UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Numerical Approximation of Hyperbolic Conservation Laws with Stiff Terms

Bjorn Engquist
Bjorn Sjogreen

March 1989

CAM Report 89-07

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

Numerical Approximation of Hyperbolic Conservation Laws with Stiff Terms

Björn Engquist[1]

Björn Sjögreen[2]

Abstract

Standard numerical shock capturing methods introduce unphysical states at discontinuities. These states do not cause any problems in the approximation of regular fluid flow. If the flow is coupled to chemical reactions there are often stiff source terms in the corresponding differential equations. It is then a substantial risk that the unphysical values will produce spurious numerical shocks. In this paper a new computational procedure is introduced which avoids this difficulty for a certain class of hyperbolic conservation laws with stiff lower order terms. The procedure is based on difference approximations with a projection step which eliminates the unphysical states. Numerical examples in one and two space dimensions are presented.

1. Introduction. The process of chemical reactions between different gases coupled to fluid motion contains many different scales in time and space. If the effect of all scales are not well represented by a numerical approximation there is a substantial risk of large errors. One type of error consists of numerical discontinuous

Research supported by ONR-URI Grant N00014-86-K-0691, NSF Grant DMS 88-11863, and NASA Consortium NCA2-372.

[1]Dept. of Math., UCLA, Los Angeles, CA 90024.

[2]Dept. of Scientific Computing, Uppsala University, Sturegat. 4B, S-75223 Uppsala, Sweden.

fronts moving with the wrong velocity, [3], [8].

We shall restrict our study to simple models of the equations of inviscid gas dynamics which are coupled to lower order source terms representing chemistry. In this paper we shall present a new class of algorithms which can overcome some of the inherent difficulties in many standard types of schemes. The focus here will be on the design and numerical tests of the algorithms for equations in one and two space dimensions.

Consider the model problem of a scalar hyperbolic conservation law with a stiff lower order term,

$$u_t + f_1(u)_x + f_2(u)_y = g_M(u). \tag{1.1}$$

One major difficulty in the numerical approximation of (1.1) occurs when the solution $u(x, y, t)$ develops discontinuities and when at the same time $|g'_M(u)| = O(M)$ is large.

Even without the lower order term, the equation,

$$u_t + f_1(u)_x + f_2(u)_y = 0, \tag{1.2a}$$

has solutions which generically develops discontinuous shock fronts. These solutions can be successfully approximated by modern shock capturing schemes, see e.g. [10].

In many practical problems the chemical reaction rate is relatively large. This rate is here represented by M. If there are no space derivatives the stiff ordinary differential equation (ode),

$$u_t = g_M(u), \tag{1.2b}$$

can be well approximated by stiff ode-solvers. The time step can be reasonably large and does not usually need to be of the order $1/M$.

It is thus natural to couple a shock capturing algorithm for (1.2a) to a stiff ode-solver for (1.2b). Even if the methods work well for the individual problems there are fundamental difficulties when they are coupled together in a fractional step algorithm. This is described in [3] by Colella, Majda, and Roytburd and in [8] by LeVeque and Yee.

In the presence of a discontinuity, a conservation form approximation of (1.2a) introduces values $\tilde{u}$ which are not in equilibrium of g ($g(\tilde{u}) \neq 0$), see Figure 1.1. The stiff solver then changes these $\tilde{u}$-values to an equilibrium. This interaction, in general, results in a solution which looks realistic but which contains fronts moving with the wrong velocity.

The above scenario can be avoided if the advection step (1.2a) does not introduce intermediate unphysical values $\tilde{u}$. Such an algorithm can, however, not be on conservation form locally for all timesteps. The random choice method avoids this dilemma by achieving conservation in average over a number of timesteps [1], [6]. In one dimension there are no unphyscial states introduced. This approach was successfully developed in our context by Chorin [2].

The so called subcell resolution methods also avoid the intermediate states $\tilde{u}$, see [7] by Harten. The fronts in $u(x, y, t)$ are kept sharp and a coupling to stiff ode-solvers is possible.

These two methods are, at present, basically one dimensional. The extensions to two dimensions have so far not been so successful.

The algorithm we shall propose here follows another approach. In [4] a projection or filter method was introduced in order to modify standard hyperbolic difference schemes to get high resolution shock capturing properties. Numerical overshoots at discontinuities were eliminated in a fractional step by an explicit projection on conservation form.

In the present algorithm the unphysical intermediate states $\tilde{u}$ are also eliminated in a separate fractional step between the approximation of (1.2a) and (1.2b). This procedure is not on conservation form at every timestep. The local conservation error is recorded and compensated for in the following timesteps. The new overall algorithm for equations of type (1.1) keeps the fronts sharp with the correct velocities. There is no restriction to one space dimension.

In the discussion above only capturing methods have been considered. Tracking the fronts by explicit curves or surfaces is, of course, also a possibility as well as the use of adaptive mesh refinement in order to resolve all the scales at the fronts. The fine grid resolution in time and space is very costly but needed if the effect of the interaction of different processes at the front is not known. The capturing procedure which we are proposing here is only appropriate for certain classes of problems. This is briefly discussed in connection to Examples 2 and 3 in Section 3, below.

2. The computational algorithm. We shall start from a given fractional step method for an equation of the type (1.1). The description of the algorithm is general and will apply to systems as well as more space dimensions,

$$u^{n+\frac{1}{2}} = F(u^n), \tag{2.1a}$$

$$u^{n+1} = G(u^n). \tag{2.1b}$$

The index n denotes the time level, ($t_n = n\Delta t$). The discrete vector u^n approximates $u(x, y, t_n)$ on a grid. The formula (2.1a) corresponds to a conservation form numerical approximation of (1.2a) and (2.1b) to a time step with a stiff ode-solver approximating (1.2b).

A third step is now added to (2.1),

$$u^{n+\frac{1}{3}} = F(u^n) \tag{2.2a}$$

$$u^{n+\frac{2}{3}} = P(u^{u+\frac{1}{3}}) \tag{2.2b}$$

$$u^{n+1} = G(u^{n+\frac{2}{3}}) \tag{2.2c}$$

As usual, the fractional indices do not correspond to actual time levels but only to different steps in the same algorithm. The projection step (2.2b) modifies the solution of (2.2a) at some mesh points in space in order to satisfy certain criteria to be discussed later. The criteria should be designed in order to eliminate unphysical values. This modification may violate the conservation property. The conservation error must be saved and compensated for in later timesteps. A more accurate formulation of (2.2b) is the following,

$$(u^{n+\frac{2}{3}},\ I^{n+1},\ C^{n+1}) = P(u^{n+\frac{1}{3}},\ I^n,\ C^n), \tag{2.3}$$

where I^n is a set of indices representing mesh points in space. These are the points for which the earlier conservation errors are stored. The vector C^n with these errors is thus defined over I^n.

The different steps for the evaluation of (2.3) are summarized as follows:

a) Find and index set J of mesh points which includes I^n and the points J' at the fronts of $u^{n+\frac{1}{3}}$,

$$J = I^n \cup J'.$$

The set J' is defined by the set of discrete inflection points of $u^{n+\frac{1}{3}}$ in each coordinate direction: the set of circles in Figure 2.1.

b) Decompose J into subsets,

$$J = \bigcup_{k=1}^{K} J_k, \quad J_i \cap J_j = \emptyset, \quad i \neq j$$

These subsets should be small and all should contain points from the boundary of J, see Figure 2.1.

c) Associate with each subset J_k a super and subfunction: $\bar{u}_k(x,y)$ and $\underline{u}_k(x,y)$.

d) Define the values of $u^{u+\frac{2}{3}}$ in each J_k at a given mesh point as either $\bar{u}_k$ or $\underline{u}_k$.

e) Save the conservation error in J_k at one or more mesh points. We shall call this set of points I_k. The vector C_k contains the sum of the previous conservation errors and the difference between $u^{n+\frac{1}{3}}$ and $u^{n+\frac{2}{3}}$, for points in J_k,

$$I^{n+1} = \bigcup_{k=1}^{K} I_k, \quad C^n = \bigcup_{k=1}^{K} C_k. \tag{2.4}$$

Let us first discuss the general purpose of the steps a) - e) and then comment on some of the details.

The overall strategy aims at localizing potential discontinuities and keeping them sharp without spurious intermediate numerical values. This will be done without any assumptions of the geometry of the discontinuities. Finally the correct conservation form is imposed by the algorithm over the time-space domain.

The set J contains all critical mesh points which are close to discontinuities or where there were conservation errors in the previous time step. This is the set in

which the unphysical $\tilde{u}$ values may appear. The set J is divided into small subsets J_k since the conservation and the definition of acceptable states should be relatively localized in space. The different sets J_k are defined recursively in b). For a given partial sequence of sets J_k, $k = 1, \ldots, p$ find a new index $(i,j) \in J$, $(i,j) \notin \bigcup_{k=1}^{p} J_k$. Either extend the definition of one of the sets J_k, $k = 1, \ldots, p$ to include (i,j) or define $J_{p+1} = \{(i,j)\}$. A penalty function restricts the number of elements per J_k.

The sets J_k are also used for the definition of $\bar{u}_k$ and $\underline{u}_k$ in c). These functions can be chosen as extrapolations from mesh points outside of J. There should be extrapolations from each side of a front close to J_k. Another possible way of defining the super and subfunctions are as local stable equilibria of g. In the numerical examples below these functions were defined from first order extrapolations.

In the step d) the values of $u^{n+\frac{2}{3}}$ in J_k are defined to be either $\bar{u}_k$ or $\underline{u}_k$ such that the sum of the conservation errors and the differences $|u^{n+\frac{1}{3}} - u^{n+\frac{2}{3}}|$ is minimal.

3. Numerical examples. Consider first the linear advection of a discontinuity in two space dimensions. The problem,

$$
\begin{aligned}
&u_t + u_x + u_y = Mu(1-u)(1+u), \\
&u(x,y,0) = u_0(x,y) = \begin{cases} 1 & x+y \le 1, \\ -1 & x+y > 1, \end{cases}
\end{aligned}
\tag{3.1}
$$

has the solution,

$$
u(x,y,t) = \begin{cases} 1 & x+y \le 1+2t, \\ -1 & x+y > 1+2t. \end{cases}
$$

In the numerical computations we used $M = 10^6$ to get a problem with a very stiff source term. The computational domain was $[0,1] \times [0,1]$ with a uniform discretization space,

$$
x_i = (i-1)\Delta,\ i = 1, \ldots, m, \quad y_j = (j-1)\Delta,\ j = 1, \ldots, m, \quad (m-1)\Delta = 1.
$$

A solution was first found using a standard fractional step method. The first step consists of solving the advection part with the upwind difference scheme,

$$u_{i,j}^{n+\frac{1}{2}} = u_{i,j}^{n} - \Delta t D_{-}^{x} u_{i,j}^{n} - \Delta t D_{-}^{y} u_{i,j}^{n},$$

where $D_{-}^{x} a_{i,j} = (a_{i,j} - a_{i-1,j})/\Delta$ and $D_{-}^{y} a_{i,j} = (a_{i,j} - a_{i,j-1})/\Delta$. The mesh function $u_{i,j}^{n}$ is supposed to approximate $u(x_i, y_j, t_n)$. Boundary values are given at inflow ($x = 0$, and $y = 0$) and extrapolation is used at outflow. In the second step the equation for the right hand side was solved with the implicit Euler method,

$$u_{i,j}^{n+1} = u_{i,j}^{n+\frac{1}{2}} + \Delta t M u_{i,j}^{n+1}(1 - u_{i,j}^{n+1})(1 + u_{i,j}^{n+1}).$$

The pointwise form makes it unnecessary to supply boundary values after this step. The method has formal order of accuracy one, which is not ideal. It is however sufficient for the purpose of this paper.

All numerical solutions for this problem are displayed at $t = 0.2$ and computed with $m = 40$. The CFL-number λ is defined as $\Delta t/\Delta$. We begin with demonstrating the incorrect dependence of wave speed on the CFL-number. In Fig. 3.1 the solution was computed with $\lambda = 0.25$ and in Fig 3.2 with $\lambda = 0.6$.

Next we show the result from the same method, but with the projection inserted after the advection step. Figs. 3.3 and 3.4 show that the wave speed obtained when the projection is used is the correct one. The front is also much sharper than in the Figures 3.1 and 3.2. Here Figure 3.3 displays a solution computed with $\lambda = 0.25$ and Figure 3.4 with $\lambda = 0.6$.

The second test problem is

$$\begin{aligned} & u_t + (u^2/2)_x + (u^2/2)_y = Mu(1-u)(u-1/2) \\ & u(x,y,0) = u_0(x,y) = \begin{cases} 1 & x^2 + y^2 \leq 0.25 \\ 0 & x^2 + y^2 > 0.25 \end{cases} \end{aligned} \tag{3.2}$$

This problem is interesting because of the nonlinear flux functions and the curved shock front. We keep the two step solution method from the linear problem, and use an upwind approximation [5] of the derivatives of the flux functions. In the exact solution, the shock front is moving with speed = 0.5. At $t = 0.4$ the front should cross the x-axis at $x = 0.7$ and the y-axis at $y = 0.7$.

The solution at $t = 0.4$ was computed, using $m = 50$. Figure 3.5 shows the solution obtained without using the projection step with $\lambda = 0.6$. The stiffness of the right hand side does not permit the shock to move. The boundary conditions were given as the exact solution at inflow and extrapolated at outflow. When the projection was inserted after the advection step the solution in Figure 3.6 was obtained. The CFL-number used there was 0.6. Again one can see how well the projection corrects the shock speed. In the projection algorithms used above, the conservation error was stored at one gridpoint per J_k. The index for this point was chosen to correspond to the largest value of $|u^{n+\frac{1}{3}} - u^{n+\frac{2}{3}}|$. If the local conservation error was distributed evenly over J_k the resulting solution was almost the same as in Figure 3.6.

If the solution of (3.1) with $M = 0$ contains a rarefaction wave the velocity of a corresponding front for large M-values depends explicitly on $g_M(u)$. The projection step can be modified to handle such a situation. This will not be considered here since this case is not of combustion type.

Next we consider a model problem for fluid flow with chemical reactions,

$$u_t + (u^2/2 - q_0 Z)_x = 0, \tag{3.3a}$$

$$Z_x = K\phi(u)Z. \tag{3.3b}$$

The initial data for u, and the value of Z as $x \to \infty$, are given,

$$u(x,0) = u_0(x)$$

$$\lim_{x\to\infty} Z = 1.$$

This problem was introduced by Majda [9]. In [3] the existence of waves travelling at one grid point per time step were proved for a certain numerical method.

The problem is a simplified version of the Euler equations for a burning gas. The function Z can be thought of as describing the fraction of unburnt gas. K and q_0 are constants and $\phi(u)$ is a function describing how the reaction rate depends on u. In our experiments we considered,

$$\phi(u) = \begin{cases} 1 & u > u_I, \\ 0 & u \le u_I, \end{cases}$$

where u_I can be thought of as an ignition temperature. The problem is one dimensional but contains some additional difficulties to the pure stiff source term problems above.

We shall approximate also this problem with a fractional step method. In the first step, the advection part,

$$u_t + (u^2/2)_x = 0,$$

is solved. This is done with an upwind approximation in the same way as in the previous example. The difficulty appears in second fractional step. We are trying to solve the equation,

$$u_t = q_0 Z_x, \tag{3.4}$$

where Z is obtained as the solution of the ode problem,

$$\begin{aligned} Z_x &= K\phi(u)Z, \\ Z &= 1, \quad x \to \infty. \end{aligned} \tag{3.5}$$

The vector Z_i is given by the following approximation of the exact solution to (3.5),

$$Z_{i-1} = Z_i e^{-K\Delta x/2(\phi(u_i)+\phi(u_{i-1}))}, \tag{3.6}$$

where $Z_i = 1$ for i sufficiently large. If the equation (3.4) is approximated by the backward Euler method and the exact expression for Z_x is used, we get,

$$u_i^{n+1} = u_i^n + \Delta t q_0 K \phi(u_i^{n+1}) Z_i^{n+1}. \tag{3.7}$$

If now $K\Delta x$ is large so that the layer where Z_i changes from 1 to 0 can not be resolved on the grid the right hand side, $\phi(u_i^{n+1})Z_i^{n+1}$, will always be zero and a shock will move with the speed given by Burger's equation, $s = 0.5$ $(u_L \neq u_R)$. The right hand side will never influence the grid values. This difficulty can not be overcome by keeping the shock free from intermediate points, as was done in the examples with stiff source terms. We can however replace the representation (3.6) of Z_x in (3.7) with a conservation form finite difference approximation.

$$u_i^{n+1} = u_i^n + \Delta t q_0 (Z_{i+1}^{n+1} - Z_{i-1}^{n+1})/(2\Delta x). \tag{3.8}$$

The transition in Z_i will now correctly contribute to the solution.

We begin with a numerical example showing the difference between the procedures in (3.7) and (3.8).

In the numerical examples we used the initial data,

$$u(x,0) = u_0(x) = \begin{cases} 1.363325 & x \leq 0 \\ -1.5 & x > 0 \end{cases}, \tag{3.9}$$

and ignition value $u_I = 0$. The constant K determines the width of the transition region in Z (Z changes from 0 to 1 like $e^{-(Kx/2)}$) and was given the value 100 or 1. The constant q_0 was 2.2.

In Table 3.1 the location of the shock at time = 4 is given for a sequence of refined grids. The exact location is $x = 2.8$. For Burger's equation and the initial data (3.4), the shock position is $x = -0.27$ at this time. We can see how the shock position becomes more accurate with decreasing Δx, and how much better the method (3.8) works. The exact solution is displayed in Figure 3.7. With (3.7) and sufficiently coarse mesh, the shock propagates, as expected, like a pure hydrodynamic shock. The projection step was introduced, as usual, after the advection step, but did not substantially affect the shock speed.

We now turn the other difficulty associated with this problem. In [3] conditions for existence of unphysical waves in the numerical solution were given. The ignition temperature was required to be sufficiently close to the right hand side state. We therefore changed u_I to -1.45. This value, together with the initial data (3.9), satisfies the conditions given in [3].

The numerical solution to this problem without using the projection step is shown in Figure 3.8. Here a stiffness problem, similar to the previous ones in examples 1 and 2, appears in the Z component of the solution. The advection step increases the gridpoint value to the right of the shock above u_I which immediately causes (due to the large value of K) Z to jump from 1 to ≈ 0. In this way the wave in Z will propagate with one gridpoint per timestep. This shows up as the rightmost shock in Figure 3.8. The discontinuity in u does not necessarily follow the discontinuity in Z, and in fact we can see in Figure 3.8 how the shock in u have separated itself from the discontinuity in Z. The solution is not the correct one. This is the type of problem which can be handled by our present technique. With the projection introduced we obtain the solution in Figure 3.9. The stiffness problem does not occur here, since the projection prevents intermediate $\tilde{u}$ values in the shock.

Bibliography

[1] A. Chorin, "Random choice solution of hyperbolic systems," J. Comput. Phys., **22** (1976), pp. 517-533.

[2] A. Chorin, "Random choice methods with applications for reacting gas flow," J. Comput. Phys., **25** (1977), pp. 253-272.

[3] P. Colella, A. Majda, and V. Roytburd, "Theoretical and numerical structure for reacting shock waves," SIAM J. Sci. Stat. Comput., **7** (1986), pp. 1059-1080.

[4] B. Engquist, P. Lotstedt, and B. Sjogreen, "Nonlinear filters for efficient shock computation," Math. Comp., **52** (1989), pp. 509-537.

[5] B. Engquist and S. Osher, "Stable and entropy condition satisfying approximations for transonic flow calculations," Math. Comp., **34** (1980), pp.45-75.

[6] J. Glimm, "Solutions in the large for nonlinear hyperbolic systems of equations," Comm. Pure Appl. Math., **18** (1965), pp. 697-715.

[7] A. Harten, "ENO schemes with subcell resolution," ICASE Report 87-65 (1987).

[8] R.J. LeVeque, and H.C. Yee, "A study of numerical methods for hyperbolic conservation laws with stiff source terms," J. Comput. Phys., to appear.

[9] A. Majda, "A qualitative model for dynamic combustion," SIAM J. Appl. Math., **41** (1981), pp. 70-93.

[10] P. Woodward, and P. Colella, "The numerical simulation of two-dimensional fluid flow with strong shocks," J. Comput. Phys., **55** (1984), pp. 115-173.

Figure 1.1: Numerical shock with equilibrium states u_1, u_2.

Figure 2.1: Decomposition of mesh points at a front (cricles $\in J^1$) into clusters J_k.

Figure 3.1: Example (3.1); upwind scheme, $\lambda = 0.25$.

Figure 3.2: Example (3.1); upwind scheme, $\lambda = 0.6$.

Figure 3.3: Example (3.1); upwind scheme with projection, $\lambda = 0.25$.

Figure 3.4: Example (3.1); upwind scheme with projection, $\lambda = 0.6$.

Figure 3.5: Example (3.2); upwind scheme.

Figure 3.6: Examples (3.2); upwind scheme with projection.

Table 3.1: Example (3.3); shock locations, $K = 100$.

Figure 3.7: Example (3.3), method (3.7), $K = 1$, $U_I = 0$.

Figure 3.8: Example (3.3), method (3.7), $K = 100$, $U_I = -1.45$.

Figure 3.9: Example (3.3), method (3.7) with projection, $K = 100$, $U_I = -1.45$.

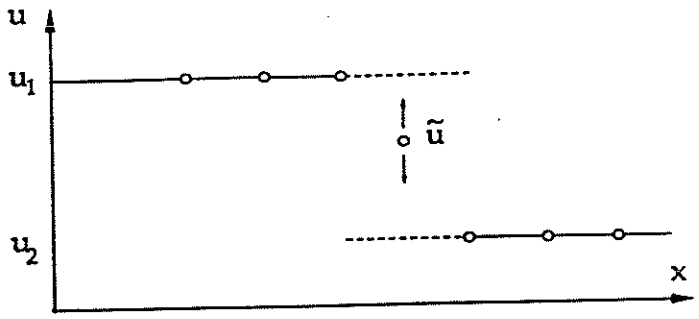

Figure 1.1

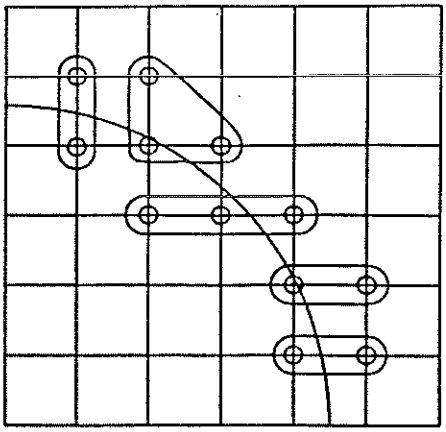

Figure 2.1

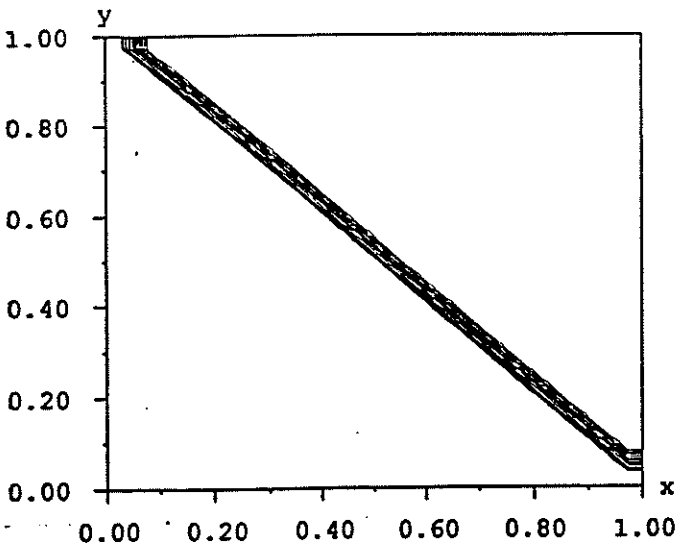

Figure 3.1

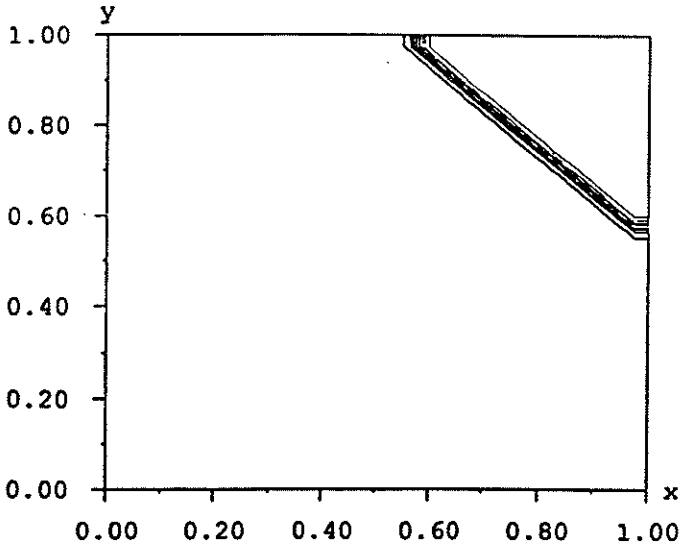

Figure 3.2

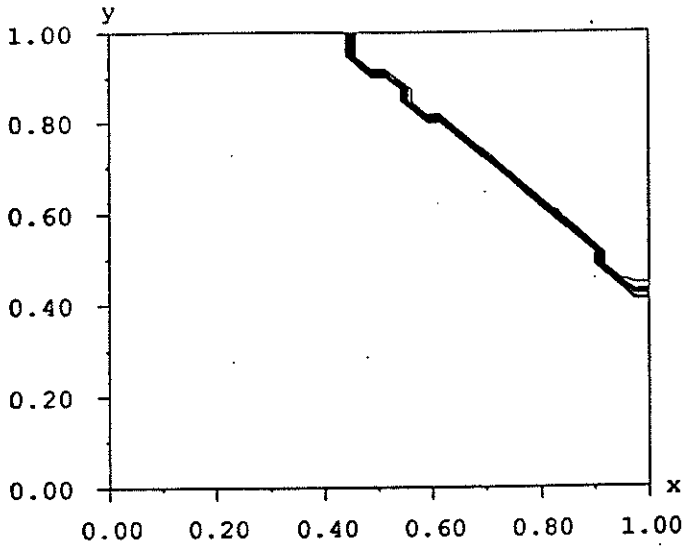

Figure 3.3

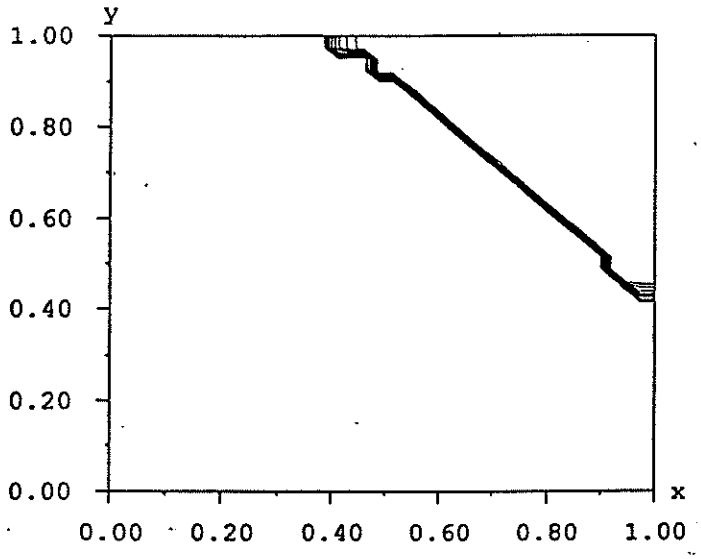

Figure 3.4

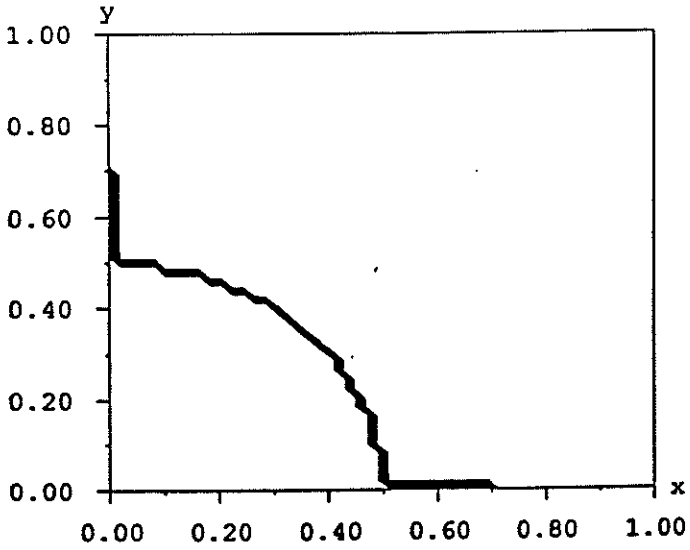

Figure 3.5

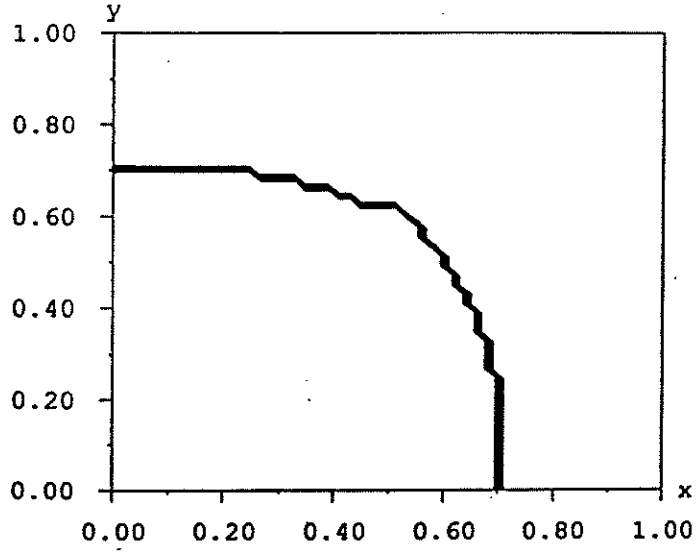

Figure 3.6

m	(3.6)	(3.8)	(3.6) + proj.
20	-0.2	2.7	-0.3
100	0.7	2.8	0.6
400	2.5	2.8	2.5
800	2.8	2.8	2.8

Table 3.1

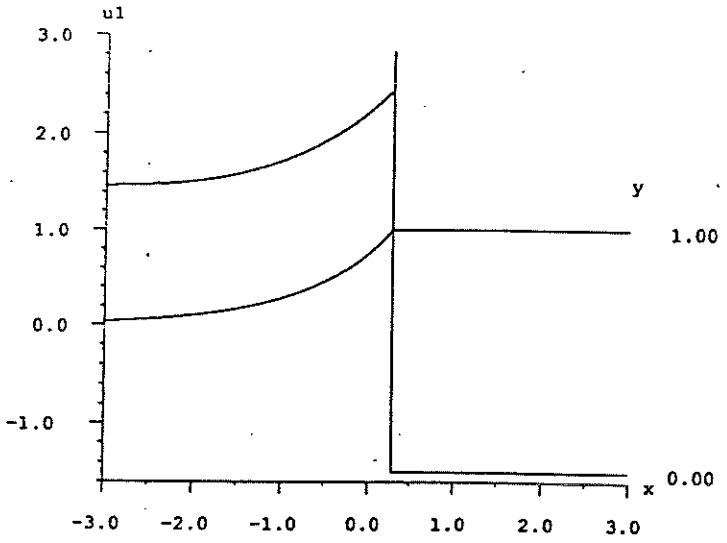

Figure 3.7

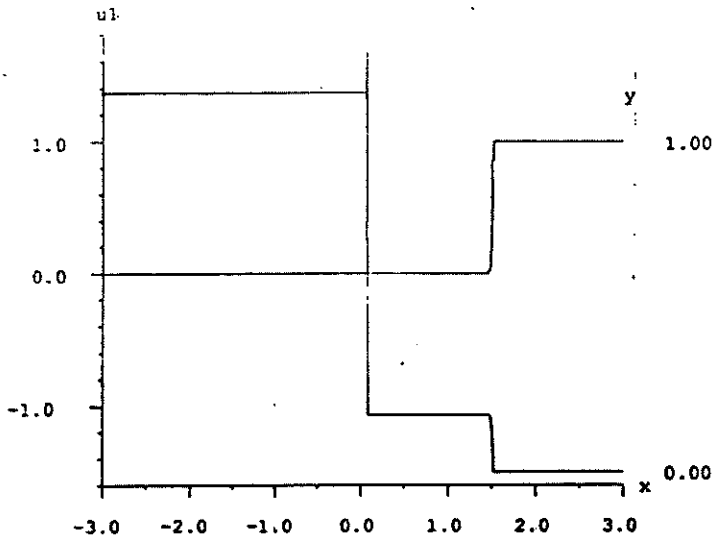

Figure 3.8

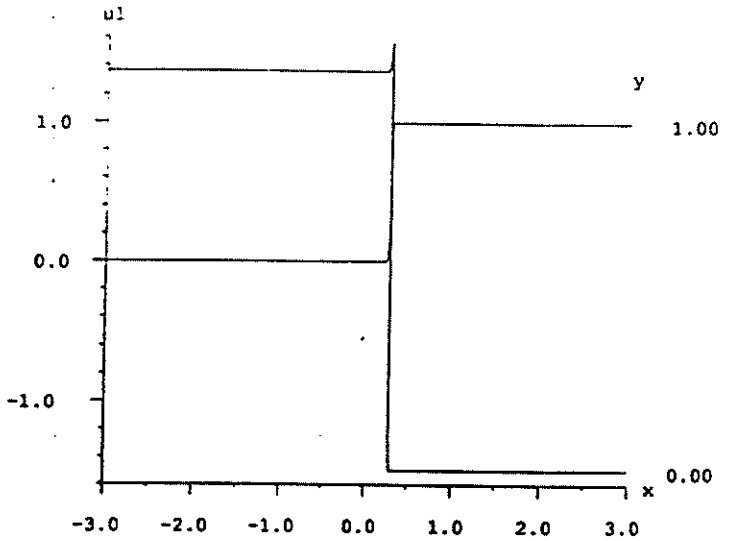

Figure 3.9