Elimination of First Order Errors in Time Dependent Shock Calculations

Malin Siklosi* and Gunilla Kreiss†

Abstract

First order errors downstream of shocks have been detected in computations with higher order shock capturing schemes in one and two dimensions. We use matched asymptotic expansions to analyze the phenomenon for one dimensional time dependent hyperbolic systems and show how to design the artificial viscosity term in order to avoid the first order error. Numerical computations verify that second order accurate solutions are obtained.

1 Introduction

In many cases, solutions of conservation laws obtained by formally higher order methods are only first order accurate downstream of shocks, see e.g. [2], [5] and [4]. Basically, errors from the shock region follow outgoing characteristics and pollute the solution downstream. Examples in one space dimension where this effect can be seen are steady state calculations for systems with a source term and time dependent calculations for systems with non-constant solution. The effect can not be seen in one dimensional Riemann problems, because the exact global conservation determines the post shock states.

The degeneration in accuracy is troublesome, even though the first order term for reasonable mesh-sizes seems to be small in many cases. In some applications, as e.g. aeroacoustics where small amplitude waves need to be computed

*Department of Numerical Analysis and Computer Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden, malins@nada.kth.se

†Department of Numerical Analysis and Computer Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden, gunillak@nada.kth.se

accurately, it is however important with a very high accuracy. It is also important to understand the phenomenon more deeply in order to be able to design new methods which do not suffer from this deficiency.

The aim of this paper is to show that the first order error can be understood by matched asymptotic analysis of the modified equation and that the analysis can be used to construct methods that yield second order accurate solutions.

We consider the case of systems with time dependent solutions. We assume that the numerical solution can be modeled by a slightly viscous equation, a so called modified equation. In the shock layer, the coefficient of the viscous term is $\mathcal{O}(h)$, where h is the grid size. We analyze the solution of the modified equation using matched asymptotic expansions. It is assumed that an inner solution is valid in the shock region, and an outer solution is valid elsewhere. The two solutions are matched in a so called matching zone. From the analysis, we see that generally, the outer solution contains a term of $\mathcal{O}(h)$ downstream of the shock. We also see that if the inner solution satisfies a certain condition, the $\mathcal{O}(h)$ term would be eliminated. Based on this observation, we design a matrix valued viscosity coefficient which gives the inner solution the right shape to eliminate the $\mathcal{O}(h)$ downstream term. We construct a numerical scheme, using this matrix viscosity coefficient, and show in numerical experiments that the first order downstream error really is eliminated. However, we do not claim to have constructed an efficient and robust numerical method which can be used in realistic computations.

Similar analysis and construction of a matrix viscosity coefficient is done for the case of a steady state solution of a systems with a source term in [8]. In [3], matched asymptotic expansions for a problem which is very similar to the problem we study in this paper is analyzed for other purposes. The phenomenon has also been studied by other methods in [5] and [2]. In [5], analytic examples are constructed where the numerical solution is only first order accurate downstream of a shock, although the numerical scheme is formally second order. It is also shown that a converging numerical method will yield solutions having the formal order of accuracy in domains where no characteristics have passed through a shock. In [2], the first order downstream error is numerically detected in solutions of a shock-sound interaction problem solved by a fourth order ENO method. A scalar, linear equation is used to model the problem. It can be seen that the solution of the model problem computed with the fourth order ENO method behaves qualitatively different depending on if the discontinuity is located on a cell interface or in the interior of a cell. In the first case, the solution is fourth order in all of the domain, but in the second case the solution is only first order downstream of the discontinuity. Based on this observation,

the numerical method is modified such that the shock position will always be on a cell interface, and fourth order accuracy of the solution of the shock-sound interaction problem is obtain both upstream and downstream.

This paper is organized as follows. In section 2, we use asymptotic analysis to explain the first order downstream error and derive a matrix viscosity coefficient that eliminates it. In section 3, we implement a numerical method using the matrix viscosity coefficient and show in computations that the first order downstream error is eliminated.

2 Analysis

2.1 The Inviscid Problem

Consider the inviscid problem

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = 0, \quad 0 \le x \le x_{\text{end}}, \tag{1}$$

$$\mathbf{u}(x,0) = \mathbf{g}(x), \tag{2}$$

where $\mathbf{u}(x,t), \mathbf{g}(x) \in \mathbf{R}^n$, $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^n$ and $\mathbf{g}$ is a piecewise smooth function. We denote the Jacobian of the flux function $\mathbf{f}'(\mathbf{u})$ by $J(\mathbf{u})$. We assume that the eigenvalues of $J(\mathbf{u})$, denoted $\lambda_i(\mathbf{u}), i = 1, 2, \ldots, n$, are real and ordered in increasing order.

The initial and boundary conditions are chosen such that a shock forms at some inner point $s(t)$. At the shock the solution satisfies the Rankine-Hugoniot condition

$$\dot{s}[\mathbf{u}] = [\mathbf{f}(\mathbf{u})].$$

Here $[\mathbf{u}] = \mathbf{u}^+ - \mathbf{u}^-$, where $\mathbf{u}^\pm = \lim_{\delta \to 0^+} \mathbf{u}(s(t) \pm \delta, t)$. Corresponding notation for other quantities will be used frequently.

We assume that the shock is a classical Lax 1-shock i.e.

$$\dot{s} < \lambda_1^-,$$
$$\lambda_1^+ < \dot{s} < \lambda_2^+,$$

and the matrix

$$D = \begin{pmatrix} S_{II}^+ & [\mathbf{u}] \end{pmatrix} \tag{3}$$

is non-singular. Here the columns of S_{II}^+ are the eigenvectors of J^+ corresponding to the eigenvalues $\lambda_2^+, \lambda_3^+, \cdots, \lambda_n^+$.

We assume that the problem is closed by suitable boundary conditions, see e.g. [9]. We call these boundary conditions mathematical boundary conditions to distinguish them from numerical boundary conditions.

Remark: We analyze a 1-shock, since for 1-shocks there is just one downstream side. Hence, the first order error appears only on one side of the shock. For other Lax-shocks, both side of the shock are downstream sides, and first order errors appear on both sides of the shock.

2.2 The Slightly Viscous Model

We intend to study the behavior of numerical solutions of (1), i.e. we want to study the behavior of discrete functions which are the solutions of difference equations. A useful technique for studying the behavior of solutions to difference equations is to model the difference equation by a differential equation. Such a differential equation is often called a modified equation, see e.g. [10], [6]. Many numerical solutions of (1) can be viewed as higher order accurate solutions of the modified equation

$$\mathbf{u}_t^\varepsilon + \mathbf{f}(\mathbf{u}^\varepsilon)_x = (\Gamma \mathbf{u}_x^\varepsilon)_x, \quad 0 \le x \le x_{\text{end}}.$$

In the shock region, the modified equation can be shown to be valid only for weak shocks, see e.g. [7]. However, our computations indicate that it applies also for strong shocks. In the neighborhood of a shock layer we must have $\Gamma = \mathcal{O}(h)$, where h is the grid size, in order to avoid oscillations in the solution. Outside the shock region Γ can be smaller. In this paper we consider methods which can be modeled by

$$\mathbf{u}_t^\varepsilon + f(\mathbf{u}^\varepsilon)_x = \varepsilon(\phi \mathbf{u}_x^\varepsilon)_x + c_2 \varepsilon^2 \mathbf{u}_{xx}^\varepsilon, \tag{4}$$

where $\varepsilon = c_1 h$ and c_1 and c_2 a scalar constants. Here ϕ is a smooth function of $(x - s(t))/\varepsilon$ satisfying

$$\phi\left(\frac{x - s(t)}{\varepsilon}\right) = \begin{cases} 1 & \text{for } |\frac{x-s(t)}{\varepsilon}| \le K_0, \\ 0 & \text{for } |\frac{x-s(t)}{\varepsilon}| \ge K_1, \end{cases}$$

where $K_0 < K_1$ are constants with K_0 sufficiently large.

We must also model the initial data. In computations, the shape of the shock profile will depend on the method. If the initial data doesn't have exactly the right shape, the profile will after a short time adjust and obtain the right shape. In this process, small waves appear and flow out of the shock region, following

the outgoing characteristics. We are not interested in studying this initial effect and consequently assume that the initial profile is exactly the right profile for the method that is modeled. We specify the initial profile in equation (13) and (14).

We consider the same mathematical boundary conditions for $\mathbf{u}^\varepsilon$ as for $\mathbf{u}$, augmented with numerical boundary conditions.

We define the position of the viscous shock layer as the smallest x-value such that $\mathbf{u}^{\varepsilon(1)}(x,t) = (\mathbf{u}^{-(1)} + \mathbf{u}^{+(1)})/2$, and denote this point x_ε, i.e. the viscous shock position is defined as the point where the first component of viscous solution $\mathbf{u}^\varepsilon$ is half way between the right and left states in the corresponding inviscid shock.

2.3 Asymptotic Expansions

We assume that the solution of (4) can be described by an inner solution, valid in the shock layer, and an outer solution, valid elsewhere. These solutions can be expanded in powers of ε and matched in a region of overlap. Also the position of the shock layer can be expanded in ε. In [3], analysis of the asymptotic expansions for a very similar problem is presented and also the existence of an asymptotic expansion is treated.

The inner solution is expressed using the variables $(\tilde{x}, \tilde{t})$ where

$$\begin{aligned} \tilde{x} &= \frac{x - s(t)}{\varepsilon}, \\ \tilde{t} &= t. \end{aligned}$$

Thus we have expansions of the form

$$\text{Outer:} \quad \mathbf{u}^\varepsilon \sim \mathbf{u}(x,t) + \varepsilon \mathbf{u}_1(x,t) + \varepsilon^2 \mathbf{u}_2(x,t) + \cdots, \tag{5}$$

$$\text{Inner:} \quad \mathbf{u}^\varepsilon \sim \mathbf{U}_0(\tilde{x},\tilde{t}) + \varepsilon \mathbf{U}_1(\tilde{x},\tilde{t}) + \varepsilon^2 \mathbf{U}_2(\tilde{x},\tilde{t}) + \cdots, \tag{6}$$

$$\text{Position:} \quad x_\varepsilon \sim s(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \cdots. \tag{7}$$

We will match the inner and the outer solution at an upstream and a downstream matching point, $x_m^-(t)$ and $x_m^+(t)$. The matching points must satisfy $\lim_{\varepsilon\to 0} |x_m^\pm - s| = 0$. We will also need $e^{\mp \tilde{x}_m^\pm} = \mathbf{o}(1)$. Choosing $x_m^\pm = s \mp \varepsilon \log(\varepsilon)$, we have $e^{\mp \tilde{x}_m^\pm} = \mathcal{O}(\varepsilon)$ and both requirements are satisfied.

The viscous problem (4) models a method which is a second order accurate approximation of (1) away from the shock region. We claim that the solution will be second order accurate upstream of the shock but in general only first

order downstream. Hence we must show that $\mathbf{u}_1 = 0$ upstream and $\mathbf{u}_1 \neq 0$ downstream. To do this we need equations, initial data and boundary conditions for $\mathbf{u}_1$. Via the boundary conditions in the shock region, the outer solution will be coupled to the inner solution. Specifically, to derive boundary conditions for $\mathbf{u}_1$ we need information about $\mathbf{U}_0$. Hence, we derive equations and boundary conditions also for $\mathbf{U}_0$.

To obtain equations for the terms in the outer and inner expansions we substitute the expansions into (4) and collect terms multiplying the same power of ε. The equations for $\mathbf{U}_0$ is

$$(\phi \mathbf{U}_{0\bar{x}})_{\bar{x}} + \dot{s}\mathbf{U}_{0\bar{x}} - \mathbf{f}(\mathbf{U}_0)_{\bar{x}} = 0, \quad -\infty < \tilde{x} < \infty, \tag{8}$$

where we have used that the relations between derivatives in x and t and derivatives in $\tilde{x}$ and $\tilde{t}$ are

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{1}{\varepsilon}\frac{\partial}{\partial \tilde{x}}, \\ \frac{\partial}{\partial t} &= -\frac{\dot{s}}{\varepsilon}\frac{\partial}{\partial \tilde{x}} + \frac{\partial}{\partial \tilde{t}}. \end{aligned}$$

The outer solution expressed in the variables $(\tilde{x}, \tilde{t})$ is

$$\mathbf{u}^{\varepsilon} \sim \mathbf{u}(s(\tilde{t}) + \varepsilon\tilde{x}, \tilde{t}) + \varepsilon \mathbf{u}_1(s(\tilde{t}) + \varepsilon\tilde{x}, \tilde{t}) + \dots$$

Taylor expansion around $x = s \pm 0$ yields

$$\mathbf{u}^{\varepsilon} \sim \mathbf{u}^{\pm} + \varepsilon\tilde{x}\mathbf{u}_x^{\pm} + \frac{1}{2}\varepsilon^2\tilde{x}^2\mathbf{u}_{xx}^{\pm} + \varepsilon(\mathbf{u}_1^{\pm} + \varepsilon\tilde{x}\mathbf{u}_{1x}^{\pm}) + \mathcal{O}(\varepsilon^2).$$

It follows that the matching conditions for $\mathbf{U}_0$ are

$$\mathbf{U}_0(\tilde{x}, \tilde{t}) = \mathbf{u}^{\pm}, \tag{9}$$

as $\tilde{x} \to \pm\infty$. Note that (8) and (9) determines the shape of $\mathbf{U}_0$, but not the exact position of the shock layer.

Define $\hat{\mathbf{U}}(\tilde{x}, \tilde{t})$ by

$$\hat{\mathbf{U}}_{\bar{x}\bar{x}} + \dot{s}\hat{\mathbf{U}}_{\bar{x}} - f(\hat{\mathbf{U}})_{\bar{x}} = 0, \quad -\infty < \tilde{x} < \infty, \tag{10}$$

$$\hat{\mathbf{U}}(\tilde{x}, \tilde{t}) = \mathbf{u}(s \pm 0, \tilde{t}) \quad \text{as } \tilde{x} \to \infty, \tag{11}$$

$$\hat{\mathbf{U}}^{(1)}(0, \tilde{t}) = (\mathbf{u}^{-(1)} + \mathbf{u}^{+(1)})/2. \tag{12}$$

Note that $\hat{\mathbf{U}}$ approaches its limit values exponentially fast. We see that $\hat{\mathbf{U}}$ differs from $\mathbf{U}_0$ in two ways. First, $\hat{\mathbf{U}}$ is independent of ϕ, which makes the equation for $\hat{\mathbf{U}}$ much easier to analyze. However, if K_0 is sufficiently large, replacing ϕ by 1 just changes the solution by exponentially small terms. Second, the position of $\hat{\mathbf{U}}$ is fixed at $\tilde{x} = 0$. Since the position of the shock layer has the expansion (7), we have, except exponentially small terms, to leading order

$$\mathbf{U}_0(\tilde{x}, \tilde{t}) = \hat{\mathbf{U}}(\tilde{x} - x_1(\tilde{t}), \tilde{t}).$$

Below we will derive an ordinary differential equation for $x_1(\tilde{t})$. The initial value of $x_1(\tilde{t})$ is determined by the initial condition $\mathbf{g}^\varepsilon$, which we now specify.

$$\text{Outer region: } \mathbf{g}^\varepsilon(x) = \mathbf{g}(x) \tag{13}$$

$$\text{Inner region: } \mathbf{g}^\varepsilon(\tilde{x}) = \hat{\mathbf{U}}(\tilde{x}, 0). \tag{14}$$

This is sufficient for our purposes. However, if one consider more terms in the inner expansion one would have to add the corresponding terms to (14). Note that (14) means that $x_1(0) = 0$.

The equation for $\mathbf{u}_1$ is

$$\mathbf{u}_{1t} + (f'(\mathbf{u})\mathbf{u}_1)_x = 0, \quad x \in \text{outer region}, \tag{15}$$

where we have used that $\phi = 0$ in the outer region. We also need initial data and boundary conditions for $\mathbf{u}_1$. The initial conditions for $\mathbf{u}^\varepsilon$, (13), gives $\mathbf{u}_1(x, 0) = 0$. By substituting the expansion for $\mathbf{u}^\varepsilon$ into the mathematical boundary conditions at $x = 0$ and $x = 1$, Taylor expand and take into account that $\mathbf{u}^\varepsilon$ and $\mathbf{u}$ satisfies the same mathematical boundary conditions we see that the mathematical boundary conditions for $\mathbf{u}_1$ are homogeneous. Since all characteristics of (15) are going into the shock, $\mathbf{u}_1$ in the upstream region is fully determined by initial data and mathematical boundary conditions at $x = 0$. Since the equation (15), initial data and the mathematical boundary conditions are homogeneous, we have $\mathbf{u}_1 \equiv 0$ in the upstream region. To determine $\mathbf{u}_1$ in the downstream region, we also need boundary conditions at $x = s^+$. We derive such boundary conditions in the next section.

2.4 Downstream Boundary Condition for the First Order Outer Term

Integration of the viscous equation (4) over the shock layer, from matching point x_m^- to matching point x_m^+ gives

$$\int_{x_m^-}^{x_m^+} \mathbf{u}_t^\varepsilon \, dx + [f(\mathbf{u}^\varepsilon)]_{x_m^-}^{x_m^+} = \mathcal{O}(\varepsilon^2), \tag{16}$$

where we have used that ϕ vanishes in the matching regions.

Using the outer expansion of $\mathbf{u}^\varepsilon$ we obtain

$$[f(\mathbf{u}^\varepsilon)]_{x_m^-}^{x_m^+} = [f(\mathbf{u})]_{x_m^-}^{x_m^+} + \varepsilon [J(\mathbf{u})\mathbf{u}_1]_{x_m^-}^{x_m^+} + \mathcal{O}(\varepsilon^2), \tag{17}$$

By integrating the inviscid (1) over the same interval, we obtain

$$[f(\mathbf{u})]_{x_m^-}^{x_m^+} = \dot{s}[\mathbf{u}] - \int_{x_m^-}^{s^-} \mathbf{u}_t \, dx - \int_{s^+}^{x_m^+} \mathbf{u}_t \, dx. \tag{18}$$

Note that $\mathbf{u}$ is discontinuous at $x = s(t)$ and the Rankine-Hugoniot condition applies across the discontinuity. After taking into account that $\mathbf{u}_1 \equiv 0$ to the left of the shock layer and introducing (17) and (18) into (16) we arrive at

$$\dot{s}[\mathbf{u}] + \varepsilon J(\mathbf{u}(x_m^+, t))\mathbf{u}_1(x_m^+, t) + I_1 = \mathcal{O}(\varepsilon^2), \tag{19}$$

where we have introduced the notation

$$I_1 = \int_{x_m^-}^{s^-} (\mathbf{u}_t^\varepsilon - \mathbf{u}_t) \, dx + \int_{s^+}^{x_m^+} (\mathbf{u}_t^\varepsilon - \mathbf{u}_t) \, dx.$$

After Taylor expansion of $\mathbf{u}$ and $\mathbf{u}_1$ around $x = s^+$, equation (19) can be rewritten as

$$\dot{s}[\mathbf{u}] + \varepsilon J(\mathbf{u}(s^+, t)\mathbf{u}_1(s^+, t)) + I_1 = \mathbf{o}(\varepsilon), \tag{20}$$

In the coordinate system $(\tilde{x}, \tilde{t})$ we have

$$I_1 = -\dot{s}A + \varepsilon I_2$$

where

$$A = \int_{\tilde{x}_m^-}^{0^-} (\mathbf{u}^\varepsilon - \mathbf{u})_{\tilde{x}} \, d\tilde{x} + \int_{0^+}^{\tilde{x}_m^+} (\mathbf{u}^\varepsilon - \mathbf{u})_{\tilde{x}} \, d\tilde{x}$$

$$I_2 = \int_{\tilde{x}_m^-}^{0^-} (\mathbf{u}^\varepsilon - \mathbf{u})_{\tilde{t}} \, d\tilde{x} + \int_{0^+}^{\tilde{x}_m^+} (\mathbf{u}^\varepsilon - \mathbf{u})_{\tilde{t}} \, d\tilde{x}$$

Evaluating the integral yields

$$A = [\mathbf{u}] + [\mathbf{u}^\varepsilon - \mathbf{u}]_{\bar{x}_m^-}^{\bar{x}_m^+}.$$

By using the outer expansion of $\mathbf{u}^\varepsilon$, taking into account that $\mathbf{u}_1$ is zero upstream and Taylor expanding $\mathbf{u}_1$ around $x = s^+$ we obtain

$$A = [\mathbf{u}] + \varepsilon \mathbf{u}_1^+ + \mathbf{o}(\varepsilon).$$

Next, consider I_2. Now using the inner expansion of $\mathbf{u}^\varepsilon$, the Taylor expansion of $\mathbf{u}$ around $x = s \pm 0$ and $\mathbf{U}_0(\tilde{x}, \tilde{t}) = \hat{\mathbf{U}}(\tilde{x} - x_1, \tilde{t})$ we obtain

$$I_2 = \int_{\bar{x}_m^-}^0 (\hat{\mathbf{U}}(\tilde{x} - x_1, \tilde{t}) - \mathbf{u}^-)_{\bar{t}}\, d\tilde{x} + \int_0^{\bar{x}_m^+} (\hat{\mathbf{U}}(\tilde{x} - x_1, \tilde{t}) - \mathbf{u}^+)_{\bar{t}}\, d\tilde{x} + \mathbf{o}(1).$$

We rewrite I_2 in two steps. First we make a substitution of variable $\hat{x} = \tilde{x} - x_1$. Next we use that $\hat{\mathbf{U}}$ approaches the limit values exponentially fast and the matching points are chosen such that $e^{\mp \bar{x}_m^\pm} = \mathcal{O}(\varepsilon)$. Hence we can extend the integration interval to infinity, still keeping the remainder term $\mathbf{o}(1)$. We obtain

$$I_2 = I_{3\bar{t}} - (x_1[\mathbf{u}])_{\bar{t}} + \mathbf{o}(1),$$

where

$$I_3(\tilde{t}) = \int_{-\infty}^0 (\hat{\mathbf{U}}(\tilde{x}, \tilde{t}) - \mathbf{u}^-)\, d\tilde{x} + \int_0^\infty (\hat{\mathbf{U}}(\tilde{x}, \tilde{t}) - \mathbf{u}^+)\, d\tilde{x}.$$

Since I_2, I_3, x_1 and $[\mathbf{u}]$ are functions of $\tilde{t}$ only, and $\tilde{t} = t$ this can be written

$$I_2(t) = \frac{\partial}{\partial t}\left(I_3(t) + x_1(t)[\mathbf{u}](t)\right) + \mathbf{o}(1).$$

Hence we have

$$I_1 = -\dot{s}[\mathbf{u}] + \varepsilon(-\dot{s}\mathbf{u}_1^+ + I_{3t} - (x_1[\mathbf{u}])_t) + \mathbf{o}(\varepsilon).$$

Substituting this into (20) and rearranging we obtain

$$(J^+ - \dot{s}I)\mathbf{u}_1^+ - (x_1[\mathbf{u}])_t + I_{3t} = \mathbf{o}(1).$$

Hence the equations for $\mathbf{u}_1^+$ and $x_1(t)$ are

$$\begin{gathered}(J^+ - \dot{s}I)\mathbf{u}_1^+(t) - (x_1(t)[\mathbf{u}])_t = -I_{3t} \\ x_1(0) = 0.\end{gathered}$$

This can be rewritten as

$$\begin{pmatrix} w_{II}^+ \\ \dot{x}_1 \end{pmatrix} = \begin{pmatrix} \Lambda_{II}^+ - \dot{s}I & 0 \\ 0 & -1 \end{pmatrix}^{-1} D^{-1} \left(-I_{3t} + x_1[\mathbf{u}]_t - S_I^+(\lambda_1^+ - \dot{s})w_I^+ \right), \tag{21}$$

$$x_1(0) = 0, \tag{22}$$

where w_{II}^+ are the characteristic variables of u_1^+ going out of the shock, w_I^+ is the characteristic variable going into the shock, $\Lambda_{II}^+ = \text{diag}(\lambda_1^+, \cdots \lambda_n^+)$, S_I^+ is the eigenvector of J^+ corresponding to the eigenvalue λ_1^+ and D is defined by (3). Hence, generally we have $u_1(x,t) \neq 0$ for $x > s$.

2.5 A Matrix Viscosity Coefficient Eliminating the $\mathcal{O}(h)$ Error

We will now investigate if it is possible to design the viscosity term such that the first order downstream error is eliminated and second order accurate solutions are obtained. We consider a method which has the modified equation

$$\mathbf{u}_t^\varepsilon + f(\mathbf{u}^\varepsilon)_x = \varepsilon(\phi(x)E(\mathbf{u}^\varepsilon)\mathbf{u}_x^\varepsilon)_x + c_2\varepsilon^2\mathbf{u}_{xx}^\varepsilon, \tag{23}$$

where $E(\mathbf{u}^\varepsilon)$ is a matrix valued function. The solution of such a method can be analyzed in the same way as in the previous sections. The only thing which will change in the analysis is the equation for $\hat{\mathbf{U}}$. The new equation for $\hat{\mathbf{U}}$ is

$$(E(\hat{\mathbf{U}})\hat{\mathbf{U}}_{\tilde{x}})_{\tilde{x}} + \dot{s}\hat{\mathbf{U}}_{\tilde{x}} - f(\hat{\mathbf{U}})_{\tilde{x}} = 0. \tag{24}$$

together with the conditions (11) and (12). The boundary condition for u_1 at $x = s^+$ is still given by (21) and (22). If $E(\hat{\mathbf{U}})$ can be chosen such that $I_{3t} = 0$, we will have $u_1(x,t) = 0$ also in the downstream region. We note that if $\hat{\mathbf{U}} = \hat{\mathbf{U}}^*$ with

$$\hat{\mathbf{U}}^* = \mathbf{u}^- + \gamma(\tilde{x})[\mathbf{u}],$$

where γ is a scalar smooth function, we obtain

$$I_3 = c_\gamma[\mathbf{u}],$$

where

$$c_\gamma = \int_{-\infty}^{0} \gamma(\tilde{x})\, d\tilde{x} + \int_0^\infty (\gamma(\tilde{x}) - 1)\, d\tilde{x}.$$

If γ is monotone and antisymmetric around $(0, 0.5)$ with

$$\gamma(-\infty) = 0, \quad \gamma'(-\infty) = 0, \quad \gamma(\infty) = 1, \quad \gamma'(\infty) = 0$$

then $c_\gamma = 0$ and $\hat{\mathbf{U}} = \hat{\mathbf{U}}^*$ has a reasonable shape. Note that there is a one to one correspondence between γ and $\hat{\mathbf{U}}^*$. We also need to be able to express γ' in terms of γ, i.e.

$$\gamma' = \psi(\gamma).$$

Hence, if it is possible to chose E such that $\hat{\mathbf{U}}^*$ is a solution of (24) with boundary conditions we would obtain a truly second order accurate approximation of $\mathbf{u}$.

Integrating (24) from $-\infty$ to $\tilde{x}$ and substituting $\hat{\mathbf{U}}^*$ gives

$$\psi(\gamma) E(\hat{\mathbf{U}}^*)[\mathbf{u}] = \mathbf{q}(\hat{\mathbf{U}}^*) \tag{25}$$

where

$$\mathbf{q}(\mathbf{U}) = f(\mathbf{U}) - f(\mathbf{u}^-) - \gamma \dot{s}(\mathbf{U} - \mathbf{u}^-).$$

Hence

$$E(\mathbf{U}) = \frac{1}{\psi(\gamma)} \frac{\mathbf{q}(\mathbf{U})\mathbf{q}^T(\mathbf{U})}{\mathbf{q}^T(\mathbf{U})[\mathbf{u}]} \tag{26}$$

satisfies (25). To ensure that $E(\mathbf{u}^\varepsilon)$ is bounded as $\tilde{x} \to \pm\infty$ we require

$$\lim_{\tilde{x}\to-\infty} \frac{\gamma(\tilde{x})}{\psi(\gamma(\tilde{x}))} = M^-, \quad \lim_{\tilde{x}\to\infty} \frac{\gamma(\tilde{x}) - 1}{\psi(\gamma(\tilde{x}))} = M^+,$$

where $|M^\pm| < \infty$. Note that in order to evaluate $E(\hat{\mathbf{U}})$ the quantities $\dot{s}$, $\mathbf{u}^-$ and $\mathbf{u}^+$ must be known or estimated.

Remark: Prescribing the viscous profile in the above way means that the solution follows a straight line in phase space between the upstream and the downstream states. Many other shapes of the solution, and hence, paths in phase space, would also be possible. The properties of $E(\mathbf{u}^\varepsilon)$ will depend on which path that is chosen. In order to obtain a stable method, it is necessary that the total viscosity coefficient of the method is positive definite. Since the term $c_2\varepsilon^2\mathbf{u}^\varepsilon_{xx}$ also is present, it is sufficient that $E(\mathbf{u}^\varepsilon)$ is positive semi definite. We have found that the choice (26) is not positive semi definite for all problems. In order to design a robust numerical method, it must be further investigated what paths in phase space to use. Probably, this will differ depending on what equations you want to solve. However, we are only interested in showing that it is possible to obtain second order accuracy also downstream, and for this purpose is good enough to use $E(\mathbf{u}^\varepsilon)$ defined by (26).

3 Numerical Experiments

In this section we test how the matrix viscosity coefficient we derived in the previous section behaves in computations, and compare with corresponding computations with a scalar viscosity coefficient.

3.1 The Test Problems

We consider two test problems. In the both problems equations, domain, initial data and the boundary condition at $x = x_{\text{end}}$ are the same, while the boundary condition at $x = 0$ differs.

We consider the Euler equations with Riemann initial data connected by a 1-shock. That is, we consider (1) and (2) with

$$\mathbf{u} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad \mathbf{f}(\mathbf{u}) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix}, \quad x_{\text{end}} = 6,$$

$$\mathbf{u}(x,0) = \begin{cases} \mathbf{u}_L & \text{for } x \le s_0, \\ \mathbf{u}_R & \text{for } x > s_0, \end{cases}$$

where E and p are connected by the equation of state for a polytropic gas

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho u^2.$$

Since $\mathbf{u}_L$ and $\mathbf{u}_R$ are connected by a 1-shock, they are fully determined if ρ_L, u_L and p_L, the initial density, velocity and pressure at $x \le s_0$, and p_R, the initial pressure at $x > s_0$, are specified. We have used $\rho_L = 3$, $u_L = 1.2$, $p_L = 2$ and $p_R = 5$. This gives a 1-shock with speed ≈ -0.26. We have not specified the initial shock position s_0 explicitly. In the computations, which will be further described below, we started the computation at $t = -1$ with the profile (28) and the shock located at $x = 1.75$. We computed for one time unit using $\mathbf{u}(0,t) = \mathbf{u}_L$. In this way we obtained a good initial profile. We do this to avoid that the numerical solution is polluted by disturbances due to non-perfect initial data.

At $x = x_{\text{end}}$ we have the boundary condition

$$R_1(x_{\text{end}}, t) = R_1(x_{\text{end}}, 0),$$

where $R_1 = u - 2c/(\gamma - 1)$ is the Riemann invariant connected to λ_1, and $c = \sqrt{\gamma p/\rho}$ is the local speed of sound.

At $x = 0$ the boundary condition is specified by

$$p(0,t) = p_L(1 + \alpha d(t)),$$
$$\rho(0,t) = \rho_L \left(\frac{p(0,t)}{p_L} \right)^{1/\gamma},$$
$$u(0,t) = u_L + \frac{2}{\gamma - 1}(c(0,t) - c_L),$$

see [2]. In both test problems we have used $\alpha = -0.2$. The two problems have different functions $d(t)$:

$$\textbf{Test problem 1} : d(t) = \sin 2t$$
$$\textbf{Test problem 2} : d(t) = \sin 2t + \frac{1}{2} e^{-4t} \sin 10t$$

3.2 The Standard Method

We solved (1) with the semidiscrete scheme

$$(\mathbf{u}_j)_t + D_0 \mathbf{f}(\mathbf{u}_j) = \kappa_1 h D_+ \phi_j D_- \mathbf{u}_j + \zeta h^2 D_+ D_- \mathbf{u}_j. \tag{27}$$

In our calculation we used $\kappa_1 = 1$ and $\zeta = 20$. We discretized in space by introducing $x_j = jh$, $h = 1/N$, $j = 0, 1, \ldots N$, where $\mathbf{u}_j(t)$ is a grid function with $\mathbf{u}_j(t) \approx \mathbf{u}^\varepsilon(x_j, t)$. The system of ODEs (27) was solved with the classical fourth-order Runge-Kutta method. The times step k was chosen $k = 0.2h$.

The switch ϕ was

$$\phi(x) = \begin{cases} 0.5 \tanh((x - s(t) + s_1 h)/s_2 h) + 0.5, & x \le s(t), \\ 0.5 \tanh((x - s(t) - s_1 h)/s_2 h) + 0.5, & x > s(t). \end{cases}$$

with $s_1 = 60$ and $s_2 = 4$, see Figure 1. Generally, there will be approximately $2s_1$ points with ϕ value above 0.5, hence, we have used a very wide switch. The parameter s_2 determines how steep the gradient of ϕ is in the transition area.

The shock position $s(t)$ was numerically determined. The approximate shock positions was taken as the x_j value where θ_j has its maximum. The function θ was defined by

$$\theta_j = \frac{|\rho_{j+1} - 2\rho_j + \rho_{j-1}|}{\rho_{j+1} + 2\rho_j + \rho_{j-1}}.$$

The function θ and similar functions are often used to detect shocks, see [1].

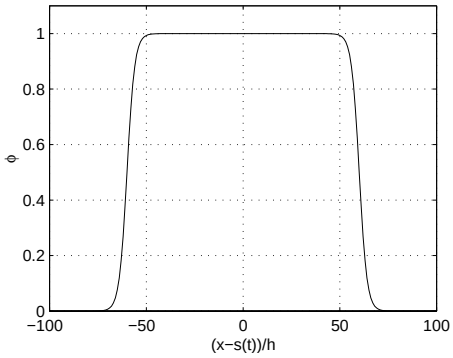


Figure 1: The function ϕ, with $s_1 = 60$ and $s_2 = 4$.

At $x = 6$ we used the mathematical boundary condition

$$R_1(6,t) = R_1(6,0),$$

and the numerical boundary conditions

$$R_{2,3}(6,t) = 2R_{2,3}(6-h,t) - R_{2,3}(6-2h,t),$$

where the Riemann invariants R_2 and R_3 are

$$R_2 = \frac{p}{\rho^\gamma}, \quad R_3 = u + \frac{2c}{\gamma - 1}.$$

The initial data was obtained in the following way. We started the computations at $t = -1$ with the profile (28) and the shock located at $x = 1.75$. We computed for one time unit using $\mathbf{u}(0,t) = \mathbf{u}_L$.

This method can be modeled by equation (4) and we will refer to it as the standard method.

3.3 The Matrix Viscosity Method

We will now introduce a method which can be modeled by (23) and we will refer to this method as the matrix viscosity method. The matrix viscosity method is the same as the standard method except that (27) is replaced by

$$(\mathbf{u}_j)_t + D_0\mathbf{f}(\mathbf{u}_j) = \kappa_2 h D_+\phi_j E_j D_-\mathbf{u}_j + \zeta h^2 D_+ D_-\mathbf{u}_j.$$

In our calculations we also here used $\zeta = 20$. We used $\kappa_2 = 9$ in order to make the number of points in the shock layer approximately equal for the standard method and the matrix viscosity method. Also, $E_j \approx E(\mathbf{u}^\varepsilon(x_j, t))$. The expression (26) needs to be modified when used in numerical computations, since both $\mathbf{q}$ and γ' tends to zero as $\tilde{x} \to \pm\infty$. For large $\tilde{x}$ we can linearize the expression for $\mathbf{q}$ and find

$$\mathbf{q} = \begin{cases} \gamma(J^- - \dot{s}I)[\mathbf{u}] & \text{as } \tilde{x} \to -\infty \\ (\gamma - 1)(J^+ - \dot{s}I)[\mathbf{u}] & \text{as } \tilde{x} \to \infty \end{cases}$$

By assumption on γ we find

$$E = \begin{cases} E^- & \text{as } \tilde{x} \to -\infty \\ E^+ & \text{as } \tilde{x} \to \infty \end{cases}$$

where

$$E^\pm = M^\pm \frac{(J^\pm - \dot{s}I)[\mathbf{u}][\mathbf{u}]^T (J^\pm - \dot{s}I)^T}{[\mathbf{u}]^T (J^\pm - \dot{s}I)^T [\mathbf{u}]}.$$

We have used

$$\gamma(\tilde{x}) = \frac{1}{2}(\tanh(\tilde{x}) + 1),$$

hence we have $M^- = 1/2$ and $M^+ = -1/2$.

The solution changes rapidly in the shock layer from being close to $\mathbf{u}^-$ to being close to $\mathbf{u}^+$. For this reason we have chosen not to compute $E(\hat{\mathbf{U}})$ from (26) in the shock layer. In the computations we have instead used

$$E_j = (1 - \gamma(\tilde{x}_j))E^- + \gamma(\tilde{x}_j)E^+.$$

in all of the computational domain, i.e. a weighted sum of E^- and E^+. As weight function we used γ. This seems to work well. The profile obtains the right shape and we obtain second order accurate solutions in the downstream region (see section 3.4 and Figure 5). The quantities $\dot{s}$, $\mathbf{u}^-$ and $\mathbf{u}^+$ were numerically determined. We used

$$\mathbf{u}^-_{\text{approx}} = \mathbf{u}_{J-20}, \quad \mathbf{u}^+_{\text{approx}} = \mathbf{u}_{J+20}$$

where J is the index of the approximate shock position. To approximate the shock states by the values of the numerical solutions 20 points away from the approximate shock position seems to work well in computations. The shock speed was approximated by

$$\dot{s}_{\text{approx}} = \sum_{k=1}^{3} [\mathbf{f}^{(k)}(\mathbf{u})] \Big/ \sum_{k=1}^{3} [\mathbf{u}^{(k)}].$$

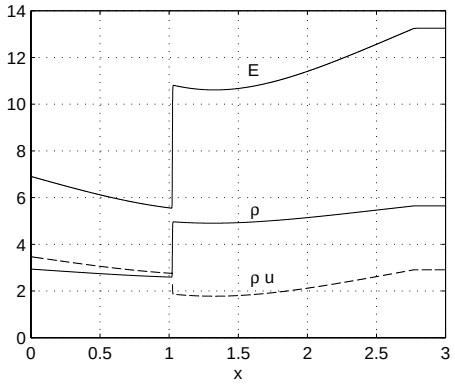


Figure 2: The solution of test problem 1 at $t = 1.5$. The solution is computed numerically using the standard method with $h = 3.125 \cdot 10^{-4}$.

By summing the jumps in the different components of $\mathbf{f}(\mathbf{u})$ and divide by the sum of the jumps in the different components of $\mathbf{u}$ we avoid that rounding errors give large errors in $\dot{s}_{\text{approx}}$. As initial profile at $t = -1$ we used

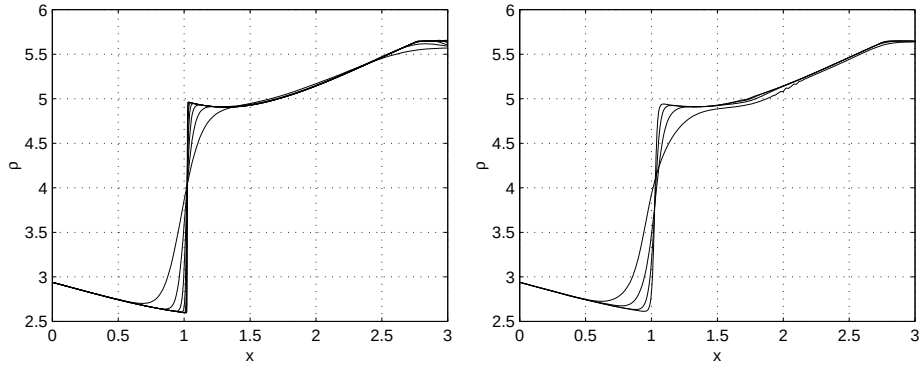
$$\mathbf{u}_j = \mathbf{u}_L + \gamma(\tilde{x}_j)(\mathbf{u}_R - \mathbf{u}_L). \tag{28}$$

3.4 Results

We have numerically investigated the speed of convergence of the standard method and the matrix viscosity method by solving the two test problems with successively refined space step.

First, consider test problem 1. In Figure 2 we see the solution of test problem 1 at $t = 1.5$. Since the solution is constant in the interval $3 \leq x \leq 6$, except for small disturbances due to non-perfect initial data, see section 2.2, we have have plotted the solution only on the interval $0 \leq x \leq 3$. We have solved test problem 1 with successively halved space step h with both methods. We started with $h = 0.02$. The matrix viscosity method converges much faster than the standard method. The solution from the matrix viscosity method with $h = 2.5 \cdot 10^{-3}$ and the solution from the standard method with $h = 3.125 \cdot 10^{-4}$ are approximately as accurate. In all, we computed four solutions with the matrix viscosity method and seven solutions with the standard method.

In Figure 3 we see an overview of how the ρ-component of the solutions con-



(a) Standard method

(b) Matrix viscosity method

Figure 3: Overview of the convergence of test problem 1. In the plots we see the ρ-component of the solution. In both cases, the most viscous solution is computed using $h = 0.02$. The following solutions are computed using successively halved space step. For the standard method we see seven different solutions and for the matrix viscosity method four solutions.

verge. Around $x = 1.75$ we see a small bump in the ρ-component of the solution from the matrix viscosity method. The bump is related to the discontinuity at $t = 0$ in the derivatives of the boundary condition at $x = 0$. The effect can also be seen, but is much weaker, in the ρu-component of the solution. In the E-component it is not visible for the eye. The bump decays as $\mathcal{O}(h)$. In the solutions with the standard method this effect is not seen at all. In Figure 4 we see a typical close up of how the solutions converges downstream (away from $x = 1.75$). It is clear that the speed of convergence is much faster for the matrix viscosity method.

The order of accuracy in the upstream region was estimated in the standard way by computing

$$\log\left(\frac{||\rho_h - \rho_{h/2}||_U}{||\rho_{h/2} - \rho_{h/4}||_U}\right) \Big/ \log 2,$$

where ρ_h denote the discrete approximation of ρ with space step h. Corresponding expressions where also used for the ρu- and the E-component of the

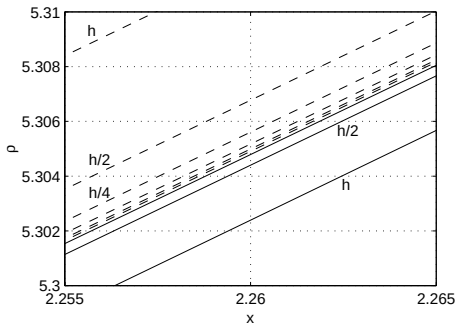


Figure 4: Close up of the ρu-component of the solution, computed with successively refined space step. Solid lines: matrix viscosity method, dashed lines: standard method.

solution. The norm $||\cdot||_U$ was defined by

$$||\cdot||_U = ||\cdot||_{L_2^h(0,0.2)},$$

for each component. Here L_2^h is the discrete L_2-norm. The order of accuracy in the downstream region was computed correspondingly, using the norm

$$||\cdot||_D = ||\cdot||_{L_2(1.8,2.6)}.$$

The integration domain $(0, 0.2)$ was chosen in order to avoid effects from the shock region also for the very viscous solutions computed with the largest time step. Because of the bump, the estimated order of accuracy in the ρ-component becomes qualitatively different if the area around $x = 1.75$ is included or not. For the other components of the solution, this effect is not seen. For this reason, all estimates of order of accuracy downstream presented here are computed using the integration domain $(1.8, 2.6)$. In Table 1 we see that the standard method is second order accurate upstream, but only first order accurate downstream of the shock. From Table 2 we see that the matrix viscosity method is second order accurate, both upstream and downstream of the shock.

By plotting the shock profile in phase space, see Figure 5, we see that the shock profile obtained by the matrix viscosity method approximately follows a straight line between the shock states. The shock profile of the standard method clearly has another shape.

Next, consider test problem 2. In Figure 6 we see the solution of test problem 2 at $t = 1.5$. Also test problem 2 was solved with successively halved space

h	Estimated order of accuracy					
	Upstream			Downstream		
	ρ	ρu	E	ρ	ρu	E
$2 \cdot 10^{-2}$	3.00	4.81	4.93	1.85	1.90	1.92
$1 \cdot 10^{-2}$	1.97	1.98	2.02	1.88	1.64	1.54
$5 \cdot 10^{-3}$	2.00	2.01	2.00	1.59	1.47	1.41
$2.5 \cdot 10^{-3}$	2.00	2.00	1.99	1.27	1.30	1.27
$1.25 \cdot 10^{-3}$	2.00	2.00	2.00	1.16	1.16	1.16

Table 1: Estimated order of accuracy, standard method, test problem 1.

h	Estimated order of accuracy					
	Upstream			Downstream		
	ρ	ρu	E	ρ	ρu	E
$2 \cdot 10^{-2}$	7.26	5.14	7.46	1.95	2.47	2.80
$1 \cdot 10^{-2}$	2.00	2.04	1.99	2.81	2.84	2.56

Table 2: Estimated order of accuracy, matrix viscosity method, test problem 1.

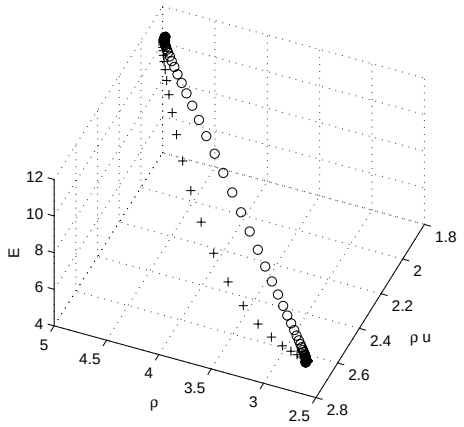


Figure 5: Numerical phase diagram of the shock profile computed by the matrix viscosity method (o) and the standard method (+). Both solutions are computed using $h = 2.5 \cdot 10^{-3}$. The shock profile computed by the matrix viscosity method follows a straight line in phase space quite closely.

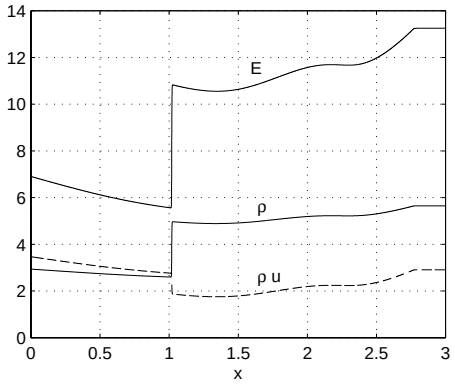
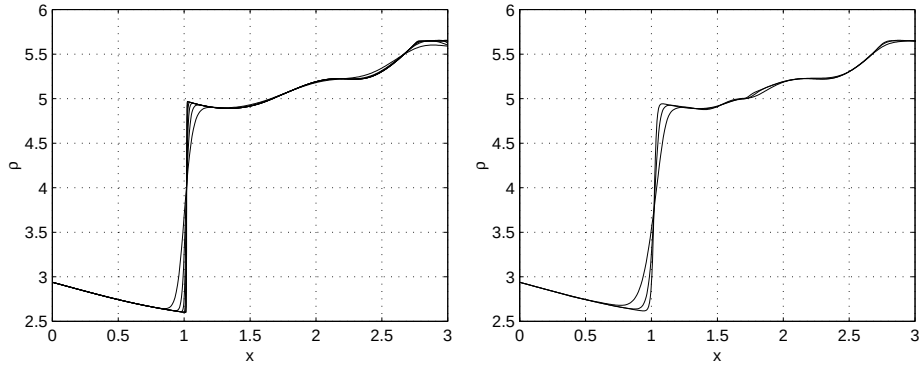


Figure 6: The solution of test problem 2 at $t = 1.5$. The solution is computed numerically using the standard method with $h = 3.125 \cdot 10^{-4}$.

step h with both methods. The matrix viscosity method failed to obtain a solution for $h = 0.02$, hence we started with $h = 0.01$ and computed in all six solutions with the standard method and three solutions with the matrix viscosity method. In Figure 7 we see an overview of how the ρ-component converges. As expected, the bump around $x = 1.75$ still remains. In Table 3 we see the estimated order of accuracy for the standard method. Upstream, the solution is second order. Downstream the convergence is slower, and the order of accuracy is slowly approaching one. The size of the second order error term depends on space derivatives of the solution. These are larger in test problem 2 than in test problem 1. This is probably the reason why the first order downstream effect is seen less clearly for test problem 2. In Table 4 we see that the matrix viscosity method is second order accurate both upstream and downstream. In phase space, the shock profiles of the solutions of test problem 2 are qualitatively the same as in Figure 5.



(a) Standard method (b) Matrix viscosity method

Figure 7: Overview of the convergence of test problem 2. In the plots we see the ρ-component of the solution. In both cases, the most viscous solution is computed using $h = 0.01$. The following solutions are computed using successively halved space step. For the standard method we see six different solutions and for the matrix viscosity method three solutions.

h	Estimated order of accuracy					
	Upstream			Downstream		
	ρ	ρu	E	ρ	ρu	E
$1 \cdot 10^{-2}$	2.04	1.98	1.95	1.72	1.64	1.61
$5 \cdot 10^{-3}$	2.06	1.99	1.90	1.91	1.86	1.83
$2.5 \cdot 10^{-3}$	2.02	1.99	1.95	1.78	1.78	1.78
$1.25 \cdot 10^{-3}$	2.00	2.00	1.99	1.64	1.64	1.64

Table 3: Estimated order of accuracy, standard method, test problem 2.

h	Estimated order of accuracy					
	Upstream			Downstream		
	ρ	ρu	E	ρ	ρu	E
$1 \cdot 10^{-2}$	2.06	2.02	1.95	2.37	2.36	2.20

Table 4: Estimated order of accuracy, matrix viscosity method, test problem 2.

References

[1] John D. Andersson, Jr. *Computational Fluid Dynamics.* McGraw-Hill, Inc., 1995.

[2] Jay Casper and Mark H. Carpenter. Computational Considerations for the Simulation of Shock-Induced Sound. *SIAM Journal of Scientific Computing*, 19(3):813–828, 1998.

[3] Gunilla Efraimsson and Gunilla Kreiss. Approximate Solutions to Slightly Viscous Conservation Laws. Technical Report TRITA-NA 0140, NADA, KTH, November 2001.

[4] Gunilla Efraimsson, Jan Nordström, and Gunilla Kreiss. Artificial Dissipation and Accuracy Downstream of Slightly Viscous Shocks. American Institute of Aeronautics and Astronautics Paper 2001-2608, June 2001.

[5] Bjorn Engquist and Björn Sjögreen. The Convergence Rate of Finite Difference Schemes in the Presence of Shocks. *SIAM Journal of Numerical Analysis*, 35:2464–2485, 1998.

[6] Jonathan Goodman and Andrew Majda. The Validity of the Modified Equation for Nonlinear Shock Waves. *Journal of Computational Physics*, 58:336–348, 1985.

[7] Smadar Karni and Sunčica Čanić. Computations of Slowly Moving Shocks. *Journal of Computational Physics*, 136:132–139, 1997.

[8] Gunilla Kreiss, Gunilla Efraimsson, and Jan Nordström. Elimination of First Order Errors in Shock Calculations. *SIAM Journal of Numerical Analysis*, 38(6):1986–1998, 2001.

[9] Heinz-Otto Kreiss and Jens Lorenz. *Initial-Boundary Value Problems and the Navier-Stokes Equations.* Academic Press, Inc., 1989.

[10] Randall J. LeVeque. *Numerical Methods for Conservation Laws.* Birkhäuser Verlag, 1992.