J}(\theta_N,\alpha_1,\alpha_2)). \quad (37)$$

This shows that the eigenvalues of the general system are simply the union of the eigenvalues of N systems of the type (34). It also shows that there will not be more than N eigenvectors, for the $2N$x$2N$ system. Hence, we have shown that the general system (23) is weakly hyperbolic.

3 Numerical Approximations

This section includes some results on the numerical treatment of (23). As was discussed in the last section, the system (23) is very sensitive, and this creates problems for the numerical methods. The sensitivity derives from two facts. Firstly, the system is only weakly hyperbolic. Secondly, since numerical errors can induce extra, unphysical, low amplitude phases, spurious delta functions can appear even where the analytic solution is smooth.

For the numerical methods we will use the following notation. Space and time is discretized uniformly with step sizes Δx, Δy and Δt. The grid function $\boldsymbol{U}_{ij}^n$ approximates the analytic solution,

$$\boldsymbol{U}_{ij}^n \approx \boldsymbol{u}(i\Delta x, j\Delta y, n\Delta t), \quad (38)$$

where $\boldsymbol{u}$ are the variables introduced in (16).

Single Phase

The point of departure for our numerical approximations is the basic first order accurate *Lax-Friedrichs* finite difference method. For the one phase system in a homogeneous media, $\eta = \text{const}$, it takes the form

$$\begin{aligned} \boldsymbol{U}_{ij}^{n+1} = \frac{1}{4}(\boldsymbol{U}_{i-1,j}^{n} + \boldsymbol{U}_{i+1,j}^{n} + \boldsymbol{U}_{i,j-1}^{n} + \boldsymbol{U}_{i,j+1}^{n}) \\ - \frac{\Delta t}{2\Delta x}\left(\boldsymbol{F}_1(\boldsymbol{U}_{i+1,j}^{n}) - \boldsymbol{F}_1(\boldsymbol{U}_{i-1,j}^{n})\right) \\ - \frac{\Delta t}{2\Delta y}\left(\boldsymbol{F}_2(\boldsymbol{U}_{i,j+1}^{n}) - \boldsymbol{F}_2(\boldsymbol{U}_{i,j-1}^{n})\right), \end{aligned} \tag{39}$$

with

$$\boldsymbol{F}_1(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1^2}{\sqrt{u_1^2+u_2^2}} \\ \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} \end{pmatrix}, \qquad \boldsymbol{F}_2(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} \\ \frac{u_2^2}{\sqrt{u_1^2+u_2^2}} \end{pmatrix}. \tag{40}$$

Even if the Lax-Friedrichs method is only of first order, it works quite well and remains stable despite the sensitivity of the equations. Most of our results are produced using this method. The purpose of the numerical experiments is just to show the feasibility of the moment closure technique and for this purpose a first order method is sufficient. The reason for the Lax-Friedrich scheme's stability is that it introduces a substantial amount of viscosity, which implies that discontinuities in the solution are smeared out.

Less smearing of shocks is obtained with the *Godunov* method (see e.g. [6]), another first order method which adds a smaller measure of viscosity than Lax-Friedrichs. The two-dimensional Godunov method is constructed by applying an ordinary splitting approach.

Even though the Godunov method applied to the single phase system converges in L_1 (see Table 1) there are large L_∞ errors also for smooth problems (see Table 2 and Figure 3).

A second order accurate scheme introduces less artificial viscosity and can therefore be expected to be more sensitive. To avoid oscillations at discontinuities, so called *TVD* methods are desirable [6]. These nonlinear methods use limiters to ensure that the method does not introduce new

Δx	Lax-Friedrichs		Godunov		Nessyahu-Tadmor	
	L_1(err)	order	L_1(err)	order	L_1(err)	order
1/10	0.0778		0.1130		0.0480	
		0.85		0.80		0.43
1/20	0.0433		0.0650		0.0357	
		0.92		0.69		0.98
1/40	0.0229		0.0404		0.0181	
		0.96		0.78		1.11
1/80	0.0118		0.0235		0.00839	
		0.98		0.85		1.11
1/160	0.00599		0.0130		0.00390	

Table 1: L_1 norm of the errors for test case A (see Section 4), using the single phase equations.

Δx	Lax-Friedrichs		Godunov		Nessyahu-Tadmor	
	L_∞(err)	order	L_∞(err)	order	L_∞(err)	order
1/10	0.949		3.038		0.278	
		1.26		0.062		0.25
1/20	0.397		2.911		0.235	
		1.21		0.022		0.30
1/40	0.171		2.867		0.191	
		1.15		0.017		1.15
1/80	0.0771		2.834		0.0857	
		1.09		0.010		0.54
1/160	0.0363		2.815		0.0589	

Table 2: L_∞ norm of the errors for test case A (see Section 4), using the single phase equations.

artificial extrema in the solution. At an extrema, TVD methods are at most first order accurate.

We have implemented the *Nessyahu-Tadmor* method with the *min-mod* limiter [7]. It is a second order TVD method based on the Lax-Friedrichs structure. To preserve second order accuracy when moving from the one-dimensional to a two-dimensional method, *Strang splitting*, [9], was used. The result is however not perfectly satisfactory. The on/off switching of the limiter seems to induce oscillations. The convergence rate for the Nessyahu-Tadmor scheme turns out to be somewhat slower than Lax-Friedrichs in L_∞ and only marginally higher in L_1 (see Table 1 and Table 2).

Two Phases

It is more difficult to get reliable calculations when solving (23) with two phases, than in the case of a single phase. The two phase equations add a few new problems to the numerical methods. In each time step a nonlinear system of equations must be solved. With proper initialization of the variables at time $t = 0$, however, it seems possible to compute solutions for most configurations.

In the two phase case, we have only used the Lax-Friedrichs method. With a homogeneous media, it can be written as

$$\begin{aligned} \boldsymbol{F}_0(\boldsymbol{U}_{ij}^{n+1}) = \frac{1}{4}\Big(&\boldsymbol{F}_0(\boldsymbol{U}_{i-1,j}^n) + \boldsymbol{F}_0(\boldsymbol{U}_{i+1,j}^n) \\ &+ \boldsymbol{F}_0(\boldsymbol{U}_{i,j-1}^n) + \boldsymbol{F}_0(\boldsymbol{U}_{i,j+1}^n)\Big) \\ &- \frac{\Delta t}{2\Delta x}\Big(\boldsymbol{F}_1(\boldsymbol{U}_{i+1,j}^n) - \boldsymbol{F}_1(\boldsymbol{U}_{i-1,j}^n)\Big) \\ &- \frac{\Delta t}{2\Delta y}\Big(\boldsymbol{F}_2(\boldsymbol{U}_{i,j+1}^n) - \boldsymbol{F}_2(\boldsymbol{U}_{i,j-1}^n)\Big), \end{aligned} \tag{41}$$

where

$$\boldsymbol{F}_0(\boldsymbol{u}) = \begin{pmatrix} u_1 + u_3 \\ u_2 + u_4 \\ \frac{u_1^2}{\sqrt{u_1^2+u_2^2}} + \frac{u_3^2}{\sqrt{u_3^2+u_4^2}} \\ \frac{u_2^2}{\sqrt{u_1^2+u_2^2}} + \frac{u_4^2}{\sqrt{u_3^2+u_4^2}} \end{pmatrix}, \tag{42}$$

$$\boldsymbol{F}_1(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1^2}{\sqrt{u_1^2+u_2^2}} + \frac{u_3^2}{\sqrt{u_3^2+u_4^2}} \\ \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} + \frac{u_3 u_4}{\sqrt{u_3^2+u_4^2}} \\ \frac{u_1^3}{u_1^2+u_2^2} + \frac{u_3^3}{u_3^2+u_4^2} \\ \frac{u_1 u_2^2}{u_1^2+u_2^2} + \frac{u_3 u_4^2}{u_3^2+u_4^2} \end{pmatrix}, \quad \boldsymbol{F}_2(\boldsymbol{u}) = \begin{pmatrix} \frac{u_1 u_2}{\sqrt{u_1^2+u_2^2}} + \frac{u_3 u_4}{\sqrt{u_3^2+u_4^2}} \\ \frac{u_2^2}{\sqrt{u_1^2+u_2^2}} + \frac{u_4^2}{\sqrt{u_3^2+u_4^2}} \\ \frac{u_1^2 u_2}{u_1^2+u_2^2} + \frac{u_3^2 u_4}{u_3^2+u_4^2} \\ \frac{u_2^3}{u_1^2+u_2^2} + \frac{u_4^3}{u_3^2+u_4^2} \end{pmatrix}.$$

We see from (41) that for each iteration, at each point, it is necessary to solve a nonlinear system of equations of the type

$$\boldsymbol{F}_0(\boldsymbol{U}_{ij}^{n+1}) = \boldsymbol{d}_{ij}. \tag{43}$$

We use the standard Newton method, which works well in most cases. Initial values for the Newton iterations can be either the previous $\boldsymbol{U}$ value, $\boldsymbol{U}_{ij}^n$, or the solution to the linear system $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{d}$, where $\boldsymbol{A}(\theta_1, \theta_2)$ is the matrix defined in (17) evaluated at the angles corresponding to $\boldsymbol{U}_{ij}^n$.

The Newton method uses the Jacobian of $\boldsymbol{F}_0$ in the iterations. One problem is that in general, this Jacobian is singular at some points in the computational domain. More specifically, the Jacobian,

$$\boldsymbol{J}_0 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 2\cos\theta_1 - \cos^3\theta_1 & -\sin\theta_1 + \sin^3\theta_1 & 2\cos\theta_2 - \cos^3\theta_2 & -\sin\theta_2 + \sin^3\theta_2 \\ -\cos\theta_1 + \cos^3\theta_1 & 2\sin\theta_1 - \sin^3\theta_1 & -\cos\theta_2 + \cos^3\theta_2 & 2\sin\theta_2 - \sin^3\theta_2 \end{pmatrix}, \tag{44}$$

is singular when

$$\cos\theta_1 = \cos\theta_2 \qquad \text{or} \qquad \sin\theta_1 = \sin\theta_2, \tag{45}$$

since then the first (second) and third (fourth) columns are equal.

Another feature of the system (43) should also be noted. It has always at least two solutions, since

$$\boldsymbol{F}_k(u_1, u_2, u_3, u_4) = \boldsymbol{F}_k(u_3, u_4, u_1, u_2), \qquad k = 0, 1, 2. \tag{46}$$

The two phases are thus interchangeable, which from a physical standpoint is quite natural. Numerically, it has the effect that we cannot be

certain which of the two roots the Newton method finds. Therefore the numerically calculated variables u_1, u_2, u_3, u_4 can be very discontinuous over the domain, even though the moments, which we get by applying the $\boldsymbol{F}_k$ to the variables, are smooth.

Another problem is how to deal with the points where $\boldsymbol{u} \equiv 0$, i.e. where no wave front has yet passed by. To avoid a division by zero error in the functions $\boldsymbol{F}_k$ we must initialize $\boldsymbol{u}$ at these points to some value. In the single phase case we have simply set $\boldsymbol{u} = (\epsilon, \epsilon)$, with ϵ equal to some very small, positive, value. For the two phase case, that method does not succeed however and a more sophisticated initialization is needed. The angles θ_1 and θ_2 easily lose regularity.

4 Results

In this section we show results from three different test cases. In all cases we have considered a homogeneous medium with the sources located outside the computational domain. We use the value of the exact solution as a Dirichlet boundary condition on all boundaries. The test cases are

A) the rectangle $0 \leq x \leq 1$ and $0 \leq y \leq 2$; one source located at coordinates $(-0.2,\ 1)$; smooth point source with exact solution $g = \max(0, t - r)^3/r$,

B) the rectangle $0 \leq x \leq 1$ and $0 \leq y \leq 1$; two sources located at coordinates $(-0.2,\ -0.2)$ and $(1.2,\ 1.2)$; smooth point sources with exact solution $g_k = \max(0, t - r_k)^2/r_k$, $k = 1, 2$,

C) the same rectangle as in B but with sources located at coordinates $(-0.3,\ 0.65)$ and $(1.3,\ 0.35)$; discontinuous point sources with exact solution $g_k = H(t - r_k)/r_k$, $k = 1, 2$.

The variable $r_k = r_k(x, y)$ is the distance to source k. General results for test case A is shown in Figure 2, where the Lax-Friedrich method was used to solve the $N = 1$ system (39, 40). The difficulties with using the Godunov and the Nessyahu-Tadmor methods for the same problem are highlighted in Figure 3 and Figure 4 respectively. Convergence for the different methods are summarized in Table 1 and Table 2.

For test case B we only used the Lax-Friedrich method. In Figure 5 the single phase system was solved, even though the physically correct solution contains two phases. A measure valued solution is suggested. In Figure 6 we used the $N = 2$ system (41, 42) for the same problem and it captures both phases.

Also for test case C, all solutions were computed using the Lax-Friedrich method. We present the results for the $N = 2$ system in Figure 7.

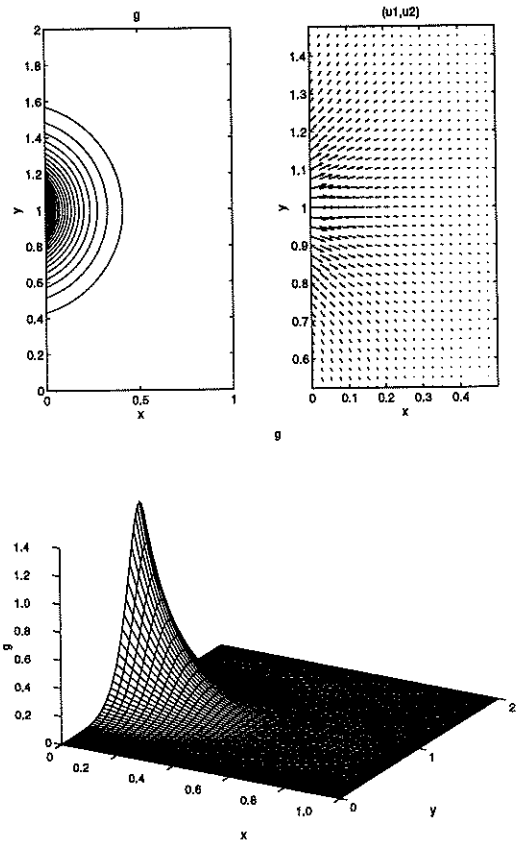

Figure 2: Solution at time $t = 0.85$ of the single phase system for test case A, using Lax-Friedrich with 40×80 points. Top left figure is a contour plot of the ray strength g. Top right figure shows the vector field $\boldsymbol{u} = (u_1, u_2)$. Bottom figure shows g.

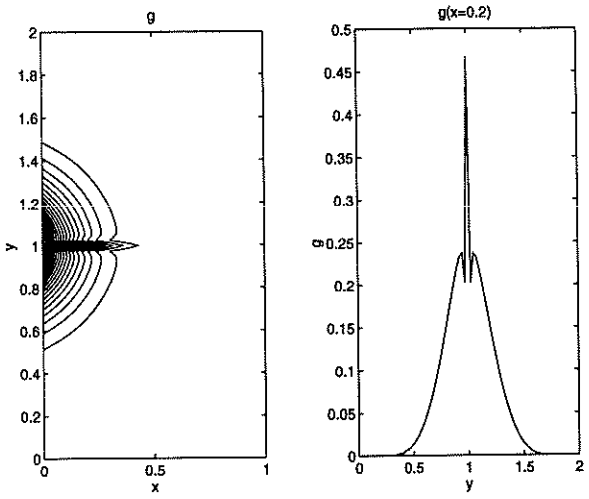

Figure 3: Solution at time $t = 0.85$ of the single phase system for test case A, using the Godunov method with 40×80 points. Left figure is a contour plot of the ray strength g. Right figure shows g in a vertical cut at $x = 0.2$.

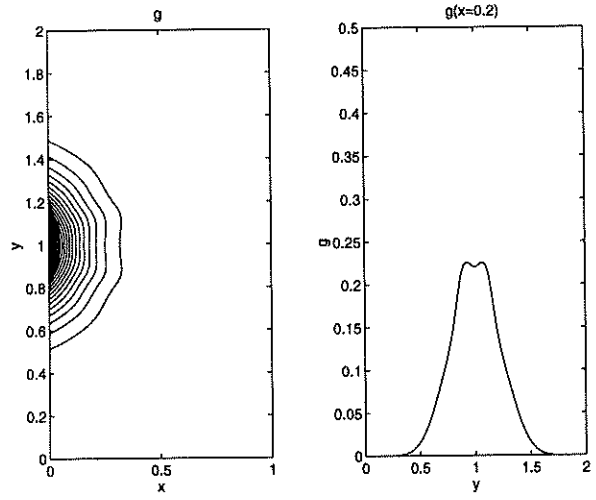

Figure 4: Solution at time $t = 0.85$ of the single phase system for test case A, using the Nessyahu-Tadmor method with 40×80 points. Left figure is a contour plot of the ray strength g. Right figure shows g in a vertical cut at $x = 0.2$.

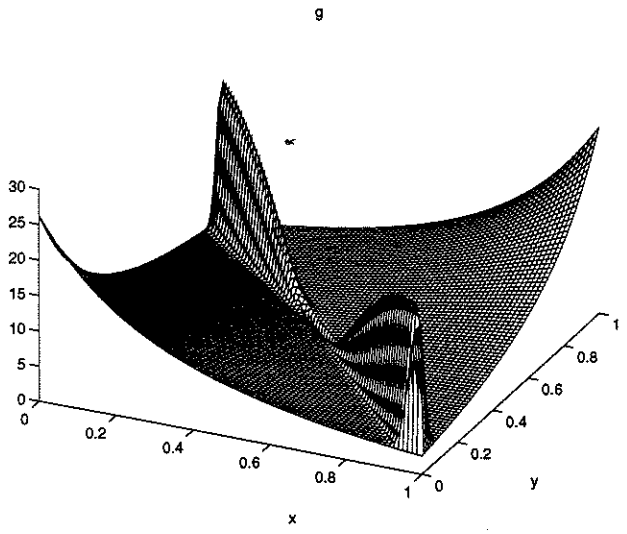

Figure 5: Solution at time $t = 3.0$ of the single phase system for test case B, using the Lax-Friedrich method with 80×80 points. Figure shows ray strength g.

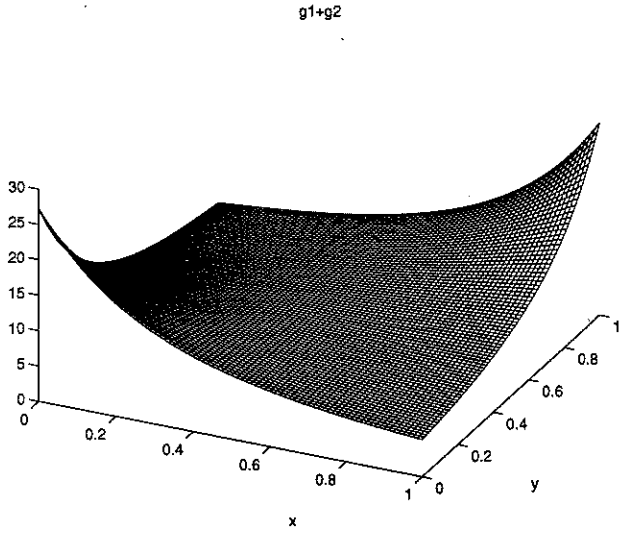

Figure 6: Solution at time $t = 3.0$ of the two phase system for test case B, using the Lax-Friedrich method with 80×80 points. Figure shows the combined ray strength $g_1 + g_2 = m_{00}$.

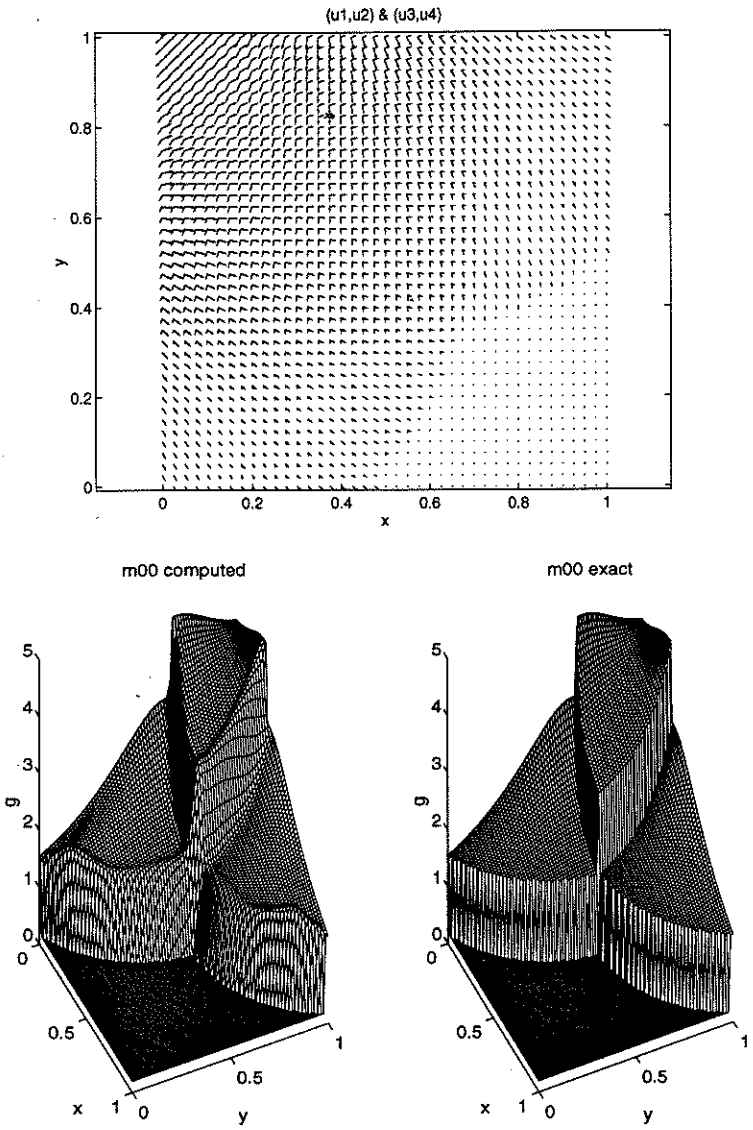

Figure 7: Solution of the two phase system for test case C, using the Lax-Friedrich method. Top figure shows the vector fields (u_1, u_2) and (u_3, u_4) superimposed at time $t = 0.85$, computed using 40×40 points. Bottom figure shows the combined strength of the two waves, $g_1 + g_2 = m_{00}$, at time $t = 0.7$, computed using 80×80 points (left) and exact (right).

References

[1] Y. Brenier and L. Corrias. Capturing multivalued solutions, to appear.

[2] Y. Brenier and E. Grenier. On the model of pressureless gases with sticky particles, to appear.

[3] M. Crandall and P. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277:1–42, 1983.

[4] W. E, Yu. G. Rykov, and Ya. G. Sinai. Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics, to appear.

[5] B. Engquist, E. Fatemi, and S. Osher. Numerical solution of the high frequency asymptotic expansion for the scalar wave equation. CAM Report 93-05, Department of Mathematics, UCLA, 1993.

[6] R. J. LeVeque. *Numerical Methods for Conservation Laws.* Birkhäuser, 1992.

[7] H. Nessyahu and E. Tadmor. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics*, 87(2):408–463, 1990.

[8] F. Qin et al. Finite-difference solution of the eikonal equation along expanding wavefronts. *Geophysics*, 57(3):478–487, March 1992.

[9] G. Strang. On the construction and comparison of difference schemes. *SIAM Journal of Numerical Analysis*, 5:506–517, 1968.

[10] J. van Trier and W. W. Symes. Upwind finite-difference calculation of traveltimes. *Geophysics*, 56(6):812–821, June 1991.

[11] J. Vidale. Finite-difference calculation of traveltimes. *Bulletin of the Seismological Society of America*, 78(6):2062–2076, December 1988.