Lectures on
Numerical Methods For Time Dependent Equations
Applications to Fluid Flow Problmes

By

P. Lascaux

Tata Institute of Fundamental Research
Bombay
1976

Lectures on
Numerical Methods For Time Dependent Equations
Applications to Fluid Flow Problmes

By

P. Lascaux

Notes By

S. Kesavan, M. Vanninathan

Tata Institute of Fundamental Research
Bombay
1976

Printed in India By
Anil D. Ved At Prabhat Printers. Bombay 4000 004

And Published By
The Tata Institute of Fundamental Research.

Foreword

The solution of time dependent equations of hydrodynamics is a subject of great importance. Except for some very particular cases, the solution cannot be obtained in an analytic form which, in passing, causes difficulties when onw wishes to test a numerical method because one only has very few solutions, chiefly related to 2-dimensional problems.

For the 1-dimensional problems, the numerical methods studied in these notes are the method of characteristics and the method of finite differences. Unfortunately, we had not much time to treat 2-dimensional problems. But the last chapter is an introduction of the method of finite elements which one can utilise for solving them.

Here we have essentially restricted ourselves to non-viscous fluids and we have especially studied the cases of propagation of shocks. To resolve this problem, we have presented a method of "shock-fitting" if one uses the method of characteristics and a method of pseudo-viscosity if one wants to use the method of finite differences.

The course is mainly concentrated on the study of the stability of the various schemes. We have considered only the stability for linearised problems. A rigorous analysis in the nonlinear case is impossible at the present moment.

In the first part of the course, we have chosen to study the schemes for the three particularly simple model equations (in one-space variable and in one time-variable): the heat equation, the wave equation, and the advection equation.

To do this, we have first introduced the mathematical notions of the hyperbolic system of equations, weak solutions of the equations, energy

inequalities and the boundary conditions for the problem to be well-posed.

Next, we have studied the consistency and stability of sufficiently large number of schemes by obtaining energy inequalities using the Fourier transform.

In the second part of the course, we have shown the practical application of these numerical methods to the solution of the equations of hydrodynamics.

Thereforem, this course covers only a very small portion of the vast subject of the Numerical approximation of the equations of Fluid Mechanics. The interested reader can refer to the various articles and works given in the references. He can also find a very complete bibliography in the book of ROACHE. Every two years, a Congress on the Numerical methoda of Fluid mechanics takes place whose Proceedings furnish precise details. Numerous articles on this subject are published in the following 2 reviews:

Physics of Fluids and especially Journal of Computational Physics.

I thank all those who have enabled me to deliver this course, in particular Professor K.G. Ramanathan and Professor R. Sridharan of T.I.F.R., Bombay. My thanks go to Messrs. Kesavan and Vanninathan who have written these notes with great clarity and within a very short time. I appreciate the discussions I had with them and with some of their collegues at the Indian Institute of Science, Bombay.

I conclude by saying that India is a wonderful country which has won my heart and I hope to have another opportunity to visit it again.

P. Lascaus

Contents

Chapter 1

The equations of fluid dynamics

1.1 Introduction

In this section we merely write down the basic equations of fluid dynamics involving one space variable and time. (This occurs, for instance, in the study of the flow of a gas in a narrow cylindrical tube where the state of flow is constant across any cross section and so depends only on the linear coordinate measured along the axis of the tube, and on time). From these general equations, we write down three simple particular equations whose properties will then be studied in the sequel. We will return to the general equations in section 9. 1

1.2 Notations

To start with we put down the various notations which will be used in writing these equations. We will denote by ρ, the density of the fluid; by $V = \frac{1}{\rho}$, the specific volume; by p, the pressure; by q, the pseudo-viscosity term; by ε, the internal energy per unit of mass; by E, the total energy per unit of mass; by u, the velocity of the fluid; by T the absolute temperature.

One has the relation

$$E = \varepsilon + \frac{1}{2}u^2 \tag{1.1}$$

We denote by μ, the coefficient of viscosity and by k, the coefficient of conductivity of the relevant fluid,

The quantities ρ, p, ε and T are thermodynamical quantities and they are related by the *equations of state:*

$$\left.\begin{aligned} p &= p\,(\rho, T) \\ \varepsilon &= \varepsilon\,(\rho, T) \end{aligned}\right\} \tag{1.2}$$

For instance, in the case of a perfect fluid the equations (1.2) assume the form

$$\left.\begin{aligned} p &= R\,\rho\,T \\ &= \frac{RT}{\gamma - 1} \end{aligned}\right\} \tag{1.3}$$

2 where R is the universal gas constant and γ the constant ratio of specific heats.

1.3 Coordinate systems

In writing down the equations of fluid dynamics we express in mathematical form the following three laws: the law of conservation of mass, the law of conservation of quantity of movement (i.e. momentum) and the law of conservation of energy.

One way write these equations in several equivalent forms. Chiefly, one uses the two types of coordinate systems described below.

(i) The Eulerian System. Here the independent variables are x and t, where t is the time and x is the position of a point in space with reference to a frame fixed in the laboratory.

(ii) The Lagrangian System. We now have the independent variables a and t where t is as in (i). Now a is the position at time $t = 0$, of the particle which is at position $x = x(a, t)$ at time t.

One assumes that the particles do not cross one another at any instant. In other words, for every t, the transformation $a \mapsto x(a,t)$ is invertible. If we denote by J the Jacobian of the transformation, i.e.

$$J = \frac{\partial x}{\partial a}. \tag{1.4}$$

then $J \neq 0$ everywhere.

Given any physical quantity in one system we can always express it in the other using this transformation. Thus we have

$$f(x,t) = f(x(a,t),t) = f(a,t). \tag{1.5}$$

The derivative of $\bar{f}$ w.r.t t is given in terms of the derivatives of f by the relation 3

$$\frac{\partial \bar{f}}{\partial t} = \frac{\partial t}{\partial t} + u\frac{\partial f}{\partial x} \tag{1.6}$$

where u, the derivative of x w.r.t. t, is the velocity of the fluid particle at time t which is at position x. The relation (1.6) leads to the following

Definition 1.1. The Lagrangian (or particular) time derivative of a function $f(x,t)$ is given by

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + u\frac{\partial f}{\partial x} \tag{1.7}$$

1.4 The equations in eulerian system

We now write down the equations of fluid dynamics in Eulerian form, in the slab symmetric case. The equations of one-dimensional cylindrical or spherical symmetric flows will assume different forms. The derivations of these equations can be found in any standard text on fluid dynamics.

E1. Conservation of Mass

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \tag{1.8}$$

E2. Conservation of Momentum

$$\frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}\left(\rho u^2 + p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right) = g\rho \tag{1.9}$$

where g is the volume acceleration applied from the exterior of the system.

E3. Conservation of Energy

$$\frac{\partial}{\partial t}(\rho E) + \frac{\partial}{\partial x}\left(\rho u E + pu - \frac{4}{3}\mu u\frac{\partial u}{\partial x}\right) = \rho g u + \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right). \tag{1.10}$$

4

Exercise 1.1. (a) Starting from (E1) and (E2) show that one can write (E2) also as

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{1}{\rho}\frac{\partial}{\partial x}\left(p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right) = g. \tag{E2′}$$

(b) Using (E1) and (E2′) show that one can rewrite (E3) as

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x}(\rho u\varepsilon) + p\frac{\partial u}{\partial x} - \frac{4}{3}\mu\left(\frac{\partial u}{\partial x}\right)^2 = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) \qquad \text{E3′}$$

or as

$$\frac{\partial \varepsilon}{\partial t} + u\frac{\partial \varepsilon}{\partial x} + \frac{1}{\rho}\left(p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right)\frac{\partial u}{\partial x} = \frac{1}{\rho}\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right). \qquad \text{E3″}$$

Remark 1.1. Equations (E3′) and (E3″) give the law of conservation of energy in terms of the *internal* energy while (E3) gives the same in terms of the total energy.

Remark 1.2. Setting

$$\bar{U} = \begin{bmatrix} \rho \\ \rho u \\ \rho E \end{bmatrix}$$

one can put the equations $E1$, $E2$ and $E3$ into a single vector equation

$$\frac{\partial \bar{U}}{\partial t} + \frac{\partial}{\partial x}\left(\bar{F}\left(\bar{U}, \frac{\partial \bar{U}}{\partial x}\right)\right) = \bar{G}(\bar{U}), \tag{1.11}$$

where

$$\bar{F} = \begin{bmatrix} \rho u \\ \rho u^2 + p - \frac{4}{3}\mu \frac{\partial u}{\partial x} \\ \rho u E + pu - \frac{4}{3}\mu u \frac{\partial u}{\partial x} - k\frac{\partial T}{\partial x} \end{bmatrix}$$

and

$$\bar{G} = \begin{bmatrix} 0 \\ \rho g \\ \rho g u \end{bmatrix}.$$

This is known as the *Conservation form* of the equations and is quite useful. 5

1.5 The equations in the lagrangian system

For problems with free surfaces or for solution with shocks, the Lagrangian form of the equations is more useful. As we will see presently, the Largrangian form of the equations does not contain advective terms like $\frac{\partial}{\partial x}(\rho u^2)$ and $\frac{\partial}{\partial x}(\rho u E)$ which have been found difficult to approximate in numerical methods.

We now give the Lagrangian form of these equations.

L1. Conservation of Mass

$$\frac{D\rho}{Dt} + \rho \frac{\partial u}{\partial x} = 0. \tag{1.12}$$

Since one can check that $\frac{1}{J}\frac{DJ}{Dt} = \frac{\partial u}{\partial x}$, we can write (1.12) equivalently as

$$\frac{D}{Dt}(\rho J) = 0. \tag{1.12′}$$

(**Note:** To be strictly Lagrangian in our formulation, we must omit usage of $\frac{\partial}{\partial x}$. One should replace it by $\frac{1}{J}\frac{\partial}{\partial a}$).

L2. Conservation of Momentum

$$\rho\frac{Du}{Dt} + \frac{\partial}{\partial x}\left(p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right) = \rho g \tag{1.13}$$

L3. Conservation of Energy

$$\frac{D\varepsilon}{Dt} + \left(p - \frac{4}{3}\mu\frac{\partial u}{\partial x}\right)\frac{D}{Dt}\left(\frac{1}{\rho}\right) = \frac{1}{\rho}\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right). \tag{1.14}$$

6 **Exercise 1.2.** Starting from the Eulerian form of the equations, derive the equations $L1$, $L2$, and $L3$.

Having written down these general equations we write down three model equations which arise out of these.

1.6 The advection equations

This is the law of conservation of mass all over again:

$$\frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = 0.$$

As a particular case, suppose u is a constat. Then the equation becomes

$$\frac{\partial\rho}{\partial t} + u\frac{\partial\rho}{\partial x} = 0. \tag{1.15}$$

Then one can readily check that

$$\rho(x,t) = \rho(x - ut, 0), \tag{1.16}$$

satisfies (1.15). Thus the value of ρ at any point (x,t) is determined by the value at the point $(x - ut, 0)$. Thus ρ is constant along the lines $x - ut =$ constant and these are the *characteristic curves* of (1.15). (Cf. Sec.2).

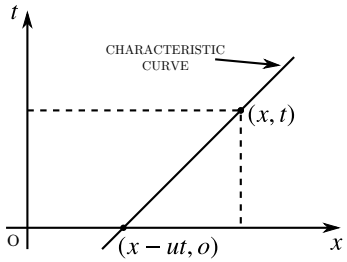


Figure 1.1:

Exercise 1.3. Find an analytic expression for $\rho(x,t)$ in terms of $\rho(x,0)$ and $u(x,t)$ when u is not longer constant. 7

Remark 1.3. As such that advection equation is not difficult to solve exactly. However when coupled with the other equations of fluid dynamics difficulties arise and one looks for an efficient numerical scheme of approximation. But one must be cautious in the choice of such a scheme or else a "diffusion process" is likely to be introduced into the approximate equation while no such thing exists in the exact case. We will se this later.

1.7 The wave equations

We specialize to the case when $\mu = k = 0$, and assume that g is negligible. Then L3 becomes

$$\frac{D\varepsilon}{Dt} + p\frac{D}{Dt}\left(\frac{1}{\rho}\right) = 0. \tag{1.17}$$

This together with the equation of state (Cf. (1.2)) can be integrated to give a relationship between p and ρ. For instance, in the case of a perfect gas, we get $p\rho^{-\gamma} = \varphi(a)$, $\varphi(a)$ being a constant if the initial state of the fluid is constant. In this case the equations L1 and L2 read as

$$\left.\begin{aligned} &\frac{D\rho}{Dt} + \rho\frac{\partial u}{\partial x} = 0 \\ &\frac{Du}{Dt} + \frac{1}{\rho}\frac{\partial}{\partial x}(p(\rho)) = 0. \end{aligned}\right\} \tag{1.18}$$

Writing in vector form, we get

$$\frac{D}{Dt}\begin{bmatrix}\rho\\ u\end{bmatrix}+\begin{bmatrix}0 & \rho\\ \frac{p'(\rho)}{\rho} & 0\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\rho\\ u\end{bmatrix}=0 \tag{1.19}$$

As such this equation is non-linear. If one wants to study small
8 perturbations around a constant state of the fluid at rest at time $t = 0$, on e can linearize the state equation as

$$p - p_\circ = C_\circ^2(\rho - \rho_\circ) \tag{1.20}$$

where $p_\circ, \rho_\circ, C_\circ$ are respectively the pressure, density and sound-speed of the constant state at $t = 0$. By neglecting second order terms one can then write

$$\frac{D}{Dt}\begin{bmatrix}\rho\\ u\end{bmatrix}+\begin{bmatrix}0 & \rho_\circ\\ C_\circ^2/\rho_\circ & 0\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\rho\\ u\end{bmatrix}=0. \tag{1.21}$$

Setting

$$\varphi = \frac{\rho}{\rho_\circ} + \frac{u}{C_\circ}$$
$$\psi = \frac{\rho}{\rho_\circ} - \frac{u}{C_\circ},$$

we get, on substitution into (1.21),

$$\left.\begin{aligned}\frac{\partial\varphi}{\partial t} + C_\circ\frac{\partial\varphi}{\varphi x} = 0\\ \frac{\partial\psi}{\partial t} - C_\circ\frac{\partial\psi}{\partial x} = 0\end{aligned}\right\} \tag{1.22}$$

which are of the advective type. We know that φ, ψ assume the form

$$\left.\begin{aligned}\varphi(x,t) = \varphi(x - C_\circ t, 0)\\ \psi(x,t) = \psi(x + C_\circ t, 0).\end{aligned}\right\} \tag{1.23}$$

We now have two characteristic curves $x - C_\circ t =$ constant and $x + C_\circ t =$ constant. The value at (x,t) is now dependent on the values over the *finite* interval $[x - C_\circ t, x + C_\circ t]$ on the real axis (Cf. Fig. 1.2).

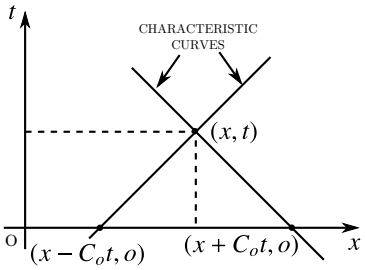


Figure 1.2:

1.8 The heat equation

We assume ρ to be constant and u to be zero. Then using an equation of state like (1.3), L3 will read as

$$\frac{\partial T}{\partial t} - \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) = 0. \tag{1.24}$$

In particular if k is a constant then we have

$$T(x,t) = \frac{1}{2\sqrt{k\pi t}}\int_{-\infty}^{+\infty} \exp\left(\frac{-(x-y)^2}{4kt}\right) T(y,0)dy.$$

This shows that unlike the advective or the wave equations, the value at (x,t) of the solution of the heat equation depends on the initial value, on the *entire* real line.

Our immediate aim is to study the mathematical properties of these three types of equations and methods of approximating them.

References: General references for the entire course are Richtmyer and Morton [32], Potter [30], Ames [2] and Mitchell [27]. For the equations of fluid dynamics and their properties one can refer to Courant and Friedrichs [9].

Chapter 2

Hyperbolic System

2.1 Introduction

In this section we define a first order hyperbolic system and its characteristic cruves. We study the characteristic form of a first order hyperbolic system and apply these ideas to the equations of hydrodynamics. **10**

Consider a system of n equations given by

$$\frac{\partial u_i}{\partial t} + \sum_{j=1}^{n} a_{ij}(u, x, t)\frac{\partial u_j}{\partial x} = 0, \quad 1 \leq i \leq n. \tag{2.1}$$

By setting

$$\left.\begin{aligned} A &= (a_{ij}),\ 1 \leq i, j \leq n \\ U^T &= (u_1(x), \ldots, u_n(x)) \end{aligned}\right\} \tag{2.2}$$

we can write (2.1) in the vector form as

$$\frac{\partial U}{\partial t} + A(U, x, t)\frac{\partial U}{\partial x} = 0. \tag{2.3}$$

Given a system of equations of the form (2.1) or equivalently (2.3) with A depending on x, U and t, we look for a curve C parametrised by $x = x(s)$, $t = t(s)$ so that along C some kind of differential relationship holds for U.

2.2 Characteristic form of a first order hyperbolic system

In the study of first order systems as in (2.1), the eigenvalues and eigenvectors of the matrix A play an important role. We start our discussion with the following:

Definition 2.1. The system of equations (2.1) is strictly hyperbolic if, and only if, for all values of its arguments, the matrix A always has n real and distinct eigenvalues.

11 We will henceforth consider only such systems which are strictly hyperbolic.

Remark 2.1. If A depends on U then the system of equations is nonlinear.

Let us now study the various cases involved in equation (2.3).

Case (i). Let us assume that the matrix A is constant. Let p^T be a left-eigen vector of A, i.e.

$$p^T A = \lambda p^T. \tag{2.4}$$

Since A is a constant matrix, λ is a constant and p^T is a constant row-vector. Using these facts and multiplying (2.3) on the left by p^T, we get

$$\frac{\partial}{\partial t}(p^T U) + \lambda \frac{\partial}{\partial x}(p^T U) = 0. \tag{2.5}$$

If we now define a plane curve C (which is easily seen to be a straight line in this case) by

$$\frac{dt}{ds} = 1; \ \frac{dx}{ds} = \lambda \tag{2.6}$$

or, equivalently,

$$\frac{dx}{dt} = \lambda \tag{2.6′}$$

We then get

$$\frac{d}{ds}(p^T U) = 0 \tag{2.7}$$

where the derivative w.r.t. s indicates differentiation along C. Thus $p^T U$ is constant along C and is called the *Riemann Invariant* of the system (2.1). The curve C is a *characteristic curve* of the system.

Case (ii). Let us assume A is purely a function of x and t. Again we choose p^T so that (2.4) holds. Note that now λ and p^T are also functions of x and t. If U is a solution of (2.3), we have **12**

$$\begin{aligned}\frac{\partial}{\partial t}(p^T U) + \lambda\frac{\partial}{\partial x}(p^T U) &= p^T\frac{\partial U}{\partial t} + \lambda p^T\frac{\partial U}{\partial x} + \frac{\partial p^T}{\partial t}U + \lambda\frac{\partial p^T}{\partial x}U \\ &= p^T\left(\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x}\right) + \frac{\partial p^T}{\partial t}U + \lambda\frac{\partial p^T}{\partial x}U \\ &= \frac{\partial p^T}{\partial t}U + \lambda\frac{\partial p^T}{\partial x}U.\end{aligned}$$

Thus if we again define C by (2.6) or ((2.6′)), (observe that C is no longer a straight line) we get the relation

$$\frac{d}{ds}(p^T U) = R^T U, \text{ along } C \tag{2.8}$$

where

$$R^T = \frac{\partial p^T}{\partial t} + \lambda\frac{\partial p^T}{\partial x}. \tag{2.9}$$

The relation (2.8) is the differential relation along the characteristic curve C.

Case (iii). Let A depend explicitly on U alone.

Let us choose p^T and λ to satisfy (2.4). Then $\lambda = \lambda(U)$, $p^T = p^T(U)$. We then get from (2.3) that

$$\begin{aligned}0 &= p^T\left(\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x}\right) \\ &= p^T\left(\frac{\partial U}{\partial t} + \lambda\frac{\partial U}{\partial x}\right) \\ &= p^T\frac{dU}{ds}, \text{ along } C\end{aligned}$$

where C is again given by (2.6). Thus we get the differential relation

$$p^T = \frac{dU}{ds} = 0 \text{ along } C. \tag{2.10}$$

13 We do not treat the most general case where $A = A(x, t, U)$.

Remark 2.2. If $n = 2$, one can always find a function $R = R(U)$ such that (2.10) takes the form

$$\frac{d}{ds}(R(U)) = 0, \text{ along } C. \tag{2.11}$$

Thus a Riemann invariant always exists in case (iii) when $n = 2$. (Cf. Exercise 2.2).

Remark 2.3. Evern though it is not always possible to get a Riemann invariant in case (iii), nevertheless a relation of the type (2.10) is always useful. There are numerical methods based on such relationships as will be seen later. Instead of working on the original set of equations, we can work on these equations in numerical schemes (Cf. Section 8).

2.3 Application to the hydrodynamic equations

Let us now turn to the hydrodynamic equations. We will work in the Lagrangian form. The parallel derivations in the Eulerian framework will be left as an exercise.

We take the special case where $\mu = 0$, $k = 0$, $g = 0$. The Lagrange equations assume the form

$$\left.\begin{aligned} &\text{(i)} && \frac{D\rho}{Dt} + \rho\frac{\partial u}{\partial x} = 0, \text{ or, } \frac{D}{Dt}(\rho J) = 0, \\ &\text{(ii)} && \frac{Du}{Dt} + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0, \text{ and} \\ &\text{(iii)} && \frac{D\varepsilon}{Dt} + p\frac{D}{Dt}\left(\frac{1}{\rho}\right) = 0. \end{aligned}\right\} \tag{2.12}$$

From (2.1) it follows that ρJ is free of t and since $J = 1$ at $t = 0$, we have

$$\rho J = \rho(a, 0) = \rho_\circ(a). \tag{2.13}$$

We now introduce a function $m = m(a)$ which gives the total mass between a and a fixed point $\bar{a}$ at time $t = 0$. In other words **14**

$$m(a) = \int_{\bar{a}}^{a} \rho_\circ(a)da, \tag{2.14}$$

or, equivalently

$$\frac{dm}{da} = \rho_\circ(a). \tag{2.15}$$

Using the relation $V = \dfrac{1}{\rho}$, we rewrite equations (2.12) with independent variables t and m. Dividing (2.12), (i) by ρ^2, we get

$$\begin{aligned} 0 &= \frac{1}{\rho^2}\frac{D\rho}{Dt} + \frac{1}{\rho}\frac{\partial u}{\partial a}\frac{1}{J} \\ &= -\frac{D}{Dt}\left(\frac{1}{\rho}\right) + \frac{1}{\rho_\circ}\frac{\partial u}{\partial m}\frac{dm}{da} \\ &= -\frac{D}{Dt}(V) + \frac{\partial u}{\partial m}. \end{aligned}$$

Thus the equations (2.12) become

$$\left.\begin{aligned} &\text{(i)} && \frac{DV}{Dt} - \frac{\partial u}{\partial m} = 0, \\ &\text{(ii)} && \frac{Du}{Dt} + \frac{\partial p}{\partial m} = 0, \text{ and} \\ &\text{(iii)} && \frac{D\varepsilon}{Dt} + p\frac{DV}{Dt} = 0, \end{aligned}\right\} \tag{2.16}$$

(In the equation (2.16), (ii) we have used

$$\frac{1}{\rho}\frac{\partial}{\partial x} = \frac{1}{\rho J}\frac{\partial}{\partial a} = \frac{1}{\rho_\circ}\frac{\partial}{\partial m}\frac{dm}{da} = \frac{\partial}{\partial m}.$$

Again we may replace $\dfrac{DV}{Dt}$ in (2.16) (iii) by $\dfrac{\partial u}{\partial m}$ by virtue of (2.16) (i). We further assume that the state equation $\varepsilon = f(p, V)$ can be inverted to give $p = g(\varepsilon, V)$.

Thus if we set **15**

$$U^T = (V, u, \varepsilon), \tag{2.17}$$

on gets the equations (2.16) in the form

$$\frac{DU}{Dt} + \begin{bmatrix} 0 & -1 & 0 \\ g_v & 0 & g_\varepsilon \\ 0 & p & 0 \end{bmatrix} \frac{\partial U}{\partial m} = 0, \tag{2.18}$$

where g_V, g_ε are the corresponding partial derivatives of g. Thus the equations have been put in a form similar to (2.3).

The characteristic polynomial of the matrix in (2.18) is

$$-\lambda(\lambda^2 + (g_V - pg_\varepsilon)) = 0. \tag{2.19}$$

From physical considerations it is known that $pg_\varepsilon - g_V > 0$. In fact one has a relation of the type

$$pg_\varepsilon - g_V = \frac{C^2}{V^2} \tag{2.20}$$

where $C^2 = \dfrac{\partial p}{\varepsilon\rho}$ along an adiabatic transformation, is the square of the velocity of sound. (Along an adiabatic we have $d\varepsilon + pdV = 0$. Since $p = g(\varepsilon, V)$, $dp = g_\varepsilon d\varepsilon + g_V dV = (g_V - pg_\varepsilon)dV$ and since $V = 1/\rho$, $\dfrac{dp}{d\rho} = \dfrac{-1}{\rho^2}\dfrac{dp}{dV}$. Thus $pg_\varepsilon - g_V = \dfrac{dp}{dV} = -\dfrac{1}{V^2}\dfrac{dp}{d\rho} = \dfrac{C^2}{V^2}$, by definition of C^2).

Thus we get three distinct and real eigen values $0, +\dfrac{C}{V}$ and $-\dfrac{C}{V}$, or, equivalently, 0, $\pm C\rho$. We then have three characteristic curves, one vertical line, corresponding to $\lambda = 0$, and two curves on either side of it, in the $x - t$ plane, (Cf. Fig. 2.1).

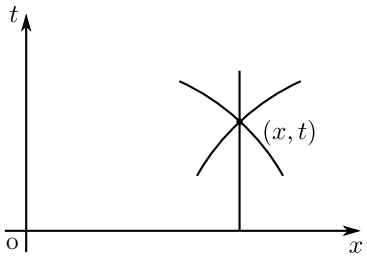


Figure 2.1:

16

Remark 2.4. In the fluid dynamic case the equations are non-linear and we cannot know beforehand the characteristic values. However, their signs are known and these already give some insight into the properties of the solution.

Exercise 2.1. Perform the same analysis in the Eulerian framework and show that the eigenvalues are u, $u \pm C$.

Remark 2.5. In fact the characteristics in the $m - t$ plane are defined by

$$\frac{dm}{ds} = \lambda; \ \frac{dt}{ds} = 1.$$

In the $x - t$ plane, the images of the characteristic curves will be the curves

$$x = x(m(s), \ t(s)); \ t = t(s),$$

and

$$\frac{dx}{ds} = \frac{\partial x}{\partial m}\frac{dm}{ds} + \frac{\partial x}{\partial t}\frac{dt}{ds} = \frac{J}{\rho_\circ}\lambda + u = \frac{\lambda}{\rho} + u$$
$$\frac{dt}{ds} = 1.$$

Hence the eigenvalues are u, $u \pm C$.

Exercise 2.2. In the euqations (2.16), assuming we can integrate (iii) to get p in terms of V, we are then left with only two equations. Thus for a **17**

perfect gas we get $pV^\gamma = a$ constant. For the system (2.16) (i) and (ii) in this case compute the eigenvalues and eigen vectors. Find the Riemann invariant for each characteristic curve. (They exist by virtue of remark 2.2).

Reference. The reader is referred to Courant and Friedrichs [9] for a detailed exposition of characteristics of a hyperbolic system and applications to hydrodynamic equations.

Chapter 3

Discontinuous solutions of hyperbolic systems-shocks

3.1 Introduction

We illustrate the notion of a solution with shocks via a very simple equation known as the Burger's equation. We derive the Rankine-Hugoniot relation to determine the curve across which a discontinuity occurs. We then generalise these ideas to a system of equations. **18**

3.2 Burger's equation

Burger's equation is given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0. \tag{3.1}$$

(In the literature one also finds at times an extra term $-\epsilon\frac{\partial^2 u}{\partial x^2}$ on the left hand side, but we consider the limiting case as $\epsilon \to 0$). This equation is trivially a hyperbolic "system" (a system with a matrix of order 1 and the single eigen value u!). If we set

$$\frac{dt}{ds} = 1;\ \frac{dx}{dt} = u(x,t), \tag{3.2}$$

then (3.1) will read as

$$\frac{du}{ds} = \frac{\partial u}{\partial t}\frac{dt}{ds} + \frac{\partial u}{\partial x}\frac{dx}{ds} = 0. \tag{3.3}$$

Hence u is constant along the characteristic curve C whose slope is given by

$$\frac{dx}{dt} = u(x,t). \tag{3.4}$$

Thus from (3.4) and the fact that u is constant along C we see that all the characteristics of (3.1) are straight lines.

In the following figure, the lower graph indicates the initial value $u(x,0)$ as a function of x. The characteristics through $(x,0)$ in the $x-t$ plane are shown in the upper graph. These are lines whose slopes are given by (3.4).

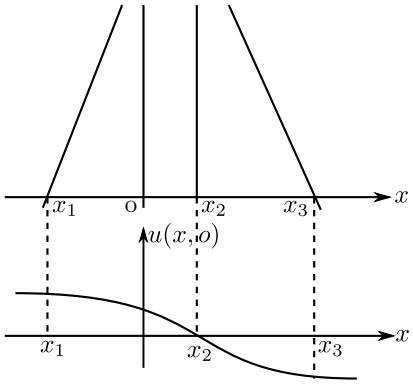


Figure 3.1:

19

(Since $u(x_1,0) > 0$, the characteristic at $(x_1,0)$ has positive slope and so on).

Note, however, that for an arbitrary initial value function $u(x,0)$, it is possible that the various characteristics intersect. Then we have a contrdiction because the point of intersection lies on two different characteristics and must possess two different values of u, which is impossible.

So we try to find a discontinuous solution to this problem. For this we need the notion of the weak form of the equation.

To do this we first rewrite the equation (3.1) in the *proper* conservative form:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = 0. \tag{3.5}$$

(The importance of the proper conservative form cannot be over- **20**
emphasized. Cf. Remark 3.2).

Let φ be a test-function (i.e. a function over $\mathbb{R}_x \times \mathbb{R}_t$ with compact support which is as smooth as we please). We multiply (3.5) by φ and integrate by parts over the domain Ω, where

$$\Omega = \mathbb{R}_x \times \{t \in \mathbb{R} \mid t > 0\}.$$

Thus we get

$$\begin{aligned} 0 &= -\int_\Omega \varphi\left[\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right)\right] dx\, dt \\ &= \int_\Omega u\frac{\partial \varphi}{\partial t} dx\, dt + \int_{-\infty}^{\infty} u(x,0)\varphi(x,0)dx + \int_\Omega \frac{u^2}{2}\frac{\partial \varphi}{\partial x} dx\, dt \end{aligned}$$

where $u(x,0)$ is given to us already. The weak form is then the problem of finding u such that

$$\int_\Omega u\frac{\partial \varphi}{\partial t} dx\, dt + \int_\Omega \frac{u^2}{2}\frac{\partial \varphi}{\partial x} dx\, dt + \int_{-\infty}^{\infty} u(x,0)\varphi(x,0)dx = 0 \tag{3.6}$$

for any test-function φ.

Note that if u is smooth and satisfies (3.6) we can reverse the integration by parts and show that u satisfies (3.5) as well.

Let us now assume that u is piecewise continuously differentiable. Let Γ be a curve across which a discontinuity occurs in u. Assume that Γ is smooth. Let Γ divide the domain Ω into two parts Ω_1 and Ω_2. Denote by $\overline{n} = (n_x, n_t)$ the unit normal along Γ directed from Ω_1 into Ω_2. Thus $u \mid \Omega_i$ is smooth ($i = 1, 2$) and u satisfies (3.6). By choosing φ to have support completely contained either in Ω_1 or in Ω_2 we see immediately that $u \mid \Omega_i$ satisfies (3.5) in $\Omega_i (i = 1, 2)$.

Now choose φ so that supp $\varphi \subset \Omega$, and supp $\varphi \cap \Gamma$ is non-empty. We set down the following notations: **21**

$$\left.\begin{aligned} D_i &= \text{ supp. } \varphi \cap \Omega_i,\ i = 1, 2. \\ F &= \text{ supp } \varphi \cap \Gamma. \end{aligned}\right\} \tag{3.7}$$

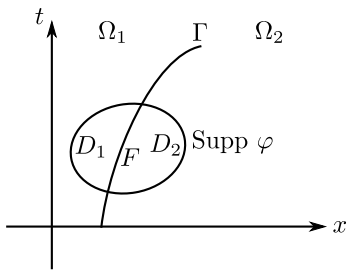


Figure 3.2:

Using this particular test-function φ in (3.6) and splitting the integrals over $D_i (i = 1, 2)$, we get

$$\begin{aligned} 0 &= \int_{D_1} \left(u\varphi_t + \frac{u^2}{2}\varphi_x \right) dx\, dt + \int_{D_2} \left(u\varphi_t + \frac{u^2}{2}\varphi_x \right) dx\, dt \\ &= -\int_{D_1} {}^1u_t \varphi dx\, dt + \int_F {}^1u\varphi n_t ds - \int_{D_1} \left(\frac{({}_1u)^2}{2} \right)_x \varphi dx\, dt \\ &\qquad + \int_F \frac{({}^1u)^2}{2} \varphi n_x dx \\ &\quad - \int_{D_2} {}^2u_t \varphi dx\, dt - \int_F {}^2u\varphi n_t ds - \int_{D_2} \left(\frac{({}^2u)^2}{2} \right)_x \varphi dx\, dt \\ &\qquad - \int_F \frac{({}^2u)^2}{2} \varphi n_x ds \end{aligned}$$

where, ds denotes integration along F, and ${}^1u = u \mid \Omega_1$, ${}^2u = u \mid \Omega_2$. The change of sign in the integrals over F for the second set of terms is due to the fact that the outer normal to Ω_2 is the negative of the to Ω_1. Since ${}^1u$ and ${}^2u$ are smooth inside their domains, they satisfy (3.5) in

22 these domains and hence in the above expression the integrals over D_1 and D_2 vanish. Thus we are left with

$$\int_F \varphi \left[({}^1u - {}^2u)n_t + \left(\frac{({}^1u)^2}{2} - \frac{({}^2u)^2}{2} \right) n_x \right] dx = 0 \tag{3.8}$$

for all test-functions φ with support in Ω. Therefore if we define the jump in a function f across Γ by

$$[f] = {}^1f - {}^2f, \tag{3.9}$$

we then get

$$[u]n_t + \left[\frac{u^2}{2} \right] n_x = 0. \tag{3.10}$$

If we assume Γ to be the curve $x = x(t)$, we can write $\frac{dx}{dt} = \frac{-n_t}{n_x}$. Hence (3.10) can be rewritten as

$$[u]\frac{dx}{dt} = \left[\frac{u^2}{2} \right]. \tag{3.11}$$

The relation (3.11) is most fundamental in determining Γ. It is known as the *Rankine-Hugoniot relation*.

We illustrate these ideas by means of examples.

Example 3.1. Let

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

We then draw the characteristics as we did in Fig. 3.1.

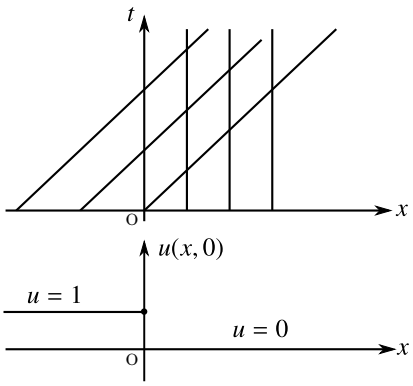


Figure 3.3:

23

Thus we look for a discontinuous (weak) solution of the Burger's equation which will be as in the following figure.

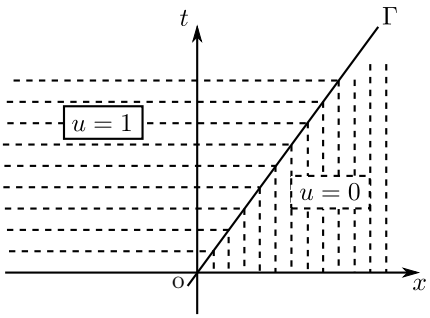


Figure 3.4:

The region shaded horizontally (to the left of Γ) has $u = 1$ and to the right of Γ (shaded vertically) has $u = 0$.

24 To determine the curve Γ we use (3.11). The jump $[u] = 1$ across Γ and $\left[\frac{u^2}{2}\right] = \frac{1}{2}$. Thus by (3.11) one has

$$\frac{dx}{dt} = \frac{1}{2} \tag{3.12}$$

as the slope of Γ. Thus Γ is the straight line through origin with slope given by (3.12).

Remark 3.1. This case has a parallel in the case of solution with shocks in fluid dynamics. In gas dynamics, Γ is the curve where the shock occurs and $\frac{dx}{dt}$ is the speed of the shock. This speed is 'supersonic' (i.e. larger than the slopes of characteristics) w.r.t. the state ahead and 'subsonic' (i.e. smaller than the slope of characteristics) w.r.t. the state behind.

Exercise 3.1. Perform a similar analysis for the solution of the Burger's equation when $u(x, 0)$ is of the following form:

$$u(x,0) = \begin{cases} 1, & \text{if } x \leq -1, \\ -x, & \text{if } -1 \leq x \leq 0, \\ 0, & \text{if } x \geq 0. \end{cases}$$

Example 3.2. Consider the solution of Burger's equation when the initial value is given by

$$u(x,0) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0. \end{cases}$$

We then draw the characteristics.

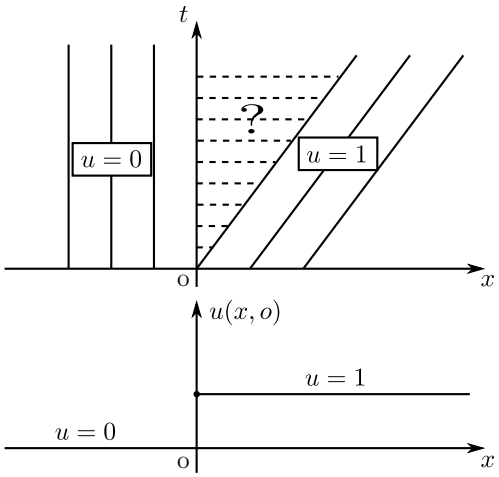


Figure 3.5:

25

It is clear that in the region to the left of the t-axis (i.e. for $x < 0$) we must have $u(x,t) = 0$. Similarly if (x,t) lies to the right of the line $x = t$ (i.e. the characteristic through the origin) we must have $u(x,t) = 1$. The problem is in the (shaded) region in between. Here we now have two possibilities.

First we may search for a straight line Γ passing throught origin so that to its left, $u = 0$ and to its right, $u = 1$. Using (3.11) it is easy to check that this line has slope given by $\dfrac{dx}{dt} = \frac{1}{2}$.

On the other hand, we see that the function x/t satisfies Burger's equation. The solution tends to zero as $x \to 0$ and to 1 as $x \to t$. Thus if we define u to be x/t in this region, we get a continuous solution to Burger's equation. Thus we now have two solutions to the equation (Cf. Fig. 3.6).

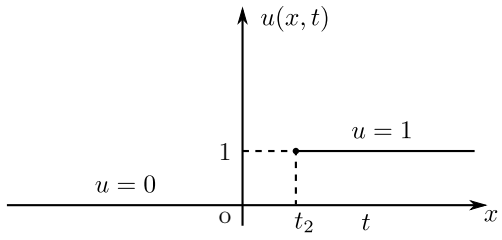
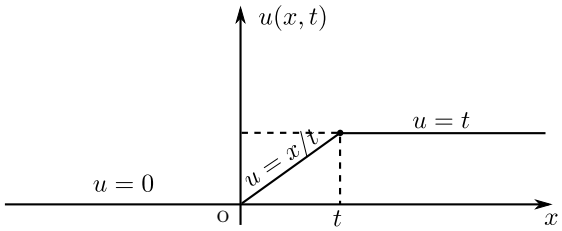


Figure 3.6: (a): *Discontinuous solution*

26



(b): *Continuous solution*

We therefore need one more condition to fix up uniquely the solution for Burger's equation in this case. Such a condition comes from analogy with physical considereations of admissibility of a solution. We demand that $\frac{\partial u}{\partial x} \neq +\infty$. Using this we see that the *continuous* solution alone is admissible. The "velocity" u in case of a shock cannot step up in the positive direction of the x-axis.

This latter example is the analogue of a rarefaction wave in fluid dynamics. Here the function u is continuous while its derivative is not, whereas in the case of a shock, u is iteself discontinuous.

27

Remark 3.2. We now reiterate the need for a cautious choice of the conservative form. As an example consider the equation (3.1). Multiplying by $2u$ throughout and setting $v = u^2$, we get

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(\frac{2}{3}v^{3/2}) = 0, \tag{3.13}$$

which is in conservative form. If we apply the relation (3.11) to get the

slope of the curve Γ for the case of example 3.1, we get

$$[v]\frac{dx}{dt} = \left[\frac{2}{3}v^{3/2}\right],$$

or

$$\left[u^2\right]\frac{dx}{dt} = \left[\frac{2}{3}u^3\right],$$

or

$$\frac{dx}{dt} = \frac{2}{3}$$

which contradicts the result of example 3.1.

Thus one must make the right choice of the conservative form. In general, changes in the *dependent* variable are inadmissible unless the solution is continuous. But one can change the independent variables; one can work with Lagrangian or Eulerian coordinates.

3.3 Rankine-Hugoniot relations for a system

We now turn to the case of a system of equations. Consider the system

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(F(U)) = 0,$$

where $U = U(x,t)$ is a column vector and the vector $F(U)$ has no explicit dependence on the derivatives of U. This is in conservative form.
28 The above system may also be written as

$$\frac{\partial u_i}{\partial t} + \frac{\partial f_j}{\partial x} = \frac{\partial u_j}{\partial t} + \sum_{j=1}^{n} \frac{\partial f_i}{\partial u_j}\frac{\partial u_j}{\partial x} = 0 \tag{3.15}$$

We assume the system to be strictly hyperbolic. In other words, the matrix $\left(\frac{\partial f_i}{\partial u_j}\right)_{1\le i,\ j\le n}$ is assumed to have n distinct and real eigen values. The weak form is now formulated in terms of "test-vectors" Φ, whose components are test-functions:

$$\left.\begin{aligned}\int_\Omega (\langle U, \Phi_t\rangle + \langle F(U), \Phi_x\rangle)dx\,dt + \\ + \int_{-\infty}^{\infty} \langle U(x,0),\ \Phi(x,0)\rangle dx = 0\end{aligned}\right\} \tag{3.16}$$

for every test-vector ϕ.

Proceeding as in the scalar case we get the Rankine-Hugoniot relation for the slope of the curve of discontinuity:

$$[U]\frac{dx}{dt} = [F(U)]. \tag{3.17}$$

This is a vector relation which holds componentwise.

Remark 3.3. In this linear case $[F(U)] = A[U]$ and thus $\dfrac{dx}{dt}$ has to be an eigenvalue of A. Thus the discontinuities can only be across characteristic curves.

3.4 Application to the hydrodynamic system

We now give the Rankine-Hugoniot relations for the hydrodynamic system of equations considered in section 2.3. The equations are

$$\left.\begin{array}{ll} \text{(i)} & \dfrac{DV}{Dt} - \dfrac{\partial u}{\partial m} = 0, \\ \text{(ii)} & \dfrac{Du}{Dt} + \dfrac{\partial p}{\partial m} = 0, \text{ and} \\ \text{(iii)} & \dfrac{D\varepsilon}{Dt} + p\dfrac{\partial u}{\partial m} = 0 \end{array}\right\} \tag{3.18}$$

This is not in conservative form. We multiply (3.18) (ii) by u and add it to (3.18) (iii). Using the fact that $E = \varepsilon + \frac{1}{2}u^2$, one has

$$\left.\begin{array}{ll} \text{(i)} & \dfrac{DV}{Dt} - \dfrac{\partial u}{\partial m} = 0, \\ \text{(ii)} & \dfrac{Du}{Dt} + \dfrac{\partial p}{\partial m} = 0, \text{ and} \\ \text{(iii)} & \dfrac{DE}{Dt} + \dfrac{\partial}{\partial m}(pu) = 0, \end{array}\right\} \tag{3.19}$$

which is in conservative form. Setting

$$\left.\begin{array}{l} U^T = (V, u, E) \\ F(U)^T = (-u, p, pu) \end{array}\right\}. \tag{3.20}$$

one gets (3.19) in the vector form

$$\frac{DU}{Dt} + \frac{\partial}{\partial m}(F(U)) = 0. \tag{3.21}$$

Hence, if we look for a discontinuous solution, the Rankine - Hugoniot relations (3.17) take the form

$$\left.\begin{array}{ll} \text{(i)} & M({}^2V - {}^1V) = -({}^2u - {}^1u) \\ \text{(ii)} & M({}^2u - {}^1u) = ({}^2p - {}^1p), \text{ and} \\ \text{(iii)} & M({}^2E - {}^1E) = ({}^2p\,{}^2u - {}^1p\,{}^1u), \end{array}\right\} \tag{3.22}$$

where $M = \dfrac{dm}{dt}$ along the line of discontinuity, called *shock*.

30 **Exercise 3.2.** Obtain the Rankine-Hugoniot relations in the Eulerian framework and show that they are equivalent to the relation (3.22).

References: The reader is referred to Lax [21],[22] and to Conway and Smoller [8].

Chapter 4

Energy Inequalities

4.1 Introduction

The question of well-posedness of a system of partial differential equations with given initial or boundary conditions is fundamental. This question is often answered using what are called energy estimates or energy inequalities. From these estimates one obtains the existence, uniqueness and continuous dependence on the data of the solution of the problem, thus establishing well-posedness. **31**

In this section, we take up the threee types of equations of section 1-the advection, wave and heat equations - and obtain energy estimates in the linear case. We do not talk about the existence of solutions as it entails some work in functional analysis. We only briefly sketch how existence is proved, at the end of this section.

Before we go forth, it is useful to put down our notations for the various norms that we will be using. We write $\|\cdot\|_p$ for the norm in $L^p(-\infty,\infty)$ where $1 \leq p \leq +\infty$. If U is a vector with n components, its Euclidean norm is denoted by

$$|U|^2 = \sum_{i=1}^{n} u_i^2. \tag{4.1}$$

We also define

$$||U||^2 = \int_{-\infty}^{\infty} |U|^2 dx = \int_{-\infty}^{\infty} \sum_{i=1}^{n} u_i^2 dx, \tag{4.2}$$

when each component u_i is in $L^2(-\infty, \infty)$. We also consider only those functions in $L^2(-\infty, \infty)$ which vanish at $\pm\infty$.

4.2 The advection equation

32 We consider the simplest case where u is a *constant*. Then we have

$$\frac{\partial \varphi}{\partial t} + \frac{\partial \varphi}{\partial x} = 0 \tag{4.3}$$

with $\varphi(x, 0) = \varphi_0(x)$, given,

Multiplying (4.3) by φ and integrating w.r.t. x over the entire real line, we get

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial t}\left(\frac{1}{2}\varphi^2\right) dx + u \int_{-\infty}^{\infty} \frac{\partial}{\partial x}\left(\frac{1}{2}\varphi^2\right) dx = 0. \tag{4.4}$$

Since we look for a solution φ such that $\varphi(\cdot, t) \in L_x^2(-\infty, \infty)$ for each $t > 0$ and which vanishes at $x = \pm\infty$, the above equation gives

$$\frac{1}{2}\frac{d}{dt}(||\varphi(\cdot, t)||_2^2) = 0 \tag{4.5}$$

which gives the *energy equality*

$$||\varphi(\cdot, t)||_2 = ||\varphi_\circ||_2. \tag{4.6}$$

Observe that the uniqueness of the solution for given $\varphi_\circ$ is immediate from (4.6). Indeed if, φ_1, φ_2 are two solutions to (4.3) then so is $\varphi_1 - \varphi_2$ with $(\varphi_1 - \varphi_2)(x, \circ) = 0$ for all x. Then (4.6) shows that $||(\varphi_1 - \varphi_2)(\cdot, t)||_2 = 0$ for all t and hence the uniqueness follows.

Exercise 4.1. Consider the equation

$$\frac{\partial \varphi}{\partial t} + \frac{\partial}{\partial x}(\varphi u) = 0,$$

where u is not a constant. Assuming that u and $\frac{\partial u}{\partial x}$ are bounded and looking for a solution $\varphi(x,t)$ such that $\varphi(\cdot,t)$ is in $L^2_x(-\infty,\infty)$ for each t, vanishing for $x = \pm\infty$, derive the energy inequality **33**

$$\|\varphi(\cdot,t)\|_2 \leq C\|\varphi(\cdot,0)\|_2, \ 0 < t < \bar{t}.$$

Remark 4.1. We have derived the energy inequality in the homogeneous case. It can be shown that in case of linear equations, the estimate in the homogeneous case also implies the existence of such an energy estimate in the non-homogeneous case. For example, consider the equation

$$\frac{\partial \varphi}{\partial t} + u\frac{\partial \varphi}{\partial x} = f, \quad u \text{ constant.} \tag{4.7}$$

Multiplying by φ and integrating w.r.t. x, we get

$$\frac{1}{2}\frac{d}{dt}(\|\varphi(\cdot,t)\|_2^2) = \int_{-\infty}^{\infty} f(x,t)\varphi(x,t)dx \leq \|f(\cdot,t)\|_2\|\varphi(\cdot,t)\|_2.$$

Hence

$$\frac{d}{dt}(\|\varphi(\cdot,t)\|_2) \leq \|f(\cdot,t)\|_2. \tag{4.8}$$

Integrating (4.8) we get the energy inequality in the non-homogeneous case as

$$\|\varphi(\cdot,t)\|_2 \leq \|\varphi(\cdot,0)\|_2 + \int_{\circ}^{2} \|f(\cdot,s)\|_2 ds. \tag{4.9}$$

Again, this is a key-step in the proof of well-posedness of the problem.

Remark 4.2. One can also seek such 'a priori' estimates in other spaces. For instance in case of equation (4.3) we know from section 1.6 that the solution is given by

$$\varphi(x,t) = \varphi_{\circ}(x - ut). \tag{4.10}$$

Thus we can get the estimate in the L^∞-norm, assuming $\varphi_\circ \in L^\infty(-\infty, \infty)$, as

$$\|\varphi(\cdot, t)\|_\infty = \|\varphi_\circ\|_\infty. \tag{4.11}$$

4.3 The wave equation

We will follow Friedrichs' argument for symmetric systems. We will deal with the hyperbolic case. The system

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0 \tag{4.12}$$

where U is an n-vector, is said to be symmetric if the matrix A is symmetric. For instance if $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, we get the wave equation.

First of all we observe that

$$\sum_{i=1}^{n} u_i \frac{\partial u_i}{\partial t} = \frac{1}{2}\sum_{i=1}^{n} \frac{\partial (u_i)^2}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}(|U|^2). \tag{4.13}$$

One also has

$$\frac{\partial}{\partial x}\left(\sum_{i,j} a_{ij}u_i u_j\right) = \sum_{i,j}\frac{\partial a_{ij}}{\partial x}u_i u_j + \sum_{i,j} a_{ij}\frac{\partial u_i}{\partial x}u_j + \sum_{i,j} a_{ij}u_i\frac{\partial u_j}{\partial x} =$$
$$= \sum_{i,j}\frac{\partial a_{ij}}{\partial x}u_i u_j + 2\sum_{i,j} u_i a_{ij}\frac{\partial u_j}{\partial x}$$

by the symmetry of A. Hence one has

$$\sum_{i,j} a_{ij}u_i\frac{\partial u_j}{\partial x} = \frac{1}{2}\left[\frac{\partial}{\partial x}\langle U, AU\rangle - \langle U, \frac{\partial A}{\partial x}U\rangle\right] \tag{4.14}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product in $\mathbb{R}^n$.

Using these, we get, by taking the scalar product with U of the equation (4.12).

$$\frac{1}{2}\frac{\partial}{\partial t}(|U|^2) + \frac{1}{2}\left[\frac{\partial}{\partial x}\langle U, AU\rangle - \langle U, \frac{\partial A}{\partial x}U\rangle\right] = 0. \tag{4.15}$$

Integrating w.r.t. x, and using the notation of section 4.1, we get

$$\frac{1}{2}\frac{d}{dt}(\|U(\cdot,t)\|^2) = \frac{1}{2}\int \langle U, \frac{\partial A}{\partial x}U\rangle dx. \tag{4.16}$$

Assume that the $\dfrac{\partial a_{ij}}{\partial x}$ are all bounded, we get the inequality

$$\frac{d}{dt}(\|U(\cdot,t)\|^2) \leq C\|U(\cdot,t)\|^2. \tag{4.17}$$

It is a simple step to get the energy inequality from (4.17). We leave it as an

Exercise 4.2. Starting from (4.17) deduce the inequality

$$\|U(\cdot,t)\| \leq \|U(\cdot,0)\| \exp(Ct).$$

4.4 The heat equation

We take the simplet case:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^3} = 0. \tag{4.18}$$

Again, multiplying by u and integrating w.r.t. x, we get

$$\frac{1}{2}\frac{d}{dt}(\|u(\cdot,t)\|_2^2) - \int_{-\infty}^{\infty} u\frac{\partial^2 u}{\partial x^2}dx = 0. \tag{4.19}$$

Integrating the second term by parts, (4.19) becomes

$$\frac{1}{2}\frac{d}{dt}(\|u(\cdot,t)\|_2^2) + \int_{-\infty}^{\infty} (\frac{\partial u}{\partial x})^2 dx = 0 \tag{4.20}$$

Since the second term is non-negative, we can write

$$\frac{d}{dt}(\|u(\cdot,t)\|_2^2) \leq 0 \tag{4.21}$$

which gives the energy inequality

$$\|u(\cdot,t)\|_2 \leq \|u(\cdot,0)\|_2. \tag{4.22}$$

4.5 Remarks on existence of solutions

A word about the existence of solutions. As is readily seen, the uniqueness of the solution and its continuous dependence on the data follows
36 easily from the energy estimates. However, for the existence of solutions more work is necessary. In the linear case we have the Galerkin method. We take a basis $\{v_1, \ldots v_n, \ldots\}$ for $L^2(-\infty, \infty)$ and then consider the finite dimensional spaces S_n, spanned by $\{v_1, \ldots, v_n\}$. We approximate the initial value function $u(x,0)$ by $u_n(x,0)$ in S_n and in the space S_n, the partial differential equations give a system of ordinary differential equations. To this we apply we existence theory available and get an approximate solution $u_n(x,t)$. We then use the energy inequalities to show that $u_n(\cdot,t)$ are bounded and we can extract a subsequence converging (weakly) to a function u which can be shown to be a solution of the equations.

In the non-linear case,

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(F(u)) = 0, \tag{4.23}$$

one not only has to show that $u_n \to u$ but also that $F(u_n) \to F(u)$ for which a single 'a priori' estimate is not enough. One does not have general techniques for non-linear systems. Work has been done only on a few specific examples.

Reference: Lions [26].

Chapter 5

Boundary conditions and well-posedness

5.1 Introduction

We have considered, so far, only initial value probelms. We now turn to the situation when we have to solve a system of equations in a domain which is bounded w.r.t. x, say $0 < x < 1$, and are given a certain initial-value function. In this case, we look for appropriate boundary conditions on $x = 0$ and $x = 1$ which will lead to the well-posedness of the problem. More precisely, we look for boundary conditions which enable us to get energy estimates. **37**

We illustrate the relevant ideas in the case of the three special equations of section 1.

5.2 The heat equation

Consider the simplest case of the heat equation given by

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1 \tag{5.1}$$

with the initial condition $u(x, 0) = u_\circ(x)$ being given.

We need boundary conditions on the lines $x = 0$ and $x = 1$. If $\bar{n}$ is the outer normal on the boundary, we shall try to maintain the same boundary conditions which work for the stationary elliptic equation $\frac{\partial^2 u}{\partial x^2} = 0$, i.e. we write

$$u + k\frac{\partial u}{\partial n} = 0,\ k \geq 0, \tag{5.2}$$

in case we look for a homogeneours boundary condition. Note that on $x = 1$, $\frac{\partial}{\partial n} = \frac{\partial}{\partial x}$ and on $x = 0$, $\frac{\partial}{\partial n} = \frac{-\partial}{\partial x}$. Thus (5.2) can also be written as

$$\left.\begin{aligned} u(1,t) + k_1\frac{\partial u}{\partial x}(1,t) = 0,\ k_1 \geq 0 \\ u(0,t) - k_\circ\frac{\partial u}{\partial x}(0,t) = 0,\ k_\circ \geq 0. \end{aligned}\right\} \tag{5.3}$$

38 We now obtain an energy inequality. Multiplying (5.1) by u and integrating w.r.t. x from 0 to 1, one has

$$\int_0^1 u\frac{\partial u}{\partial t}dx - \int_0^1 u\frac{\partial^2 u}{\partial x^2}dx = 0$$

or,

$$\frac{1}{2}\frac{d}{dt}\int_0^1 u^2dx - u(1,t)\frac{\partial u}{\partial x}(1,t) + u(0,1)\frac{\partial u}{\partial x}(0,t) + \int_0^1 (\frac{\partial u}{\partial x})^2dx = 0.$$

Incorporating the conditions (5.3), we get

$$\frac{1}{2}\frac{d}{dt}\int_0^1 u^2dx + \int_0^1 \left(\frac{\partial u}{\partial x}\right)^2 dx + \frac{1}{k_1}u(1,t)^2 + \frac{1}{k_\circ}u(0,t)^2 = 0,$$

and, since the last three terms are non-negative, one gets,

$$\frac{d}{dt}\int_0^1 u^2dx \leq 0,$$

which leads to the energy estimate

$$\|u(\cdot,t)\|_2 \leq \|u(\cdot,0)\|_2 \tag{5.4}$$

where $\|\cdot\|_2$ denotes the norm in $L^2(0,1)$.

5.3 The advection equation

We study the linear case where u is a constant. The equation is

$$\frac{\partial \varphi}{\partial t} + u\frac{\partial \varphi}{\partial x} = 0, \quad 0 < x < 1 \tag{5.5}$$

with given initial condition $\varphi(x, 0) = \varphi_\circ(x)$.

We now ask ourselves if we can impose boundary conditions on $x = 0$ and $x = 1$ freely. The answer comes from a consideration of the characteristics.

Let us first consider the case where $u > 0$. Then the characteristics have positive slope. (Cf. Fig. 5.1).

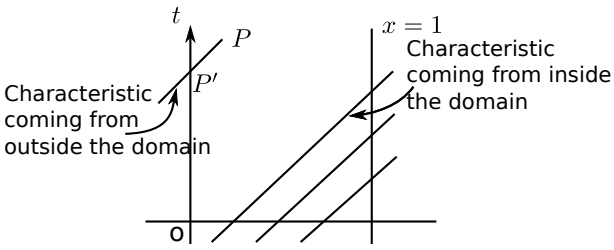


Figure 5.1: ($u > 0$)

Since we know that $\varphi(x, t) = \varphi_\circ(x-ut)$, if we take any point on $x = 1$, with $0 < t < \frac{1}{u}$, then the value of φ at this point is already determined, by the corresponding value on $t = 0$, on the same characteristic and we have no freedom of imposing boundary conditions on $x = 1$.

On the other hand, if we impose a boundary condition on $x = 0$, then for a point P as in Fig. 5.1, we can determine $\varphi(P) = \varphi(P')$. Thus φ will be completely determined in the domain starting with the initial value $\varphi_\circ$ and the boundary condition on $x = 0$.

If $u < 0$, the roles of $x = 0$ and $x = 1$ are reversed.

More generally, for a point on the boundary, if the characteristic through that point comes from inside the domain, we cannot impose any condition there. If it comes from outside (i.e. cuts $t = 0$ outside the domain) we can impose a suitable condition there.

By our above arguments we showed that a boundary condition on $x = 0$ determined φ uniquely when $u > 0$. We now show how to obtain an energy estimate as well. Let us set the boundary condition

$$\varphi(0, t) = g(t) \tag{5.6}$$

40 on the line $x = 0$.

Multiplying (5.5) by φ and integrating w.r.t. x over $[0, 1]$, we get

$$\frac{1}{2}\frac{d}{dt}\int_0^1 \varphi^2 dx + \frac{u}{2}(\varphi^2(1, t) - \varphi^2(0, t)) = 0.$$

Integrating this w.r.t. t, we get

$$\int_0^1 \varphi^2(x, t)dx - \int_0^1 \varphi^2(x, 0)dx + u\int_0^t \varphi^2(1, s)ds - u\int_0^t \varphi^2(0, s)ds = 0.$$

Thus

$$\int_0^1 \varphi^2(x, t)dx + u\int_0^t \varphi^2(1, s)ds = \int_0^1 \varphi^2(x, 0)dx + u\int_0^t \varphi^2(0, s)ds$$

where the right-hand side is known and so is bounded. The second term on the left is non-negative and one can omit it to get an energy inequality for $\|\varphi(\cdot, t)\|_2$ in terms of known quantities.

5.4 The wave equation-method of characteristics

We write down the wave equation as a system of two first order equations:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial v}{\partial x} = 0 \\ \dfrac{\partial v}{\partial t} + \dfrac{\partial u}{\partial x} = 0 \end{cases} \quad \text{in} \quad 0 < x < 1 \tag{5.7}$$

with the initial conditions, $u(x,0) = u_\circ(x)$, and $v(x,0) = v_\circ(x)$. Adding these two equations we get

$$\frac{\partial}{\partial t}(u+v) + \frac{\partial}{\partial x}(u+v) = 0, \tag{5.8}$$

and subtracting, one has

$$\frac{\partial}{\partial t}(u-v) - \frac{\partial}{\partial x}(u-v) = 0. \tag{5.9}$$

Thus, through each point we have two characteristics along which **41** $u+v$ and $u-v$ are respectively constants, (Cf. Fig. 5.2).

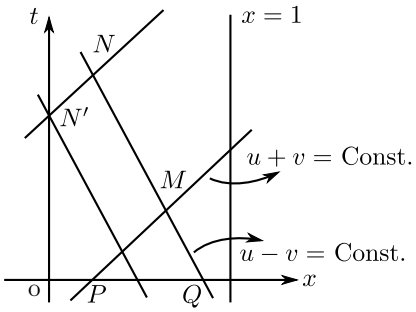


Figure 5.2:

For a point M as in the above figure, both $u+v$ and $u-v$ are known from the corresponding values at P and Q, and hence u, v can also be uniquely computed at M. On the other hand consider the points N and N'. For these one only knows $u-v$ from the initial conditions. If we impose a boundary condition on $x = 0$ so that we can solve for $u+v$ and $u-v$, then u and v can also be uniquely computed.

To this end we impose a boundary condition of the type

$$u(0,t) + \alpha v(0,t) = 0, \quad \alpha \neq -1 \tag{5.10}$$

on $x = 0$. Then knowing $u-v$ at N', together with (5.10) we can compute u and v at N'. Hence we know $u+v$ and $u-v$ at N' and at N, from which we calculate u and v at N, uniquely.

Similar arguments show that the boundary condition on $x = 1$ could be

$$u(1,t) + \beta v(1,t) = 0, \beta \neq 1. \tag{5.11}$$

Thus the conditions (5.10) and (5.11) help us to compute the solu-
42 tion uniquely in the entire domain.

More generally, to get the boundary conditions by the *method of characteristics*, we take any point on the boundary and draw the n characteristics (for a hyperbolic system of n equations) through the point. If p of them come from outside, we may impose p boundary conditions in such a way as to be independent of each other, as well as the $n - p$ relations already given by the characteristics from inside so that, we can solve for the solution in the entire domain.

However it must be noted that by this method we do not get any energy estimate which is a key step in the study of well-posedness.

5.5 The wave equation-Friedrichs' method

In order to get an energy estimate to study the well-posedness, we turn to Friedrichs' method for symmetric systems to find the appropriate boundary conditions. We specialize to the simplest case, and refer the interested reader to Friedrichs [10].

Consider the system of n equations

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0, \quad 0 < x < 1 \tag{5.12}$$

with the initial condition $U(x,0) = U_\circ(x)$, and such that the matrix $A(x,t)$ is symmetric.

Define a matrix $B(t)$ by

$$\left.\begin{aligned} B_1(t) &= A(1,t) \\ B_\circ(t) &= -A(0,t) \end{aligned}\right\} \tag{5.13}$$

on the boundary. We introduce also a matrix $M(t)$ (which is $M_1(t)$ on $x = 1$ and $M_\circ(t)$ on $x = 0$) which has the following properties:

$$\left.\begin{aligned} &\text{(i)} \quad M + M^T \text{ is positive semi-definite,} \\ &\text{(ii)} \quad \mathrm{Ker}(B+M) \oplus \mathrm{Ker}(B-M) = \mathbb{R}^n. \end{aligned}\right\} \tag{5.14}$$

Then the Friedrichs' boundary condition is that $U \in \text{Ker}(B - M)$ on the boundary. Assuming A to be a constant matrix, we derive an energy inequality. Taking the scalar product of (5.12) with U and integrating w.r.t. x over $[0, 1]$, one has

$$\int_0^1 \left(\left\langle U, \frac{\partial U}{\partial t} \right\rangle + \left\langle U,\ A\frac{\partial U}{\partial x} \right\rangle \right) dx = 0. \tag{5.15}$$

Using the symmetry of A and the fact that A is constant, (5.15) can be written as

$$\frac{1}{2}\frac{d}{dt}\int_0^1 |U|^2 dx + \frac{1}{2}\int_0^1 \frac{\partial}{\partial x}\langle U, AU\rangle dx = 0.$$

Or,

$$\frac{d}{dt}\int_0^1 |U|^2 dx + \langle U, AU\rangle_{x=1} - \langle U, AU\rangle_{x=0} = 0. \tag{5.16}$$

Using the relations (5.13) and the fact that $U \in \ker(B - M)$, one has

$$\langle U, AU\rangle_{x=1} - \langle U, AU\rangle_{x=0} = \langle U, B_1 U\rangle + \langle U, B_\circ U\rangle$$
$$\langle U, M_1 U\rangle + \langle U, M_\circ U\rangle.$$

But

$$\begin{aligned} \langle U, MU\rangle &= \langle M^T U, U\rangle = \langle U, M^T U\rangle \\ &= \frac{1}{2}\langle U, MU\rangle + \frac{1}{2}\langle U, M^T U\rangle \\ &= \frac{1}{2}\langle U, (M + M^T)U\rangle \\ &\geq 0 \end{aligned}$$

by virtue of (5.14) (i).

Using this in (5.16) (i), we get the condition

$$\frac{d}{dt}\int_0^1 |U|^2 dx \leq 0$$

44 or, following the notation of Sec. 4.1,

$$\|U(\cdot,t)\|^2 \leq \|U(\cdot,0)\|^2, \tag{5.17}$$

which is an energy inequality.

Remark 5.1. We have not used the condition (5.14) (ii) at all. This is used in proving the existence of solution, for which one has to work with the adjoint problem.

5.6 Comparison of the preceding methods

Observe that the case of the wave equation (5.7) falls within the framework of Friedricns' theory if we take

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Exercise 5.1. Find all matrices M which satisfy the conditions (5.14) w.r.t. the above matrix A and compare the Friedrichs' boundary condition with the boundary conditions (5.10) and (5.11).

The solution of exercise 5.1 will reveal that the Friedrichs' conditions are more restrictive. The advantage of Friedrichs' method lies in the fact that we get an energy estimate in this case while we could not get one by the method of characteristics.

Another interesting question is whether Friedrich's boundary conditions, though restrictive, ar equal in number to those got by the method of characteristics.

We show that this is the case in a very particular situation. (The
45 general case is still open, to the best of the knowledge of the author).

Assume that B is symmetric and regular. Diagonalizing it by an orthogonal matrix Q, one has

$$B_1 = QDQ^T.$$

Choose $M = Q|D|Q^T$. Thus,

$$B_1 + M = Q(D + |D|)Q^T$$

$$B_1 - M = Q(D - |D|)Q^T.$$

If $D = \text{ diag } \{\lambda_1, \ldots \lambda_k, -\lambda_{k+1}, \ldots, -\lambda_n\}$ where $\lambda_i > 0$ for all i, then

$$\begin{aligned}\text{Ker}(B_1 - M) &= Q\,\text{Ker}(D - |D|)Q^T \\ &= \left\{U | V = Q^T U \text{ with } V_{k+1} = \ldots = V_n = 0\right\}.\end{aligned}$$

Similarly,

$$\begin{aligned}\text{Ker}(B_1 + M) &= Q\,\text{Ker}(D + |D|)Q^T \\ &= \left\{U | V = Q^T U, \text{ with } V_{k+1} = \ldots = V_n = 0\right\}.\end{aligned}$$

On the boundary $x = 1$, $B(= +A)$ has k eigenvalues > 0. Thus $n - k$ characteristics come from outside and we should have $n - k$ conditions. One the other hand $U \in \text{Ker}(B - M) = Ker(D - |D|)$ also gives $n - k$ conditions.

Similarly one can argue for $B_\circ = -B_1$ on $x = 0$.

A final word on the wave equation. One can also write the equation as one of second order in the form

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} = 0, \quad 0 < x < 1. \tag{5.18}$$

By the substitutions

$$u = \frac{\partial w}{\partial t};\ v = -\frac{\partial w}{\partial x}, \tag{5.19}$$

we can retrieve the system (5.7). **46**

One of the types of boundary conditions one can impose on (5.18) is the same as for the stationary elliptic problems, i.e.

$$w + k\frac{\partial w}{\partial n} = 0,\ k \geq 0, \tag{5.20}$$

where $\bar{n}$ is the outer normal on the boundary.

Exercise 5.2. Find an energy estimate for the solution of (5.18) with the boundary condition (5.20).

However there is one other boundary condition which we state in

Exercise 5.3. Using the boundary condition

$$\frac{\partial w}{\partial t} + k\frac{\partial w}{\partial n} = 0,\ k \geq 0, \tag{5.21}$$

find an energy estimate for the solution of (5.18). Using the substitutions (5.19) compare this boundary condition with that of Friedrichs or that got by the method of characteristics for the system (5.7).

All this work has been done only when the equations are linear and involve one space variable. Kreiss has done some work in the 2-dimensional case.

For non-linear problems, one linearizes the problem around the boundary to find out appropriate boundary conditions. As usual, the number of characteristics from outside give as many boundary conditions which must be chosen independent of themselves as well as of the relations given by those characteristics from inside.

Chapter 6

Finite Difference Schemes, Stability

6.1 Introduction

In this section we define what we mean by the stability of difference schemes and view some conditions which are necessary and/or sufficient for stability. In the study of stability, the Fourier transform is a useful tool and so is it also in the study of well-posedness. 47

6.2 The Fourier Transform

For a well-posed problem we have a correspondence between the initial data $\varphi_\circ(x)$ and the solution $\varphi(x,t)$. For a given t, define

$$\varphi(t) : x \mapsto \varphi(x,t). \tag{6.1}$$

If the problem is linear, then there exists a linear operator $G(t)$ such that

$$\varphi(t) = G(t)\varphi(0). \tag{6.2}$$

The operator $G(t)$ acts on the space where we seek the solution and where the initial data is also given. For instance we may have $\varphi(0) \in$

$L^2_x(\mathbb{R})$ and we may look for the solution in the same space. Then

$$G(t) : L^2_x(\mathbb{R}) \to L^2_x(\mathbb{R}). \tag{6.3}$$

A fundamental question arising in the study of well-posedness is whether the induced norm $\|G(t)\|_{L(X,X)}$ (where X is the space on which $G(t)$ operates) is bounded or not.

One way to answer this question is via the energy inequality. A second method, which is essentially the same but more convenient (especially when working in $L^2(\mathbb{R})$), is that of the Fourier transform.

The Fourier transform, $\hat{\varphi}$, of a function $\varphi \in L^1\mathbb{R})$ is defined by

$$\hat{\varphi}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mathrm{i}x\xi} \varphi(x) dx. \tag{6.4}$$

48 That this can be extended to the space $L^2(\mathbb{R})$ and that the map

$$\mathscr{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$$

which maps φ to $\hat{\varphi}$ is an isometry of $L^2(\mathbb{R})$ are all well-known results of functional analysis. (See, for instance, Rudin: *Functional Analysis*, McGraw-Hill). The invers. Fourier transform is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\mathrm{i}x\xi} \hat{\varphi}(\xi) d\xi. \tag{6.5}$$

Now consider $\varphi(\cdot, t)$ and $\varphi(\cdot, 0)$ to be in $L^2_x(\mathbb{R})$. Then $\mathscr{F}$ maps them to $\hat{\varphi}(\cdot, t)$ and $\hat{\varphi}(\cdot, 0)$ respectively in $L^2_\xi(\mathbb{R})$. The correspondence defined by (6.2) induces a relation between their Fourier transforms which we denote by

$$\hat{\varphi}(t) = \hat{G}(t)\hat{\varphi}(0). \tag{6.6}$$

Since $\mathscr{F}$ is an isometry from $L^2_x(\mathbb{R})$ onto $L^2(\mathbb{R})$ we see that $\|G(t)\| = \|\hat{G}(t)\|$. Thus it is equivalent to checking either the boundedness of $\|\hat{G}(t)\|$ or that of $\|G(t)\|$. The former is often easier to apply when we are in the case of partial differential equations with constant coefficients.

Example 6.1. Consider the advection equation with u constant:

$$\varphi_t + u\varphi_x = 0. \tag{6.7}$$

Applying the Fourier transform w.r.t. x, we have, for fixec $\xi \in \mathbb{R}$,

$$\frac{d}{dt}(\hat{\varphi}(\xi, t)) - \mathrm{iu}\xi\hat{\varphi}(\xi, t) = 0. \tag{6.8}$$

Integrating this equation, we get

$$\hat{\varphi}(\xi, t) = \hat{\varphi}(\xi, 0)\exp(\mathrm{iu}\xi t). \tag{6.9}$$

Thus $G(t)$ is merely the multiplication by $\exp(\mathrm{iu}\,\xi t)$.

Example 6.2. In the heat equation

$$\varphi_t - \varphi_{xx} = 0, \tag{6.10}$$

we get, on applying the Fourier transform,

$$\frac{d}{dt}(\hat{\varphi}(\xi, t)) + \xi^2\hat{\varphi}(\xi, t) = 0, \tag{6.11}$$

which gives

$$\hat{\varphi}(\xi, t) = \hat{\varphi}(\xi, 0)\exp(-\xi^2 t). \tag{6.12}$$

Thus $\hat{G}(t)$ is just multiplication by $\exp(-\xi^2 t)$.

Exercise 6.1. Find $\hat{G}(t)$ for the wave equation system

$$\begin{cases} u_t + v_x &= 0 \\ v_t + u_x &= 0. \end{cases}$$

To compute the norms of these we need two simple results:

Lemma 6.1. *Let $a(\xi)$ be a bounded function on $\mathbb{R}$ and let $A : L^2_\xi(\mathbb{R}) \to L^2_\xi(\mathbb{R})$ be multiplication by $a(\xi)$. Then*

$$\|A\|_{L(L^2_\xi, L^2_\xi)} = \sup \mid a(\xi) \mid . \tag{6.13}$$

Proof. Let $u(\xi) \in L^2_\xi(\mathbb{R})$. Then

$$Au(\xi) = a(\xi)u(\xi).$$

One has

$$\begin{aligned} \|Au\|_2^2 &= \int_{-\infty}^{\infty} |a(\xi)|^2 \ |u(\xi)|^2 d\xi \\ &\leq (\sup_\xi |a(\xi)|)^2 \int_{-\infty}^{\infty} |u(\xi)|^2 d\xi. \end{aligned}$$

50 Thus $\|Au\|_2 \leq (\sup_\xi |a(\xi)|)\|u\|_2$. Hence A is bounded and its norm is

$$\leq \sup_\xi |a(\xi)| = \|a\|_\infty.$$

To complete the proof, we show that for any $\varepsilon > 0$, $\|A\| > \|a\|_\infty - \varepsilon$. Consider the set

$$E = \{\xi \in \mathbb{R} \mid |a(\xi)| \geq \|a\|_\infty - \varepsilon\}.$$

By definition of $\|\cdot\|_\infty$, the above set has positive Lebesgue measure. By properties of the Lebesgue measure one can find a subset F of E which is measurable and such that

$$0 < \mu(F) < \mu(E)$$

where μ is the Lebesgue measure. Now take $u = \chi_F / \sqrt{\mu(F)}$ where χ_F is the characteristic function of F. This is clearly in $L^2(\mathbb{R})$ and $\|u\|_2 = 1$. Hence

$$\begin{aligned} \|A\|^2 \geq \|Au\|_2^2 &= \int_{\mathbb{R}} |a(\xi)|^2 \frac{(\chi_F(\xi))^2}{\mu(F)} d\xi, \\ &= \frac{1}{\mu(F)} \int_F |a(\xi)|^2 d\xi \end{aligned}$$

$$\geq \frac{1}{\mu(F)} \int_F (\|a\|_\infty - \varepsilon)^2 d\xi$$
$$\geq (\|a\|_\infty - \varepsilon)^2.$$

Thus $\|A\| \geq \|a\|_\infty - \varepsilon$ for each $\varepsilon > 0$ and this establishes (6.13). □

We can extend this to the case of a finite product of $L^2_\xi(\mathbb{R})$ with itself. We omit the proof and merely state the result;

Lemma 6.2. *If $U^T = (u_1(\xi), \ldots, u_n(\xi))$ where $u_i(\xi) \in L^2_\xi(\mathbb{R})$ and B is defined by*

$$BU(\xi) = A(\xi)U(\xi), \tag{6.14}$$

$A(\xi)$ being an $n \times n$ matrix, then the norm of B induced by the vector 51
norm

$$\|U\|^2 = \int_{-\infty}^{\infty} |U(\xi)|^2 d\xi \tag{6.15}$$

is given by

$$B = \sup_\xi |A(\xi)| \tag{6.16}$$

where $|A(\xi)|$ is the Euclidean norm of the matrix $A(\xi)$.

By virtue of these lemmas it is easy to see that in examples 6.1 and 6.2, $\|\hat{G}(t)\| = 1$. The case of exercise 6.1 for the wave equation is again left as an exercise.

6.3 Stability of two-level schemes

We now turn to finite difference schemes. Given a system of differential equations over a domain, we discretize the system by establishing a *mesh* of discrete points over the region and replacing the differential operators by difference operators involving these points.

Let us consider a *uniform* mesh of step Δx in the x-direction and step Δt in the t-direction. The nodes of this mesh are thus the points $(j\Delta x,\ k\Delta t)$ where $j, k \in \mathbb{Z}$, the set of integers with $k \geq 0$. The aim

of a difference scheme would be to express the value of the solution at $u(x, n\Delta t)$ in terms of $u(y,\ k\Delta t)$ where $k < n$. Let us denote by $(u^n(x))$ the value $u(x, n\Delta t)$. Then a general *2-level* finite difference scheme will take the form

$$\sum_{j\in\mathbb{Z}} b_j u^{n+1}(x + j\ \Delta x) = \sum_{j\in\mathbb{Z}} c_j u^n(x + j\ \Delta x). \tag{6.17}$$

Remark 6.1. Though these summations range over all $\mathbb{Z}$ in theory, we only have, in practice, j ranging over a finite set of values.

52 If $b_j = 0$ for all $j \neq 0$, then we can explicitly compute $u^{n+1}(x)$ in terms of u^n. Such a scheme is called explicit. Otherwise the scheme is implicit.

Remark 6.2. Though, on the face of it, it looks as if an explicit scheme is more desirable compared to an implicit one, this is not always the case. As we shall see in later examples, explicit schemes are not always "unconditionally stable" (i.e. stable for all values of $(\Delta x,\ \Delta t)$) while implicit schemes may have this important property.

In general, starting with a 2-level scheme, we can write, at least in theory.

$$u^{n+1}(x) = G(\Delta x, \Delta t)u^n(x), \tag{6.18}$$

where $G(\Delta x, \Delta t) : X \to X$, X being the space where the u^n,s belong. Recursively, one has

$$u^{n+1}(x) = G^{n+1}u^\circ(x). \tag{6.19}$$

Then we have

Definition 6.1. The scheme given by (6.17) is stable w.r.t. a given norm $\|\cdot\|_X$ if and only if

$$\|G^n\|_{L(X,X)} \leq \text{ a constant}, \tag{6.20}$$

the constant being independent of n for all $n > 0$.

The simplest case, because of the use of the Fourier transform, is the study of the L^2-stability of a scheme.

Starting from equation (6.17) and applying the Fourier transform, we get

$$\sum_j \int_{-\infty}^{\infty} b_j u^{n+1}(x + j\Delta x) e^{i\xi x} dx = \sum_j \int_{-\infty}^{\infty} c_j u^n(x + j\Delta x) e^{i\xi x} dx.$$

Replacing $x + j\Delta x$ by y_j, we have

$$\sum_j b_j e^{-i\xi j\Delta x} \hat{u}^{n+1}(\xi) = \sum_j c_j e^{-i\xi j\Delta x} \hat{u}^n(\xi).$$

Setting

$$\left.\begin{aligned} b(\xi) &= \sum_j b_j e^{-i\xi j\Delta x} \\ c(\xi) &= \sum_j c_j e^{-i\xi j\Delta x} \end{aligned}\right\} \tag{6.21}$$

we have

$$b(\xi)\hat{u}^{n+1}(\xi) = c(\xi)\hat{u}^n(\xi). \tag{6.22}$$

Thus $\hat{G}$ is merely multiplication by $a(\xi) = c(\xi)/b(\xi)$, known as the *coefficient of amplification* of the scheme, and $\hat{G}^n$ is multiplication by $(a(\xi))^n$.

Thus the scheme (6.17) is stable if and only if there exists a constant C, independent of n, such that

$$\max_\xi |(a(\xi))^n| \le C$$

$$\text{i.e.} \qquad (\max_\xi |a(\xi)|)^n \le C.$$

or equivalently,

$$\max_\xi |a(\xi)| \le 1. \tag{6.23}$$

Thus (6.23) is a *necessary and sufficient* condition for the L^2 - stability when we have a single *scalar equation*, and a scheme given by (6.17).

6.4 Extension of systems

In the case of a system of n equations, the U^k are all n-vectors. The
54 quantities b_j and c_j of (6.17) must now be replaced by matrices and hence we will have, on applying the Fourier transform, matrices $B(\xi)$ and $C(\xi 0$ playing the roles of $b(\xi)$ and $c(\xi)$ in (6.22). Thus the condition for stability is

$$\max_{\xi} |(B^{-1}(\xi)C(\xi))^n| \leq C, \tag{6.24}$$

C being a constant independent of n.

One cannot reduce this to a neat condition which is necessary and sufficient as in the scalar case. However, using the fact that $|\cdot|$ is a matrix norm, a *sufficient* condition would be that

$$\max_{\xi} |B^{-1}C(\xi)| \leq 1. \tag{6.25}$$

One can obtain a necessary condition by argueing with the spectral radius. Since, we have for any matrix A,

$$(\rho(A))^n \leq |A^n|,$$

where $\rho(A)$ is the spectral radius, a *necessary* condition would be

$$\max_{\xi} \rho(B^{-1}C(\xi)) \leq 1. \tag{6.26}$$

Since for a normal matrix, the spectral radius is equal to the Euclidean norm, we see that, if $B^{-1}C(\xi)$ is *normal* then the condition (6.26) is *necessary and sufficient*.

A host of necessary and/or sufficient conditions under various hypotheses can be found in Richtmyer and Morton [32].

Thomee [36] has studied the L^p-stability for $2 \leq p \leq +\infty$ using the Fourier transform.

Remark 6.3. It must be observed that one can use Fourier transform
55 only when the coefficients of the partial differential equation are constants and when the mesh is uniform.

Our subsequent study will be of numerical schemes for the heat and advection equations. We will then turn to the general equations of hydrodynamics in Section 9.

Chapter 7

Finite Difference Schemes for the Heat Equation

7.1 Introduction

In this , we view various examples of finite difference schemes for the heat equation, $u_t - u_{xx} = 0$. We study the stability and consistency of these schemes. We sketch the proof of convergence, using stability and consistency. Finally we sketch briefly how to deal with variable coefficients and with non-linearity. **56**

7.2 Four Schemes for the Heat Equation

Let us assume henceforth a uniform mesh of steps Δx and Δt. We use the notation

$$u_i^n = u(i\Delta x,\ n\Delta t). \tag{7.1}$$

We now proceed to give four different schemes for the heat equation.

Example 7.1. *An explicit scheme*.

The scheme reads as

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{1}{\Delta x^2}(u_{i+1}^n - 2u_i^n + u_{i-1}^n) = 0. \tag{7.2}$$

This scheme is clearly explicit. Applying the Fourier transform as in Sec. 6.3, we get

$$b(\xi) = \frac{1}{\Delta t}$$
$$c(\xi) = \frac{1}{\Delta t} - \frac{2}{\Delta x^2} + \frac{1}{\Delta x^2}(e^{i\xi\Delta x} + e^{-i\xi\Delta x})$$

which gives the cofficient of amplification

$$a(\xi) = \frac{c(\xi)}{b(\xi)} = 1 - \frac{4\Delta t}{\Delta x^2}\sin^2(\frac{\xi\Delta x}{2}). \tag{7.3}$$

57 Hence

$$\hat{u}^{n+1}(\xi, t) = a(\xi)\hat{u}^n(\xi, t) \tag{7.4}$$

with $a(\xi)$ as in (7.3).

Using the stability criteria of Sec. 6.3, viz. $|a(\xi)| \leq 1$ for all ξ, we see that this scheme is stable if and only if

$$\frac{2\Delta t}{\Delta x^2} \leq 1. \tag{7.5}$$

Expanding the left-hand side of (7.2), for the exact solution u of the heat equation, by means of a Taylor expansion about $(i\Delta x.n\Delta t)$, we get

$$\frac{\partial u}{\partial t} + \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2} + \ldots - \left[\frac{\partial^2 u}{\partial x^2} + \frac{\Delta x^2}{12}\frac{\partial^4 u}{\partial x^4} + \ldots\right]$$

and since u satisfies the heat equation, we see that, for the explicit scheme, the *error of discretization* is the order $O(\Delta t + \Delta x^2)$.

Note that by (7.5), in order to get a stable scheme, one needs Δt to be of the order of Δx^2 and then the overall error of discretization is of order $O(\Delta t)$. However, when we combine the heat equation with other equations, one needs Δt to be of the order of Δx. Thus one feels the need for other schemes. As already remarked in Sec. 6.3, explicit schemes are not generally unconditionally stable while implicit schemes are. Thus one generally uses an implicit scheme.

Example 7.2. *An implicit scheme*

This scheme reads as

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{1}{\Delta x^2}\left[u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}\right] = 0 \tag{7.6}$$

Again one has the relation (7.4) with $a(\xi)$ defined by **58**

$$a(\xi) = \left[1 + \frac{4\Delta t}{\Delta x^2}\sin^2\left(\frac{\xi \Delta x}{2}\right)\right]^{-1} \tag{7.7}$$

and as this always has absolute value ≤ 1, the scheme is *unconditionally stable*, i.e. there is no relation between Δx and Δt for stability. Once again, using a Taylor expansion, one finds the error of discretization to be of order $O(\Delta x^2 + \Delta t)$.

Though this scheme is unconditionally stable, one would prefer an error of discretization of order $O(\Delta x^2 + \Delta t^2)$. To this end, we present two such schemes

Example 7.3. *Richardson's Scheme.* This is, in truth, a 3-level scheme given by

$$\frac{u_i^{n+1} - u_i^{n-1}}{\Delta t} - \frac{1}{\Delta x^2}\left[u_{i+1}^{n} - 2u_i^{n} + u_{i-1}^{n}\right] = 0. \tag{7.8}$$

One reduces this to a *system* of two equations by the substitution $v_i^n = u_i^{n-1}$ for all i. By applying the Fourier transform, one finds that the spectral radius of the matrix $B^{-1}C(\xi)$ is always larger than 1, (Exercise !) and hence the scheme is *always unstable*.

Example 7.4. The Crank-Nicolson Scheme. The scheme is given by

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \nabla^2\left(\frac{u^{n+1} + u^n}{2}\right) = 0 \tag{7.9}$$

where

$$\nabla^2 v = \frac{1}{\Delta x^2}(v_{i+1} - 2v_i + v_{i-1}). \tag{7.10}$$

Once again one has the relation (7.4) with

$$a(\xi) = \frac{1 - \dfrac{2\Delta t}{\Delta x^2}\sin^2\left(\dfrac{\xi\Delta x}{2}\right)}{1 + \dfrac{2\Delta t}{\Delta x^2}\sin^2\left(\dfrac{\xi\Delta x}{2}\right)} \tag{7.11}$$

and this scheme is seen to be unconditionally stable. The order of the
59 error of discretization is easily checked to be $O(\Delta x^2 + \Delta t^2)$.

7.3 Consistency

The consideration of the order of the error of discretization leads to the following definitions.

Definition 7.1. Let L be a finite difference operator approximating a partial differential equation. Then, if u is the exact solution of the equation, the quantity

$$Lu = \varepsilon$$

is called the error of discretization.

Definition 7.2. A finite difference scheme is said to be consistent with the (partial) differential equation it approximates if, for the exact solution u, the error of discretization ε satisfies

$$\lim_{\substack{\Delta x\to 0\\ \Delta t\to 0}} \varepsilon = 0 \tag{7.12}$$

Observe that by virtue of the orders of errors of discretization computed for the examples 7.1 to 7.4 we see that all those schemes are consistent with the heat equation. We now give an example of a scheme which is not always consistent.

Example 7.5. *The Du Fort and Frankel Scheme*. Again this is a 3-level scheme defined by

$$Lu = \frac{1}{2\Delta t}\left[u_i^{n+1} - u_i^{n-1}\right] - \frac{1}{\Delta x^2}\left[u_{i+1}^n - u_i^{n+1} - u_i^{n-1} + u_{i-1}^n\right] = 0. \tag{7.13}$$

This scheme is explicit and is unconditionally stable. However the error of discretization is of order $O\left(\left(\frac{\Delta t}{\Delta x}\right)^2\right)$ and hence will be consistent only if $\Delta t \to 0$ faster than Δx (for instance $\Delta t = O(\Delta x^2)$). This scheme **60** is not very much in use now.

7.4 The coefficient of amplification

For the schemes described above one gets, on applying the Fourier transform, the coefficient of amplification $a(\xi)$. It is interesting to compare this with the exact case. Applying the Fourier transform to the exact equation, one has (Cf. Example 6.2)

$$a_{ex}(\xi) = \exp .(-\xi^2 \Delta t). \tag{7.14}$$

It can be proved that if one imposes a relation of the form $\Delta t = \Delta x^2$ in case of the heat equation and the error of discretization is of order p in Δt, then

$$a_{ex}(\xi) - a(\xi) = O(\Delta t^{p+1}).$$

Thus in the explicit and implicit schemes the error of discretization is of order $O(\Delta t)$ and $p = 1$. In the Crank-Nicolson scheme $p = 2$.

If we plot a against $\xi \Delta x$, we get the graphs shown in Fig. 7.1.

If we have a small wave number, i.e. ξ, then we see that $a(\cdot)$ for any of these systems is almost the same. However, close to $\xi \Delta x = \pi$, i.e. for large wave number, we have wide differences and through the Crank-Nicolson scheme is of second order one cannot use it here since we will get a wrong solution when $\frac{2\Delta t}{\Delta x^2} >> 1$.

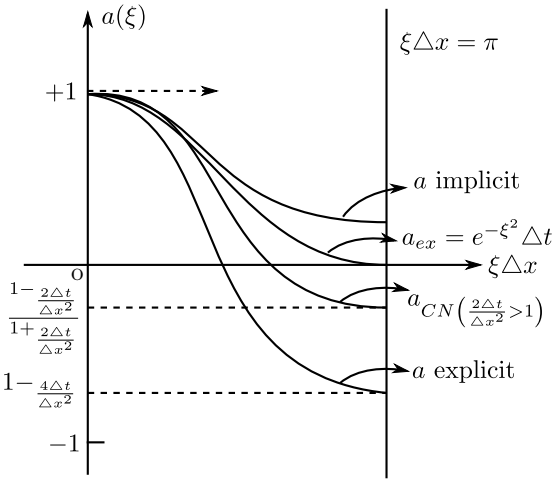


Figure 7.1:

61

7.5 Convergence

Given a finite difference scheme we would like to study the convergence of the approximate solutions to the true solution as the mesh is made more and more fine. In other words, one has to study how the error between the approximate and true solutions behaves. Let L be the approximation of the differential operator, U the approximate solution and u the exact solution. One then has

$$\left.\begin{aligned} LU &= 0 \\ Lu &= \varepsilon, \end{aligned}\right\} \tag{7.15}$$

62 where ε is the error of discretization. The error in the solution is $e = u - U$ and by virtue of (7.15) on has

$$Le = \varepsilon \tag{7.16}$$

Let us assume that the scheme defined by L is stable, i.e. $\|G^n\|$ is bounded for all n by a constant, $C(\Delta x, \Delta)$. We are interested in the case where C is freee of Δx and Δt as well. This leads us to

Definition 7.3. A given scheme is said to be uniformly stable, if the constant C while bounds $\|G^n\|$ for all n, is independent of Δx and Δt when $\Delta x \leq \overline{\Delta x}$ and $\Delta t \leq \overline{\Delta t}$ and $(\Delta x, \Delta t)$ belonging to the subspace of $\mathbb{R}^2$ which gives the stability of the scheme.

Let us assume that L gives a uniformly stable scheme in the sense above and that it is consistent as well. Let us fix any time t. Then divide $[0, t]$ into n equal parts so that $t = n\Delta t$. Let e^n be the error at time level $n\Delta t$. (Since initial condition is assumed to be given, one has $e^\circ = 0$). One can then prove that

$$\|e^n\| \leq C\|\varepsilon\|. \quad (^1)$$

Let us make $\Delta x, \Delta t \to 0$. Then automatically, $n \to \infty$ for $n\Delta t = t$ fixed. Further $\varepsilon \to 0$, by consistency. This implies that $\|e^n\| \to 0$. In other words for each time t, the approximate solutions converge to the exact solution in whatever norm we have stability.

This is a sketch of the proof of the fact that (uniform) stability and consistency imply the convergence of the scheme. This is part of the Lax's equivalence theorem which states that stability and consistency are together equivalent to convergence of a scheme.

For details of this theorem and for more difference schemes, the reader can refer to Richtmyer and Morton [32].

7.6 The energy method

We now describe another technique for studying stability of difference schemes. This is know as the *energy method*.

Example 7.6. Consider the following scheme for the heat equation:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} - \frac{1}{\Delta x^2}$$

[1] This is the analogue of remark 4.1 about the fact that an energy inequality in the homogieneous case always implies the same in the inhomogeneous case. Here energy inequality has to be replaced by stability.

$$\left[\theta(u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}) + (1-\theta)(u_{i+1}^n - 2u_i^n + u_{i-1}^n)\right] = 0 \quad (7.17)$$

for $0 \le \theta \le 1$. (If $\theta = 0$, we get the explicit scheme; for $\theta = 1$ we get the implicit scheme and for $\theta = \frac{1}{2}$ we have the Crank-Nicolson scheme).

We study stability on the space ℓ^2, of square summable sequences. On this space we define the inner product

$$\langle u, v\rangle = \Delta x \sum_{i=-\infty}^{\infty} u_i v_i \quad (7.18)$$

where $u = (u_i)$, $v = (v_i)$. Then $A : \ell^2 \to \ell^2$ is defined by

$$(Au)_i = -\frac{1}{\Delta x^2}\left[u_{i+1} - 2u_i + u_{i-1}\right]. \quad (7.19)$$

We may write the equation (7.17) in vector form as

$$\frac{u^{n+1} - u^n}{\Delta t} + A(\theta u^{n+1} + (1-\theta)u^n) = 0. \quad (7.20)$$

Let us denote by u^* the vector $\theta u^{n+1} + (1-\theta)u^n$. Then we notice that

$$2u^* = (u^{n+1} + u^n) + (2\theta - 1)(u^{n+1} - u^n). \quad (7.21)$$

One also recalls the familiar identity

$$\langle a + b,\ a - b\rangle = |a|^2 - |b|^2 \quad (7.22)$$

64 where $|\cdot|$ is the norm induced by the scalar product.

Taking the scalar product of (7.20) with $2u^*$, one has, using (7.21) and (7.22)

$$\frac{|u^{n+1}|^2 - |u^n|^2}{\Delta t} + \left(\frac{2\theta - 1}{\Delta t}\right)|u^{n+1} - u^n|^2 + 2\langle Au^*, u^*\rangle = 0.$$

Again, using (7.20) we get

$$\frac{|u^{n+1}|^2 - |u^n|^2}{\Delta t} + (2\theta - 1)\Delta t|Au^*|^2 + 2\langle Au^*, u^*\rangle = 0. \quad (7.23)$$

A moment's reflection will show that

$$\langle Au, u\rangle = \Delta x \sum_i \left(\frac{u_{i+1} - u_i}{\Delta x}\right)^2 \geq 0 \tag{7.24}$$

Thus if $\theta \geq \frac{1}{2}$ i.e. $2\theta - 1 \geq 0$, we get from (7.23) that $|u^{n+1}| \leq |u^n|$ which gives ℓ^2-stability of the scheme unconditionally. In particular, this is the case of the implicit and the Crank-Nicolson schemes. In case $0 \leq \theta < \frac{1}{2}$, the middle term of (7.22) is negative and thus one cannot expect unconditional stability.

At this stage one needs the following

Lemma 7.1. $|Au|^2 \leq \dfrac{4}{\Delta x^2}\langle Au, u\rangle$.

Proof.

$$\begin{aligned}
|Au|^2 &= \Delta x \sum_{-\infty}^{\infty} \frac{1}{(\Delta x)^4}(u_{i+1} - 2u_i + u_{i-1})^2 \\
&= \Delta x \sum_{-\infty}^{\infty} \frac{1}{(\Delta x)^2}\left(\frac{u_{i+1} - u_i}{\Delta x} - \frac{u_i - u_{i-1}}{\Delta x}\right)^2 \\
&= \frac{\Delta x}{\Delta x^2}\left[\sum_{-\infty}^{\infty}\left(\frac{u_{i+1} - u_i}{\Delta x}\right)^2 + \sum_{-\infty}^{\infty}\left(\frac{u_i - u_{i-1}}{\Delta x}\right)^2 \right.\\
&\qquad\qquad \left. +2\sum_{-\infty}^{\infty}\left|\left(\frac{u_{i+1} - u_i}{\Delta x}\right)\left(\frac{u_i - u_{i-1}}{\Delta x}\right)\right|\right].
\end{aligned}$$

Applying the Cauchy-Schwarz inequality to the last term, one has

$$\begin{aligned}
|Au|^2 &\leq \frac{4}{\Delta x^2}\sum_{-\infty}^{\infty} \Delta x(\frac{u_{i+1} - u_i}{\Delta x})^2 \\
&= \frac{4}{\Delta x^2}\langle Au, u\rangle, \qquad \text{by (7.23)} .
\end{aligned}$$

This proves the lemma. □ **65**

Using this lemma we find the condition for stability. In order to have stability, one needs $|u^{n+1}|^2 - |u^n|^2 \leq 0$. In other words, we need

$$(1 - 2\theta)\Delta t|Au^*|^2 \leq 2\langle Au^*, u^*\rangle$$

But $(1-2\theta)\Delta t|Au^*|^2 \leq (1-2\theta)\dfrac{4\Delta t}{\Delta x^2}\langle Au^*, u^*\rangle$, by lemma 7.1. Thus the stability condition is

$$\frac{4\Delta t(1-2\theta)}{\Delta x^2} \leq 2. \tag{7.25}$$

Remark 7.1. If $\theta = 0$, we get back, from (7.25), our original L^2-stability condition for the explicit scheme. Thus we see that for all our schemes the L^2 and ℓ^2 stability coditions are the same. However, this is not surprising, for under the correspondence $\{u_i\} \longleftrightarrow$ (piecewise linear function with value u_i at $i\Delta x$), the ℓ^2- norm is equivalent to the L^2-norm.

All our preceding work has been under the basic assumption that the mesh is uniform. The case where the mesh is non-uniform in the x-direction is described in the following exercises.

Exercise 7.1. Let $\{x_i\}$ be the nodes of the mesh. Let $x_{i+\frac{1}{2}}$ denote the midpoint of $[x_i, x_{i+1}]$ and $x_{i-\frac{1}{2}}$ that of $[x_{i-1}, x_i]$. Define A by

$$(Au)_i = \frac{-1}{x_{i-\frac{1}{2}} - x_{i-\frac{1}{2}}}\left[\frac{u_{i+1}-u_i}{x_{i+1}-x_i} - \frac{u_i - u_{i-1}}{x_i - x_{i-1}}\right].$$

If $\Delta x = \max_i(x_{i+1} - x_i)$, show that

$$(Au)_i \sim \frac{\partial^2 u}{\partial x^2}(x_i) + 0(\Delta x).$$

Exercise 7.2. Define the ℓ^2-inner product by

$$\langle u, v\rangle = \sum_i (x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}})u_i v_i.$$

Then show that

$$|Au|^2 \leq \frac{4}{\min(x_{i+1} - x_i)^2}\langle Au, u\rangle.$$

Remark 7.2. In the case of an explicit scheme, viz. $\theta = 0$, we get the stability condition for the non-uniform mesh as $2\Delta t \leq \min(x_{i+1} - x_i)^2$ by virtue of exercise 7.2. In general it is found that when looking for stability criteria in a non-uniform mesh, one can adopt the followig rule of the thumb:

Find out the condition for the uniform mesh case. Then imposing this condition locally on each interval of the non-uniform mesh, pick out the strongest of these as the required stability criterion.

Thus for the explicit scheme, one has $2\Delta t \leq (x_{i+1} - x_i)^2$ starting from a uniform mesh. This leads to the 'worst' condition

$$2\Delta t \leq \min_i (x_{i+1} - x_i)^2$$

which was deduced from the preceding exercises.

A word of caution! This method is purely heuristic, but generally works. In some cases one can rigorously prove this heuristic stability criterion (as in the case of the explicit scheme above) but this is not always possible.

Before closing our discussion of the heat equation, we say a few words about the variable coefficient case and the non-linear case.

7.7 Heat equation with variable coefficients

The heat equation with variable coefficients reads as **67**

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(\sigma(x)\frac{\partial u}{\partial x}\right) =, \tag{7.26}$$

where $\sigma(x) \geq \alpha > 0$.

This may be also written as

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x}\frac{\partial \sigma}{\partial x} - \sigma\frac{\partial^2 u}{\partial x^2} = 0. \tag{7.27}$$

To set up a discrete scheme, the problem is essentially to approximate the terms other than $\dfrac{\partial u}{\partial t}$. To do this we have two approaches, based on (7.26) and (7.27).

The first is the 'conservative' scheme based on the form (7.26). We approximate the terms involving derivatives w.r.t. x by

$$(A^c u)_i = \frac{-1}{\Delta x^2}\left[\sigma_{i+\frac{1}{2}}(u_{i+1} - u_i) - \sigma_{i-\frac{1}{2}}(u_i - u_{i-1})\right] \tag{7.28}$$

and this is of order $O(\Delta x^2)$.

The second scheme uses (7.27) to approximate the derivatives w.r.t. x. This is the 'non-coservative' scheme given by

$$(Au^{Nc})_i = \frac{-1}{\Delta x^2}\left[\frac{(u_{i+1} - u_{i-1})}{2}\,\frac{(\sigma_{i+1} - \sigma_{i-1})}{2} + \sigma_i(u_{i+1} - 2u_i + u_{i-1})\right] \tag{7.29}$$

which is also of second order accuracy.

Remark 7.3. The equation (7.26) is essentially a conservation law. Integrating between a and b w.r.t. x, we have

$$\int_a^b -\frac{\partial}{\partial x}\left(\sigma\frac{\partial u}{\partial x}\right)dx = -\sigma(b,t)\frac{\partial u}{\partial x}(b,t) + \sigma(a,t)\frac{\partial u}{\partial x}(a,t)$$

and replacing the term involving σ by summation, one has

$$\Delta x \sum_{i=i_\circ}^{i_1}(A^c u)_i = -\frac{1}{\Delta x}\left[\sigma_{i+\frac{1}{2}}(u_{i+1} - u_i)\right]_{i=i_1} + \frac{1}{\Delta x}\left[\sigma_{i-\frac{1}{2}}(u_i - u_{i-1})\right]_{i=i_0} \tag{7.30}$$

68 which turns our to be the discreate analogue of (7.30). If we use the non-conservative scheme to replace the terms involving σ, the summation will not be 'telescopic' to resemble equation (7.30). Thus the conservation scheme gives the discrete analogue of the continuous case and is used in general.

One can prove a lemma analogous to lemma 7.1 and exercise 7.2. We state it below.

Lemma 7.2. *If $A : \ell^2 \to \ell^2$ is defined by (7.28), then,*

$$|Au|^2 \le \frac{4}{\Delta x^2}\max_i |\sigma_{i+\frac{1}{2}}|\langle Au, u\rangle \tag{7.31}$$

Remark 7.4. If $\sigma(x) \equiv 1$, we get back lemma 7.1. We observe that the stability conditions can be made to follow from a heuristic argument similar to that enunciated in remark 7.2. The stability condition for the explicit scheme will then be

$$\frac{2\Delta t}{\Delta x^2} \max_i |\sigma_{i+\frac{1}{2}}| \leq 1$$

which, in this case, can be proved using lemma 7.2.

7.8 A non-linear example

The non-linear heat equation is of the form

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(\sigma(u)\frac{\partial u}{\partial x}\right) = 0. \tag{7.32}$$

If $\sigma \geq 0$, $\sigma \in C^1$ and $0 < \alpha \leq \sigma(u) \leq \beta$, then one can imitate the analysis of Section 7.7. of the linear case. However, if $\alpha = 0$, then such techniques do not extend to the non-linear equation. Nevertheless, particular forms of $\sigma(u)$ have been studied. **69**

We consider the case where $\sigma(u) = mu^{m-1}$, $u \geq 0$, $m > 1$. The condition $u \geq 0$ is the one that is encountered in physics since u is the temperature. Then the equation can be written as

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u^m}{\partial x^2} = 0, \tag{7.33}$$

with the initial condition, say, $u(x, 0) = 0$, and boundary condition given by prescribing either $u(0, t)$ or $\frac{\partial u^m}{\partial n}(0, t) = -\frac{\partial u^m}{\partial x}(0, t)$ (the normal derivative on $x = 0$), both non-negative so as to ensure, by the maximum principle, that the solution u will be ≥ 0 everywhere. For a given t, the solution will take the form as shown in Fig. 7.2.

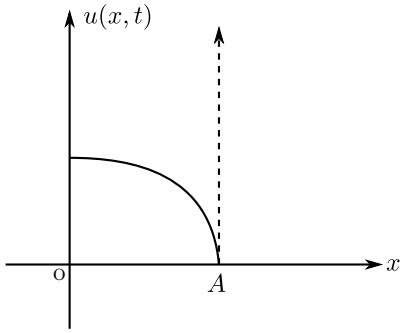


Figure 7.2:

In the case $m \geq 2$, the derivative of u at A is infinite. However there is a result due to Aronson [3] which states that u^{m-1} has bounded derivatives. Hence to set up an approximate scheme we rewrite (7.33) as

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(\frac{mu}{m-1}\frac{\partial u^{m-1}}{\partial x}\right) = 0. \tag{7.34}$$

70 Now we approximate the term involving derivatives w.r.t. x by

$$(Au)_i = \frac{-m}{\Delta x(m-1)}\left[\left(\frac{u_{i+1}^{m-1}+u_i^{m-1}}{2}\right)^{\frac{1}{m-1}}\left(\frac{u_{i+1}^{m-1}-u_i^{m-1}}{\Delta x}\right)+\right.$$
$$\left.+\left(\frac{u_i^{m-1}+u_{i-1}^{m-1}}{2}\right)^{\frac{1}{m-1}}\left(\frac{u_i^{m-1}-u_{i-1}^{m-1}}{\Delta x}\right)\right]. \tag{7.35}$$

Notice that we use again the "almost linearity" of u^{m-1} to define a value $u_{i+\frac{1}{2}}$ interpolated between u_i and u_{i+1}. From this we can generate various implicit and explicit schemes. For example, the implicit scheme will read as

$$\frac{u_i^{n-1}-u_i^n}{\Delta t} + (Au^{n+1})_i = 0 \tag{7.36}$$

which is, w.r.t. the unknown at time $(n+1)\Delta t$, of the form

$$F_i(u_{i+1}^{n+1}, u_i^{n+1}, u_{i-1}^{n+1}) = 0. \tag{7.37}$$

So we get a non-linear ("tridiagonal") system to solve and this could be done by Newton's method of linearising locally, so that at each iteration one has to solve a linear tridiagonal system, which is easy.

If one wants to use the explicit scheme (given by $\frac{u_i^{n+1}-u_i^n}{\Delta t} + (Au^n)_i = 0$), in order to avoid solving linear systems one would get the heuristic stability condition

$$2m \ \max |u|^{m-1} \frac{\Delta t}{\Delta x^2} \leq 1 \tag{7.38}$$

which can be very drastic.

For non-linear examples see also Graveleau and Jamet [14].

Chapter 8

Numerical Methods for the Advection Equation and Hyperbolic systems

8.1 Introduction

We now study numerical schemes for the advection equation which is a scalar equation. We consider the equation with constant coefficients in order to apply Fourier transform techniques to study stability. We do this for the pure initial value problems and give modifications for problems with boundary conditions. Later we shall extend these ideas to simple non-linear cases like Burger's equation and also to systems of linear equations. **71**

The advection equation with constant coefficients is given by

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0. \tag{8.1}$$

To start with we establish a uniform mesh of step Δx over $\mathbb{R}$ and step Δt in time. As usual u_i^n will denote the value $u(i\Delta x, n\Delta t)$. We now proceed to give various numerical schemes for the equation (8.1).

8.2 Explicit Schemes for the Advection Equation

Example 8.1. The simplest shceme to approximate equation (8.1) is given by

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0. \tag{8.2}$$

It is obvious from expanding by a Taylor series about $(i\Delta x,\ n\Delta t)$ this scheme has an error of discretization of order $O(\Delta x^2 + \Delta t)$. However setting

$$\alpha = a\frac{\Delta t}{\Delta x} \tag{8.3}$$

and using the Fourier transform, the coefficient of amplification turns out to be

$$a(\xi) = 1 + i\alpha \sin(\xi\Delta x). \tag{8.4}$$

72 Hence $|a(\xi)| \geq 1$, which means that this extremely simple explicit shceme is always unstable and is thus never used!

Example 8.2. *Lax's Scheme*. This is an explicit scheme of first order accuracy, given by

$$\frac{u_i^{n+1} - \frac{1}{2}(u_{i+1}^n + u_{i-1}^n)}{\Delta t} + a\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta x} = 0. \tag{8.5}$$

Again, from the Fourier transform, we get

$$a(\xi) = \mathrm{Cos}(\xi\Delta x) + i\alpha \sin(\xi\Delta x) \tag{8.6}$$

where α is as in (8.3). The criterion $|a(\xi)| \leq 1$ for all ξ, gives the stability condition, $|\alpha| \leq 1$, which is

$$|a|\frac{\Delta t}{\Delta x} \leq 1. \tag{8.7}$$

This scheme has an error of discretization of order $O(\Delta x^2 + \Delta t + \frac{\Delta x^2}{\Delta t})$.

Example 8.3. *One-sided scheme.* The scheme is defined by

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \begin{cases} a\dfrac{(u_i^n - u_{i-1}^n)}{\Delta x} = 0, \text{ if } a > 0 \\ a\dfrac{(u_{i+1}^n - u_i^n)}{\Delta x} = 0, \text{ if } a < 0. \end{cases} \tag{8.8}$$

To motivate this, one can argue with characteristics. Assume that $a > 0$. Then the characteristics have positive slope. The value u_i^{n+1} is then the value at the point where the characteristic through $(i\Delta x, (n+1)\Delta t)$ meets the level $n\Delta t$, (Cf. Fig. 8.1) which we assume lies between $((i-1)\Delta x, n\Delta t)$ and $((i+1)\Delta x,\ n\Delta t)$ (Cf. Remark 8.2 on stability condition).

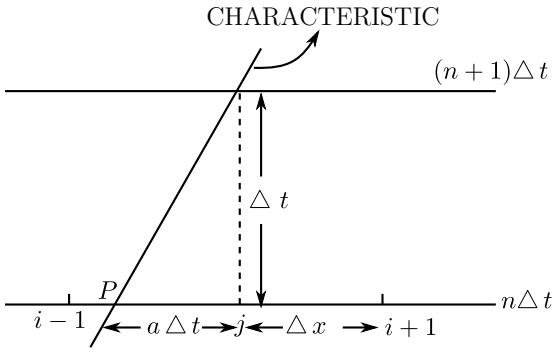


Figure 8.1:

To get the value of P, we interpolate it linearly between the values u_i^n **73**
and u_{i-1}^n. Writing this out we get precisely (8.8) (multiplied throughour by Δt). Similarly we can treat the case $a < 0$.

This scheme has error of discretization of order $O(\Delta x + \Delta t)$.

Exercise 8.1. Show that the stability condition (8.7) holds in the case of the 1-sided scheme as well.

Example 8.4. *The Lax-Wendroff Scheme.* This is a second order scheme. We motivate the scheme by the following arguments. Expanding u_i^{n+1}

by a Taylor expansion w.r.t. t, we get

$$u_i^{n+1} = u_i^n + \Delta t \frac{\partial u}{\partial t}\Big|_i^n + \frac{\Delta t^2}{2}\frac{\partial^2 u}{\partial t^2}\Big|_i^n + \cdots . \tag{8.9}$$

Using the fact that u satisfies the advection equation, we get

$$\begin{aligned} \frac{\partial u}{\partial t} &= -a\frac{\partial u}{\partial x} \\ \frac{\partial^2 u}{\partial t^2} &= \frac{\partial}{\partial t}\left(-a\frac{\partial u}{\partial x}\right) = a^2\frac{\partial^2 u}{\partial x^2}. \end{aligned}$$

Substituting, we get

$$u_i^{n+1} = u_i^n - a\Delta t \frac{\partial u}{\partial x}\Big|_i^n + a^2\frac{\Delta t^2}{2}\frac{\partial^2 u}{\partial x^2}\Big|_i^n + \cdots \tag{8.10}$$

74 We use (8.10) as the guideline for forming the scheme which is given by

$$u_i^{n+1} = u_i^n - a\Delta t\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta t} + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{i-1}^n - 2u_i^n + u_{i-1}^n).$$

Equivalently, we can write

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta x} - \frac{a^2\Delta t}{2\Delta x^2}(u_{i+1}^n - 2u_i^n + u_{i-1}^n) = 0 \tag{8.11}$$

The error of discretization is of order $O(\Delta x^2 + \Delta t^2)$.

Exercise 8.2. Find the coefficient of amplification $a(\xi)$ for the Lax-Wendroff scheme and show that the stabilitity condition is again given by (8.7).

Remark 8.1. We may rewrite the scheme of Lax and the one-sided scheme as follows:

Lax's shceme:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta x} - \frac{\Delta x^2}{2\Delta t}\left(\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2}\right) = 0 \tag{8.12}$$

One-Sided scheme:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta x} - \frac{|a|\Delta x}{2}\left(\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2}\right) = 0 \qquad (8.13)$$

In writing both (8.12) and (8.13) we see that we have essentially approximated to second order accuracy, the perturbed equation

$$\frac{\partial u}{\partial t} + \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2} + a\frac{\partial u}{\partial x} - \epsilon\frac{\partial^2 u}{\partial x^2} = 0. \qquad (8.14)$$

where ϵ is the coefficient occuring in the last term of (8.12) or (8.13). Since the solution satisfies (8.1), we can rewrite (8.14) as 75

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} - \left(\epsilon - \frac{a^2\Delta t}{2}\right)\frac{\partial^2 u}{\partial x^2} = 0. \qquad (8.14)'$$

A criterion of stability due to Hirt [15] and Yanenko is the $\epsilon - \dfrac{a^2\Delta t}{2} \geq 0$. This last term involving ϵ is called the dissipative term. In all cases $\epsilon \to 0$ as $\Delta x, \Delta t \to 0$ (and, if $\dfrac{\Delta x^2}{\Delta t} \to 0$ in case of Lax's scheme).

Remark 8.2. *Interpretations with characteristics.* Just as the one-sided scheme was interpreted to be a linear interpolation between x_{i+1} and $x_i (a < 0)$ or between x_i and $x_{i-1} (a > 0)$, we can interpret the Lax scheme as linear interpolation between x_{i+1} and x_{i-1}. The Lax-Wendroff scheme is the quadratic interpolation between x_{i+1}, x_i and x_{i-1}. All these three interpolations are for the same point P of Fig. 8.1.

We may also interpret the stability condition in terms of characteristics.

Consider the mesh given in Fig. 8.2.

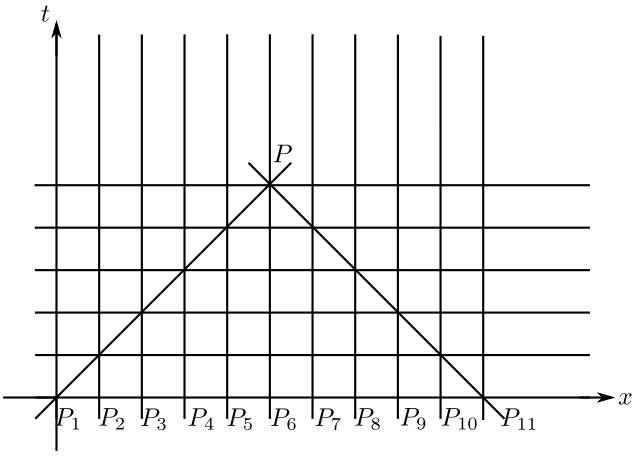


Figure 8.2:

76 The stability condition implies that Δt cannot be too large compared to Δx. Note that for the approximate schemes given above the value at P depends on the values on the 3 nodes immediately below and they in terms of the 5 nodes below them and so on. Thus we define the approximate domain of dependence for P at any time level to be that portion of the grid between P_1P and $P_{11}P$. However, if the stability condition is violated, then the characteristic through P will lie outside the region between these two lines and the exact domain of dependence, which is a single point for the advection equation, (if $P = (x, t)$ then the exact domain of dependence at $t = 0$ is the point $(x - ut, 0)$, Cf. 1.6), will lie outside this region. This will violate the Courant-Friedrichs-Lewy convergence condition that *the exact domain of dependence must be contained in the approximate domain of* dependence and we will not get convergence when $\Delta x, \Delta t \to 0$ keeping $\dfrac{\Delta t}{\Delta x}$ constant.

8.3 Implicit Schemes for the Advection Equation

We will now look at a few implicit schemes for the advection equation.

Example 8.5. *The SNG-Scheme*. This scheme was devised by Carlson for the neutron transport equation. It is a method of first order of accuracy and is essentially an extension of the oen-sided scheme.

Let us assume $a > 0$. (The details when $a < 0$ follow a parallel line of thought).

When $\frac{a\Delta t}{\Delta x} \leq 1$, one uses the one-sided scheme which was only a linear interpolation between the points i and $i+1$ at time $n\Delta t$.

If $\frac{a\Delta t}{\Delta x} > 1$, then the characteristic through $(i\Delta x, (n+1)\Delta t)$ meets the line $x = (i-1)\Delta x$ before it meets $t = n\Delta t$ (See Fig. 8.3).

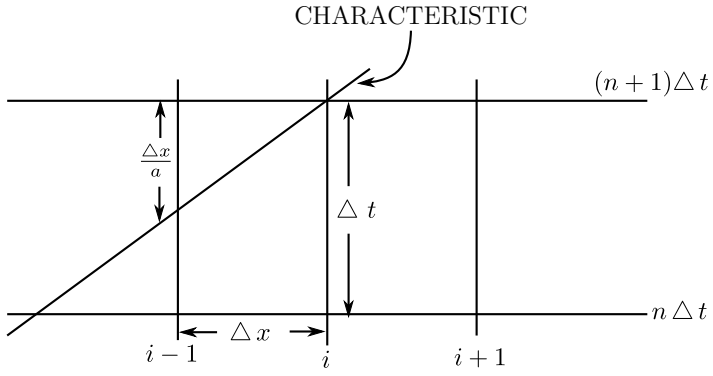


Figure 8.3:

Hence we now have the value u_i^{n+1} equal to the value of u at P which 77
we interpolate linearly between $((i-1)\Delta x,\ n\Delta t)$ and $((i-1)\Delta x,\ (n+1)\Delta t)$, to get

$$u_i^{n+1} = (1 - \frac{\Delta x}{a\Delta t})u_{i-1}^{n+1} + \frac{\Delta x}{a\Delta t}u_{i-1}^n$$

or equivalently,

$$\frac{u_{i-1}^{n+1} - u_{i-1}^n}{\Delta t} + a\frac{(u_i^{n+1} - u_{i-1}^{n+1})}{\Delta x} = 0 \qquad (8.15)$$

The SNG scheme consists of (8.15) as well as the one-sided scheme depending on the value of $\alpha = \frac{a\Delta t}{\Delta x}$. In case of (8.15) being used one

has

$$a(\xi) = (1 + \alpha(e^{-i\xi\Delta x} - 1))^{-1} \tag{8.16}$$

and $|a(\xi)| \leq 1$ if $\alpha > 1$, which is indeed true. Thus the scheme is *unconditionally stable*.

Suppose we have to work in a bounded interval, say, $0 < x < 1$. Then, when $a > 0$, one can impose a boundary condition on $x = 0$.

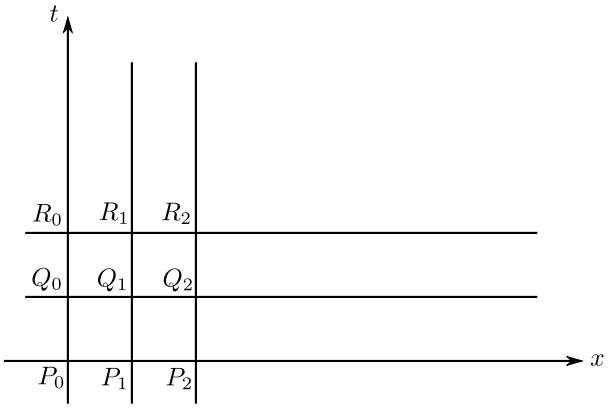


Figure 8.4:

78 Then to calculate Q_1, we know the values at $Q_\circ$, $P_\circ$ and P_1 and hence the value at Q_1 can be calculated explicitly from the scheme. Now for Q_2, since we know the values at Q_1, P_1 and P_2, we can calculate the value at Q_2 and so on. This can be done at any level $n\Delta t$. Thus though the scheme is *formally implicit*, one can, with the aid of the initial value and the boundary condition, solve for the value at each node *explicitly, step by step*. Such a scheme is called quasiexplicit.

Remark 8.3. In case of the neutron transport equation, one has to solve the equation (8.1) when a take its values over an interval $[-A, A]$, using the *same* time step Δt for all a in this interval. Hence it is here that the SNG scheme is very useful, it being unconditionally stable.

Example 8.6. *The Crank-Nicolson Scheme*. The scheme is given by

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{a}{2}\left[\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2\Delta x} + \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x}\right] = 0 \qquad (8.17)$$

This scheme is of second order and is also unconditionally stable. However, it is purely an implicit scheme unlike the SNG-scheme. **79**

In this linear case, this leads to a system with a tridiagonal matrix which is easy to solve by Gauss' method adapted to a tridiagonal system. But in using this scheme for the non-linear equation, with u's in the second term of (8.17) being replaced by a function $f(u)$, the solution becomes more complicated. Thus one devises iterative methods. Let $u_i^{n,p}$ denote the value of u_i^n at the p^{th}-iteration. We can use then the following methods.

$$\frac{u_i^{n+1,p+1} - u_i^n}{\Delta t} + \frac{a}{2}\left[\frac{u_{i+1}^{n+1,p} - u_{i-1}^{n-1,p}}{2\Delta x} + \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x}\right] = 0$$

where we assume all values u_i^n to be known. If $U^T = (u_1^{n+1}, \ldots u_1^{n+1})$, we get

$$U^{(p+1)} = -\frac{a\Delta t}{4\Delta x}\begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & -1 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & 0 & -1 & 0 \end{pmatrix} U^{(p)} + F \qquad (8.18)$$

where F is known. The convergence of this iterative method implies and is implied by $\rho(H) < 1$, where H is the matrix occurring as the coefficient of $U^{(p)}$ in (8.18). Here $\rho(H) \sim \frac{|a|}{2}\frac{\Delta t}{\Delta x}$. Thus in the non-linear case we have to resort to an iterative method which ends up with a condition similar to the stability condition and this undoes all our advantages of achieving unconditional stability.

Of course, one can devise better iterative methods but considerations like computer time etc. do not make this worthwhile. If to solve the Crank-Nicolson scheme with $a\Delta t$, N times larger than the time step

allowed for an explicit scheme $\left(\Delta t < \dfrac{\Delta x}{|a|}\right)$ one needs $M > N$ iterations, then obviously the scheme is not worthwhile. Moreover consideration of accuracy usually prohibits taking Δt large compared to $\dfrac{\Delta x}{|a|}$.

80 **Example 8.7.** This scheme was used by Robert and Weiss [34] in fluid dynamics. The scheme is of second order accuracy and is quasi-explicit when we have a boundary condition. It is also unconditionally stable. (Exercise: check these assertions!) To put down the scheme, we approximate the derivative w.r.t. x by using the mid-points of the diagonals, shown in Fig. 8.5. when $a > 0$.

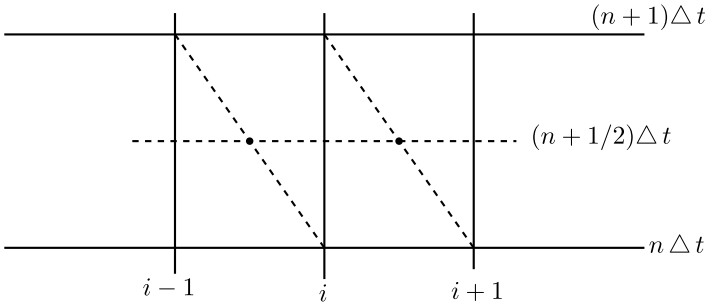


Figure 8.5:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{a}{\Delta x}\left[\frac{u_i^{n+1} + u_{i+1}^n}{2} - \frac{u_{i-1}^{n+1} + u_i^n}{2}\right] = 0. \qquad (8.19)$$

when $a > 0$.

If $a < 0$, we use the other two diagonals of these rectangles. This scheme is rarely used.

Example 8.8. *The DSN–Scheme.* This was also devised by Carlson for the neutron transport equation. In each rectangle of the grid, one evaluates the values of u at the four mid-points of the sides and at the centroid.

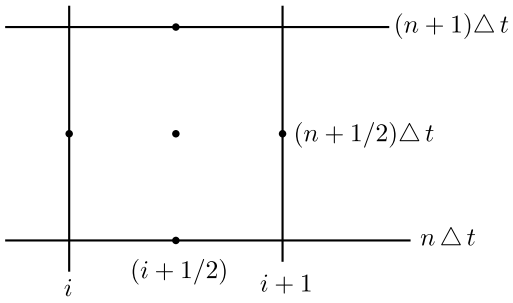


Figure 8.6:

For this we use the set of three equations given below: **81**

$$\begin{cases} \text{(i)} & \dfrac{u_{i+\frac{1}{2}}^{n+1} - u_{i+\frac{1}{2}}^{n}}{\Delta t} + a\dfrac{(u_{i+1}^{n+\frac{1}{2}} - u_{i}^{n+\frac{1}{2}})}{\Delta x} = 0, \\ \text{(ii)} & 2u_{i+\frac{1}{2}}^{n+\frac{1}{2}} = u_{i+\frac{1}{2}}^{n+1} + u_{i+\frac{1}{2}}^{n} = u_{i+1}^{n+\frac{1}{2}} + u_{i}^{n+\frac{1}{2}} \end{cases} \tag{8.20}$$

This scheme is quasi explicit for the following reasons. (our notations are based on Fig. 8.7).

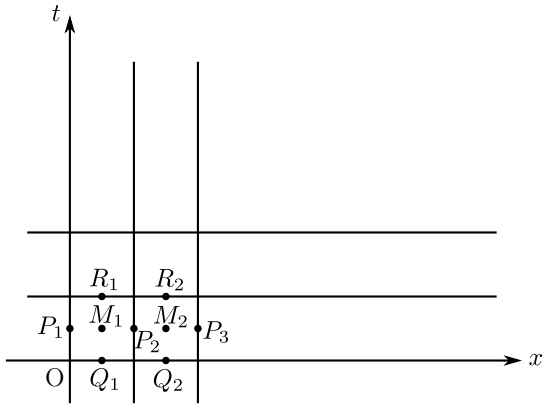


Figure 8.7:

Assume $a > 0$. We then can have a boundary condition on $x = 0$. Hence from the boundary and initial conditions, we know the values of

82 u at P_1 and Q_1. Using equations (8.20) (ii) we can express the values at R_1 and P_2 in terms of that at M_1. Now (8.20) (i) reduces to an equation in one unknown, viz., the value at M_1 and we can solve for this explicitly and from this get back the values at R_1 and P_2.

Thus step by step we can evaluate, explicitly, all the unknowns. The scheme is therefore quasi-explicit. The scheme is of second order accuracy. However, for the stability one cannot use the Fourier transform. But one can get an energy inequality and we leave this as an exercise.

Exercise 8.3. Using an energy inequality show that the DSN-scheme is unconditionally stable.

8.4 Comparison of the Schemes Above

We summarize the main features of the preceding schemes in table 8.1.

Type	Name of the scheme	(L^2) Stability	Order of accuracy	Possibility of extension to systems	Comments
Explicit	Centred scheme.	Always unstable	-	-	Never used
	Lax's Scheme.	$\lvert a\rvert\frac{\Delta t}{\Delta x} \leq 1$ for stability	1	Yes	L^∞-stable as well
	One-sided scheme	〃	1	No	
	Lax-Wendroff scheme.	〃	2	Yes	This is the best among the explicit schemes. But it is not L^∞-stable.
Implicit	Crank-Nicolson scheme	Unconditionally stable	2	Yes	Rarely used
Quasi-explicit	SNG-Scheme	〃	1	By method of characteristics	Used for neutron transport equation
	DSN-Scheme	〃	2	Yes, but extension is purely implicit	〃
	Robert-Weiss' scheme	〃		Rarely used	

Remark 8.4. As mentioned in Table 8.1, the Lax-Wendroff scheme is not L^∞-stable. (Cf. Themée [36]). If the initial value of the solution of the advection equation has a shock, so has $u(x,t)$ for any time t. If one uses the Lax-Wendroff scheme, the approximate solution has a lot of oscillations about such a discontinuity. This behaviour is related to the L^∞-instability of the scheme. The Lax and the one-sided scheme, on the other hand, are L^∞-stable but their approximate solutions are too smooth around the point of discontinuity of the true solution and this does not represent the ture situation either. In practice one has to exercise one's own judgement as to which behaviour is preferable! One can also add a non-linear dissipative term to the equation (8.1) which is negligible everywhere except at a shock where it has to smooth out the oscillations. See Boris and Book [4] and Van de Leer [37].

8.5 The non-linear equations

The non-linear equation takes the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u)) = 0. \tag{8.21}$$

If one wants to imitate the Lax-Wendroff scheme, one again motivates this by the Taylor expansion:

$$\begin{aligned} u^{n+1} &= u^n + \Delta t u_t + \frac{\Delta t^2}{2} u_{tt} + \dots \\ &= u^n - \Delta t (f(u))_x + \frac{\Delta t^2}{2} \frac{\partial}{\partial x} \left(f'(u) \frac{\partial f(u)}{\partial x} \right) + \dots \end{aligned}$$

84 again using (8.21). In case of a non-linear *system*, the matrix $f'(u)$ might be very difficult to compute. Hence one modifies the Lax - Wendroff scheme to avoid this difficulty and permit generalization to systems. This is called the *2-step Lax-Wendroff scheme* which is described as follows:

Step I. Using u^n_{i+1} and u^n_i, one obtains, say, by the Lax's scheme, $u^{n+\frac{1}{2}}_{i-\frac{1}{2}}$. Using u^n_i, u^n_{i-1} one obtains $u^{n+\frac{1}{2}}_{i-\frac{1}{2}}$ in the same fashion.

Step II. Using the results of step I, we write the scheme

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{f\left(u_{i-\frac{1}{2}}^{n+\frac{1}{2}}\right) - f\left(u_{i-\frac{1}{2}}^{n+\frac{1}{2}}\right)}{\Delta x} = 0. \tag{8.22}$$

In the linear case when $f(u) = au$, this reduces to the usual Lax-Wendroff scheme (check!).

The scheme we have just described is a memeber of class of schemes known as S_α^β-schemes, all of which reduce to the Lax-Wendroff scheme in the linear case. These have been studied by Lerat and Peyret [25].

All these schemes give oscillations about a shock. To treat this we can add a dissipative term to the differential equation so that this term is small where the gradient of the solution is small and acts only where the gradient is large i.e. at a shock. One such term is the pseudo-viscosity term and in this case the equation reads as

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(f(u)) - \epsilon \frac{\partial}{\partial x}\left(\left|\frac{\partial f'(u)}{\partial x}\right| \frac{\partial u}{\partial x}\right) = 0. \tag{8.23}$$

One uses the Lax-Wendroff scheme for the first two terms and add an explicit approximation to the dissipative term. **85**

There is no rigorous rule for the choice of the dissipative term. However, one can apply a dimension analysis to get some idea about it.

Raviart [31] has studied the convergence when $\Delta x, \Delta t \to 0$ of three schemes for the equation (8.23) with $f(u) = \frac{1}{2}u^2$, but with ϵ fixed. This of course, is not exactly the state of affairs since we require ϵ to be small and $\to 0$. How ever, this is a step in the right direction.

8.6 Boundary conditions

Let us consider the following problem with boundary conditions:

$$\left.\begin{aligned} &\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0, \quad a > 0, \quad 0 < x < 1 \\ &u(x, 0) = u_\circ(x), \text{ (initial value)} \\ &u(0, t) = 0 \text{ (boundary value)} \end{aligned}\right\} \tag{8.24}$$

We have seen that when there is no condition on $x = 1$, this problem is well posed. (Cf. Sec. 5.3).

The various numerical schemes cited for the purely initial-value problems are all three point schemes involving $i-1$, i and $i+1$ of the previous level. However, in case of the bounded domain, if $0 \leq i \leq I$, then the equation for u_I^{n+1} will involve u_{I+1}^n which lies outside the domain and hence we do not know it. Thus one feels the need for extrapolating u^n to the point $I + 1$ as well. One such interpolation is provided by

$$u_{I+1} = u_I. \tag{8.25}$$

However, this is not a sufficiently accurate choice. A much better choice

$$u_{I+1} = 2u_I - u_{I-1}. \tag{8.26}$$

86

One could, of course, give better extrapolation formulae, but as all schemes are of atmost second order accuracy, the formula (8.26) is quite sufficient for our purposes.

Exercise 8.4. Apply the formulae (8.25) and (8.26) to the Lax-Wendroff scheme and show that (8.25) gives an inconsistent scheme while that given by (8.26) is consistent.

Remark 8.5. The stability of problems with boundary conditions is usually studied by the energy method. This problem has been studied deeply by kreiss [19].

8.7 The leap-frog scheme

This scheme is given by

$$\frac{u_i^{n+1} - u_i^{n-1}}{2\Delta t} + a\frac{(u_{i+1}^n - u_{i-1}^n)}{2\Delta x} = 0. \tag{8.27}$$

This is a 3-level scheme which, though not widely used for the advection equation, is very useful for systems of hyperbolic equations,

especially for the wave equation. It is easy to see that the scheme has an error of discretization, of order $O(\Delta x^2 + \Delta t^2)$ and hence is of second order of accuracy.

Setting $v_i^n = u_i^{n-1}$, we get the system

$$\begin{bmatrix} u_i^{n+1} \\ v_i^{n+1} \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{i-1}^n \\ v_{i-1}^n \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_i^n \\ v_i^n \end{bmatrix} + \begin{bmatrix} -\alpha & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{i+1}^n \\ v_{i+1}^n \end{bmatrix}$$

which gives

$$\begin{bmatrix} \hat{u}^{n+1}(\xi) \\ \hat{v}^{n+1}(\xi) \end{bmatrix} = \begin{bmatrix} 2iA & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{u}^n(\xi) \\ \hat{v}^n(\xi) \end{bmatrix}$$

where $A = a\dfrac{\Delta t}{\Delta x} \sin \xi \Delta x$. **87**

The eigenvalues are given by the roots of the equation

$$\lambda^2 - 2iA\lambda - 1 = 0.$$

If $1 - A^2 \geq 0$, then

$$\lambda = iA \pm \sqrt{1 - A^2}$$

and if $1 - A^2 \leq 0$, then

$$\lambda = iA \pm i\sqrt{A^2 - 1}$$

In the latter case $|\lambda| > 1$ for at least one λ and in the former case $|\lambda| = 1$ for both eigen values. Thus we get the stability condition $|A| \leq 1$ for all ξ, which is the condition

$$|a|\frac{\Delta t}{\Delta x} \leq 1. \tag{8.28}$$

Note that this condition is *necessary*.

We can get a *sufficient* condition if the matrix of amplification is diagonalizable. Let

$$A(\xi) = S(\xi)D(\xi)S^{-1}(\xi).$$

Then $A^n(\xi) = SD^nS^{-1}(\xi)$. By the preceding necessary condition for stability, one has $\|D^n\|$ bounded. Thus a sufficient condition is

$$\max_{\xi}(\|S(\xi)\| + \|S^{-1}(\xi)\|) \leq \text{ constant}. \tag{8.29}$$

Applying this to the Leap-frong scheme, if $\alpha = |a|\frac{\Delta t}{\Delta x} < 1$, then $|A|^2 < 1$ and the eigenvalues are distinct. The sufficient condition (8.29) is indeed satisfied. (Check!).

On the other hand, if $\alpha = 1$ and $\xi\Delta x = \pi/2$, then $\lambda_1 = \lambda_2$ and A is not diagonalizable. Thus the sufficient condition is not satisfied. Though
88 this is no proof of instability, we give below an example of this situation where the scheme is indeed unstable.

Exercise 8.5. In the leap-frog scheme with $\alpha = 1$, given $u_j^\circ = \exp\left(\frac{i\pi j}{2}\right)$ and $u_j^1 = \exp\left(\frac{i\pi(j+1)}{2}\right)$, show that

$$u_j^n = (2n-1)\exp.\left(\frac{i\pi}{2}(j+2-n)\right), \quad n \geq 1,$$

and hence the scheme is unstable.

Finally, when dealing with bounded domains, we need to extrapolate at the point $I+1$, where $0 \leq i \leq I$ are the nodes of the mesh. The following procedure of extrapolation can be proved (by the energy method) to be stable although it is probably not the best one w.r.t. accuracy:

$$u_{I+1}^n = \frac{1}{2}(u_I^{n+1} + u_I^{n-1}). \tag{8.30}$$

One can show, by energy methods, the stability of the scheme under the sufficient condition

$$|a|\frac{\Delta t}{\Delta x} < 1. \tag{8.31}$$

We demonstrate this when the domain is $\mathbb{R}$ and leave the boundary condition case as an exercise.

Multiplying the equation (8.27) by $\Delta x(u_i^{n+1} + u_i^{n-1})$ and summing over all i, we get

$$\frac{|u^{n+1}|^2 - |u^{n-1}|^2}{2\Delta t} + \frac{a}{2}\left(\sum_i (u_{i+1}^n - u_{i-1}^n)u_i^{n+1} + \sum_i (u_{i+1}^n - u_{i-1}^n)u_i^{n-1}\right) = 0, \tag{8.32}$$

where $|\cdot|$ is the ℓ^2-norm induced by the innerproduct defined in sec. 7.6 (Cf. equation 7.18). Now, one can easily verify that the following holds (by a simple)

$$\sum_i (u_{i+1}^n - u_{i-1}^n) u_i^{n-1} = -\sum_i (u_{i+1}^{n-1} - u_{i-1}^{n-1}) u_i^n. \tag{8.33}$$

We now define **89**

$$X^{n+\frac{1}{2}} = |u^{n+1}|^2 + |u^n|^2 + \frac{a\Delta t}{\Delta x} \cdot \Delta x \sum_i (u_{i+1}^n - u_{i-1}^n) u_i^{n+1}. \tag{8.34}$$

Using (8.33) and (8.34), (8.32) becomes

$$X^{n+\frac{1}{2}} = X^{n-\frac{1}{2}}.$$

Proceeding recursively, one gets

$$X^{n+\frac{1}{2}} = X^{n-\frac{1}{2}} = \ldots = X^{\frac{1}{2}}. \tag{8.35}$$

Note that $X^{\frac{1}{2}}$ is expressed in terms of u° and u^1 and hence is known. Further $X^{n+\frac{1}{2}}$ is bounded.

Observe that

$$\begin{aligned} |\Delta x \sum_i (u_{i+1}^n - u_{i-1}^n) u_i^{n+1}| &\leq \sqrt{\Delta x \sum_i |u_{i+1}^n - u_{i-1}^n|^2 \cdot |u^{n+1}|} \\ &\leq 2|u^n| \cdot |u^{n+1}| \\ &\leq |u^n|^2 + |u^{n+1}|^2, \end{aligned}$$

by using the Cauchy-Schwarz and the Minkowski inequalities for the innerproduct and norm respectively and also the fact that non-negative a and b, $2ab \leq a^2 + b^2$. Thus

$$\begin{aligned} |X^{n+\frac{1}{2}}| &\leq |u^{n+1}|^2 + |u^n|^2 + |\alpha|(|u^{n+1}|^2 + |u^n|^2) \\ &= (1 + |\alpha|)(|u^{n+1}|^2 + |u^n|^2). \end{aligned}$$

Also

$$|X^{n+\frac{1}{2}}| \geq |(|u^{n+1}|^2 + |u^n|^2) - |\alpha| \, |\Delta x \sum_i (u_{i+1}^n - u_{i-1}^n) u_i^{n+1}||$$

But **90**

$$
\begin{aligned}
&|u^{n+1}|^2 + |u^n|^2 - |\alpha||\Delta x \sum_i (u^n_{i+1} - u^n_{i-1})u^{n+1}_i| \\
&\geq |u^{n+1}|^2 + |u^n|^2 - |\alpha|(|u^{n+1}|^2 + |u^n|^2) \\
&= (1 - |\alpha|)(|u^{n+1}|^2 + |u^n|^2) \geq 0 \text{ if } |\alpha| \leq 1.
\end{aligned}
$$

Thus one has

$$(1 - |\alpha|)(|u^{n+1}|^2 + |u^n|^2) \leq |X^{n+\frac{1}{2}}| \leq (1 + |\alpha|)(|u^{n+1}|^2 + |u^n|^2). \quad (8.36)$$

If $|\alpha| < 1$ (as in (8.31)), we get

$$|u^{n+1}|^2 + |u^n|^2 \leq \frac{1}{(1 - |\alpha|)}|X^{n+\frac{1}{2}}| = \frac{1}{(1 - |\alpha|)}|X^{\frac{1}{2}}| \leq \left(\frac{1 + |\alpha|}{1 - |\alpha|}\right)(|u^1|^2 + |u^\circ|^2).$$

Since $\dfrac{1 + |\alpha|}{1 - |\alpha|}$ is a constant, we get the energy inequality

$$|u^n|^2 \leq (\text{const.})\ (|u^\circ|^2 + |u^1|^2).$$

for all n. This implies the stability of the scheme.

Exercise 8.6. Generalize this to the case when the domain is $0 < x < 1$, with $a > 0$ and the boundary condition $u(\circ, t) = 0$ on $x = 0$.

8.8 The phase error

We saw in Sec. 7.4 the importance of comparing the coefficient of amplification $a(\xi)$ with the exact coefficient of amplification of the given equation. In general these are complex numbers and one can compare either their moduli or their arguments.

The error in the modulus of the coefficient of amplification is related to the dissipativity of the scheme.

The *phase error* is the error in the argument of the coefficient of amplification and this is related to the error in the velocity of propagation of the wave. For instance, in case of the advection equation if $u_\circ$ is the

91 initial value, then there is a phase lag of - at in the solution (when a is a constant) at time t. When we set up a numerical scheme one would like to know how much the error in the phase lag will be. Generally, given a particular scheme it may turn out that it is satisfactory w.r.t. one of these errors but not w.r.t. the other and we seek to modify the scheme so as not to spoil the good behaviour w.r.t. one error and at the same time improve the behaviour w.r.t. the other. We illustrate this with the following example.

Given the advection equation, one has the exact coefficient of amplification (Cf. equation (6.9))

$$a_{ex}(\xi, t) = \exp . (ia\xi t). \tag{8.37}$$

In the case of the Lax-Wendroff scheme one has

$$a(\xi) = 1 + \alpha^2(\cos(\xi\Delta x) - 1) + i\alpha \sin(\xi\Delta x) \tag{8.38}$$

where $\alpha = \dfrac{a\Delta t}{\Delta x}$ and for stability one must have $|\alpha| \leq 1$. We note that

$$\left.\begin{aligned} |a_{ex}(\xi)| &= 1 \\ |a(\xi)|^2 &= 1 - 4\alpha^2(1-\alpha^2)\sin^4\left(\frac{\xi\Delta x}{2}\right) \end{aligned}\right\} \tag{8.39}$$

and the error $|a_{ex}(\xi)| - |a(\xi)|$ at time Δt is of third order (i.e. of order $O((\xi\Delta x)^4)$ for small ξ). Hence the Lax-Wendroff scheme is satisfactory as far as the error in modulus is concerned.

However, the phase in the exact case is a $\xi\Delta t$ while in the scheme it is

$$\tan^{-1}\left[\frac{\alpha \sin(\xi\Delta x)}{1 + \alpha^2(\cos(\xi\Delta x) - 1)}\right].$$

Hence the phase error is

$$\begin{aligned} &a\xi\Delta t - \tan^{-1}\left[\frac{\alpha \sin(\xi\Delta x)}{1 + \alpha^2(\cos(\xi\Delta x) - 1)}\right] \\ &= \frac{\alpha(1-\alpha^2)}{6}(\xi\Delta x)^3 + 0(\xi\Delta x)^4 > 0 \end{aligned} \tag{8.40}$$

when $|\alpha| < 1$, which is only of second order accuracy and is not very **92** satisfactory. One thus tries to devise schemes which reduce this phase error.

A method due to Fromme [11, 12] is to set up a scheme with the same order of phase error, but which is < 0 and then take a linear combination of these two schemes.

We take S1 as the Lax-Wendroff scheme and the scheme S2 defined analogously as follows:

By Remark 8.2. on the interpretation of the scheme S1 via characteristics, one saw that it was got by quadratic interpolation between the points $(i-1)\Delta x$, $i\Delta x$ and $(i+1)\Delta x$ of the point where the characteristic through $(i\Delta x, (n+1)\Delta t)$ meets the level $n\Delta t$. To get the scheme S2, we interpolate this same point quadratically between $(i-2)\Delta x$, $(i-1)\Delta x$ and $i\Delta x$. Explicitly, the scheme reads as

$$u_i^{n+1} = u_i^n - \frac{a\Delta t}{2\Delta x}(u_{i-2}^n - 4u_{i-1}^n + 3u_i^n) + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{i-2}^n - 2u_{i-1}^n + u_i^n). \quad (8.41)$$

Then on computing as before we get the phase error,

$$\frac{-\alpha(\alpha-1)(\alpha-2)}{6}(\xi\Delta x)^3 + 0((\xi\Delta x)^4) < 0$$

when $|\alpha| < 1$.

Using these two shcemes, Fromm defined the *zero average phase error scheme (SO) by*

$$SO = \frac{1}{2}(S1) + \frac{1}{2}(S2). \quad (8.42)$$

93 More generally one can devise the scheme

$$SO \equiv \left(\frac{2-\alpha}{3}\right)(S1) + \frac{(1+\alpha)}{3}(S2). \quad (8.43)$$

Note that when α takes the average value of its range viz. $\alpha = \frac{1}{2}$, we get (8.42) (8.43) to give the same scheme.

The scheme (8.43) is of third order of accuracy and can be considered (when a is a constant) as a cubic interpolation between the points

$i-2$, $i-1$, i and $i+1$ of the point we are interested in. By taking this linear combination, we cancel the $(\xi\Delta x)^3$ term in the phase error and improve our accuracy.

This scheme is also L^∞-stable. Rusanow [35] and Burstein and Mirin [5] have given third order schemes in the non-linear case.

8.9 Hyperbolic systems

We now describe briefly how some of these schemes could be extended to a hyperbolic system of equations

If $U^T = (u_1(x,t), \ldots, u_n(x,t))$ is a vector and A is an $n \times n$ matrix, the equation reads as

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0. \tag{8.44}$$

We assume, for simplicity, that A is a constant matrix.

In this case the schemes of Lax, Crank-Nicolson and Lax-Wendroff all generalize very easily. For instance, the Lax-Wendroff scheme will be

$$U_i^{n+1} = U_i^n - \Delta t.A\frac{(U_{i+1}^n - U_{i-1}^n)}{2\Delta x} + \frac{\Delta t^2}{2\Delta x^2}A^2(U_{i+1}^n - 2U_i^n + U_{i-1}^n). \tag{8.45}$$

This scheme is of second order of accuracy.

In general, 2-level schemes assume the form

$$\sum_j P_j(A)U^{n+1}(x + j\Delta x) = \sum_j Q_j(A)U^n(x + j\Delta x) \tag{8.46}$$

where the matrices $P_j(A)$ and $Q_j(A)$ are all polynomials in the matrix A. For instance, in the Lax-Wendroff scheme, **94**

$$\begin{aligned} P_\circ(A) &= I; P_j(A) = 0, . \quad j \neq 0 \\ Q_{-1}(A) &= \frac{\Delta t^2}{2\Delta x^2}A^2 + \frac{\Delta t}{2\Delta x}A. \\ Q_\circ(A) &= I - \frac{\Delta t}{\Delta x^2}A^2, \end{aligned}$$

$$Q_1(A) = \frac{\Delta t^2}{2\Delta x^2}A^2 - \frac{\Delta t}{2\Delta x}A.$$

To study the stability of the scheme (8.46) we assume A is diagonalizable, i.e.

$$A = SDS^{-1}, \tag{8.47}$$

with D diagonal. Defining $V = S^{-1}U$, and multiplying (8.44) on the left by S^{-1}, we get

$$\frac{\partial V}{\partial t} + D\frac{\partial V}{\partial x} = 0 \tag{8.48}$$

from which we get the n scalar equations

$$\frac{\partial v_i}{\partial t} + \lambda_i \frac{\partial v_i}{\partial x} = 0, \quad 1 \le i \le n. \tag{8.49}$$

The stability condition for each of these n equations would be

$$|\lambda_i|\frac{\Delta t}{\Delta x} \le 1, \tag{8.50}$$

for those schemes (8.46) which are generalised from the scalar case. Taking the most restrictive of these, we get

$$\rho(A)\frac{\Delta t}{\Delta x} \le 1. \tag{8.51}$$

To see that we do indeed get the same condition starting from the scheme (8.46) we set $V^n = S^{-1}U^n$ and multiply (8.46) on the left by S^{-1}. Then using the fact that

$$S^{-1}P_j(A)S = P_j(D).$$

95 we get

$$\sum_j P_j(D)V^{n+1}(x + j\Delta x) = \sum_j Q_j(D)V^n(x + j\Delta x)$$

which again splits into n scalar scheme equations

$$\sum_j P_j(\lambda_i)v_i^{n+1}(x + j\Delta x) = \sum_j Q_j(\lambda_i)v_i^n(x + j\Delta x), \quad 1 \le i \le n.$$

Once again the scalar case demands that $|\lambda_i|\dfrac{\Delta t}{\Delta x} \leq 1$, which gives the stability condition (8.51).

The one-sided or the SNG-schemes do not generalise in a straightforward manner to the case of a system. However, as in these schemes one can set up schemes involving characteristics. This is known as the *Method of characteristics*.

If the system is strictly hyperbolic, we can find the left-eigen vectors p_k corresponding to the eigenvalue λ_k of A such that

$$p_k^T A = \lambda_k p_k^T, \quad 1 \leq k \leq n. \tag{8.52}$$

If C_k is the characteristic defined by $\dfrac{dx}{dt} = \lambda_k$, and if $\dfrac{d}{ds_k}$ stands for the differentiation along C_k, we have

$$p_k^T \frac{dU}{ds_k} = 0, \quad 1 \leq k \leq n. \tag{8.53}$$

(Cf. Sec. 2.2., equation (2.10)). Our subsequent notations will be based on Fig. 8.8.

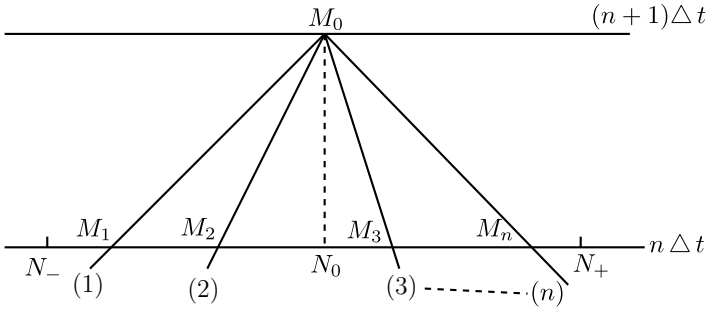


Figure 8.8:

We choose a grid such that all the characteristics through $M_\circ = (i\Delta x, (n+1)\Delta t)$ meet the line $n\Delta t$ between $(i-1)\Delta x$ and $(i+1)\Delta x$. These two nodes have been denoted by N_- and N_+. The points where the characteristics $(1), \ldots, (n)$ meet $n\Delta t$ are denoted by $M_1, \ldots M_n$, respectively. Finally $N_\circ = (i\Delta x,\ n\Delta t)$. **96**

From the relation (8.53), we write the approximate scheme

$$p_k^T(U(M_\circ) - U(M_k)) = 0, \quad 1 \leq k \leq n. \tag{8.54}$$

To compute $U(M_k)$, we use linear interpolation between $N_\circ$ and N_+ or between $N_\circ$ and N_- according as λ_k is < 0 or > 0. Thus we get n equations for the n-components of $U(M_\circ)$ and we can solve this, in principle. This generalises the SNG-scheme.

Remark 8.6. If we use quadratic interpolation between N_-, $N_\circ$ and N_+ we get an analogue of the Lax-Wendroff scheme.

The stability condition is the same as condition (8.51). The interpretation of this condition is that $[M_1, M_n] \subset [N_-, N_+]$ got from the Courant-Friedrichs-Lewy convergence condition, viz., the exact domain
97 of dependence must lie within the approximate domain of depnedence. Again to get condition (8.51) we take the case for each characteristic and choose the strongest inequality amongst them.

8.10 Non-linear systems-method of characteristics

If we have a pure dependence on U of the matrix A i.e. $A = A(U)$, we generalise the method of characteristics. The approximation of the equation (8.53) will have to be centred between $M_\circ$ and M_k for each k. We write

$$\frac{p_k^T(U(M_k)) + p_k^T(U(M_\circ))}{2}(U(M_\circ) - U(M_k)) = 0. \tag{8.55}$$

and

$$M_kN_\circ = \frac{\Delta t}{2}[\lambda_k(U(M_\circ)) + \lambda_k(U(M_k))]. \tag{8.56}$$

The unknowns are x_k, the abscissa of M_k, which can be got from the evaluation of $M_kN_\circ$, and also the n-components of $U(M_\circ)$ occurring in equation (8.55).

We usually solve the system (8.55) - (8.56) by iterative methods. We assume $U^\circ(M_\circ)$ to start with. We assume all values at level $n\Delta t$. Then

setting

$$\left.\begin{aligned}\lambda_k^{(\nu)} &= \frac{1}{2}(\lambda_k(U^{(\nu)}(M_\circ)) + \lambda_k(U(M_k^{(\nu)})))\\ p_k^{T(\nu)} &= \frac{1}{2}(p_k^T(U^{(\nu)}(M_\circ)) + p_k^T(U(M_k^{(\nu)})))\end{aligned}\right\} \tag{8.57}$$

the iterative method could be

$$\left.\begin{aligned}&M_k^{(\nu+1)}N_\circ = \lambda_k^{(\nu)}\Delta t.\\ &p_k^{T(\nu)}(U^{(\nu+1)}(M_\circ) - U(M_k^{(\nu+1)})) = 0.\end{aligned}\right\} \tag{8.58}$$

Now we use this to get $U^{(\nu+1)}(M_\circ)$ and $M_k^{(\nu+1)}N_\circ$. Thus we can again get $\lambda_k^{(\nu+1)}$ and $p_k^{T(\nu+1)}$ and proceed.

The convergence of these iterations depends upon the nature of the **98** non-linearity but is usually valid for small enough Δt.

Remark 8.7. When $n = 2$, and when we have one positive and one negative eigenvalue, we can modify the method of characteristics.

We fix a Δx and draw both the characteristics through each point $i\Delta x$. Through the intersections of these we draw two more and so on. (Cf. Fig. 8.9).

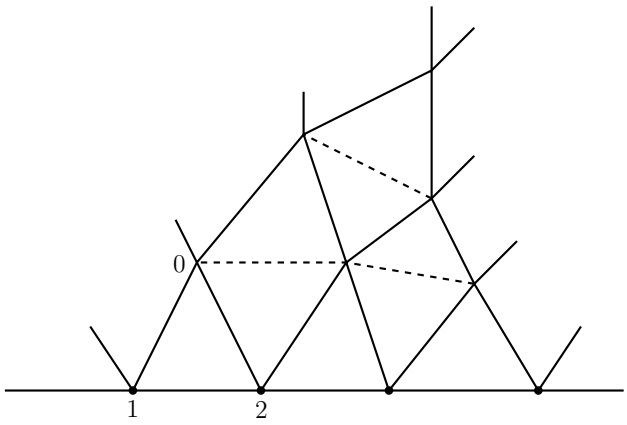


Figure 8.9:

If $U^T = (u, v)$, then the unknowns at 0, are its coordinates (x, t) and the values of u and v. Then we use the following set of equations to obtain these values.

$$\frac{x_\circ - x_1}{t_\circ - t_1} = \frac{\lambda_1(0) + \lambda_1(1)}{2} \tag{8.59}$$

$$\frac{x_\circ - x_2}{t_\circ t_2} = \frac{\lambda_2(0) + \lambda_2(2)}{2} \tag{8.60}$$

$$\frac{1}{2}(p_1^T(0) + P_1^T(1)).(U(0) - U(1)) = 0, \tag{8.61}$$

$$\frac{1}{2}(p_2^T(0) + p_2^T(2)).(U(0) - U(2)) = 0 \tag{8.62}$$

We solve these, step by step at each node. Note that the nodes no longer form a regular mesh if A is not constant.

99 It is this form of the method of characteristics which has been widely used in supersonic flow calculation. Its main advantage is that it does not need interpolation of the values of u and v which might be inaccurate.

REFERENCES: Apart from the references cited in the text, the reader could also look at the following papers: Lax [21]. Lax and Wendroff [23, 24], Boris and Book [4], Hirt [15], Hoskin [16], Kot [18], Kasahara [17] and Gourlay and Morris [13].

Chapter 9

Numerical Methods for the System of Equations of Hydrodynamics - Lagrangian Coordinates

9.1 Introduction

We now study the approximation of the system of equations of hydrodynamics in the slab-symmetric, 1-dimensional case, in the Lagrangian framework. We will esentially use the leap-frog scheme in discretizing these equations. We will also discuss the question of boundary conditions and the pseudo-viscosity term and present a heuristic discussion of stability criteria. **100**

With $\mu = 0$ and $g = 0$, the equations fo hydrodynamics assume the form

$$\frac{D\rho}{Dt} + \rho\frac{\partial u}{\partial x} = 0 \tag{9.1}$$

or, equivalently,

$$\frac{D}{Dt}(\rho J) = 0 \tag{9.1'}$$

$$\frac{Du}{Dt} + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0, \text{ and} \tag{9.2}$$

$$\frac{D\epsilon}{Dt} + p\frac{D}{Dt}\left(\frac{1}{\rho}\right) - \frac{1}{\rho}\frac{1}{\rho}\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) = 0 \tag{9.3}$$

(Note: We recall our comments on the Lagrangian form of the equation made in section 1.5. To be strictly Lagrangian in formulation, x must be expressed in terms of a).

9.2 Leap-Frog scheme for the isentropic case

We assume for the time being that $k = 0$ and also that (9.3) can be integrated to get p in terms of ρ, i.e. $p = p(\rho)$. (Example: The isentropic case of a perfect gas gives $p = c\rho^{\gamma}$) Then we are left with (9.1) and (9.2) to discretize.

101 Recalling our definition of m (Cf. Sec. 2.3) one can take as space variable either a or m. In using the leap-frog scheme one notes an essential property of this scheme: when dealing with a hyperbolic system such as (9.1) and (9.2) one need not compute all the unknowns at all the nodes of the mesh. One case compute certain unknowns on some of the nodes and the others at the remaining nodes. Thus here one computes ρ and p at the nodes $\left(i + \frac{1}{2}, n\right)$ and the velocity u at the nodes $(i, n + \frac{1}{2})$.

We approximate the equation defining u as $\dfrac{dx}{dt}$ by.

$$x_i^{n+1} = x_i^n + u_i^{n+\frac{1}{2}}\Delta t \tag{9.4}$$

where x_i°, is the same as the a_i chosen at the horizontal axis. (Note that (9.4) merely gives discretization of the relation between the Eulerian and Lagrangian coordinates). Then the equation ((9.1′)) says that the mass contained between x_{i+1}^{n+1} and x_i^{n+1} is independent of time, i.e. n. Hence we can write

$$\rho_{i+\frac{1}{2}}^{n+1}(x_{i+1}^{n+1} - x_i^{n+1}) = \text{ Const. } = \Delta m_{i+\frac{1}{2}} = \rho_{i+\frac{1}{2}}^{n}(x_{i+1}^{n} - x_i^{n}) \tag{9.5}$$

as a discretization of ((9.1′)) assuming ρ constant in this interval.

Actually (9.4) and (9.5) which are the discretizations of ((9.1′)) imply the following which can be thought as a discretization of (9.1)

$$\frac{\rho_{i+\frac{1}{2}}^{n+1} - \rho_{i+\frac{1}{2}}^{n}}{\Delta t} + \frac{1}{2}(\rho_{i+\frac{1}{2}}^{n+1} + \rho_{i+\frac{1}{2}}^{n}) \frac{(u_{i+1}^{n+\frac{1}{2}} - u_i^{n+\frac{1}{2}})}{(\frac{1}{2}(x_{i+1}^{n+1} + x_{i+1}^{n}) - \frac{1}{2}(x_i^{n+1} + x_i^{n}))} = 0. \quad (9.6)$$

If we define

$$\Delta m_i = \frac{1}{2}(\Delta m_{i+\frac{1}{2}} + \Delta m_{i-\frac{1}{2}}) \quad (9.7)$$

one can discretize (9.2) by

$$\frac{u_i^{n+\frac{1}{2}} - u_i^{n-\frac{1}{2}}}{\Delta t} + \frac{p_{i+\frac{1}{2}}^{n} - p_{i-\frac{1}{2}}^{n}}{\Delta m_i} = 0 \quad (9.8)$$

102

This completes the discretization of the first two equations of hydrodynamics using the leap-frog scheme.

9.3 Boundary conditions

Let us examine the question of boundary conditions. One usually encounters one of the following types of boundary conditions:

(i) u is known on the boundary,

or (ii) p is known on the boundary.

For instance, if we consider a gas enclosed in a tube on one side of which a piston is working, then we encounter the problem of a moving boundary. However, since the same particles remain on the piston, this boundary is fixed in the Lagrangian system.

Let us assume that the velocity is known on the boundary, i.e. $u_I^{n+\frac{1}{2}}$ is known for all n. Since by equation (9.8), knowing all quantitites upto and inclusive of level n, one can compute $u_i^{n+\frac{1}{2}}$ for all $i < I$, we have complete knowledge of u_i^{n+1}, for all i. Knowing these, one immediately uses (9.4) to get x_i^{n+1} for all i. Then one can use either (9.5) or (9.6) to get

$\rho_{i+\frac{1}{2}}^{n+1}$ for all i, and since p is a known function of ρ, we can compute $p_{i+\frac{1}{2}}^{n+1}$ as well. Thus we can sucessfully deal with the first type of boundary condition.

Given the pressure on the boundary, the situation is not so straightforward. Of course, knowing all quantities upto level n, one can compute $u_i^{n+\frac{1}{2}}$ for $i < I$ using (9.8). But in order to compute $u_I^{n+\frac{1}{2}}$, i.e. the velocity on the boundary, we are obliged to use a one-sided difference quotient to approximate $\dfrac{\partial p}{\partial m}$. Thus for the index I above, we discretize (9.2) by

$$\frac{u^{n+\frac{1}{2}} - u_I^{n-\frac{1}{2}}}{\Delta t} + \frac{p_I^n - p_{I-\frac{1}{2}}^n}{m_I - m_{I-\frac{1}{2}}} = 0. \tag{9.9}$$

103

Knowing p_I^n (given by boundary condition) we can compute $u_I^{n+\frac{1}{2}}$. Now the rest of the unknowns are calculated as in the previous case.

9.4 Discretization of the energy equation

Let us now return to the original equations of hydrodynamics. Let us continue to assume that $k = 0$, but now suppose that we connot integrate equation (9.3) to get p as a function of ρ. Then we bring in the state equation $\epsilon = f(p, \rho)$ and discretize this together with (9.3). We compute ϵ also at $(i + \frac{1}{2}, n)$ like p and ρ. Thus our discretization will read as

$$\frac{\epsilon_{i+\frac{1}{2}}^{n+1} - \epsilon_{i+\frac{1}{2}}^n}{\Delta t} + \frac{1}{2}\left(p_{i+\frac{1}{2}}^{n+1} + p_{i+\frac{1}{2}}^n\right)\left(\frac{1}{\rho_{i+\frac{1}{2}}^{n+1}} - \frac{1}{\rho_{i+\frac{1}{2}}^n}\right)\frac{1}{\Delta t} = 0. \tag{9.10}$$

and

$$\epsilon_{i+\frac{1}{2}}^{n+1} = f\left(p_{i+\frac{1}{2}}^{n+1}, \rho_{i+\frac{1}{2}}^{n+1}\right). \tag{9.11}$$

Now assuming knowledge of all quantities upto level n, one can get $u_i^{n+\frac{1}{2}}$, x_i^{n+1} and $\rho_{i+\frac{1}{2}}^{n+1}$ for all i. Then substituting in (9.10) and (9.11), we get a system of two non-linear equations to solve for $\epsilon_{i+\frac{1}{2}}^{n+1}$ and $p_{i+\frac{1}{2}}^{n+1}$ for each i.

However, if the temperature term is also present in (9.3), (i.e. $k \neq 0$), one can approximate the term $-\dfrac{\partial}{\partial x}\left(k\dfrac{\partial T}{\partial x}\right)$ by a 3-point formula. Again, T is also computed at $\left(i+\frac{1}{2}, n\right)$ and $T = T(\epsilon, \rho)$ is a known function of ϵ and ρ. The discretization of (9.3) assumes the form

$$\frac{\epsilon^{n+1}_{i-\frac{1}{2}} - \epsilon^{n}_{i+\frac{1}{2}}}{\Delta t} + \frac{1}{2}(p^{n+1}_{i+\frac{1}{2}} + p^{n}_{i+\frac{1}{2}})\frac{1}{\Delta t}\left(\frac{1}{\rho^{n+1}_{i+\frac{1}{2}}} - \frac{1}{\rho^{n}_{i+\frac{1}{2}}}\right) - \frac{2}{\rho^{n+1}_{i+\frac{1}{2}} + \rho^{n}_{i+\frac{1}{2}}}\nabla^2\frac{1}{2}(T^{n+1} + T^{n}) = 0 \tag{9.12}$$

where, $\nabla^2(T)$ is the usual approximation for $-\dfrac{\partial}{\partial x}(k\dfrac{\partial T}{\partial x})$ involving the values $T_{i+3/2}$, $T_{i+\frac{1}{2}}$ and $T_{i-\frac{1}{2}}$. **104**

$$\nabla^2(T) = \frac{1}{(x_{i+1} - x_i)}\left[k_{i+1}\frac{(T_{i+3/2} - T_{i+\frac{1}{2}})}{x_{i+3/2} - x_{i+\frac{1}{2}}} - k_i\frac{(T_{i-\frac{1}{2}} - T_{i-\frac{1}{2}})}{x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}}\right]. \tag{9.13}$$

Now equations (9.11), (9.12) get coupled and we have to solve a system which is non-linear and "tridiagonal" in the following sense: if $Z^T_{i+\frac{1}{2}} = (\epsilon, T, \rho)^{n+1}_{i+\frac{1}{2}}$, then the system of equations is in the form

$$F_{i+\frac{1}{2}}(Z_{i+3/2}, Z_{i+\frac{1}{2}}, Z_{i-\frac{1}{2}}) = 0.$$

These will have to be linearized by Newton's method or solved by some type of Gauss-Seidel iteration techniques.

We now say a few words about the use of internal and total energies in our equations. We gave (9.3) in terms of the internal energy ϵ. However as was done in Sec. 3.4 we can multiply (9.2) by u and add it to (9.3) to get

$$\frac{D}{Dt}\left(\epsilon + \frac{1}{2}u^2\right) + \frac{\partial}{\partial m}(pu) = 0 \tag{9.14}$$

which, in view of the relation $E = \frac{1}{2}u^2 + \epsilon$, gives the equation in terms of the total energy. One can do the same thing in the discrete case as well. One can discretize (9.2) and (9.3) and by multiplying the former by "something like u" and adding to the latter, we can get a proper discretization of (9.14).

We show how this is done only one the semi-discrete form of the equation (i.e. discretization w.r.t. the space variable only and keeping all derivatives w.r.t. time). Descritizing only w.r.t. m, we can write the semi-discrete analogue of (9.8) as

$$\frac{du_i}{dt} + \frac{p_{i+\frac{1}{2}} - p_{i-\frac{1}{2}}}{\Delta m_i} = 0. \tag{9.15}$$

105 We can also write

$$\frac{d\epsilon_{i+\frac{1}{2}}}{dt} + p_{i+\frac{1}{2}} \frac{(u_{i+1} - u_i)}{\Delta m_{i+\frac{1}{2}}} = 0 \tag{9.16}$$

because from (9.4) and (9.5),

$$\frac{1}{\Delta t}\left(\frac{1}{\rho^{n+1}_{i+\frac{1}{2}}} - \frac{1}{\rho^{n}_{i+\frac{1}{2}}}\right) = \frac{u^{n+\frac{1}{2}}_{i+1} - u^{n+\frac{1}{2}}_{i}}{\Delta m_{i+\frac{1}{2}}},$$

thus giving (9.16) as the semi-discrete analogue of (9.10).

Writing (9.15) again at $i+1$ as well and multiplying the former by $\frac{u_i}{2}$ and the latter by $\frac{u_{i+1}}{2}$ and adding to (9.16) one gets

$$\frac{d}{dt}(\epsilon_{i+\frac{1}{2}} + \frac{1}{4}(u_i^2 + u_{i+1}^2)) + \frac{p^*_{i+1}u_{i+1} - p^*_i u_i}{\Delta m_{i+\frac{1}{2}}} = 0, \tag{9.17}$$

as the semi-discrete form of (9.14), where

$$p^*_i = \frac{p_{i-\frac{1}{2}}\Delta m_{i+\frac{1}{2}} + p_{i+\frac{1}{2}}\Delta m_{i-\frac{1}{2}}}{\Delta m_{i+\frac{1}{2}} + \Delta m_{i-\frac{1}{2}}} \tag{9.18}$$

One can now use a time discretization to get a numerical scheme in terms of E.

All this was just to show that the discrete form of the energy equation in terms of one of ϵ or E, implies the discrete form in terms of the other.

9.5 The pseudo-viscous term

The pseudo-viscous term is very important when one wants to compute solutions with shocks. However, it is not obvious as to what type of term should be added. We discuss here, the pseudo-viscous term of Richtmyer and von Neumann (See [31]).

We add the pseudo-viscous term in the following form to the equation of conservation of momentum, so that the modified equation new reads as **106**

$$\rho\frac{Du}{Dt} + \frac{\partial p}{\partial x} - \frac{\partial}{\partial x}\left(\sigma\frac{\partial u}{\partial x}\right) = 0. \tag{9.19}$$

For the problem to be well-posed in the time increasing case, we need $\sigma \geq 0$. Also the role of this term being "killing" oscillations about a shock, it must be small when the gradient is small and act only where the gradient is large. Generally, one takes σ as a function of $|\frac{\partial u}{\partial x}|$. The preudo-viscous term of Richtmyer and von Neumann is

$$\sigma = \rho 1^2 \max\left(0, -\frac{\partial u}{\partial x}\right). \tag{9.20}$$

Remark 9.1. Note that $\sigma > 0$ when $\dfrac{\partial u}{\partial x} < 0$ i.e. at a compression or a shock and is zero when $\dfrac{\partial u}{\partial x} > 0$, i.e. in case of a rarefaction wave.

We generally compute σ at the points $(i + \frac{1}{2}, n)$. Thus

$$\sigma^n_{i+\frac{1}{2}} = \rho^n_{i+\frac{1}{2}} a^2 (x^n_{i+1} - x^n_i)^2 \left|u^{n-\frac{1}{2}}_{i+1} - u^{n-\frac{1}{2}}_i\right| \tag{9.21}$$

where a^2 is a coefficient which determines the amount of reduction of the oscillations. (See Richtmyer and Morton [32]).

Thus the discretization of (9.19) reads as

$$\frac{u_i^{n+\frac{1}{2}} - u_i^{n-\frac{1}{2}}}{\Delta t} + \frac{p^n_{i+\frac{1}{2}} - p^n_{i-\frac{1}{2}}}{\Delta m_i} -$$

$$-\frac{1}{\Delta m_i}\left[\sigma^n_{i+\frac{1}{2}}\frac{u^{n-\frac{1}{2}}_{i+1}-u^{n-\frac{1}{2}}_i}{x^{n-\frac{1}{2}}_{i+1}-x^{n-\frac{1}{2}}_i}-\sigma^n_{i-\frac{1}{2}}\frac{\left(u^{n-\frac{1}{2}}_i-u^{n-\frac{1}{2}}_{i-1}\right)}{x^{n-\frac{1}{2}}_i-x^{n-\frac{1}{2}}_{i-1}}\right]=0 \tag{9.22}$$

Since this is a viscous term, one must add the work done by the viscous stress to the energy equation. We write

$$q=-\sigma\frac{\partial u}{\partial x}$$

107 so (9.19) reads as

$$\rho\frac{Du}{Dt}+\frac{\partial}{\partial x}(q+p)=0. \tag{9.23}$$

Thus, in discretising the energy equation, we must replace p in equations (9.10) by $p+q$. Only then will we get the correct equivalence between the usage of the internal energy and the total energy similar to what we saw at the end of Section 9.4.

9.6 Stability

In the Lagrangian case the mesh lines are not uniform. However, by locally linearising the problem we deduce some heuristic conditions for stability.

In the continuous case one has the two equations (9.1) and (9.19). We now assume that we linearise the problem locally and take

$$p-\bar{p}=\bar{c}^2(\rho-\bar{\rho}). \tag{9.24}$$

Then differentiating (9.19) w.r.t. t and using (9.1) and (9.24) one gets (assuming ρ locally),

$$\frac{\partial^2 u}{\partial t^2}-c^2\frac{\partial^2 u}{\partial x^2}-\frac{\partial^2}{\partial x\partial t}(\frac{\sigma}{\rho}(\frac{\partial u}{\partial x}))=0. \tag{9.25}$$

We imitate this in the discrete case. One has the discretization (9.6) (or, equivalently, (9.5)) of the equation (9.1) and (9.22) of the equation (9.19). Writing (9.22) at times n and $n-1$ subtracting these and dividing

by Δt, we perform the discrete analogue of differentiating (9.19) w.r.t. t. Now using (9.6) and linearising locally and assuming a uniform mesh (locally) of step Δx, and constant values of σ and ρ (again, locally), we get the discretization of (9.25) as

$$\frac{u_i^{n+\frac{1}{2}} - 2u_i^{n-\frac{1}{2}} + u_i^{n-\frac{3}{2}}}{\Delta t^2} - c^2\nabla^2 u^{n-\frac{1}{2}} - \frac{\sigma}{\rho}\nabla^2\frac{(u^{n-\frac{1}{2}} - u^{n-3/2})}{\Delta t} = 0 \quad (9.26)$$

where **108**

$$\nabla^2 u = \frac{1}{\Delta x^2}(u_{i+1} - 2u_i + u_{i-1}). \quad (9.27)$$

This is a 3-level scheme. Setting $v^{n+\frac{1}{2}} = u^{n-\frac{1}{2}}$, we get a 2-level system of two equations whose matrix of amplification is

$$A(\xi, \Delta t) = \begin{pmatrix} 2 - 4(\alpha^2 + \beta)\sin^2(\frac{1}{2}\xi\Delta x) & -1 + 4\sin^2(\frac{1}{2}\xi\Delta x) \\ 1 & 0 \end{pmatrix} \quad (9.28)$$

where

$$\alpha = \frac{c\Delta t}{\Delta x},\ \beta = \frac{\sigma}{\rho}\frac{\Delta t}{\Delta x^2} \quad (9.29)$$

The characteristic polynomial of this matrix is

$$p(\lambda) \equiv (\lambda^2 - 2\lambda + 1) + 4\alpha^2\sin^2\left(\frac{1}{2}\xi\Delta x\right)\lambda + 4\beta\sin^2\left(\frac{1}{2}\xi\Delta x\right)(\lambda - 1).$$

If we have $\alpha + \beta/\alpha \leq 1$, then the equation $p(\lambda) = 0$ will have two roots which are complex conjugates of each other. Then the product of the roots (which is merely the square of the modulus of either) can also be checked to be ≤ 1 and hence $\rho(A) \leq 1$ giving stability.

However, in order to have $\alpha + \beta/\alpha \leq 1$, one sees that one must have $1 - 4\beta \geq 0$. Thus the heuristic stability criteria are

$$\left.\begin{array}{ll} \text{(i)} & \beta \leq \dfrac{1}{4} \\ \text{(ii)} & \alpha + \beta/\alpha \leq 1 \end{array}\right\} \quad (9.30)$$

Remark 9.2. The condition (9.30) (i) resembles the stability condition for the heat equation and the condition on α viz. (9.30) (ii) is like the wave equation condition.

Remark 9.3. If $\sigma = 0$, this gives the usual Courant-Friedrichs-Lewy condition. **109**

Remark 9.4. In the Richtmyer-von Neumann pseudo-viscosity term, β is of the order of $a^2\Delta t\left|\frac{\partial u}{\partial x}\right|$. Hence by (9.30) (i), in case of solutions with shocks, one cannot take Δt to be very large.

Remark 9.5. One can use other schemes such as the Lax-Wendroff schemes. One should note that in each case, to "kill" the oscillations, a pseudo-viscous term is dependent not only on the equation but also on the scheme used. For the term to be used with the Lax-Wendroff scheme, see Richtmyer and Morton [32].

9.7 The Method of Characteristics (without shocks)

Setting $V = \frac{1}{\rho}$, the equations of hydrodynamics can be written as (Cf. See. 2.3),

$$\left.\begin{aligned} &\text{(i)} && \frac{DV}{Dt} - \frac{\partial u}{\partial m} = 0 \\ &\text{(ii)} && \frac{Du}{Dt} + \frac{\partial p}{\partial m} = 0 \\ &\text{(iii)} && \frac{D\epsilon}{Dt} + p\frac{DV}{Dt} = 0. \end{aligned}\right\} \tag{9.31}$$

We have seen in Sec. 2.3 that the slopes of the characteristics are, 0, $\pm c/V$. If $C_\circ$, C_+ and C_- are the corresponding characteristic curves through a point, one can write the differential relations along these curves (Cf. Sec. 2.2). We get

$$\left(\frac{Du}{Dt} + \epsilon\frac{c}{V}\frac{\partial u}{\partial m}\right) + \epsilon\frac{V}{c}\left(\frac{Dp}{Dt} + \epsilon\frac{c}{V}\frac{\partial p}{\partial m}\right) = 0, \tag{9.32}$$

as the differential relation along C_+ or C_- according as $\epsilon = +1$ or -1. The differential relation along $C_\circ$ turns out to be exactly (9.31) (iii).

To give a numerical scheme based on characteristics, in the case **110** where the points at which the variables are computed are not known in

advance (Cf. Sec. 8.10), we proceed as usual. We take an initial mesh along the m-axis. Through each node we draw the two characteristics C_+ and C_-. We then do the same for the point of intersection of these two characteristics and so on. Thus we get various levels of the mesh defined by the characteristics.

Our subsequent discussion is based on notation of Fig. 9.1.

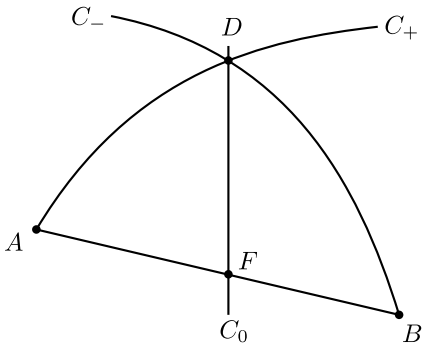


Figure 9.1:

Let A, B be two nodes at which all quantities are known. We wish to calculate the various quantities at D where the C_+ of A and C_- of B meet.

First we assume that $C = C(p, V)$ is a known function of p and V. Also observe that if (m_x, t_x) denotes the position of a point X then, $m_D = m_F$.

We now discretize the differential relations (9.32) and (9.31) (iii) and also approximate the slopes of the characteristics C_+ and C_- to get the following equations

$$\left.\begin{aligned} &\text{(i)} \quad (u_D - u_A) + \frac{1}{2}\left(\frac{V_A}{C_A} + \frac{V_D}{C_D}\right)(p_D - p_A) = 0 \\ &\text{(ii)} \quad \frac{m_D - m_A}{t_D - t_A} = \frac{1}{2}\left(\frac{C_A}{V_A} + \frac{C_D}{V_D}\right) \end{aligned}\right\} \tag{9.33}$$

for the characteristic C_+;

$$\left.\begin{aligned} &\text{(i)} \quad (u_D - u_B) + \frac{1}{2}\left(\frac{V_B}{C_B} + \frac{V_D}{C_D}\right)(p_D - p_B) = 0 \\ &\text{(ii)} \quad \frac{m_D - m_B}{t_D - t_B} = \frac{1}{2}\left(\frac{C_B}{V_B} + \frac{C_D}{V_D}\right) \end{aligned}\right\} \tag{9.34}$$

111 for the characteristic C_-;

$$\epsilon_D - \epsilon_F + \frac{1}{2}(p_D + p_F)(V_D - V_F) = 0 \tag{9.35}$$

for the characteristic $C_\circ$; and,

$$\epsilon_D = f(V_D, p_D) \tag{9.36}$$

from the equation of state.

The equations (9.33) to (9.36) give six equations to determine the six unknowns viz., $m_D, t_D, u_D, p_D, V_D, \epsilon_D$. This non-linear system can be solved iteratively. Assuming $V_D^{(0)}$, $p_D^{(0)}$ and hence knowing $C_D^{(0)}$ as well, equations (9.33) (ii) and (9.34) (ii) give $m_D^{(1)}$ and $t_D^{(1)}$. Simultaneously, (9.33) (i) and (9.34) (i) give $u_D^{(1)}$ and $p_D^{(1)}$. We now use (9.35) and (9.36) to get $\epsilon_D^{(1)}$ and $V_D^{(1)}$. Using $V_D^{(1)}$ and $p_D^{(1)}$ we proceed to the next iteration and so on.

One can also use the 'variant' form where the points at which the solution is sought are known in advance (Cf. Sec. 8.9).

9.8 The Method of Characteristics (with shocks)

Here we shall use the variant form where the points are known in advance. We assume that there is only one shock travelling with a positive velocity of propagation through the medium. Then to solve the problem one must not only compute the values of the various quantitites as the grid points but also immediately before and after the shock at each level $n\Delta t$. The shock is pictured as the curve Γ in fig. 9.2.

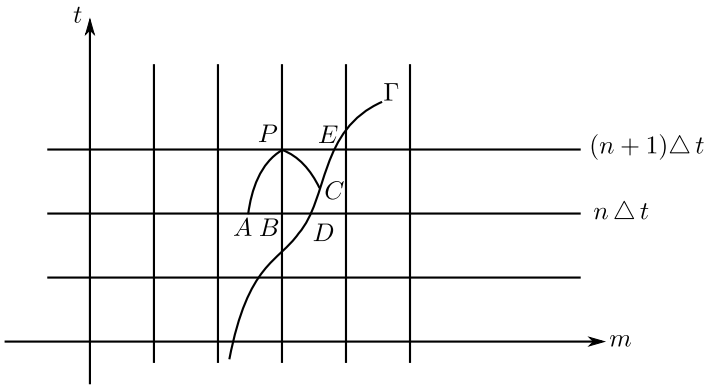


Figure 9.2:

112

To compute the various quantities at an interior grid point one draws the three characteristics through this point. If none of them meets the shock before meeting the level $n\Delta t$ then we can use our previous methods for these points. If we have a point like P, on the other hand, where the characteristic C_- meets the shock at a point C, then one has to use the points A, B and C in the methods of Section 9.7. For the values at C we can interpolate these between D and E and thus all the grid points are tackled.

To tackle the point E where the shock meets $(n+1)\Delta t$ we proceed as follows. Our notations are now based on Fig. 9.3.

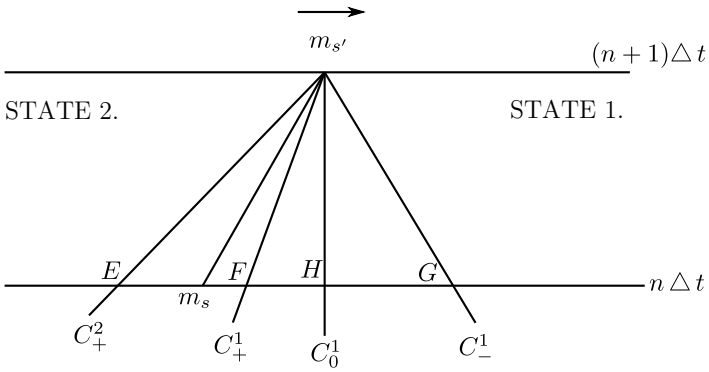


Fig. 9.3

113 The shock meets the levels $n\Delta t$ and $(n+1)\Delta t$ at m_s and $m_{s'}$ respectively. The state shead is the state 1 and the state behind is the state 2. Through $m_{s'}$ the only characteristic in state 2 is c_+^2. In state 1, on the other hand, all the three characteristics $C_+^1, C_\circ^1$ and C_-^1 exist. This is a feature of the shock.

We then have the differential relations

$$\frac{du}{ds} + \epsilon \frac{V}{C}\frac{dp}{ds} = 0 \tag{9.37}$$

along C_+ and C_- according as $\epsilon = +1$ or -1, and

$$\frac{d\epsilon}{dt} + p\frac{dV}{dt} = 0 \tag{9.38}$$

along $C_\circ$. We now discretize these equations.

We use the superscript 1 or 2 on the left to indicate the state in which we evaluate the various quantities. Along C_-^1 we have

$$\left.\begin{aligned} &\text{(i)} \quad \frac{m_G - m_{s'}}{\Delta t} = -\frac{1}{2}\left[{}^1\left(\frac{C}{V}\right)_{s'} + \left(\frac{C}{V}\right)_G\right] \\ &\text{(ii)} \quad ({}^1u_{s'} - u_G) - \left[{}^1\left(\frac{\nu}{C}\right)_{s'} + \left(\frac{V}{C}\right)_G\right]({}^1p_{s'} - p_G) = 0. \end{aligned}\right\} \tag{9.39}$$

114 Along C_+^1 we have

$$\left.\begin{aligned} &\text{(i)} \quad \frac{m_{s'} - m_F}{\Delta t} = \frac{1}{2}\left[{}^1\left(\frac{C}{V}\right)_{s'} + \left(\frac{C}{V}\right)_F\right] \\ &\text{(ii)} \quad ({}^1u_{s'} - u_F) + \frac{1}{2}\left[{}^1\left(\frac{V}{C}\right)_{s'} + \left(\frac{V}{C}\right)_F\right]({}^1p_{s'} - p_F) = 0. \end{aligned}\right\} \tag{9.40}$$

Along $C_\circ^1$ we have

$${}^1\epsilon_{s'} - \epsilon_H + \frac{1}{2}(p_H + {}^1p_{s'})({}^1V_{s'} - V_H) = 0. \tag{9.41}$$

Along C_+^2 we have

$$\left.\begin{aligned} &\text{(i)} \quad \frac{m_{s'} - m_E}{\Delta t} = \frac{1}{2}\left[{}^2\left(\frac{C}{V}\right)_{s'} + \left(\frac{C}{V}\right)_E\right] \\ &\text{(ii)} \quad {}^2u_{s'} - u_E + \frac{1}{2}\left[{}^2\left(\frac{V}{C}\right)_{s'} + \left(\frac{V}{C}\right)_E\right]({}^2p_{s'} - p_E) = 0. \end{aligned}\right\} \tag{9.42}$$

If $M = \dfrac{dm}{dt}$ is the speed of the shock, we have the Rankine-Hugoniot relations (Cf. Sec. 3.4)

$$\left.\begin{array}{ll} \text{(i)} & M[V] + [u] = 0 \\ \text{(ii)} & M[u] - [p] = 0 \\ \text{(iii)} & M[\epsilon + \dfrac{1}{2}u^2] - [pu] = 0 \end{array}\right\} \tag{9.43}$$

where $[\varphi] = {}^2\varphi_{s'} - {}^1\varphi_{s'}$ for any function φ. Also the slope of the shock is approximated by

$$\frac{m_{s'} - m_s}{\Delta t} = \frac{1}{2}[M_s + M_{s'}] \tag{9.44}$$

We also have the state equations

$$\left.\begin{array}{ll} \text{(i)} & {}^1\epsilon_{s'} = f({}^1p_{s'}, {}^1V_{s'}) \\ \text{(ii)} & {}^2\epsilon_{s'} = f({}^2p_{s'}, {}^2V_{s'}). \end{array}\right\} \tag{9.45}$$

The relations (9.39) to (9.45) give 13 non-linear equations. The number of unknowns is also 13, viz., the positions $m_E, m_F, m_G, m_{s'}$, the velocity of the shock $M_{s'}$, and the variables u, p, V, ϵ at s' in states 1 and 2. This system can be solved iteratively knowing all qualities at time $n\Delta t$. We assume $M_{s'}^{(0)}$ and all the quantities $(\frac{C}{V})^{(0)}$. Then (9.44) gives $m_{s'}^{(1)}$. Then (9.39) (i) and (9.40) (i) give $m_G^{(1)}$ and $m_F^{(1)}$. We can interpolate between the grid points at time $n\Delta t$ for the values $u_G^{(1)}$ and $p_G^{(1)}$. Using these in (9.39) (ii) and (9.40) (ii) we get ${}^1u_{s'}^{(1)}$ and ${}^1p_{s'}^{(1)}$. Then (9.41) and (9.45) (i) give ${}^1\epsilon_{s'}^{(1)}$ and ${}^1V_{s'}^{(1)}$. The relation (9.42) (i) gives $m_E^{(1)}$ and then by interpolation we get $p_E^{(1)}$, $u_E^{(1)}$ which we use in (9.42) (ii) and (9.43) (ii) to get ${}^2p_{s'}^{(1)}$ and ${}^2u_{s'}^{(1)}$. Then (9.43) (iii) gives ${}^2\epsilon_{s'}^{(1)}$ and finally (9.43) (i) gives $M_{s'}^{(1)}$. Now that we have $M_{s'}^{(1)}$ and both states of $(\frac{C}{V})^{(1)}$, we can use these in the next iteration and so on.

Remark 9.6. We have dealt with the case of only one shock travelling through the medium with a positive velocity. One can perform similar analysis on other types of shocks but these become very complicated. If

there is an interface of two media, then a shock is transmitted into the second medium from the first and, depending on the ratio of the densities of the media, a shock may be reflected back into the first medium. In such a situation to keep track of all the shocks becomes very complicated from the logical point of view of a computer programme.

REFERENCE. The reader is referred to the papers of Hoskin in the Proceedings of the Conferences on Numerical Methods in Fluid Dynamics (1969, 1971, 1973 and 1975). See Roache [33] for a detailed bibliography.

Chapter 10

Numerical Methods for the System of Equations of Hydrodynamics-Eulerian Coordinates

10.1 Introduction

In the slab symmetric, one dimensional case, the Eulerian form the equa- **116**
tions of hydrodynamics is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \tag{10.1}$$

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial}{\partial x}(\rho u^2 + p + q) = 0 \tag{10.2}$$

$$\frac{\partial}{\partial t}(\rho \epsilon) + \frac{\partial}{\partial x}(\rho u \epsilon) + (p + q)\frac{\partial u}{\partial x} = 0 \tag{10.3}$$

assuming μ, k and g to be zero. Here q is the pseudo-viscous term given by

$$q = 2\rho \Delta x^2 \max\left(0, -\frac{\partial u}{\partial x}\right)\frac{\partial u}{\partial x}, \tag{10.4}$$

for a mesh of step Δx. Note that the equation (10.3) can be written in terms of the total energy E as

$$\frac{\partial}{\partial t}(\rho E) + \frac{\partial}{\partial x}(\rho u E) + \frac{\partial}{\partial x}((p+q)u) = 0 \qquad (10.3')$$

The only interesting problem in the 1-dimensional Eulerian case is that of a moving boundary (i.e. due to a piston or free surface etc.). Unlike the Lagrangian system where, due to the same particles lying on the free surface, the boundary is fixed, in the Eulerian case the boundary moves with time. We will discuss this presently. To start with we see how we can discretize the equation (10.1) to (10.3) at an interior node of the mesh. One could use characteristic, methods or the 2-step Lax-Wendroff scheme. We study here the problems connected with the use of the leap-frog scheme.

10.2 Discretization at interior nodes

Assuming a uniform mesh with steps Δx and Δt, we compute as usual
117 the quantities p, ρ and ϵ at the points $(i+\frac{1}{2}, n)$. We try to compute the quantity ρu at $(i, n+\frac{1}{2})$ so that we can discretize the equation (10.1) by

$$\frac{\rho_{i+\frac{1}{2}}^{n+1} - \rho_{i+\frac{1}{2}}^{n}}{\Delta t} + \frac{1}{\Delta x}\left[(\rho u)_{i+1}^{n+\frac{1}{2}} - (\rho u)_{i}^{n+\frac{1}{2}}\right] = 0 \qquad (10.5)$$

Now we set $\bar{p} = p + q$. Then the equation (10.3) can be discretized by

$$\frac{(\rho\epsilon)_{i+\frac{1}{2}}^{n+1} - (\rho\epsilon)_{i+\frac{1}{2}}^{n}}{\Delta t} + \frac{1}{2}\left(\bar{p}_{i+\frac{1}{2}}^{n+1} + \bar{p}_{i+\frac{1}{2}}^{n}\right)\left(\frac{u_{i+1}^{n+\frac{1}{2}} - u_{i}^{n+\frac{1}{2}}}{\Delta x}\right) + \\ + \frac{1}{\Delta x}\left[u_{i+1}^{n+\frac{1}{2}}(\rho\epsilon)_{i+1}^{*} - u_{i}^{n+\frac{1}{2}}(\rho\epsilon)_{i}^{*}\right] = 0, \qquad (10.6)$$

where we set

$$(\rho\epsilon)_{i}^{*} = \begin{cases} (\rho\epsilon)_{i+\frac{1}{2}}^{n}, & \text{if } u < 0 \\ (\rho\epsilon)_{i-\frac{1}{2}}^{n}, & \text{if } u > 0. \end{cases} \qquad (10.7)$$

This is a 1-sided scheme and hence only of first order. However it is not as dissipative as Lax's scheme.

Notice however that due to the second term in (10.6) one needs $u_i^{n+\frac{1}{2}}$ as well. This means that we need $\rho_i^{n+\frac{1}{2}}$ also (since we know $(\rho u)_i^{n+\frac{1}{2}}$). Thus effectively this method demands computation of ρ not only at the points $(i+\frac{1}{2}, n)$ but also at the points $(i, n+\frac{1}{2})$.

We now discretise equation (10.2). We write

$$\frac{(\rho u)_i^{n+\frac{1}{2}} - (\rho u)_i^{n-\frac{1}{2}}}{\Delta t} + \frac{1}{\Delta x}(\rho^n_{i+\frac{1}{2}}(u^*_{i+\frac{1}{2}})^2 - \rho^n_{i-\frac{1}{2}}(u^*_{i-\frac{1}{2}})^2) + \frac{\bar{p}^n_{i+\frac{1}{2}} - \bar{p}^n_{i-\frac{1}{2}}}{\Delta x} = 0, \tag{10.8}$$

where

$$u^*_{i+\frac{1}{2}} = \begin{cases} u_{i+1}^{n-\frac{1}{2}}, \text{ if } u < 0 \\ u_i^{n-\frac{1}{2}}, \text{ if } u > 0, \end{cases} \tag{10.9}$$

is again a one-sided approximation.

We also have **118**

$$q^n_{i+\frac{1}{2}} = f(\rho^n_{i+\frac{1}{2}}(u_{i+1}^{n-\frac{1}{2}} - u_i^{n-\frac{1}{2}})). \tag{10.10}$$

As already remarked, the major problem is the additional computation of the values $\rho_i^{n+\frac{1}{2}}$. Of course, one has the immediate (but rather crude) approximation

$$\rho_i^{n+\frac{1}{2}} = \frac{1}{2}(\rho^n_{i-\frac{1}{2}} + \rho^n_{i+\frac{1}{2}}). \tag{10.11}$$

One could use a more sophisticated formula by applying Lax' scheme to the equation of conservation of mass (i.e. (10.1)) to define $\rho_i^{n+\frac{1}{2}}$. Thus we

$$\rho_i^{n+\frac{1}{2}} = \frac{1}{2}(\rho^n_{i+\frac{1}{2}} + \rho^n_{i-\frac{1}{2}}) - \frac{\Delta t}{2\Delta x}(\rho^n_{i+\frac{1}{2}} u^n_{i+\frac{1}{2}} - \rho^n_{i-\frac{1}{2}} u^n_{i-\frac{1}{2}}). \tag{10.12}$$

But then, we are faced with the need of computing $u^n_{i+\frac{1}{2}}$ as well. Thus we end up by computing both u and ρ at all nodes!

Noh [29] computed $\rho_i^{n+\frac{1}{2}}$ from $\rho_i^{n-\frac{1}{2}}$ and $\rho_{i+1}^{n-\frac{1}{2}}$ or $\rho_i^{n-\frac{1}{2}}$ and $\rho_{i-1}^{n-\frac{1}{2}}$ (according as $u < 0$ or > 0) using the one-sided scheme. The only draw back here is that $\rho_{i+\frac{1}{2}}^n$ is computed on the nodes $(i+\frac{1}{2}, n)$ and $\rho_i^{n+\frac{1}{2}}$ on the nodes $(i, n+\frac{1}{2})$ and these two grids are unconnected. On each grid one could get a good solution. But unless the mesh is very fine, when we put these values together the resulting function is not a good approximation of ρ.

However, all these methods give fairly satisfactory results.

10.3 Treatment of boundary nodes

We will illustrate the discretization of the equation of conservation of mass ((10.1)) in case of a moving boundary. One can do the same for the other equations as well. It can be seen that the schemes in this case are fairly complicated.

119 To start with, let us assume that the moving node at $(n+1)\Delta t$ falls within the same grid-intervals as that at $n\Delta t$ (See Fig. 10.1).

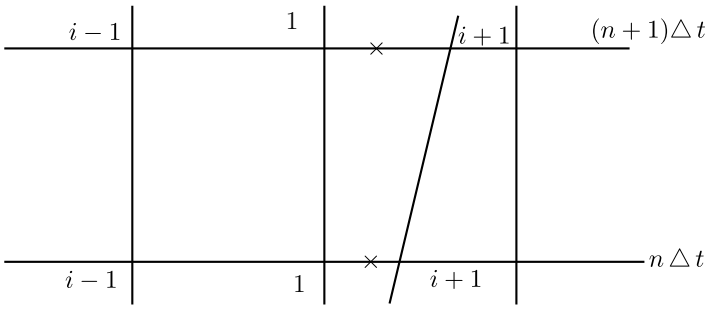


Figure 10.1:

One then has to compute ρ at the mid-point of $(i, i+1)$ in the last interval. To discretise (10.1) we imitate the following procedure for the continuous case:

Integrate ρ between x_i and $x_{i+1}(t)$ and differentiate w.r.t. t. Thus

$$\frac{d}{dt}\int_{x_i}^{x_{i+1}(t)} \rho dx = \int_{x_i}^{x_{i+1}(t)} \frac{\partial \rho}{\partial t} dx + \rho_{i+1}\frac{dx_{i+1}(t)}{dt} - \rho_i \frac{dx_i}{dt}.$$

Since x_i is fixed and S_{i+1} moves with velocity u_{i+1}, we get, on using equation (10.1),

$$\begin{aligned}\frac{d}{dt}\int_{x_i}^{x_{i+1}} \rho dx &= \rho_{i+1}u_{i+1} - \int_{x_i}^{x_{i+1}} \frac{\partial}{\partial x}(\rho u)dx \\ &= \rho_{i+1}u_{i+1} - \rho_{i+1}u_{i+1} + \rho_i u_i = \rho_i u_i.\end{aligned}$$

We use this to discretise (10.1). We write (at the boundary)

$$\frac{1}{\Delta t}((x_{i+1}^{n+1} - x_i)\rho_{i-\frac{1}{2}}^{n+1} - (x_{i+1}^n - x_i)\rho_{i+\frac{1}{2}}^n) = u_i \rho_i^* \tag{10.13}$$

where **120**

$$\rho_i^* = \begin{cases}\rho_{i+\frac{1}{2}}^n, & \text{if } u < 0 \\ \rho_{i-\frac{1}{2}}^n, & \text{if } u > 0.\end{cases} \tag{10.14}$$

Let us now consider the case where the moving node is not in the same grid interval at times $n\Delta t$ and $(n+1)\Delta t$. Again, this splits into two cases one where the boundary has a forward slope and thus there are more nodes at $(n+1)\Delta t$ than at $n\Delta t$ (Cf. Fig. 10.2 (a)) and the other where the boundary has a backward slope and there are less points at $(n+1)\Delta t$ than at $n\Delta t$ (Cf. Fig. 10.2 (b))

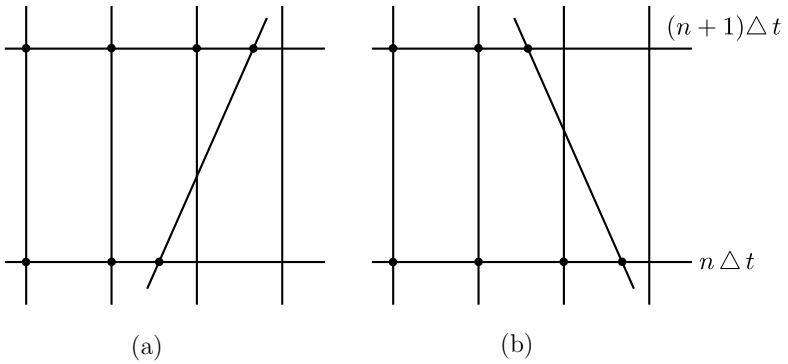


Figure 10.2:

We observe that as a result of these, the final mesh length may be too small and for stability reasons, this is not a satisfactory state of affairs. Hence we choose our nodes such that the final mesh length satisfies the condition

$$\frac{1}{2}\Delta x \leq (x_{I+1} - x_I) \leq \frac{3}{2}\Delta x. \tag{10.15}$$

Let us consider the case of Fig. 10.2 (a) where we have more nodes at $(n+1)\Delta t$. We base our discussion on Fig. 10.3.

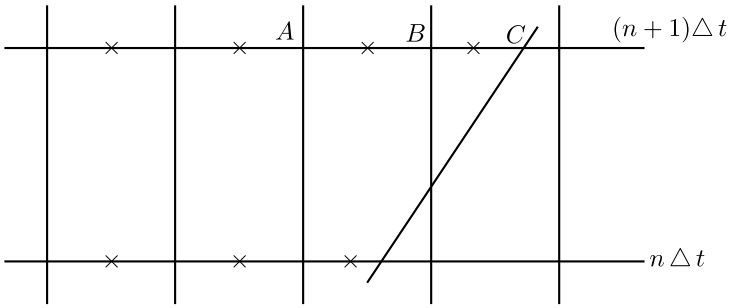


Figure 10.3:

121

In order to discretise equation (10.1) one must compute ρ at the mid-points of the mesh lengths. One can do this for the interior points as before. Now one can ignore the node B and treat AC as the last interval

and using previous methods one can compute ρ at the mid-point of AC. Let us call this $\rho^{\text{old}}_{i+\frac{1}{2}}$. However, the length $x_C - x_A > \frac{3}{2}\Delta x$ and so one must split this as AB and BC and compute ρ at the mid-points of these two intervals. If we call these values $\rho^{\text{new}}_{i+\frac{1}{2}}$ and $\rho^{\text{new}}_{i+3/2}$ respectively, we write

$$\rho^{\text{old}}_{i+\frac{1}{2}}(x_C - x_A) = \rho^{\text{new}}_{i+\frac{1}{2}}(x_B - x_A) + \rho^{\text{new}}_{i+3/2}(x_C - x_B). \tag{10.16}$$

This together with another equation will help us to compute $\rho^{\text{new}}_{i+\frac{1}{2}}$ and $\rho^{\text{new}}_{i+3/2}$. For instance, one can take $\rho^{\text{new}}_{i+\frac{1}{2}}$ to be linearly interpolated between $\rho^{\text{old}}_{i-\frac{1}{2}} (= \rho^{\text{new}}_{i-\frac{1}{2}})$ and $\rho^{\text{old}}_{i+\frac{1}{2}}$.

In the other case (Cf. Fig. 10.4) we have $x_C - x_B < \frac{1}{2}\Delta x$.

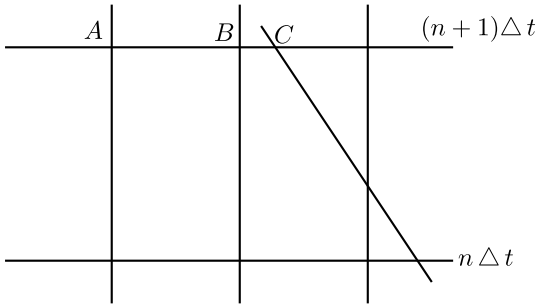


Figure 10.4:

122

Here one can treat BC as a separate mesh length and thus we have equal number of nodes at n and at $(n + 1)$. Hence we can compute ρ at the midpoints of AB and BC. However as $x_C - x_B < \frac{1}{2}\Delta x$, one must take AC as a whole interval. It $\rho^{\text{new}}_{i+\frac{1}{2}}$ is the value of ρ at the mid-point of AC, and $\rho^{\text{old}}_{i+\frac{1}{2}}$ and $\rho^{\text{old}}_{i+3/2}$ those at the mid-points of AB and BC, we write

$$\rho^{\text{new}}_{i+\frac{1}{2}}(x_C - x_A) = \rho^{\text{old}}_{i+\frac{1}{2}}(x_B - x_A) + \rho^{\text{old}}_{i+3/2}(x_C - x_B) \tag{10.17}$$

and we immediately get $\rho^{\text{new}}_{i+\frac{1}{2}}$.

Note that (10.16) and (10.17) merely approximate the fact that the mass contained in AC is the same as the total mass contained in AB and BC.

The problem of a moving boundary is complicated essentially because (i) the mesh is not longer uniform and (iii) the number of nodes involved in the computation varies as time proceeds. Hence if we are able to choose a coordinate system, i.e. rewrite the equation in terms of new variables, so that one can choose the nodes to move with the boundary, the situation will improve. We discuss this possibility in the next subsection.

10.4 The ale-method[1]

123 In this method, we rewrite the equation in terms of a new type of coordinates which is neither Lagrangian nor Eulerian. The technique is comparatively new and has not yet been widely used.

Let us assume that for each t we have a homeomorphism $\varphi_t : \mathbb{R} \to \mathbb{R}$ with the following properties: the mapping $a \mapsto x(a,t)(= \varphi_t(a))$ (where $x = \varphi_\circ(a)$ at time 0) is such that the Jacobian

$$J = \frac{\partial x}{\partial a} \tag{10.19}$$

is defined and is non-zero. We denote the derivative of x w.r.t. t, when a is fixed, by

$$v = \frac{Dx}{Dt}. \tag{10.20}$$

That is to say, a point initially at position a moves with velocity v and its position at time t is given by $x = x(a,t)$. The various 'trajectories' do not cross one another and this is meaning of the condition that $J \neq 0$. Note that for the Lagrangian coordinate system we have $v = u$, the velocity of the fluid. In case of the Eulerian system x never changes and hence $J = 1$ and $v = 0$. Thus this type of coordinate system contains both the Lagrangian and Eulerian systems as particular cases.

[1] Arbitrarily Lagrangian-Eulerian

If one knows a function $(x,t) \mapsto \varphi(x,t)$ we define $\bar{\varphi}(a,t)$ by

$$\bar{\varphi}(a,t) = \varphi(x(a,t).t) \tag{10.21}$$

and one has

$$\frac{D\bar{\varphi}}{Dt} = \frac{\partial \varphi}{Dt} + v\frac{\partial \varphi}{\partial x}, \tag{10.22}$$

Here $\frac{D\bar{\varphi}}{Dt}$ means the derivative of $\bar{\varphi}$ with respect to t when a is kept fixed. **124**
We now rewrite the equation (10.1) in terms of this systems: c

$$\begin{aligned}
\frac{D}{Dt}(\rho J) &= \rho\frac{DJ}{Dt} + J\frac{D\rho}{Dt} \\
&= \rho\frac{\partial v}{\partial a} + J(\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial x}) \\
&= \rho\frac{\partial v}{\partial a} + J\left(v\frac{\partial \rho}{\partial x} - \frac{\partial}{\partial x}(\rho u)\right) \text{ (using (10.1))} \\
&= \rho\frac{\partial v}{\partial a} + v\frac{\partial \rho}{\partial a} - \frac{\partial}{\partial a}(\rho u) = \frac{\partial}{\partial a}(\rho(v-u)).
\end{aligned}$$

Thus (10.1) now takes the form

$$\frac{D}{Dt}(\rho J) + \frac{\partial}{\partial a}(\rho(u-v)) = 0. \tag{10.23}$$

By doing the same thing for the other two equations one has

$$\frac{D}{Dt}(\rho u J) + \frac{\partial}{\partial a}(\rho u(u-v) + p) = 0, \tag{10.24}$$

and

$$\frac{D}{Dt}(\rho \epsilon J) + \frac{\partial}{\partial a}(\rho\xi(u-v)) + p\frac{\partial u}{\partial a} = 0. \tag{10.25}$$

Remark 10.1. In view of our comments made previously, the above equations contain the Lagrangian and Eulerian equations as particular cases.

Let us now take a fixed $(a-t)$ grid. We define x at the points (i,n), u at the points $(i, n+\frac{1}{2})$, and J, ρ at the points $(i+\frac{1}{2}, n)$, thus we write the discrete equations

$$x_i^{n+1} - x_i^n = v_i^{n+\frac{1}{2}}\Delta t. \tag{10.26}$$

$$J^{n+1}_{i+\frac{1}{2}} = \frac{x^{n+1}_{i+1} - x^{n+1}_{i}}{a_{i+1} - a_i}, \tag{10.27}$$

and

$$\frac{1}{\Delta t}\left[\rho^{n+1}_{i+\frac{1}{2}} J^{n+1}_{i+\frac{1}{2}} - \rho^{n}_{i+\frac{1}{2}} J^{n}_{i+\frac{1}{2}}\right] + \\ + \frac{1}{(a_{i+1} - a_i)}\left[(u-v)^{n+\frac{1}{2}}_{i+1}\rho^{*}_{i+1} - (u-v)^{n+\frac{1}{2}}_{i}\rho^{*}_{i}\right] = 0 \tag{10.28}$$

125 with, for example,

$$\rho^{*}_{i+1} = \begin{cases} \rho^{n}_{i+3/2}, \text{ if } u - v < 0 \\ \rho^{n}_{i+\frac{1}{2}}, \text{ if } u - v > 0 \end{cases} \tag{10.29}$$

if one uses a one-sided scheme. This discretizes the equation (10.23). One can similarly discretise the other equations of hydrodynamics as well.

Remark 10.2. The equations (10.27) and (10.28) give

$$\left[\rho^{n+1}_{i+\frac{1}{2}}(x^{n+1}_{i+1} - x^{n+1}_{i}) - \rho^{n}_{i+\frac{1}{2}}(x^{n}_{i+1} - x^{n}_{i})\right] \\ = \Delta t\left[-(u-v)^{n+\frac{1}{2}}_{i+1}\rho^{*}_{i+1} + (u-v)^{n+\frac{1}{2}}_{i}\rho^{*}_{i}\right].$$

This merely states that if $M^{n+1}_{i+\frac{1}{2}}$ is the mass at time $n+1$ between the nodes i and $i+1$, then

$$M^{n+1}_{i+\frac{1}{2}} - M^{n}_{i+\frac{1}{2}} = \\ = [\text{Inflow into the cell } (i, i+1)] - [\text{Outflow from the cell } (i, i+1)].$$

The use of the one-sided scheme to define ρ^{*}_{i} then says, that one ought to define the density at the boundary of the cell to be the one inside the cell from which the fluid is flowing through that boundary. This is why the above equation is some times called the *Donor-cell equation*.

At each time $n + 1$, we have to define $v_i^{n+\frac{1}{2}}$ or, equivalently x_i^{n+1}. One can, for example, define x such that the mesh is always uniform. To avoid the free surface crossing the grid one can take v on the surface to be the velocity of the surface itself. This is the merit of this method.

For most of the 1-dimensional problems, the Lagrangian coordinate system is good enough (i.e. the preceding method with $v = u$), be- **126**
cause there is no problem of distortion of the mesh. This is no longer true in the 2-dimensional problems where the generalization of the ALE method may come in very useful when dealing with boundaries moving w.r.t. a frame fixed in the laboratory.

As for comparing the Lagrangian and Eulerian systems, it is clear that the Lagrangian method is preferable. In pseduo-viscosity methods Eulerian methods involve fine meshes about a shock. In case of Lagrangian systems the mesh will be automatically fine near a compression or shock without increasing the number of nodes. Thus it is more feasible from the point of view of computers.

This brings us to a close of the discussion of one dimensional time dependent equations. In the next section we will take up two - dimensional problems.

Chapter 11

The 2-Dimensional Problem

11.1 Introduction

We now consider the approximation of the equations of fluid dynamics in 2 dimensions, in the plane symmetric case. Once again we have two coordinate systems-the Lagrangian coordinates (a, b) and the Eulerian coordinates (x, y). These are connected by the transformation

$$x = x(a, b, t),\ y = y(a, b, t) \tag{11.1}$$

where $x(a, b, t)$ and $y(a, b, t)$ are the coordinates at time t of the fluid particle located at (a, b) at time $t = 0$. We denote by J the Jacobian of the transformation (11.1). One defines the time derivative $\frac{Df}{Dt}$ (so called "particular derivative") by

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + u\frac{\partial f}{\partial x} + v\frac{\partial t}{\partial y} \tag{11.2}$$

where u and v are the velocity components. The Lagrangian form of the equation of motion is as follows:

Conservation of mass:

$$\frac{D\rho}{Dt} + \rho div.\ \vec{u} = 0,\ \vec{u} = (u, v) \tag{11.3}$$

or, equivalently,

$$\frac{D}{Dt}(\rho J) = 0. \tag{11.3'}$$

Conservation of momentum.

$$\left.\begin{array}{ll} \text{(i)} & \rho\dfrac{Du}{Dt} + \dfrac{\partial}{\partial x}(p+q) = 0 \\ \text{(ii)} & \rho\dfrac{Du}{Dt} + \dfrac{\partial}{\partial y}(p+q) = 0. \end{array}\right\} \tag{11.4}$$

where q is the pseudo-viscous term given by

$$q = \sigma(-div.\vec{u}) \tag{11.5}$$

128 with

$$\sigma = \begin{cases} 0, \text{ if } div\vec{u} > 0 \\ \rho\ell^2|div\vec{u}|, \text{ if } div\vec{u} < 0. \end{cases} \tag{11.6}$$

Conservation of energy.

$$\frac{D\epsilon}{Dt} + (p+q)\frac{D}{Dt}(\frac{1}{\rho}) = 0. \tag{11.7}$$

One also has the equation of state

$$\epsilon = f(p,\rho). \tag{11.8}$$

Remark 11.1. One reiterates our comments of Sec. 1.5, regarding the Lagrangian form of the equations. In the equations above $\dfrac{\partial}{\partial x}$ and $\dfrac{\partial}{\partial y}$ must be expressed in terms of derivatives w.r.t. a and b via the Jacobian J.

11.2 The weak form

We use the finite element method to discretise the equations of hydrodynamics w.r.t. the space variables. For this we need to write these equations in the weak form.

Let Ω be the domain of consideration and $\partial\Omega$ its boundary. Let $\vec{\nu} = (\nu_x, \nu_y)$ be the unit outer normal along $\partial\Omega$.

To write the equations in the weak form, we multiply our equations by suitable test-functions and integrate over Ω. We write equations (11.3) in the form

$$\int_\Omega \left(\frac{D\rho}{Dt} + \rho div.\vec{u}\right)\varphi dx\, dy = 0. \tag{11.9}$$

for all suitable φ.

If (φ, ψ) is a test-vector, then an multiplying (11.4) (i) by φ and (11.4) (ii) by ψ and integrating by parts, we get the weak form:

$$\left.\begin{aligned} &\text{(i)} \quad \int_\Omega \rho\frac{Du}{Dt}\varphi dx\, dy + \int_{\partial\Omega} p\varphi\nu_x ds - \int_\Omega p\frac{\partial\varphi}{\partial x}dx\, dy = 0 \\ &\text{(ii)} \quad \int_\Omega \rho\frac{Dv}{Dt}\psi dx\, dy + \int_{\partial\Omega} p\psi\nu_y\, ds - \int_\Omega p\frac{\partial\psi}{\partial y}dx\, dy = 0. \end{aligned}\right\} \tag{11.10}$$

for all suitable φ and ψ. Thus we have got rid of all the derivatives of p in (11.4). Since the third equation viz., (11.7) does not involve derivatives w.r.t. space variables we keep it as it is. **129**

From these equations it is clear that it suffices to take $\rho,\ p,\ \epsilon \in L^2(\Omega)$ while we need $x, y, u, v, \varphi, \psi \in H^1(\Omega)$. We shall use the finite element method with quadrilateral elements and trial functions which are piecewise polynomials. Thus we need to have these trial functions continuous across the inter element boundaries (Cf. Ciarlet [6]) in order that these functions may be in $H^1(\Omega)$.

We now proceed to describe the simplest element known.

11.3 An isoparametric quadrilateral element

Let us assume that the domain Ω is such that it can be subdivided into quadrilaterals. Since we only need $\rho,\ p,\ \epsilon \in L^2(\Omega)$, one can take as trial functions the space $V_\circ$ of piecewise constant (which are, in particular,

discontinuous) functions. Thus if χ_Q is the characteristic function of the quadrilateral Q, then one can write

$$V_\circ = \left\{ \sum_Q \alpha_Q \chi_Q \mid \alpha_Q = \alpha_Q(t), \text{ a constant w.r.t. } a \text{ and } b \right\} \subset L^2(\Omega).$$

Now to achieve a space of approximants contained in $H^1(\Omega)$, as already mentioned, we use continuous piecewise polynomials.

This is most easily achieved if we can define these functions in each quadrilateral Q in such a way so that they depend only on their values at the four nodes of Q, and along each boundary their restriction is a linear interpolation of the values at the end vertices. Then, obviously, given
130 two adjacent elements Q and Q^1, if the values of a piecewise polynomial, f, are prescribed at the nodes then the restriction of f to that common edge from both Q and Q^1 will be the same and hence continuity is established.

If Q is a rectangle whose sides are parallel to the axes (say. the unit square) then the definition of such polynomials is easy (Fig. 11.1).

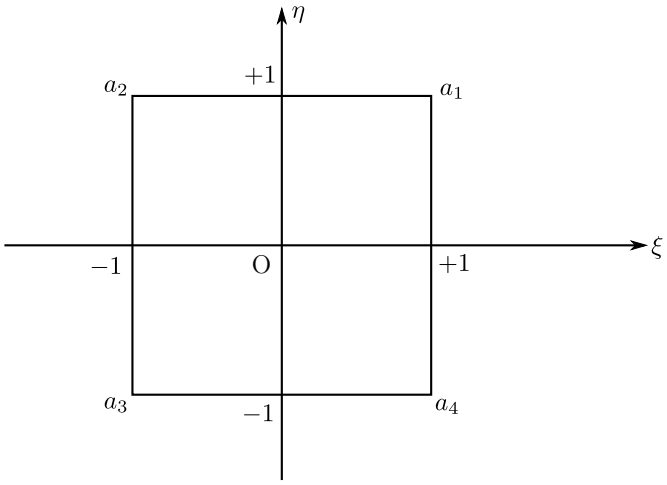


Figure 11.1:

We define the space of polynomials to be

$$Q_1 = \{p(\xi, \eta) \mid p(\xi, \eta) = a + b\xi + c\eta + d\xi\eta\}\,.$$

Then restricted to each side, $p(\xi, \eta)$ is a polynomial of degree ≤ 1 either in ξ or in η alone. Further it is linearly interpolated on each edge between the values at the end-points. In fact one can give a basis for Q_1 by four polynomials $\ell_i (i = 1, 2, 3, 4)$ such that $\ell_i \in Q_1$ and ℓ_i takes the value 1 at the vertex a_i and zero at the other vertices. Thus

$$\left.\begin{aligned} \ell_1(\xi, \eta) &= \frac{1}{4}(1 + \xi)(1 + \eta) \\ \ell_2(\xi, \eta) &= \frac{1}{4}(1 - \xi)(1 + \eta) \\ \ell_3(\xi, \eta) &= \frac{1}{4}(1 - \xi)(1 - \eta) \\ \ell_4(\xi, \eta) &= \frac{1}{4}(1 + \xi)(1 - \eta). \end{aligned}\right\} \tag{11.11}$$

If $p \in Q_1$, one has **131**

$$p(\xi, \eta) = \sum_{i+1}^{4} p(a_i)\ell_i(\xi, \eta).$$

Thus Q_1 has all the properties we need.

Let us go to a general quadrilateral with vertices A_1, A_2, A_3 and A_4 in the (a, b)-plane.

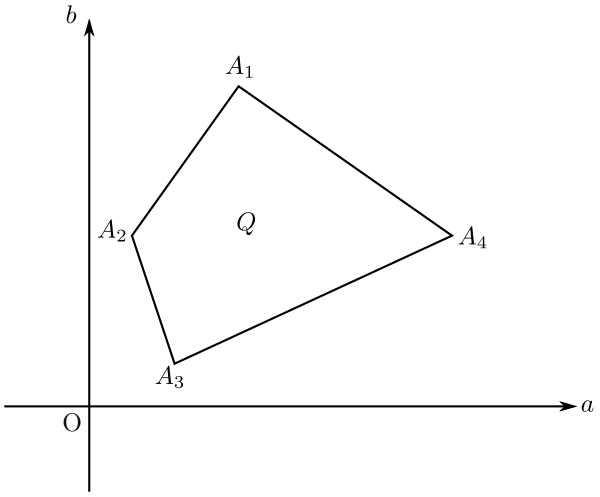


Figure 11.2:

We now define a transformation

$$\left.\begin{aligned} a(\xi,\eta) &= \sum_{i=1}^{4} a_i \ell_i(\xi,\eta) \\ b(\xi,\eta) &= \sum_{i=1}^{4} b_i \ell_i(\xi,\eta) \end{aligned}\right\} \tag{11.12}$$

so that $a_1 \mapsto A_1$, $a_2 \mapsto A_2$, $a_3 \mapsto A_3$ and $a_4 \mapsto A_4$. It is routine checking to see that the four edges of the square map linearly into the corresponding edges of the quadrilateral, i.e. if $0 \leq \lambda \leq 1$, then the point $\lambda a_i + (1-\lambda)a_{i+1}$ maps to the point $\lambda A_i + (1-\lambda)A_{i+1}$. Further, the transformation can be inverted. i.e. for every point $(a,b) \in Q$, there exists a unique point (ξ,η) in the unit square which maps into (a,b) under the transformation (11.12). We use this correspondence to define a space V_Q of polynomials over Q:

$$V_Q = \{u(a,b) \mid u(a,b) = \sum_{i=1}^{4} u(A_i)\ell_i(\xi,\eta)\}. \tag{11.13}$$

Then it is easy to see that u is completely determined by its values on the nodes and that on each boundary it is a linear interpolation of the values at the end-points.

Now given a subdivision of Ω into quadrilaterals, we define V_1 by

$$V_1 = \{u \mid u \mid Q \in V_Q\}. \tag{11.14}$$

Then V_1 satisfies the continuity condition and hence is a subspace of $H^1(\Omega)$. Also every function in V_1 is completely described by its values at all the nodes. Indeed if we number all the nodes of the subdivision suitably and if φ_i is that function of V_1 whose value at the i^{th} node is 1 and it takes the value 0 at other nodes, for $1 \leq i \leq I$, then such functions form a basis for V_1. Every function $u \in V_1$ may be written as

$$u = \sum_{i=1}^{I} u_i \varphi_i \tag{11.15}$$

where $u_i = u_i(t)$ is the value of u at the i-th node.

Another important property of the functions φ_i is, that φ_i is non-zero only on atmost four quadrilaterals of which the i^{th} node is a vertex. (We will come to the question of boundary nodes later).

Such finite elements as described in this section are called isoparametric because we use the *same* types of functions both for the space of approximants as well as for the transformations from the unit square. (See Ciarlet and Raviart [7] for a complete discussion on isoparametric finite elements).

11.4 Discretization of the equations

We now discretise the weak forms of the equations. Thus we substitute **133**
in the equations (11.9), (11.10) and (11.7) the trial functions and look for solutions in spaces of these trial functions. Thus we take $p, \rho, \epsilon \in V_\circ$ and $u, v, x, y, \varphi, \psi \in V_1$ and demand that the equations (11.10) are satisfied for all $\varphi, \psi \in V_1$ and (11.9) for all χ_Q. Of course, it is enough to satisfy (11.10) for the basis functions φ_i. In order to make these statements

precise and to obtain the discrete equations, we first write

$$\left.\begin{aligned} \rho &= \sum_Q \rho_Q(t)\chi_Q,\ \epsilon = \sum_Q \epsilon_Q(t)\chi_Q \\ p &= \sum_Q p_Q(t)\chi_Q,\ u = \sum_{i=1}^{I} u_i(t)\varphi_i,\ v = \sum_{i=1}^{I} v_i(t)\varphi_i \end{aligned}\right\} \tag{11.16}$$

where the nodes are numbered from 1 to I. Now equation (11.9) becomes

$$\int_\Omega \left(\frac{D\rho}{Dt} + \rho div\vec{u}\right)\chi_Q dx\ dy = 0 \text{ for all } Q.$$

or

$$\int_Q \left(\frac{D\rho}{Dt} + \rho div\vec{u}\right) dx\ dy = 0 \text{ for all } Q.$$

Using (11.16) we get (by setting S_Q to be the area of the quadrilateral Q), the discretization of (11.9) w.r.t. the space variables as

$$S_Q \frac{D\rho_Q}{Dt} + \rho_Q \sum_i u_i \int_Q \frac{\partial\varphi_i}{\partial x} dx\ dy + \rho_Q \sum_i v_i \int_Q \frac{\partial\varphi_i}{\partial y} dx\ dy = 0 \tag{11.17}$$

for all Q.

Since in the equation (11.7) everything is constant w.r.t. the space variables. we get

$$\frac{D\epsilon_Q}{Dt} + \left(p_Q \frac{D}{Dt}\right)\left(\frac{1}{\rho_Q}\right) = 0 \text{ for all } Q. \tag{11.18}$$

We now show how to discretise (11.10) (i). (The method for (11.10) (ii) is identical).

134 Substituting in (11.10) (i) from (11.16), we get

$$\sum_j \frac{Du_j}{Dt} \int_\Omega \rho\varphi_j\varphi_i dx\ dy + \int_{\partial\Omega} p\varphi_i \nu_x ds - \sum_Q p_Q \int_\Omega \chi_Q \frac{\partial\varphi_i}{\partial x} dx\ dy = 0 \tag{11.19}$$

for all $1 \le i \le I$.

Using the fact that $\rho J = \rho_\circ$, we get

$$\sum_j \frac{Du_j}{Dt} \int_\Omega \rho_\circ \varphi_i \varphi_j da\, db + \int_{\partial\Omega} p\varphi_i \nu_x ds - \sum_Q p_Q \int_Q \frac{\partial \varphi_i}{\partial x} dx\, dy \quad (11.20)$$
$$= 0,\ 1 \le i \le I.$$

We see that the last term in (11.20), which is similar to the last terms of (11.17), will be non-zero only if the i^{th} node is a vertex of Q. Thus for every i, the last term expounds into at most four non-zero terms.

The middle term of (11.20) survives only if φ_i corresponds to a boundary node. We will turn to the question of boundary nodes in Sec. 11.5.

Coming to the first term we see that $\varphi_i\varphi_j$ is non-zero only if i and j are vertices of the same quadrilateral. Thus the matrix $\int_\Omega \rho_\circ \varphi_i \varphi_j da\, db$ has got at most 9 non-zero terms in each row. However in solving a numerical scheme inverting such a matrix is still expensive. So we replace this term by an approximation which yields a diagonal matrix.

We set

$$\int_Q f\rho_\circ da\, db \sim \sum_{k=1}^{4} f(A_k)\alpha_k \int_Q \rho_\circ da\, db \quad (11.21)$$

where $\{A_k\}_{k=1}^4$ are the four nodes of the quadrilateral Q. We define the α_k so that on replacing f by the basis functions corresponding to the vertices of Q, the relation (11.21) is an equality. Thus if φ^k is the basis function corresponding to A_k, we have

$$\alpha_k = \frac{\int_Q \varphi^k \rho_\circ da\, db}{\int_Q \rho_\circ da\, db}. \quad (11.22)$$

We now have **135**

$$\int_Q \varphi_i \varphi_j \rho_\circ da db \sim \sum_{i=1}^{4} (\varphi_j \varphi_i)(A_k)\alpha_k \int_Q \rho_\circ da\, db.$$

The left-hand-side terms will be non-zero only if $i = j$ and $\varphi_i = \varphi^k$, the function which takes the value 1 at the node A_k.

Now to get the matrix in the coefficient of $\frac{DU}{Dt}$ where $U^T = (u_1, \ldots, u_I)$ we compute it over each Q and assemble these together to get a diagonal matrix M.

Thus in case we do not have the boundary term, the discretization of (11.10) (i) reads as

$$M\frac{DU}{Dt} - A^T P = 0, \tag{11.23}$$

where, if we number the quadrilaterals by $Q_1, \ldots, Q_N$, $P^T = (p_{Q_1}, \ldots p_{Q_N})$ and A^T is the $I \times N$ matrix whose element in the i^{th} row and n^{th} column is

$$\int_{Q_n} \frac{\partial \varphi_i}{\partial x} dx\, dy.$$

Similarly, if B^T is the matrix of order $I \times N$ whose (i, n)th-element is

$$\int_{Q_n} \frac{\partial \varphi_i}{\partial y} dx\, dy,$$

the discretization of (11.10) (ii) is

$$M\frac{DV}{Dt} - B^T P = 0. \tag{11.24}$$

11.5 The Boundary terms

Let us assume that Ω is a bounded domain. One essentially encounters
136 two types of boundary conditions. viz., (i) with p prescribed on the boundary or (ii) with the normal velocity $\vec{u}\cdot\vec{\nu}$ prescribed on the boundary. Note that we impose only one boundary condition and not one each on both u and v as would be the case for a viscous fluid.

In the first case where p is given, one has no problem with the term $\int_{\partial\Omega} p\varphi_i \nu_x ds$ of the equation (11.20).

Let us come to the second case. Let us assume that $\vec{u}\cdot\vec{\gamma} = g$ is given. We now define a new unknown p_s, the pressure on the boundary. We

choose this p_s to be approximated by trial functions in the same space H. For example, H could be the space $H_\circ$ of piecewise constants on the boundary or H_1, the space of continuous piecewise linear functions on the boundary, where, by the work "piecewise" we mean w.r.t. the subdivision of the boundary induced by the subdivision of Ω itself. Let $\{\chi_k\}$ be a basis for H. Then we write

$$\int_{\partial\Omega} (\vec{u}\cdot\vec{\nu} - g)\chi_k ds = 0 \text{ for all } k.$$

This expands as

$$\sum_i \int_{\partial\Omega} (u_i\varphi_i\nu_x + v_i\varphi_i\nu_y - g)\chi_k ds = 0, \tag{11.25}$$

for each k and this gives us K equations where K is the dimension of H.

Also writing

$$p_s = \sum_{k=1}^{k} (p_s)_k\chi_k, \tag{11.26}$$

we substitute in (11.20) to get the equations

$$\left.\begin{aligned} M\frac{DU}{Dt} - A^T P + CP_S &= 0 \\ M\frac{DV}{Dt} - B^T P + DP_S &= 0. \end{aligned}\right\} \tag{11.27}$$

where P_S has as its components the $(p_s)_k$ indexed by k.

One also checks easily that (11.25) takes the form 137

$$C^T U + D^T V - G = 0 \tag{11.28}$$

where G is a known vector.

11.6 Time discretization

As regards the time discretization, we evaluate U, V at $(n+\frac{1}{2})\Delta t$ and p, ρ, ϵ at time $n\Delta t$. For instance, equation (11.3′) can be discretized by

$$\rho_Q^{n+1} \int_Q J^{n+1} da\ db = \rho_Q^n \int_Q J^n da\ db \tag{11.29}$$

for all Q.

The equation (11.7) can be discretized exactly as in the 1 - dimensional case. The discretization for the momentum equations involves the details described before. For instance, in the case of the boundary pressure being zero one has

$$\left.\begin{aligned} M\frac{U^{n+\frac{1}{2}} - U^{n-\frac{1}{2}}}{\Delta t} - A^T P^n = 0 \\ M\frac{V^{n+\frac{1}{2}} - V^{n-\frac{1}{2}}}{\Delta t} - B^T P^n = 0. \end{aligned}\right\} \tag{11.30}$$

11.7 Stability Criteria

We now sketch the procedure to get heuristic stability criteria.

Let us assume that (11.7) and (11.8) together can be integrated to give p as a function of ρ. Let us also have

$$\frac{\partial p}{\partial \rho} = \overline{C}^2, \tag{11.31}$$

by linearising about some constant state defined by $(\bar{c}, \bar{\rho})$. Then (11.3) gives

$$\frac{1}{\bar{\rho}\bar{c}^2}\frac{Dp}{Dt} + div\vec{u} = 0 \tag{11.32}$$

138 where $\vec{u} = (u, v)$. Also the equations (11.4) give

$$\bar{\rho}\frac{D\bar{u}}{Dt} + \text{ grad. } (p + q) = 0. \tag{11.33}$$

Locally the derivatives in t and in x, y commute and hence one gets from (11.32) and (11.33)

$$\frac{1}{\bar{c}^2}\frac{D^2 p}{Dt^2} - div \text{ grad. } (p + q) = 0 \tag{11.34}$$

where by (11.3) and (11.5)

$$q = \frac{\sigma}{\rho}\frac{D\rho}{Dt} = \frac{\sigma}{\rho\bar{c}^2}\frac{Dp}{Dt} \tag{11.35}$$

We imitate this procedure in the discrete case. One gets a discretization of (11.34) from those of (11.3) and (11.4) and then studies the stability conditions. Now that we have div grad, which is nothing but the Laplacian, we are in a situation similar to that of the wave equation.

Let us assume that we do not have the surface pressure.

Proceeding as we did in Sec. 11.4, we get the discretization of (11.32) as

$$\mathbb{N}\frac{P^{n+1}-P^n}{\Delta t} + AU^{n+\frac{1}{2}} + BV^{n+\frac{1}{2}} = 0 \tag{11.36}$$

for all n where A, B are defined as in Sec. 11.4 and the matrix $\mathbb{N}$ is an $(N \times N)$ matrix defined by

$$N_Q = \frac{M_Q}{\rho_Q^2 C_Q^2} \tag{11.37}$$

where the suffix Q denotes the diagonal element of $\mathbb{N}$ corresponding to the quadrilateral Q and M_Q is given by

$$M_Q = \int_Q \rho_\circ da\ db. \tag{11.38}$$

The discretization of (11.38) gives **139**

$$\left.\begin{aligned} &\text{(i)} \quad M\frac{U^{n+\frac{1}{2}}-U^{n-\frac{1}{2}}}{\Delta t} - A^T\left[P^n + \sum \cdot \frac{P^n - P^{n-1}}{\Delta t}\right] = 0 \\ &\text{(ii)} \quad M\frac{V^{n+\frac{1}{2}}-V^{n-\frac{1}{2}}}{\Delta t} - B^T\left[P^n + \sum \cdot \frac{P^n - P^{n-1}}{\Delta t}\right] = 0 \end{aligned}\right\} \tag{11.39}$$

where M is the diagonal matrix of Sec. 11.4 and $\sum$ is a $N \times N$ diagonal matrix whose diagonal element corresponding to the quadrilateral Q is given by

$$\sum_Q = \frac{\sigma_Q}{\rho_Q C_Q^2} \tag{11.40}$$

Now (11.39) (i) and (ii) give

$$\left.\begin{aligned} &\text{(i)} \quad A\frac{(U^{n+\frac{1}{2}} - U^{n-\frac{1}{2}})}{\Delta t} = AM^{-1}A^T\left[P^n + \sum \frac{P^n - P^{n-1}}{\Delta t}\right] \\ &\text{(ii)} \quad B\frac{(V^{n+\frac{1}{2}} - V^{n-\frac{1}{2}})}{\Delta t} = BM^{-1}B^T\left[P^n + \sum \frac{P^n - P^{n-1}}{\Delta t}\right] \end{aligned}\right\} \tag{11.41}$$

Substituting the difference of the equation (11.36) between $n + \frac{1}{2}$ and $n - \frac{1}{2}$ and assuming A and B constant (locally) w.r.t. time we get

$$\mathbb{N}\frac{P^{n+1} - P^n + P^{n-1}}{\Delta t^2} + \left[AM^{-1}A^T + BM^{-1}B^T\right]\left[P^n + \sum \cdot \frac{P^n - P^{n-1}}{\Delta t}\right] = 0 \tag{11.42}$$

which is a discretization of equation (11.34).

Remark 11.2. The matrix $K = AM^{-1}A^T + BM^{-1}B^T$ must give an approximation of the Laplacian. On doing these calculations on a regular mesh one finds that instead of getting the usual 5-point formula for the Laplacian (involving the points $(i - 3/2,\ j + \frac{1}{2})$, $(i + \frac{1}{2},\ j + \frac{1}{2})$, $(i + 3/2,\ j + \frac{1}{2})$), for the x-derivative and $(i + \frac{1}{2},\ j - 3/2)$, $(i + \frac{1}{2},\ j + \frac{1}{2})$ and $(i + \frac{1}{2},\ j + 3/2)$ for the y-derivative) we get a 9-point formula in-
140 volthough the 5-point formula is sufficiently accurate and the 9-point formula does nothing to improve it on a regular mesh, the latter has the advantage of extensions to arbitrary meshes while the former does not.

In case $\sum = 0$, one can perform an analysis similar to the Fourier transform. Let $\{\mu_\alpha\}$ be the spectrum of $\mathbb{N}$ relative to K with eigenvectors $\{\psi_\alpha\}$ i.e.

$$K\psi_\alpha = \mu_\alpha \mathbb{N}\psi_\alpha. \tag{11.43}$$

Decomposing P^n over the eigen spaces, we can write

$$P^n = \sum_\alpha P^n_\alpha \psi_\alpha. \tag{11.44}$$

We are now reduced to studying the stability of the scalar equations

$$\frac{1}{\Delta t^2}(P^{n+1}_\alpha - 2P^n_\alpha + P^{n-1}_\alpha) + \mu_\alpha P^n_\alpha = 0 \tag{11.45}$$

for each α.

One knows that a necessary and sufficient condition for stability is that both roots of the equation

$$\frac{1}{\Delta t^2}(r^2 - 2r + 1) + \mu_\alpha r = 0 \tag{11.46}$$

have moduli ≤ 1 for each α.

When $\sum \neq 0$, we cannot do this. However, if we assume $\sum$ to be a scalar matrix, i.e. $\sum = \sigma_\circ I$, then since it commutes with $\mathbb{N}$ and K one uses (11.44) and is then reduced to studying the stability of

$$\frac{1}{\Delta t^2}(P_\alpha^{n+1} - 2P_\alpha^n + P_\alpha^{n-1}) + \mu_\alpha(P_\alpha^n + \frac{\sigma_\circ}{\Delta t}(P_\alpha^n - P_\alpha^{n-1})) = 0 \tag{11.47}$$

for which a necessary and sufficient condition is that the roots of

$$\frac{r^2 - 2r + 1}{\Delta t} + \mu_\alpha \left[r + \frac{\sigma_\circ}{\Delta t}(r - 1) \right] = 0 \tag{11.48}$$

have moduli ≤ 1. **141**

We now quote a lemma due to Lascaux [20].

Lemma 11.7.1. *Let Q be any quadrilateral of the subdivision. Let M_Q be its mass (Cf. (11.38)). Let the vertices be numbered by $\{k \mid 1 \leq k \leq 4\}$. Let M_k be the diagonal term of M corresponding to the vertex k. Let $L_{Q,k}$ be the length of the diagonal of Q opposite to the vertex k. Define*

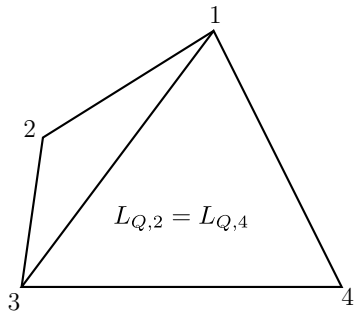


Figure 11.3:

$$L_Q^2 = \sum_{k=1}^{4} \frac{1}{4} \frac{M_Q}{M_k} L_{QK}^2$$

and

$$T_Q = \frac{C_Q^2}{S_Q^2} L_Q^2,$$

where S_Q is the area of Q. Then the eigenvalues $\{\mu_\alpha\}$ defined previously obey the inequality.

$$\mu_\alpha \leq 4 \max_Q . T_Q \tag{11.49}$$

Remark 11.3. The number S_Q/L_Q defines the "thickness" of the cell to be used in the Courant-Friedrichs-Lewy condition.

Remark 11.4. The bound (11.49) is quite realistic. In the case of a regular rectangular mesh, one can compute exactly the μ_α, s and the
142 exact bound can be shown to be

$$4 \max \left(\frac{1}{\Delta x^2}, \frac{1}{\Delta y^2} \right).$$

By the lemma the bound appears as

$$4 \left(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} \right).$$

These two are generally of the same order.

Now if $\sum = 0$, we get a necessary and sufficient condition for stability as $\mu_\alpha \Delta t^2 \leq 4$ (which implies that the roots are complex conjugates of each other with modulus 1) or equivalently,

$$T_Q^{\frac{1}{2}} \cdot \Delta t \leq 1 \text{ for all } Q. \tag{11.50}$$

If $\sum \neq 0$, but a scalar matrix as above, then we repeat what we did in Sec. 9.6 for the 1-dimensional case and we only give sufficient

conditions for the roots to be complex conjugates of each other and for their modulus to be ≤ 1; namely

$$\alpha_Q + \beta_{Q/\alpha_Q} \leq 1 \text{ for all } Q \tag{11.51}$$

(hence $\beta_Q \leq \frac{1}{4}$), where

$$\left.\begin{aligned} \alpha_Q &= \frac{C_Q \Delta t}{(S_Q/L_Q)} \\ \beta_Q &= \frac{(\sigma_Q)\Delta t}{(S_Q/L_Q)^2} \end{aligned}\right\} \tag{11.52}$$

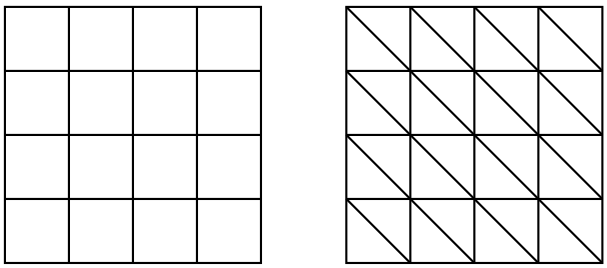
Thus again $\beta_Q \leq \frac{1}{4}$ resembles the condition for the heat equation.

For a detailed discussion of stability criteria see Lascaux [20].

11.8 Concluding Remarks

Remark 11.5. In using the finite element method for the space discretization we use quadrilateral elements and not triangular elements. We illustrate the difficulties involved when using a triangular mesh by a very particular example. **143**

Consider an incompressible fluid in a square domain where the domain has been subdivided into N^2 equal squares. (Fig. 11.4 (a). One can also have a triangular mesh by subdividing each square above into two triangles by drawing a diagonal of each sqaure. (Fig. 11.4 (b)).



(a) (b)

Figure 11.4:

In either case we have $(N+1)^2$ nodes and hence computing u and v at these nodes gives $2(N+1)^2$ unknowns. Since the fluid is incompressible, we get the equation of conservation of mass as

$$\text{div } \vec{u} = 0 \tag{11.53}$$

which is discretised as

$$\int_Q \text{div } \vec{u} dx\, dy = 0 \text{ for all } Q \tag{11.54}$$

Since we have N^2 squares this gives N^2 equations in the first case. However in the second case, we have $2N^2$ equations and then to approx-
144 imate one equation alone we have taken up $2N^2$ unknowns out of the available $2(N^2 + 1)$. This does not leave many more unknowns for the equation of conservation of momentum.

The finite element method which has been used can be named a 'mixed' method because one approximates both the displacements (x, y) and the stresses (p). In order for such a method to work, there must be some compatibility conditions between the corresponding spaces of approximates. (Cf. Raviart, to be published). We have used the space $V_\circ$ for p, ρ, ϵ and a different space for $\vec{X}$ and $\vec{u}$. For quadrilateral elements we use the space V_1 and for triangular elements we have to use another space V_1'. It can be shown that the spaces $V_\circ$ and V_1' do not work together while $V_\circ$ and V_1 are compatible.

These are some of the reasons for using quadrilateral elements instead of triangular ones.

Remark 11.6. We remarked earlier (Cf. Remark 11.2) that our mode of approximation yielded the 9-point formula for the Laplacian. One asks immediately whether ther is a finite element or finite difference scheme which gives the 5-point formula instead. The answer is yes. But this is very complicated when using Lagrangian variables (Cf. Harlow). In case of the Eulerian variable we have the MAC method which has just been devised for this reason. Here p, ρ, ϵ are computed at the centre of each cell while u is computed at the mid-points of two opposite edges and v at the mid-points of the other two edges. (CF. Fig. 11.5).

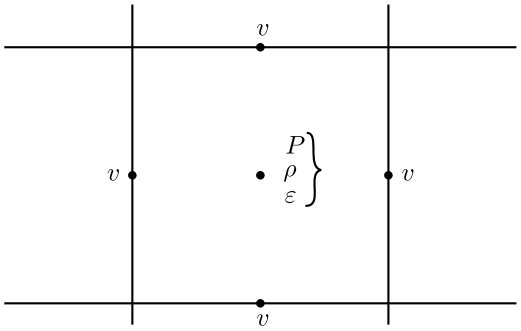


Figure 11.5:

Remark 11.7. Our final comments are on the advantages of the Lagrangian coordinates over the Eulerian coordinates. The principal advantage lies in the tackling of the moving boundary problem or the interface problem. To illustrate how messy the Eulerian equations can become we give an example of an interface between two media.

Given a fixed Eulerian grid, let us examine a cell through which an interface passes and which contains both the media (1 and 2). Then, we must not only know now the boundary moves and where the boundary meets the grid, but also write different equations for the two media. Thus the energy and state equations may be written for the medium $i (i = 1, 2)$

as

$$({}^i\epsilon^{n+1} - {}^i\epsilon^n) + \left(\frac{{}^ip^{n+1} + {}^ip^n}{2}\right)\left(\frac{1}{i\rho^{n+1}} - \frac{1}{i\rho n}\right) = 0 \tag{11.55}$$

$${}^i\epsilon^{n+1} = f({}^ip^{n+1}, {}^i\rho^{n+1}) \tag{11.56}$$

where the superscript i indicates the medium for which the equations are written. We also have the pressure equality

$${}^1p^{n+1} = {}^2p^{n+1}. \tag{11.57}$$

146

If Vol.1 and Vol.2 are the volumes occupied by the media and M_i their masses, we have

$$\left.\begin{aligned} M_1^{n+1} &= {}^1\rho^{n+1}(\text{Vol.1})^{n+1} \\ M_2^{n+1} &= {}^2\rho^{n+1}(\text{Vol.2})^{n+1} \\ (\text{Vol.1})^{n+1} + (\text{Vol.2})^{n+1} &= \text{Vol. of cell} = \\ &= \Delta x \Delta y \end{aligned}\right\} \tag{11.58}$$

The equations (11.55) to (11.58) are eight in number and there are as many unknowns (p, ρ, ϵ, Vol for each material, assuming M_1 and M_2 are known) and we must solve such a system.

Moreover the reader is asked to think of how to define the motion of the interface on the grid and how to provide tests to know when a cell contains both the media or when it does not.

A frailty of the Lagrangian method is that when the motion is too distorted, the quadrilaterals lose their shape and the subsequent equations will not be meaningful compared to the true situation. In this regard the ALE method is very useful in two dimensions. One has a lot of results on this method published by the Los Alamos group (Harlow et al). The reader is referred to Roache [33] for its extensive bibliography where references to these papers can be found.

Bibliography

[1] ALDER. B., FERNBACH. S., and ROTENBERG, M: (Editors), *Fundamental Methods in Hydrodynamics*. Methods in Computational Physics, Vol. 3, 1964, Academic Press, New York. **147**

[2] AMES, W.F. : *Numerical Methods for Partial Differential Equations*, 1969 Barnes and Noble, Inc., New York.

[3] ARONSON, D.G. : Regularity properties of flows through porous media, SIAM Journal of Applied Mathematics, Vol. 17, No. 2, 1969, pp. 461-467.

[4] BORIS, J.P. and BOOK, D.L.: Flux Corrected Transport. I SHASTA A fluid transport algorithm that works, Journal of Comput. Physics Vol. (1973) pp. 38-69.

[5] BURSTEIN, S.L. and MIRIN, A.A. : Third order difference methods for hyperbolic equations, Journal of Computational Physics, Vol. 5, 1970, pp. 547-571.

[6] CIARLET, P.G. : *Lectures on the finite element method*, Tata Institute of Fundamental Research, Bombay, 1975.

[7] CIARLET, P.G. and RAVIART, P.A. : Interpolation theory over curved elements with applications to finite elements methods, Computer methods in Applied Mechanics and Engineering, Vol. 1, 1972, pp. 217-249.

[8] CONWAY and SMOLLER: Global solutions of the Cauchy problem for quasi linear first order equations in several space variables,

Communications on Pure and Applied Mathematics, Vol. 19, 1966, pp. 95-105.

[9] COURANT, R. and FRIEDRICHS, K.O. : *Supersonic Flow and Shock Waves*, Interscience, 1948.

[10] FRIEDRICHS, K.O. : Symmetric positive linear differential equations, Communications on Pure and Applied Mathematics, Vol. 7, 1958, pp. 333-418.

[11] FROMM, J.E. : Practical investigation of convection difference approximations of reduced dispersion, High Speed Computing in Fluid Dynamics - The physics of fluids. Supplement II (1969).

[12] FROMM, J.E. : A method for reducing dispersion inconvective difference schemes, Journal of Computational Physics, Vol. 3, 1968. pp. 176-189.

[13] GOURLAY, A.R. and MORRIS, J.L. : One the comparison of multistep formulation of the optimized Lax-Wendroff method for non-linear hyperbolic systems in two space variables. Journal of Computational Physics, Vol. 5. 1970, pp. 229-243.

148 [14] GRAVELEAU, J.L. and JAMET, P. : A finite difference approach to some degenerate non-linear parabolic equations, SIAM Journal of Applied Mathematics, Vol.20, No.2, 1971, pp. 199-223.

[15] HIRT, C.W. : Heuristic stability theory for finite difference equations, Journal of Computational Physics, Vol.2, 1968, pp. 339-355.

[16] HOSKIN, N.E. : Solution by characteristics of the equations of the dimensional unsteady flow, Methods of Computational Physics, Vol.3, 1964, pp. 265-294.

[17] KASAHARA. A. : On certain finite difference methods for fluid dynamics, U.S. Monthly Weather Review, Vol.93, No.1, January 1965, pp.27-31.

[18] KOT, C.A. : An improved constant time technique for the method of characteristics, *Proceedings of the 3rd International Conference in Numerical Methods in Fluid Dynamics*, 1972, Springer-Verlag, Lecture Notes in Physics.

[19] KREISS, H.O. : Stability theory for difference approximations of mixed initial-boundary value problems - I, Mathematics of Computation, Vol.22, No.104, 1968, pp.703-714.

[20] LASCAUX, P. : Stabilite de la discretisation des equations de l'hydrodynamique lagrangienne 2D, A Report, July 1975.

[21] LAX, P.D. : Weak solutions of non-linear hyperbolic equations and their numerical computation, Communications on Pure and Applied Mathematics, Vol.7, 1954, pp.159-193.

[22] LAX, P.D. : Non-linear partial differential equations and computing, SIAM Review, Vol.11, No.1, 1969, pp.7-19.

[23] LAX, P.D. and WENDROFF, B. : Difference schemes with high order accuracy for solving hyperbolic equations, Communications on Pure and Applied Mathematics. Vol.17, 1964, pp.381.

[24] LAX, P.D. and WENDROFF, B. : Systems of conservation laws, Communications on Pure and Applied Mathematics, Vol.13, 1969, pp.217-237.

[25] LERAT, A. and PEYRET, R. : The problem of spurious oscillations in the numerical solution of the equations of gas dynamics, *Proceedings of the 4th International Conference on Numerical Methods in Fluid Dynamics.* June, 1974, Springer-Verlag, Lecture Notes in Physics.

[26] LIONS, J.L. : *Quelques Methodes de resolution de Problemes aux Limites Nonlineaires.* 1969, Dunod, Paris.

[27] MITCHELL, A.R. : *Computational Methods in Partial Differential Equations*, 1969, John Wiley and Sons

[28] MORTON, K.W. : Stability and convergence in fluid flow problems, Proceedings of the Royal Society, London, A 323, 1971, pp.237-253. **149**

[29] NOH, W.F. : *A general theory for the numerical solution of the equations of hydrodynamics, Numerical solutions of Non-linear Differential Equations* (Ed. D. Greenspan), John Wiley.

[30] POTTER, D. : *Computational Physics*, 1973. John Wiley and Sons.

[31] RAVIART. P.-A. : Sur la résolution numérique de l'équation $u_t + u\,u_x - \epsilon(u_x u_x)_x = 0$, Journal of Differential Equations, Vol.8, 1970, pp. 56-94.

[32] RICHTMYER, R.D. and MORTON, K.W. : *Difference Methods for Initial Value Problems*, 1967, Interscience Publishers, New York.

[33] ROACHE: *Computational Fluid Dynamics*, Hermosa Publishers, Albuquerque.

[34] ROBERTS, K.V, and WEISS, N.O. : Convective difference schemes, Mathematics of Computation, Vol.20. No.94, 1966, pp.272-229.

[35] RUSANOV, V.V. : On difference schemes of third order accuracy for non-linear hyperbolic systems, Journal of Computational Physics, Vol. 5, 1970, pp. 507-516.

[36] THOMEE, V. : Stability of difference schemes in the maximum norm, Journal of Differential Equations, Vol.1, 1965, pp.273-292.

[37] B. VAN LEER. : Towards the ultimate conservative difference scheme. II Monotonicity and conservation combined in a second order scheme Journal of Comput. Physics Vol.14, (1974), pp.361-370.