Multigrid Solution of the Steady Euler Equations

S.P. Spekreijse

Multigrid Solution of the Steady Euler Equations

MULTIGRID SOLUTION OF THE STEADY EULER EQUATIONS

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft
op gezag van de Rector Magnificus,
Prof.Dr. J.M. Dirken,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op donderdag 5 november 1987
te 16.00 uur

door

STEPHANUS PETRUS SPEKREIJSE
Wiskundig ingenieur

geboren te Borne

Centrum voor Wiskunde en Informatica
Amsterdam
1987

Dit proefschrift is goedgekeurd door de promotor:

Prof.Dr.Ir. P. Wesseling

Dr. P.W. Hemker heeft als begeleider in hoge mate bijgedragen aan het totstandkomen van het proefschrift. Het College van Decanen heeft hem als zodanig aangewezen.

STELLINGEN

1. Beschouw het Riemann probleem voor de 1-dimensionale Euler vergelijkingen met een willekeurige linker en rechter toestand. Dan geldt dat de exakte oplossing bestaat d.e.s.d.a. de benaderende oplossing bestaat volgens de *P*-variant van het Osher schema. Dit is niet het geval voor de oorspronkelijke *O*-variant.

2. De in [1] gestelde bewering dat de in dit proefschrift toegepaste *P*-variant van het Osher schema leidt tot overshoot bij het benaderen van een stationaire schok is onjuist.

   [1] S. OSHER and F. SOLOMON (1982). *Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws.* Math. Comp. 38, 339-374.

3. De in dit proefschrift ontwikkelde discretisatie en oplossingsmethode voor de stationaire Euler vergelijkingen vereist bij benadering het volgende aantal elementaire operaties ($+$, $-$, $\times$, $\div$) per grid punt:

   eerste-orde residu berekening: 200 el. op. per grid punt;
   tweede-orde residu berekening: 350 el. op. per grid punt;
   Gauss-Seidel relaxatie: 1000 el. op. per grid punt.

   De oplossingsmethode vereist per FAS-V-cycle bij benadering 6000 el. op. per grid punt.

4. De in dit proefschrift ontwikkelde methode is niet geschikt voor simulatie van subsone stromingen met een uniform laag Mach getal.

5. Beschouw een eerste- en tweede-orde plaatsdiscretisatie van een scalaire hyperbolische behoudswet op een equidistant grid gebaseerd op cel gecentreerde eindige volumes en een flux splitting methode. Neem aan dat in de tweede-orde discretisatie gebruik gemaakt wordt van een limiter zoals beschreven in [2].
   Dan geldt voor een stuksgewijs constante prolongatie en restrictie dat een Galerkin approximatie van de tweede-orde discretisatie op een fijn grid identiek is aan de eerste-orde discretisatie op het grovere grid. Dit is onafhankelijk van de keuze van de limiter.

   [2] S.P. SPEKREIJSE (1987). *Multigrid Solution of Monotone Second-Order Discretizations of Hyperbolic Conservation Laws.* Math. Comp. 49, 135-155.

6. Zij gegeven een zekere norm $\| \ \|$ op $\mathbb{R}^m$. Laat A een reële $m \times m$ matrix zijn met de eigenschap dat

$$\| A^n \| < c \quad \forall n \, .$$

Beschouw de recursie

$$\begin{cases} x_0 = x_0 \\ x_n = A x_{n-1} + y \quad n = 1,2,... \end{cases}$$

met $x_0, y \in \mathbb{R}^m$ willekeurig.
Dan geldt

$$\sup \| x_n \| < \infty \Leftrightarrow y \in \text{Range}\,(I - A) \, .$$

7. Algol 68 is een ideale programmeertaal indien men programmatuur wil ontwikkelen welke niet overdraagbaar is.

8. Bij bezuinigingen op kunst is het consequent om takken van wetenschap met een sterk "l'art pour l'art" beginsel niet te ontzien.

9. Faculteiten voor toegepast technisch wetenschappelijk onderzoek dienen bij het aanstellen van docenten een voorkeur te hebben voor wetenschappers met industriële werk-ervaring.

10. Ter bevordering van het gebruik van de fiets in woon-werk verkeer verdient het aanbeveling dat in iedere werkomgeving douche gelegenheid aanwezig is.

S.P. Spekreijse

Delft, 5 november 1987.

ter nagedachtenis aan mijn vader

CONTENTS

Chapter 0

Introduction

Since the invention of the computer, computational fluid dynamics has influenced the science of aerodynamics considerably. In the sixties, panel methods were introduced to compute potential flows around airfoils. In the seventies, major advances were achieved in the simulation of transonic flows by the full potential approximation with finite volume methods. Nowadays, we see rapid developments in methods for solving the Euler and compressible Navier-Stokes equations.

Euler flow simulation is especially valuable for flows where the potential hypothesis is no longer valid, e.g. flows which contain strong shocks and /or vorticity. For example, Euler flow simulation is important for transonic flow, which is the principal operating regime of both civil and military aircraft. The Euler equations describe inviscid compressible gas flows. In practice, the viscosity of air is so low that viscous effects are confined to thin boundary layers adjacent to the surface of bodies present in the flow. Such flows are usually well described by the Euler equations. But there are cases of steady transonic flow over a two-dimensional airfoil where the shock wave location is very sensitive to the boundary layer thickness distribution. A striking example of this is given in [4]. Because the lift to drag ratio of an airfoil in a transonic flow is very sensitive to the shock wave position, viscosity cannot be neglected in such cases. Then the compressible Navier-Stokes equations should be used.

At very high Reynolds numbers, the flow in the boundary layer becomes turbulent. Adequate modelling of turbulence at acceptable cost poses a challenge that will have to be met in the future. We may regard the solution of the Euler equations as a preparatory stage for the development of solution methods for Navier-Stokes equations with or without turbulence modelling.

The objective of this work is to contribute to the development of efficient numerical methods for the computation of steady solutions of the Euler equations. The major considerations for the computation of Euler flows are the capability to treat flows in complex geometrical configurations, with proper representation of shock waves and contact discontinuities, with high order of accuracy in the smooth parts of the flow, and with computational efficiency and robustness.

In this work we do not use complex geometrical configurations, to avoid grid-generation problems. Furthermore, we restrict ourselves to the Euler equations in two dimensions (2D). However, all techniques used can be extended in a straightforward way to the 3D Euler equations. In this study, we

focus on the space discretization, in order to combine second-order accuracy in the smooth parts of the flow field with a proper representation of discontinuities. Furthermore, much attention is payed to computational efficiency and robustness.

A conservative finite volume scheme is used for the space discretization. The scheme is a so-called 'shock capturing' scheme, i.e. the same numerical scheme is used everywhere in the flow; no adaptions are made in the neighbourhood of discontinuities. This is possible because the scheme is a finite volume scheme, i.e. it is based on the integral form rather than the differential form of the Euler equations. The integral form is applicable everywhere, the differential form is not valid where the solution is not differentiable. Nowadays, finite volume schemes are almost universally used for shock capturing codes.

In a finite volume scheme, flux-computation must be carried out at the boundaries of the volumes. A flux at a cell boundary is the amount of mass, momentum and energy transported per unit of time across the cell boundary. We use a cell-centered finite volume scheme, i.e. the numerical approximations are stored inside the volumes. The equations are obtained by demanding that the total flux is zero for each volume. At each cell boundary a flux is computed by approximately solving a local one-dimensional Riemann problem. As a consequence, the scheme is characteristic-based or upwind. The approximate Riemann solver used is as proposed by Osher [5]. The implementation of Osher's scheme is not so complex as is generally believed, provided that the proper dependent variables are used and that the local Riemann problems are solved approximately by using an ordering of the constituent parts of the integral path in state space which is the reverse of that proposed by Osher.

One of the merits of Osher's approach to solve the Riemann problem is that boundary conditions can be discretized in a way which is completely consistent with the discretization of the steady Euler equations in the interior of the domain. This is a consequence of the fact that Osher's scheme is based on Riemann invariants, just as proper boundary condition treatments. Osher's scheme is based on a sound mathematical theory. The scheme fulfils an entropy condition and therefore unphysical solutions are excluded. In its original form Osher's scheme is first-order accurate. Shocks and contact discontinuities are captured very well (in at most two interior grid points) as long as they are aligned to the grid. But oblique (with respect to the grid) shocks and contact discontinuities are smeared out disastrously. Furthermore, in smooth parts of the flow, first-order accuracy is too low for practical purposes. Therefore we wish to improve the order of accuracy and to steepen oblique discontinuities.

This is done by the so-called MUSCL (Monotone Upwind Schemes for Conservation Laws) -approach as proposed by Van Leer [6]. In that approach, the data is first prepared and modified (limited) before a Riemann solver is applied. The limiting is done to prevent spurious oscillations in the neighbourhood of discontinuities. The limiting must be nonlinear even when applied to linear problems. This approach allows monotonicity to be achieved

simultaneously with second-order accuracy.

In chapter II, these topics are studied thoroughly. The study concerns Riemann solvers, Osher's scheme, boundary condition treatments, linearization, the MUSCL-approach, limiters etc. Chapter I is an introductory chapter in order to prepare the material necessary for the succeeding chapters. At the end of chapter II, a first- and second-order accurate discretization of the steady Euler equations has been determined completely. The two succeeding chapters III and IV describe respectively the solution methods for the first- and second-order discretization. The first-order discretization is solved by a Nonlinear Multigrid Method (NMG), also called FAS (Full Approximation Scheme), see Brandt [1]. Nested iteration, also called FMG (Full Multigrid Method), is used to obtain a good initial approximation on the finest grid. The multigrid method is very straightforward. A Collective Symmetric Gauss-Seidel (CSGS) ralaxation procedure is used as a smoothing method. The numerical examples given in section 3.3 show that the characteristic features of a successful multigrid method are obtained: robustness, efficiency (about 3 NMG iterations are sufficient to surpass truncation error accuracy) and grid independency of the convergence rate (at least for transonic and supersonic flow). The numerical examples cover channel flows, resolution of contact discontinuities, and a blunt body (circle cylinder) in a supersonic flow.

A defect correction method is used to improve the accuracy of the first-order solutions. The defect correction method, which is the topic of chapter IV, makes use in a very effective way of the excellent multigrid solver for the solution of first-order discretizations. In fact, the second-order discretization is used only to construct appropriate source terms, and the solution of the first-order discretization of the steady Euler equations with these source terms are obtained by the multigrid solver. This process is repeated iteratively. In section 4.3, the numerical solutions of the second-order discretization obtained by the defect correction method are given and comparison with the first-order solutions given in section 3.3 show clearly the improvement in accuracy and resolution of discontinuities.

Finally, we refer to the work of B. Koren [2,3] who used the discretizations and solution methods as described in this work for airfoil flow computations. His results clearly show the feasibility of the method for such applications.

This thesis is based on the following publications:

[A] P.W. HEMKER, S.P. SPEKREIJSE (1985). *Multigrid Solution of the Steady Euler Equations.* In: Advances in Multi-Grid Methods. (D. BRAESS, W. HACKBUSH, U. TROTTENBERG, eds.). Notes on Numerical Fluid Mechanics, Volume 11, 33-44. Vieweg, Braunschweig.

[B] P.W. HEMKER, S.P. SPEKREIJSE (1986). *Multiple Grid and Osher's scheme for the Efficient Solution of the Steady Euler Equations.* Appl. Num. Math. 2, 475-493.

[C] S.P. SPEKREIJSE (1986). *Second-Order Accurate Upwind Solutions of the 2D Steady Euler Equations by the Use of a Defect Correction Method.* In:

Multigrid Methods II. (W. HACKBUSH, U. TROTTENBERG, eds.). Lecture Notes in Mathematics 1228, 285-300, Springer Verlag, Berlin.

[D] S.P. SPEKREIJSE (1987). *Multigrid Solution of Monotone Second-Order Discretizations of Hyperbolic Conservation Laws.* Math. Comp. 49, 135-155.

[E] B. KOREN, S.P. SPEKREIJSE (1987). *Multigrid and Defect Correction for the Efficient Solution of the Steady Euler Equations.* In: Research in Numerical Fluid Dynamics/Proceedings of the 25th Meeting of the Dutch Association for Numerical Fluid Dynamics. (P. WESSELING, ed.). Notes on Numerical Fluid Mechanics 17, 87-100, Vieweg, Braunschweig.

REFERENCES

1. A. BRANDT (1982). *Guide to Multigrid Development.* In: Multigrid Methods. (W. HACKBUSH, U. TROTTENBERG, eds.). Lecture Notes in Mathematics 960, 220-321, Springer Verlag, Berlin.
2. P.W. HEMKER, B. KOREN (1986). *A Non-Linear Multigrid Method for the Steady Euler Equations.* Report NM-R6821, Centre for Mathematics and Computer Science, Amsterdam. To appear in Proceedings Gamm-Workshop on the Numerical Simulation of Compressible Euler Flows, Rocquencourt, 1986. Notes on Numerical Fluid Mechanics, Vieweg Verlag, Braunschweig.
3. B. KOREN (1986). *Evaluation of Second-Order Schemes and Defect Correction for the Multigrid Computation of Airfoil Flows with the Steady Euler Equations.* Report NM-R8616, Centre for Mathematics and Computer Science, Amsterdam. To appear in J. Comput. Phys.
4. H. MCDONALD, S.J. SHAMROTH, W.R. BRILEY (1981). *Transonic Flows with Viscous Effects.* In: Transonic, Shock, and Multidimensional Flows: Advances in Scientific Computing (R.E. MEYER, ed.), 219-240. Academic Press, New York.
5. S. OSHER, F. SOLOMON (1982). *Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws.* Math. Comp. 38, 339-374.
6. B. VAN LEER (1974). *Towards the Ultimate Conservative Difference Scheme, II.* Monotonicity and Conservation Combined in a Second-Order Scheme. J. Comput. Phys. 14, 361-376.

Chapter I

The Euler Equations

1.1. Derivation of the Euler equations.

In this section the Euler equations are introduced. We consider the Euler equations in two dimensions only. The restriction to two dimensions is only for practical reasons and is not fundamental (see the introductory chapter 0).
Let there be given a Cartesian coordinate system (x,y). Let t denote the time. With $\rho=\rho(x,y,t)$, $u=u(x,y,t)$, $v=v(x,y,t)$ and $p=p(x,y,t)$ we denote density, velocity components in the x- and y-direction and pressure. These quantities are the so-called primitive variables. Consider an arbitrary simply-connected region $\Omega\subset\mathbb{R}^2$ and let $n=(n_1,n_2)^T$ be the outward unit normal on the boundary $\partial\Omega$. The region Ω is a so-called control volume; we will apply the physical laws of conservation of mass, momentum and energy to the fluid flow in Ω.

I. Conservation of mass.
The law of conservation of mass is given by

$$\frac{d}{dt}\int_{\Omega}\rho dv=-\int_{\partial\Omega}\rho(\underline{n}\,.\,\underline{v})d\sigma \tag{1.1.1}$$

where $v=(u,v)^T$ is the velocity, dv is a volume element and $d\sigma$ a surface element. With $n\,.\,v$ we denote the innerproduct: $n\,.\,v=n_1u+n_2v$. Using Gauss's theorem for a vector-field, we may write, assuming that $\rho\underline{v}$ is differentiable,

$$\int_{\partial\Omega}\rho(\underline{n}\,.\,\underline{v})d\sigma=\int_{\Omega}\operatorname{div}(\rho\underline{v})dv$$

where $\operatorname{div}(\rho\underline{v})=\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v)$ is the divergence.
Using the fact that $\Omega\subset\mathbb{R}^2$ is arbitrary we find the equation of continuity:

$$\frac{\partial}{\partial t}\rho+\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v)=0 \tag{1.1.2}$$

Equations (1.1.1) and (1.1.2) are the equation of continuity in integral and differential form respectively. Both forms are very important.

II. Conservation of momentum.
Assuming frictionless flow and absence of body forces, the law of conservation of momentum is given by

$$\frac{d}{dt}\int_{\Omega}\rho \underline{v}dv = -\int_{\partial\Omega}\rho \underline{v}(\underline{n}\cdot\underline{v})d\sigma - \int_{\partial\Omega}p\underline{n}d\sigma\,. \tag{1.1.3}$$

Equation (1.1.3) is a vector equation. The x-component is

$$\begin{aligned}\frac{d}{dt}\int_{\Omega}\rho u dv &= -\int_{\partial\Omega}\rho u(\underline{n}\cdot\underline{v})d\sigma - \left[\int_{\partial\Omega}p\underline{n}d\sigma\right]_x \\ &= -\int_{\Omega}\operatorname{div}(\rho u\underline{v})dv - \left[\int_{\Omega}\nabla p dv\right]_x \end{aligned}\tag{1.1.4}$$

where we have used the theorem of Gauss for both a vector- and scalar-field, $\nabla p = \left[\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}\right]^T$ is the gradient.

The y-component of equation (1.1.3) is

$$\begin{aligned}\frac{d}{dt}\int_{\Omega}\rho v dv &= -\int_{\partial\Omega}\rho v(\underline{n}\cdot\underline{v})d\sigma - \left[\int_{\partial\Omega}p\underline{n}d\sigma\right]_y \\ &= -\int_{\Omega}\operatorname{div}(\rho v\underline{v})dv - \left[\int_{\Omega}\nabla p dv\right]_y\,. \end{aligned}\tag{1.1.5}$$

Because Ω is arbitrary, we obtain from equations (1.1.4) and (1.1.5)

$$\frac{\partial}{\partial t}(\rho u)+\frac{\partial}{\partial x}(\rho u^2+p)+\frac{\partial}{\partial y}(\rho uv)=0 \tag{1.1.6a}$$

$$\frac{\partial}{\partial t}(\rho v)+\frac{\partial}{\partial x}(\rho uv)+\frac{\partial}{\partial y}(\rho v^2+p)=0 \tag{1.1.6b}$$

which are the momentum equations in differential form.

III. Conservation of energy.

Let e denote the internal energy of the fluid. The energy of the fluid consists of internal and kinetic energy and is equal to $\rho e+\frac{1}{2}\rho(u^2+v^2)$ per unit of volume. We define the total energy as

$$E=\rho e+\tfrac{1}{2}\rho(u^2+v^2)\,. \tag{1.1.7}$$

Under the assumptions of a nonviscous, nonconducting fluid and absence of body forces, the law of conservation of energy is given by

$$\frac{d}{dt}\int_{\Omega}E dv = -\int_{\partial\Omega}E(\underline{n}\cdot\underline{v})d\sigma - \int_{\partial\Omega}p(\underline{n}\cdot\underline{v})d\sigma \tag{1.1.8}$$

and, with similar reasoning as before, the energy equation in differential form is obtained as

$$\frac{\partial}{\partial t}E+\frac{\partial}{\partial x}(E+p)u+\frac{\partial}{\partial y}(E+p)v=0 \tag{1.1.9}$$

Combining the equations (1.1.2), (1.1.6) and (1.1.9) we find the Euler equations:

$$\frac{\partial}{\partial t}\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ E\end{bmatrix}+\frac{\partial}{\partial x}\begin{bmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ (E+p)u\end{bmatrix}+\frac{\partial}{\partial y}\begin{bmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ (E+p)v\end{bmatrix}=0\,. \tag{1.1.10}$$

These equations are valid for a nonviscous, non-heat-conducting fluid without body forces. Notice that there are 5 unknowns in the 4 equations. Another equation is provided by the thermodynamical equation of state, which can be written in general as

$$p=p(\rho,e)\,. \tag{1.1.11}$$

For a perfect gas we have

$$p=\rho RT,\ e=c_vT \tag{1.1.12}$$

where T is the temperature, $R=c_p-c_v$ the gasconstant, and c_v,c_p the specific heat at constant volume and constant pressure, respectively. Define the ratio of specific heats $\gamma=c_p/c_v$. For a perfect gas the thermodynamical equation of state gives

$$p=\frac{R}{c_v}\,.\ \rho e=(\gamma-1)\rho e=(\gamma-1)\,(E-\tfrac{1}{2}\rho(u^2+v^2))\,. \tag{1.1.13}$$

For almost all aerodynamical problems one can assume that the non-dimensional quantity γ is constant ($\gamma=1.4$ for air).
Important physical quantities and relations are listed in the appendix.

1.2. Some general properties of solutions of the Euler equations.

In this section we introduce two important quantities: the total enthalpy H and the entropy s. We show, under certain rather general circumstances, that these quantities are constant along streamlines. Furthermore, we investigate what kind of discontinuities are possible in solutions of the steady Euler equations. It turns out that there are two types: shock waves and contact discontinuities.

I. The total enthalpy.
The enthalpy h and the total enthalpy H are defined by

$$h=e+\frac{p}{\rho} \tag{1.2.1}$$

$$H=h+\tfrac{1}{2}(u^2+v^2)\,. \tag{1.2.2}$$

Using the definition (1.1.7) of the total energy E, we find

$$H=\frac{E+p}{\rho}\,. \tag{1.2.3}$$

Hence, the energy equation in integral form (1.1.8) can be written as

$$\frac{d}{dt}\int_{\Omega}\rho H dv = -\int_{\partial\Omega}\rho H(\underline{n}\,.\,\underline{v})d\sigma + \int_{\Omega}\frac{\partial p}{\partial t}dv \tag{1.2.4}$$

and this equation becomes in differential form

$$\frac{\partial}{\partial t}(\rho H)+\frac{\partial}{\partial x}(\rho u H)+\frac{\partial}{\partial y}(\rho v H)=\frac{\partial p}{\partial t}\,. \tag{1.2.5}$$

Combining this equation with the continuity equation (1.1.2) we find that (1.2.5) can be written as

$$\frac{\partial H}{\partial t}+u\frac{\partial H}{\partial x}+v\frac{\partial H}{\partial y}=\frac{1}{\rho}\frac{\partial p}{\partial t}\,. \tag{1.2.6}$$

In this equation we recognize the material derivative

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}\,. \tag{1.2.7}$$

The material derivative (or total derivative) expresses the rate of change of a property of a fluid particle. Combining (1.2.6) and (1.2.7) we find for steady flow (all time derivatives are 0)

$$\frac{DH}{Dt}=0\,. \tag{1.2.8}$$

Thus, in the case of steady flow, the total enthalpy H is constant along streamlines. We shall see that the total enthalpy remains also constant when a streamline passes a discontinuity (shock wave). When H is uniformly constant the fluid is called isenthalpic or isoenergetic.

II. The entropy.

In the same way as we have combined the energy equation with the continuity equation, it is possible to combine the momentum equations (1.1.6) with the continuity equation. Then we find

$$\begin{aligned}\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}&=-\frac{1}{\rho}\frac{\partial p}{\partial x}\\ \frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}&=-\frac{1}{\rho}\frac{\partial p}{\partial y}\,,\end{aligned} \tag{1.2.9}$$

which can be written as

$$\frac{D}{Dt}\underline{v}=-\frac{1}{\rho}\nabla p\,. \tag{1.2.10}$$

One easily derives from (1.1.2), (1.1.7) and (1.1.9) that

$$\frac{D}{Dt}(e+\tfrac{1}{2}(\underline{v}\,.\,\underline{v}))=-\frac{1}{\rho}\operatorname{div}(p\underline{v})\,. \tag{1.2.11}$$

Combining the last two equations we find

$$\frac{De}{Dt}=-\frac{p}{\rho}\operatorname{div}\underline{v}\,. \tag{1.2.12}$$

The continuity equation (1.1.2) can be written as

$$\frac{D\rho}{Dt}+\rho \operatorname{div} \underline{v}=0 . \tag{1.2.13}$$

Thus, we also have

$$\frac{De}{Dt}=\frac{p}{\rho^2}\frac{D\rho}{Dt} . \tag{1.2.14}$$

We assume that the fluid is a perfect gas, hence (see 1.1.13)

$$p=(\gamma-1)\rho e \tag{1.2.15}$$

which gives

$$\frac{De}{Dt}=\frac{1}{\gamma-1}\left\{\frac{1}{\rho}\frac{Dp}{Dt}-\frac{p}{\rho^2}\frac{D\rho}{Dt}\right\} . \tag{1.2.16}$$

Combining (1.2.14) and (1.2.16), we conclude that

$$\frac{Dp}{Dt}-\frac{\gamma p}{\rho}\frac{D\rho}{Dt}=0 . \tag{1.2.17}$$

The entropy s is defined as

$$s=c_v \ln \frac{p}{\rho^\gamma} . \tag{1.2.18}$$

Using (1.2.17) we find

$$\frac{Ds}{Dt}=\frac{c_v}{p}\left\{\frac{Dp}{Dt}-\frac{\gamma p}{\rho}\frac{D\rho}{Dt}\right\}=0 . \tag{1.2.19}$$

Hence, we have found the important result that the entropy of a fluid particle remains constant in the fluid. This result is only true for an inviscid non-conducting gas. If we take into account viscosity, a similar derivation [4] shows that

$$\frac{Ds}{Dt}\geqslant 0 \tag{1.2.20}$$

Hence, in the case of viscous flow, the entropy of a fluid particle cannot decrease.

In the derivation of (1.2.19) we have assumed that the flow is smooth. Therefore, from (1.2.19) one cannot conclude that the entropy of a fluid particle remains constant when the particle crosses a discontinuity (shock wave). Indeed, we shall show that there is an entropy jump at a shock wave. Because of the fact that in a viscous flow the entropy of a fluid particle cannot decrease and because all real fluids are, in fact, viscous, we demand that the entropy of a fluid particle, which passes a shock wave does not decrease. This is the so-called entropy condition. Thus, the idea behind the entropy condition is that a solution of the inviscid Euler equations is the limit of a sequence of solutions of the viscous Navier-Stokes equations with vanishing viscosity.

When s is uniformly constant, the fluid is called isentropic.

III. Discontinuous solutions of the steady Euler equations.
Now we investigate under which conditions a flow field composed of two uniform flows, separated by a straight line l, is a solution of the steady Euler equations. We choose the x-axis perpendicular to l and the y-axis along l. Hence, the question becomes when is:

$$q(x,y)=\begin{cases} q_L & x<0 \\ q_R & x>0 \end{cases}$$

a solution of the steady Euler equations. Because $q(x,y)$ is discontinuous we have to apply the Euler equations in integral form. Take a control volume with infinitesimal width but finite length across the discontinuity (see fig. 1.2a).

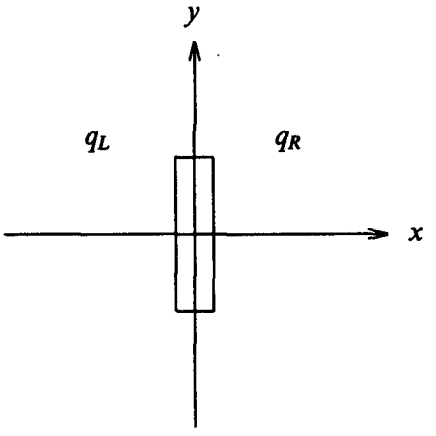


FIGURE 1.2a. Discontinuous steady flow field with a control volume.

Application of the equations of continuity, momentum and energy gives:

$$\rho_L u_L = \rho_R u_R \tag{1.2.21a}$$

$$\rho_L u_L^2 + p_L = \rho_R u_R^2 + p_R \tag{1.2.21b}$$

$$\rho_L u_L v_L = \rho_R u_R v_R \tag{1.2.21c}$$

$$\rho_L u_L H_L = \rho_R u_R H_R \tag{1.2.21d}$$

Consider two posibilities:

A. Contact discontinuity.
Suppose $u_L=0$. Then (1.2.21) is fulfilled if $u_R=0$ and $p_L=p_R$. Hence, a flow field composed of two uniformly constant flows with the same flow direction and pressure but different densities and speeds is a solution of the steady Euler equations. The discontinuity at the interface between the two flows is called a

contact discontinuity (or slip line). Notice that a fluid particle does not cross a contact discontinuity.

B. Shock wave.
Suppose $u_L \neq 0$. Then $u_R \neq 0$, $v_L = v_R$ and $H_L = H_R$. Because $u_L = u_R$ implies $q_L = q_R$ i.e. a uniformly constant flow field, we may assume that $u_L \neq u_R$. From the continuity equation (1.2.21a) it follows that u_L and u_R have the same sign. Therefore, without loosing generality, we suppose that $u_L > 0$, $u_R > 0$. This kind of discontinuity is called a shock wave. Notice that a fluid particle crosses a shock wave. From equation (1.2.21) we shall derive several jump relations. First, we introduce the speed of sound c:

$$c = \sqrt{\frac{\gamma p}{\rho}} . \tag{1.2.22}$$

One easily derives that the total enthalpy H can be expressed as

$$H = \frac{c^2}{\gamma - 1} + \frac{1}{2}(u^2 + v^2) \tag{1.2.23}$$

From (1.2.21a,b) we see that

$$u_L + \frac{p_L}{\rho_L u_L} = u_R + \frac{p_R}{\rho_R u_R}$$

or

$$u_L - u_R = \frac{c_R^2}{\gamma u_R} - \frac{c_L^2}{\gamma u_L} . \tag{1.2.24}$$

Because $H_L = H_R$, $v_L = v_R$, we also have

$$\frac{c_L^2}{\gamma - 1} + \frac{1}{2}u_L^2 = \frac{c_R^2}{\gamma - 1} + \frac{1}{2}u_R^2 \equiv \frac{1}{2}\frac{\gamma + 1}{\gamma - 1}c^{*2} \tag{1.2.25}$$

with c^* a constant. Combining these last two equations, one easily derives

$$u_L u_R = c^{*2} \tag{1.2.26}$$

which is known as the Prandtl relation.
Introducing $M = u/c$, $M^* = u/c^*$, the Prandtl relation becomes

$$M_R^* M_L^* = 1 \tag{1.2.27}$$

and the relation

$$\frac{c^2}{\gamma - 1} + \frac{1}{2}u^2 = \frac{1}{2}\frac{\gamma + 1}{\gamma - 1}c^{*2}$$

results in a relation between M and M^*:

$$M^{*2} = \frac{(\gamma + 1)M^2}{(\gamma - 1)M^2 + 2} . \tag{1.2.28}$$

Now, the jump relations are easily derived. For instance

$$\frac{\rho_R}{\rho_L}=\frac{u_L}{u_R}=\frac{u_L^2}{u_L u_R}=M_L^{*2}=\frac{(\gamma+1)M_L^2}{(\gamma-1)M_L^2+2}\,. \tag{1.2.29}$$

The jump relation for the pressure becomes:

$$p_R-p_L=\rho_L u_L^2-\rho_R u_R^2=\rho_L u_L(u_L-u_R)=\rho_L u_L^2(1-\frac{u_R}{u_L})$$

thus

$$\frac{p_R}{p_L}=1+\frac{2\gamma}{\gamma+1}(M_L^2-1)\,. \tag{1.2.30}$$

The difference between the entropy in front of and behind the shock is

$$s_R-s_L=c_v\ln\left\{\left[1+\frac{2\gamma}{\gamma+1}(M_L^2-1)\right]\cdot\left[\frac{(\gamma-1)M_L^2+2}{(\gamma+1)M_L^2}\right]^{\gamma}\right\}. \tag{1.2.31}$$

This is an important relation, because according to the entropy condition $s_R\geqslant s_L$. Using $s_R>s_L$ ($s_R=s_L\Rightarrow M_L=1$, $u_R=u_L$), we can conclude that $M_L>1$, hence, $M_L^*>1$, $M_R^*<1$ and $M_R<1$. In the case of a normal shock ($v_L=v_R=0$), M is the Mach number and the important conclusion can be drawn that in front of a normal shock the flow is supersonic ($M>1$) and behind the shock the flow is subsonic ($M<1$). Notice that this conclusion is a consequence of the entropy condition. This is an example of the importance of the entropy condition.
It is easily derived that, in the case of a weak shock i.e. $M_L=1+\epsilon$, $0<\epsilon\ll 1$, we have

$$s_R-s_L=O(\epsilon^3)\;;\;\frac{\Delta p}{p_L}=O(\epsilon)\;;\;\frac{\Delta\rho}{\rho_L}=O(\epsilon)\;;\;\frac{\Delta u}{u_L}=O(\epsilon) \tag{1.2.32}$$

with $\Delta p=p_R-p_L$; $\Delta\rho=\rho_R-\rho_L$, $\Delta u=u_R-u_L$.
Thus, a small but finite change of pressure, for which there are corresponding first-order changes of density and velocity, causes only a third-order change in entropy. Therefore, a weak shock produces a nearly isentropic change of state. This is an important result because the assumption that the fluid is isentropic (and isenthalpic) leads to a drastic simplification of the Euler equations (see section 1.4).

1.3. Hyperbolic systems.

1.3.1. General theory.

In this section we study a general first-order system of quasi-linear equations in two independent variables of the form

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)=0 \tag{1.3.1.1}$$

where $q=(q_1,\ldots,q_n)^T\in\mathbb{R}^n$ and $(x,t)\in\mathbb{R}\times\mathbb{R}^+$. We assume that the vector-

valued function $f:\mathbb{R}^n \mapsto \mathbb{R}^n$ is C^1. We define the $n \times n$ matrix $A(q)=\frac{df}{dq}(q)$.

DEFINITION 1.3.1a.
System (1.3.1.1) is called hyperbolic if there exists a real diagonal matrix $D(q)$ and a non-singular real matrix $R(q)$ such that

$$A(q)R(q)=R(q)D(q) \quad \forall q \in \mathbb{R}^n . \tag{1.3.1.2}$$

The column vectors of $R(q)$ are eigenvectors of $A(q)$ and the diagonal entries of $D(q)$ are the corresponding eigenvalues. We shall denote by $R_k(q)$ the kth column vector of $R(q)$ and with $\lambda_k(q)$ the corresponding eigenvalue: $\lambda_k(q)=D_{kk}(q)$. Furthermore we shall assume that the eigenvalues $\lambda_k(q)$ have been labeled in increasing order i.e. $\lambda_1(q) \leqslant \lambda_2(q) \leqslant \cdots \leqslant \lambda_n(q)$.

EXAMPLE 1.3.1a (The linear case).
Suppose $f(q)=Aq$ where A is a constant $n \times n$ matrix. Hence, (1.3.1.1) simplifies to

$$\frac{\partial q}{\partial t}+A\frac{\partial q}{\partial x}=0 . \tag{1.3.1.3}$$

A solution $q=q(x,t)$ of (1.3.1.3) can be expressed with respect to the basis $\{R_1, \cdots, R_n\}$ i.e.

$$q=q(x,t)=\sum_{i-1}^{n} \alpha_i(x,t)R_i$$

where $\alpha_i:\mathbb{R} \times \mathbb{R}^+ \mapsto \mathbb{R}$. Substitution of this expression in (1.3.1.3) leads to

$$\frac{\partial q}{\partial t}+A\frac{\partial q}{\partial x}=\sum_{i=1}^{n}\left\{\frac{\partial \alpha_i}{\partial t}(x,t)+\lambda_i\frac{\partial \alpha_i}{\partial x}(x,t)\right\}R_i=0$$

and because the eigenvectors R_i are independent

$$\frac{\partial \alpha_i}{\partial t}(x,t)+\lambda_i\frac{\partial \alpha_i}{\partial x}(x,t)=0 \quad i=1, \cdots, n$$

The general solution of this equation is

$$\alpha_i(x,t)=\alpha_i^0(x-\lambda_i t) \quad i=1, \cdots, n$$

with $\alpha_i^0:\mathbb{R} \mapsto \mathbb{R}$. Hence, we have found that the general solution of (1.3.1.3) is

$$q(x,t)=\sum_{i=1}^{n}\alpha_i^0(x-\lambda_i t)R_i .$$

The solution of the pure initial value problem on $\mathbb{R}^+ \times \mathbb{R}$

$$\begin{cases} \dfrac{\partial q}{\partial t}+A\dfrac{\partial q}{\partial x}=0 \\ \\ q(x,0)=q_0(x) \quad x \in \mathbb{R} \end{cases} \tag{1.3.1.4}$$

becomes

$$q(x,t)=\sum_{i=1}^{n}\alpha_i^0(x-\lambda_i t)R_i \tag{1.3.1.5}$$

with

$$\sum_{i=1}^{n}\alpha_i^0(x)R_i=q_0(x)\,. \tag{1.3.1.6}$$

Hence, the solution of (1.3.1.4) is obtained after representing the function $q_0:\mathbb{R}\mapsto\mathbb{R}^n$ with respect to the basis $\{R_1,\cdots,R_n\}$.

Now, we shall introduce the Riemann problem. The Riemann problem is very important because it forms the underlying physical model of many upwind schemes for the Euler equations. For instance, the famous Godunov upwind scheme uses the exact solution of the Riemann problem for the numerical solution of the Euler equations [2]. Other well known upwind schemes use approximate solutions of the Riemann problem.

DEFINITION 1.3.1b.
The Riemann problem for a general hyperbolic system is the following initial value problem

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)=0 \tag{1.3.1.7}$$

with

$$q(x,0)=\begin{cases}q_L & x<0\\ q_R & x>0\end{cases}$$

where q_L and q_R are constant states.

THEOREM 1.3.1a.
Suppose there exists a unique solution $q=q(x,t)$ of the Riemann problem (1.3.1.7). Then the solution $q=q(x,t)$ can be written in similarity form $q(x,t)=\tilde{q}(x/t)$.

PROOF.
Define $q_\alpha(x,t)=q(\alpha x,\alpha t)$ with $\alpha\in\mathbb{R}^+$. Then it is easily verified that $q_\alpha(x,t)$ is also a solution of the Riemann problem. Hence, $q(x,t)=q(\alpha x,\alpha t)\ \forall\alpha\in\mathbb{R}^+$, so $q(x,t)=\tilde{q}(x/t)$. □

EXAMPLE 1.3.1b (The linear case).
Consider the Riemann problem for a linear hyperbolic system (see example 1.3.1a). Following the solution method outlined in examples 1.3.1.a, suppose $q_L=\sum_{i=1}^{n}\alpha_i R_i$, $q_R=\sum_{i=1}^{n}\beta_i R_i$. Hence,

$$q(x,0)=\sum_{i=1}^{n}\{\beta_i H(x)+\alpha_i(1-H(x))\}R_i$$

with $H:\mathbb{R}\mapsto\mathbb{R}$ the Heavyside function ($H(x)=1$ if $x>0$, $H(x)=0$ if $x<0$). The solution of the Riemann-problem becomes

$$q(x,t)=\sum_{i=1}^{n}\{\beta_i H(x-\lambda_i t)+\alpha_i(1-H(x-\lambda_i t))\}R_i\ .$$

This solution is illustrated in fig. 1.3.1a for $n=3$. We have assumed that $\lambda_1<0<\lambda_2<\lambda_3$. The solution is represented by a triple (α,β,γ) i.e. (α,β,γ) corresponds with $q=\alpha R_1+\beta R_2+\gamma R_3$.

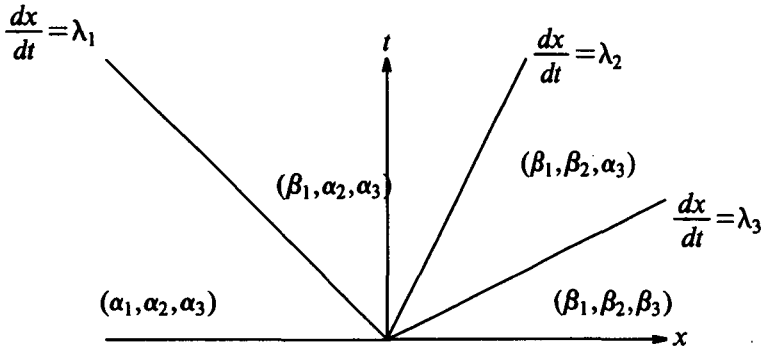


FIGURE 1.3.1a. Illustration of the solution of the Riemann problem for a linear hyperbolic system ($n=3$).

The solution of the Riemann-problem for a nonlinear hyperbolic system is hard to obtain in general. But for certain pairs (q_L,q_R) the solution of the Riemann problem may become simple. In the remainder of this section we show how to obtain these simple solutions. For this purpose we introduce the following concept:

DEFINITION 1.3.1c.
Consider the hyperbolic system (1.3.1.1). Let $R_k(q)$ be an eigenvector of $A(q)=\frac{df}{dq}(q)$ i.e. $A(q)R_k(q)=\lambda_k(q)R_k(q)$ with $\lambda_k(q)$ the corresponding eigenvalue.
We call $R_k(q)$ genuinely nonlinear if

$$(\nabla\lambda_k(q),\ R_k(q))\neq 0 \quad \forall q\in\mathbb{R}^n\ . \tag{1.3.1.8}$$

We call $R_k(q)$ linearly degenerate if

$$(\nabla\lambda_k(q),\ R_k(q))=0 \quad \forall q\in\mathbb{R}^n\ . \tag{1.3.1.9}$$

Here $(\ ,\ )$ denotes the usual inner product in $\mathbb{R}^n$ and

$\nabla\lambda_k(q)=(\frac{\partial\lambda_k}{\partial q_1},\cdots,\frac{\partial\lambda_k}{\partial q_n})^T$. We shall show in section 1.3.2 that for the Euler equations, each eigenvector $R_k(q)$ is either genuinely nonlinear or linearly degenerate.

To construct certain simple solutions of the Riemann problem, we shall show that a genuinely nonlinear eigenvector $R_k(q)$ corresponds with a so-called simple wave solution while a linearly degenerate eigenvector $R_k(q)$ corresponds with a contact discontinuity. (To avoid confusion, it should be mentioned that in this context a contact discontinuity differs from the concept of a contact discontinuity as introduced in section 1.2; here we are concerned with time dependent problems while in section 1.2 we were concerned with the time independent (steady) Euler equations).

Simple wave solution of the Riemann problem.

Suppose $R_k(q)$ is a genuinely nonlinear eigenvector. Then $R_k(q)$ can be normalized such that

$$(\nabla\lambda_k(q),\ R_k(q))=1 \quad \forall q\ . \tag{1.3.1.10}$$

For an arbitrary state q_L we consider the following ordinary differential equation

$$\begin{cases} \dfrac{dq}{d\xi}(\xi)=R_k(q(\xi)) \\ \\ q(0)=q_L \end{cases} \tag{1.3.1.11}$$

and suppose $q=\tilde{q}(\xi)$, $0\leqslant\xi\leqslant\xi_R$ is the solution. Define $q_R=\tilde{q}(\xi_R)$.
Because

$$\frac{d}{d\xi}\lambda_k(\tilde{q}(\xi))=\ \nabla\lambda_k(\tilde{q}(\xi))\ .\ R_k(\tilde{q}(\xi))=1$$

we have

$$\lambda_k(\tilde{q}(\xi))=\xi+\ \text{const}\ =\xi+\lambda_k(q_l)$$

and

$$\lambda_k(q_R)=\xi_R+\lambda_k(q_L)\ .$$

Notice that $\lambda_k(q_R)-\lambda_k(q_L)=\xi_R>0$.
Define

$$q(x,t)=\begin{cases} q_L & x/t<\lambda_k(q_L) \\ \tilde{q}(x/t-\lambda_k(q_L)) & \lambda_k(q_L)<x/t<\lambda_k(q_R) \\ q_R & x/t>\lambda_k(q_R) \end{cases} \tag{1.3.1.12}$$

We shall verify that $q(x,t)$ is the solution of the Riemann-problem. If

$\lambda_k(q_L)<x/t<\lambda_k(q_R)$ then

$$\lambda_k(q(x,t)) = \lambda_k(\tilde{q}(x/t-\lambda_k(q_L)))$$
$$= x/t-\lambda_k(q_L)+\lambda_k(q_L)=x/t\,. \qquad (1.3.1.13)$$

Hence, if $\lambda_k(q_L)<x/t<\lambda_k(q_R)$ then

$$\frac{\partial q}{\partial t}(x,t)+\frac{\partial}{\partial x}f(q(x,t))=\frac{\partial q}{\partial t}(x,t)+A(q(x,t))\frac{\partial q}{\partial x}(x,t,)$$
$$=-\frac{x}{t^2}R_k(q(x,t))+\frac{1}{t}A(q(x,t))R_k(q(x,t))$$
$$=-\frac{x}{t^2}R_k(q(x,t))+\frac{1}{t}\lambda_k(q(x,t))R_k(q(x,t))=0\,.$$

So $q(x,t)$ is indeed the solution of the Riemann problem with initial states (q_L,q_R). This solution is called a kth simple wave (or rarefaction wave). An illustation of this solution is given in fig. 1.3.1.6.

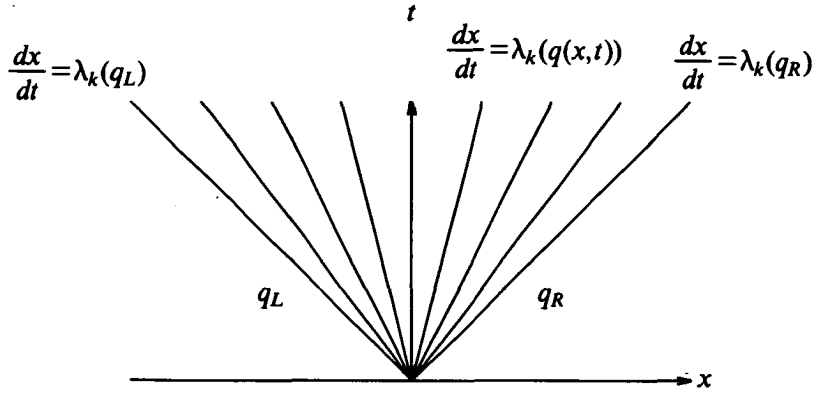


FIGURE 1.3.1b. Illustration of a kth simple wave solution of a Riemann inital value problem.

Contact discontinuity solution of the Riemann problem.
Suppose $R_k(q)$ is a linearly degenerate eigenvector. Hence

$$(\nabla\lambda_k(q),\ R_k(q))=0 \quad \forall q\in\mathbb{R}^n\,.$$

Let $\tilde{q}(\xi)$ be the solution of (1.3.1.11) and define $q_R=\tilde{q}(\xi_R)$. Because

$$\frac{d}{d\xi}(\lambda_k(\tilde{q}(\xi))=(\nabla\lambda_k(\tilde{q}(\xi)),\ R_k(\tilde{q}(\xi)))=0$$

we have $\lambda_k(\tilde{q}(\xi))=\lambda_k(q_L)=\lambda_k(q_R)\ \forall\xi\in(0,\xi_R)$.
Define

$$q(x,t)=\begin{cases}q_L & x/t<\lambda_k(q_L)=\lambda_k(q_R)\\ q_R & x/t>\lambda_k(q_L)=\lambda_k(q_R)\end{cases} \qquad (1.3.1.14)$$

Hence, $q(x,t)$ is discontinuous. We shall show that $q(x,t)$ is a solution of the Riemann problem. An illustration of this solution is given in fig. 1.3.1c.

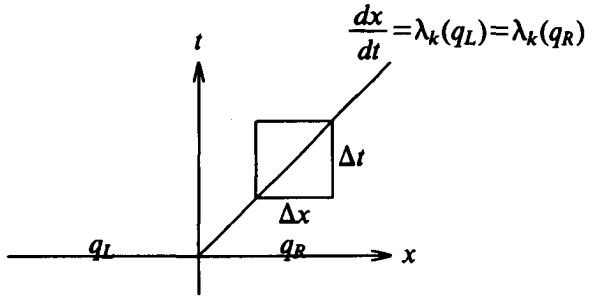


FIGURE 1.3.1c. Illustation of a kth contact discontinuity solution of a Riemann problem.

Because $q(x,t)$ is discontinuous, the integral form of (1.3.1.1) has to be employed

$$\int_{\partial\Omega}\left\{qn_t+f(q)n_x\right\}d\sigma=0 \tag{1.3.1.15}$$

with Ω an arbitrary volume in the (x,t) space and $n=(n_x,n_t)$ is the outward unit normal an $\partial\Omega$. It suffices to consider an infinitesimal rectangular volume with sides Δx and Δt straddling the discontinuity, cf. fig. 1.3.1c. Equation (1.3.1.15) results in

$$f(q_R)-f(q_L)+\lambda_k(q_L)\,(q_L-q_R)=0 \tag{1.3.1.16}$$

(notice that $\frac{\Delta x}{\Delta t}=\lambda_k(q_L)=\lambda_k(q_R)$). These are so called jump relations or Rankine-Hugoniot relations [3]. We will show that (1.3.1.16) is satisfied. Because

$$\begin{aligned}\frac{d}{d\xi}\left[f(\tilde{q}(\xi))-\lambda_k(q_L)\tilde{q}(\xi)\right] &= A(\tilde{q}(\xi))\frac{d\tilde{q}(\xi)}{d\xi}-\lambda_k(\tilde{q}(\xi))\frac{d\tilde{q}}{d\xi}(\xi)\\ &= A(\tilde{q}(\xi))R_k(\tilde{q}(\xi))-\lambda_k(\tilde{q}(\xi))R_k(\tilde{q}(\xi))\\ &= \lambda_k(\tilde{q}(\xi))R_k(\tilde{q}(\xi))-\lambda_k(\tilde{q}(\xi))R_k(\tilde{q}(\xi))=0\end{aligned}$$

we have

$$f(q_R)-\lambda_k(q_L)q_R=f(q_L)-\lambda_k(q_L)q_L\,.$$

Thus the jump relations (1.3.1.16) are indeed satisfied.

In addition to simple waves and contact discontinuities there exists another elementary type of solutions of the Riemann problem, namely shock waves. Shock wave solutions satisfy the Rankine-Hugoniot relations and the entropy condition. We refer to [3,7] for a detailed description of shock wave solutions. The general solution of a Riemann problem is, under rather general circumstances, composed by simple waves, contact discontinuities and shock waves (see [7]). Here, we can omit a detailed description of shock wave solutions because this would not contribute very much to the understanding of the numerical solution methods that will be discussed.
In this work we shall use an upwind scheme proposed by Osher [5] that is based on an approximate solution of the Riemann problem, obtained by replacing shock waves by compression waves. A compression wave is the reverse of a rarefaction wave and leads to a multi-valued solution. For more details we refer to the next section and chapter 2.
Finally, we introduce the concept of Riemann-invariants.

DEFINITION 1.3.1d.
Consider the hyperbolic system (1.3.1.1). Let $R_k(q)$ be the kth eigenvector of $A(q)=\frac{df}{dq}(q)$. A k-Riemann invariant is a smooth function $\psi_k:\mathbb{R}^n\mapsto\mathbb{R}$ such that

$$(\nabla\psi_k(q),\ R_k(q))=0\quad\forall q\in\mathbb{R}^n\ .$$

Notice that if $R_k(q)$ is linearly degenerate, the corresponding eigenvalue $\lambda_k(q)$ is a Riemann invariant (see (1.3.1.9)). In general there are $n-1$ k-Riemann invariants whose gradients are linearly independent in $\mathbb{R}^n$. Riemann-invariants are useful for the construction of simple wave and contact discontinuity solutions of Riemann problems. For the construction of simple waves or contact discontinuities we have to solve (see (1.3.1.11))

$$\begin{cases}\dfrac{dq}{d\xi}(\xi)=R_k(q(\xi))\\[2ex] q(0)=q_L\end{cases}\tag{1.3.1.17}$$

Suppose $q=\tilde{q}(\xi)$, $0\leqslant\xi\leqslant\xi_R$ is the solution. Then

$$\frac{d}{d\xi}\psi_k(\tilde{q}(\xi))=(\nabla\psi_k(\tilde{q}(\xi)),\ R_k(\tilde{q}(\xi)))=0$$

hence a k-Riemann invariant is constant along the curve described by 1.3.1.17. If there are $n-1$ k-Riemann invariants $\psi_k^1,\ldots,\psi_k^{n-1}$, then it is easily seen that the curve described by (1.3.1.17) is part of the curve described by

$$\left\{q\in\mathbb{R}^n\mid\psi_k^1(q)=\psi_k^1(q_L),\ldots,\psi_k^{n-1}(q)=\psi_k^{n-1}(q_L)\right\}.$$

In the case of the Euler equations, the Riemann invariants are very simple functions and, as we shall see in the next section, formulas describing simple waves or contact discontinuities are easily obtained.

1.3.2. Application to the Euler equations.

The Euler equations (1.1.10) can be written as

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)+\frac{\partial}{\partial y}g(q)=0 \tag{1.3.2.1}$$

where

$$\begin{aligned} q &= (\rho,\rho u,\rho v,E)^T=(q_1,q_2,q_3,q_4)^T \\ f(q) &= (\rho u,\rho u^2+p,\rho uv,(E+p)u)^T \\ &= (q_2,q_2^2/q_1+p,q_2q_3/q_1,(q_4+p)q_2/q_1)^T \\ g(q) &= (\rho v,\rho uv,\rho v^2+p,(E+p)v)^T \\ &= (q_3,q_2q_3/q_1,q_3^2/q_1+p,(q_4+p)q_3/q_1)^T \end{aligned} \tag{1.3.2.2}$$

and (see 1.1.13)

$$p=(\gamma-1)\,(E-\tfrac{1}{2}\rho(u^2+v^2)=(\gamma-1)\,(q_4-(q_2^2+q_3^2)/(2q_1))\,. \tag{1.3.2.3}$$

First, we notice the rotational invariance of the Euler equations.

Theorem (1.3.2a).
The Euler equations are rotationally invariant i.e.

$$\cos\phi\, f(q)+\sin\phi\, g(q)=T(\phi)^{-1}f(T(\phi)q) \tag{1.3.2.4}$$

for all $\phi\in\mathbb{R}$ and admissible states $q\in\mathbb{R}^4$, where $T(\phi)$ is the following rotation matrix:

$$T(\phi)=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1.3.2.5}$$

Proof.
The calculations to verify (1.3.2.4) are straightforward. □

In section 1.3.1 we have introduced the concept of hyperbolicity for a first-order system of quassi linear equations in only two independent variables. Similarly, we define:

Definition (1.3.2a).
System (1.3.2.1) is called hyperbolic (with respect to t) if there exist, for all

$\phi \in \mathbb{R}$ and admissible states $q \in \mathbb{R}^4$, a real diagonal matrix $D(q,\phi)$ and a non-singular matrix $R(q,\phi)$ such that

$$A(q,\phi)R(q,\phi) = R(q,\phi)D(q,\phi) \tag{1.3.2.6}$$

where

$$A(q,\phi) = \cos\phi \frac{df}{dq}(q) + \sin\phi \frac{dg}{dq}(q) \equiv \cos\phi A(q) + \sin\phi B(q) . \tag{1.3.2.7}$$

LEMMA (1.3.2a).
If there exists a real diagonal matrix $D(q)$ and a non-singular real matrix $R(q)$ such that

$$A(q)R(q) = R(q)D(q) \tag{1.3.2.8}$$

for all admissable states $q \in \mathbb{R}^4$, then the Euler equations (1.3.2.1) are hyperbolic.

PROOF.
By differentiating the rotational invariance relation (1.3.2.4) with respect to q, we have

$$A(q,\phi) = \cos\phi A(q) + \sin\phi B(q) = T(\phi)^{-1} A(T(\phi)q) T(\phi) .$$

Hence,

$$\begin{aligned} A(q,\phi) &= T(\phi)^{-1} R(T(\phi)q) D(T(\phi)q) R^{-1}(T(\phi)q) T(\phi) \\ &= (T(\phi)^{-1} R(T(\phi)q)) D(T(\phi)q) \, (T(\phi)^{-1} R(T(\phi)q))^{-1} . \end{aligned}$$

By taking

$$R(q,\phi) = T(\phi)^{-1} R(T(\phi)q), \; D(q,\phi) = D(T(\phi)q)$$

we see that (1.3.2.6) is valid. □

Thus, the Euler equations are hyperbolic if the matrix $A(q) = \frac{df}{dq}(q)$ has 4 linearly independent eigenvectors. Using the relation

$$\nabla p = \left[\frac{\partial p}{\partial q_1}, \frac{\partial p}{\partial q_2}, \frac{\partial p}{\partial q_3}, \frac{\partial p}{\partial q_4} \right]^T = (\gamma - 1)\left(\tfrac{1}{2}(u^2+v^2), -u, -v, 1\right) \tag{1.3.2.9}$$

and the relations for total enthalpy H:

$$H = \frac{E+p}{\rho} = \frac{c^2}{\gamma - 1} + \tfrac{1}{2}(u^2+v^2) \tag{1.3.2.10}$$

it is easily verified that

$$A(q) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -u^2 + \frac{1}{2}(\gamma-1)(u^2+v^2) & (3-\gamma)u & -(\gamma-1)v & \gamma - 1 \\ -uv & v & u & 0 \\ u(\frac{1}{2}(\gamma-1)(u^2+v^2) - H) & H - (\gamma-1)u^2 & -(\gamma-1)uv & \gamma u \end{bmatrix} \tag{1.3.2.11}$$

Notice that $A(q)$ only depends on u,v and c. The eigenvalues and corresponding eigenvectors of $A(q)$ are

$$\lambda_1(q)=u-c,\ \lambda_2(q)=u,\ \lambda_3(q)=u,\ \lambda_4(q)=u+c \tag{1.3.2.12}$$

and

$$\begin{aligned} R_1(q)&=(1,\ u-c,\ v,\ H-cu)^T\\ R_2(q)&=(1,\ u,\ v,\ \tfrac{1}{2}(u^2+v^2))^T\\ R_3(q)&=(0,\ 0,\ 1,\ v)^T\\ R_4(q)&=(1,\ u+c,\ v,\ H+cu)^T \end{aligned} \tag{1.3.2.13}$$

The eigenvectors are linearly independent and therefore we have found the following theorem:

THEOREM (1.3.2b).
The Euler equations (1.3.2.1) are hyperbolic with respect to t.

Now, we shall consider the Riemann problem for the Euler equations in one space dimension:

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)=0 \tag{1.3.2.14a}$$

with

$$q(x,0)=\begin{cases} q_L & x<0\\ q_R & x>0 \end{cases} \tag{1.3.2.14b}$$

where q_L and q_R are constant states, and q, $f(q)$ are defined in (1.3.2.2).
We assume that there is a unique solution. As we have already seen $q(x,t)=\tilde{q}(x/t)$.
Although trivial, it is worth noticing that $q(x,y,t)=\tilde{q}(x/t)$ also obeys the Euler equations in two dimensions

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)+\frac{\partial}{\partial y}g(q)=0 \tag{1.3.2.15a}$$

with initial values given by

$$q(x,y,0)=\begin{cases} q_L & x<0,\ y\in\mathbb{R}\\ q_R & x>0,\ y\in\mathbb{R} \end{cases}. \tag{1.3.2.15b}$$

We shall seek certain pairs (q_L,q_R) for which the solution of the Riemann-problem (1.3.2.14) becomes very simple, e.g. a simple wave or contact discontinuity. Therefore, we shall first investigate whether the eigenvectors $R_k(q)$, $k=1,2,3,4$, given in (1.3.2.13) are genuinely nonlinear or linearly degenerate.

THEOREM (1.3.2c).
The eigenvectors $R_k(q)$ given in (1.3.2.13) are genuinely nonlinear for $k=1$ and 4, and linearly degenerate for $k=2$ and 3.

PROOF.
Because

$$\nabla\lambda_2(q)=\nabla\lambda_3(q)=\nabla u=\frac{1}{\rho}(-u,1,0,0) \tag{1.3.2.16}$$

it is immediately clear that $R_2(q)$ and $R_3(q)$ are linearly degenerate. Using definition (1.2.22) of the speed of sound c, we see that

$$\nabla c=\frac{1}{2\rho c}\{\gamma\nabla p-c^2\nabla\rho\} \tag{1.3.2.17}$$

where $\nabla\rho=(1,0,0,0)^T$ and ∇p is given by (1.3.2.9).
Thus

$$\begin{aligned}(\nabla\lambda_1(q),R_1(q)) &= (\nabla u,R_1(q))-\frac{\gamma}{2\rho c}(\nabla p,R_1(q))+\frac{c}{2\rho}(\nabla\rho,R_1(q))\\ &= -\frac{c}{2\rho}(1+\gamma)\neq 0\end{aligned}$$

and

$$\begin{aligned}(\nabla\lambda_4(q),R_4(q)) &= (\nabla u,R_4(q))+\frac{\gamma}{2\rho c}(\nabla p,R_4(q))-\frac{c}{2\rho}(\nabla\rho,R_4(q))\\ &= +\frac{c}{2\rho}(1+\gamma)\neq 0\,.\end{aligned}$$

Hence, $R_1(q)$ and $R_4(q)$ are genuinely nonlinear □ .

Thus $R_1(q)$ and $R_4(q)$ correspond with simple waves while $R_2(q)$ and $R_3(q)$ correspond with contact discontinuities.
Riemann invariants are very useful for the construction of a simple wave solution or a contact discontinuity solution of the Riemann problem (1.3.2.14). In the following theorem, the Riemann invariants corresponding with the eigenvectors $R_k(q)$ are given.

THEOREM (1.3.2d).
The functions

$$\alpha_1(q)=u+\frac{2}{\gamma-1}c,\ \alpha_2(q)=v,\ \alpha_3(q)=s \tag{1.3.2.18}$$

are Riemann-invariants corresponding with the eigenvector $R_1(q)$.
The functions

$$\beta_1(q)=u,\ \beta_2(q)=p \tag{1.3.2.19}$$

are Riemann-invariants corresponding with the eigenvectors $R_2(q)$ and $R_3(q)$.

The functions

$$\gamma_1(q)=u-\frac{2}{\gamma-1}c,\ \gamma_2(q)=v,\ \gamma_3(q)=s \qquad (1.3.2.20)$$

are Riemann-invariants corresponding with the eigenvector $R_4(q)$.

PROOF.
From (1.2.18) we deduce that

$$\nabla s=\frac{c_v}{p}(\nabla p-c^2\nabla\rho)\,. \qquad (1.3.2.21)$$

Furthermore

$$\nabla v=\frac{1}{\rho}(-v,0,1,0)\,. \qquad (1.3.2.22)$$

With these expressions and with the expressions for $\nabla p, \nabla u$ and ∇c (see 1.3.2.9, 16 and 17) the calculations to verify this theorem become straightforward. □

Thus, if the pairs (q_L,q_R) of the Riemann problem (1.3.2.14) are such that

$$u_L+\frac{2}{\gamma-1}c_L=u_R+\frac{2}{\gamma-1}c_R\ ;\ v_R=v_L\ ;\ s_R=s_L \qquad (1.3.2.23)$$

and

$$u_L-c_L<u_R-c_R \qquad (1.3.2.24)$$

then a simple wave solution, corresponding with $R_1(q)$, exists, given by

$$\begin{aligned}
q &= q_L && \text{if } x/t<u_L-c_L\\
u+\frac{2}{\gamma-1}c &= u_L+\frac{2}{\gamma-1}c_L\\
v &= v_L && \text{if } u_L-c_L<x/t<u_R-c_R\\
s &= s_L\\
u-c &= x/t\\
q &= q_R && \text{if } x/t>u_R-c_R
\end{aligned} \qquad (1.3.2.25)$$

Notice that $u-c=x/t$ follows from (1.3.1.13). This solution is also called a 1-rarefaction wave. If (q_L,q_R) are such that (1.3.2.23) holds, while $u_L-c_L>u_R-c_R$, the solution given by (1.3.2.25) corresponds with a multi-valued solution. Then we speak of a compression wave. Allthough such a compression wave has no physical meaning, it will be shown in chapter II that allowing compression waves, an approximate solution of the Riemannn problem can be obtained which leads to an excellent upwind scheme for the Euler equations. This scheme was introduced by Osher in 1982 [5].

Similarly, if q_L and q_R are such that

$$u_L - \frac{2}{\gamma-1}c_L = u_R - \frac{2}{\gamma-1}c_R,\ v_L = v_R,\ s_R = s_L \tag{1.3.2.26}$$

and

$$u_L + c_L < u_R + c_R \tag{1.3.2.27}$$

then a simple wave solution, corresponding with $R_4(q)$ exists, given by

$$\begin{aligned} q &= q_L && \text{if } x/t < u_L + c_L \\ u - \frac{2}{\gamma-1}c &= u_L - \frac{2}{\gamma-1}c_L \\ v &= v_L \\ s &= s_L && \text{if } u_L + c_L < x/t < u_R + c_R \\ u + c &= x/t \\ q &= q_R && \text{if } x/t > u_R + c_R \end{aligned} \tag{1.3.2.28}$$

This solution is also called a 4-rarefaction wave. If $u_L + c_L > u_R + c_R$, we have a multivalued 4-compression wave.
Finally, if q_L and q_R are such that $u_L = u_R$, $p_L = p_R$ then a contact discontinuity solution, corresponding with $R_2(q)$, $R_3(q)$ exists, given by

$$\begin{aligned} q &= q_L \quad \text{if } x/t < u_L = u_R \\ q &= q_R \quad \text{if } x/t > u_L = u_R \end{aligned} . \tag{1.3.2.29}$$

1.4. Simplifications of the Euler equations: the TFP and TSP equations.

I. The Transonic Full Potential (TFP) equation.
The TFP equation is derived from the Euler equations by assuming that the flow is steady, isenthalpic and irrotational. Before deriving the TFP equation we shall show that a consequence of these assumptions is that the flow must also be isentropic. This follows from the Crocco-theorem, which is derived in the following way.
Choose the primitive variables u,v,p and ρ as dependent variables in the Euler equations. Then the entropy s and the internal energy e are functions of p and ρ. Assuming an ideal gas, these functions are given by:

$$s = s(p,\rho) = c_v \ln \frac{p}{\rho^\gamma}$$

$$e = e(p,\rho) = \frac{1}{\gamma-1}\frac{p}{\rho} .$$

After some algebraic manipulations, it is easily seen that

$$de=\frac{p}{\rho^2}d\rho+\frac{1}{\gamma-1}\cdot\frac{1}{\rho}\{dp-c^2d\rho\}$$

$$Tds=\frac{1}{\gamma-1}\cdot\frac{1}{\rho}\{dp-c^2d\rho\}$$

hence

$$Tds=de-\frac{p}{\rho^2}d\rho$$

which implies

$$T\nabla s=\nabla e-\frac{p}{\rho^2}\nabla\rho\,.$$

From the definition of total enthalpy H (see (1.2.1), (1.2.2)), it follows that

$$\begin{aligned}\nabla H &= \nabla e+\frac{1}{\rho}\nabla p-\frac{p}{\rho^2}\nabla\rho+\tfrac{1}{2}\nabla(u^2+v^2)\\ &= T\nabla s+\frac{1}{\rho}\nabla p+\tfrac{1}{2}\nabla(u^2+v^2)\,.\end{aligned}$$

Using equation (1.2.10), we find

$$T\nabla s=\nabla H+\frac{D}{Dt}v-\tfrac{1}{2}\nabla(u^2+v^2)$$

thus

$$\begin{aligned}T\frac{\partial s}{\partial x}&=\frac{\partial H}{\partial x}+\frac{\partial u}{\partial t}-v\zeta\\ T\frac{\partial s}{\partial y}&=\frac{\partial H}{\partial y}+\frac{\partial v}{\partial t}+u\zeta\end{aligned}\qquad(1.4.1)$$

with ζ the vorticity $\zeta=-\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}$.

This equation is known as the Crocco theorem and it tells us that, in case of a steady isenthalpic flow, the assumption that the flow is isentropic is equivalent with the assumption that the flow is irrotational. Equation (1.4.1) is only applicable in smooth flow field regions. In shocks the Crocco theorem does not hold. But we know that in fluid particles passing through a shock the entropy increases, depending on the strength of the shock. Hence, eq. (1.4.1) (which holds again behind the shock) tells us that unless the shock strength is uniform, the vorticity behind the shock will not be zero. Hence, the irrotational flow assumption is not compatible with the appearance of shocks. However, if the component normal to the shock of the upstream Mach number M_n is sufficiently close to unity, the flow is almost isentropic; the entropy variation is of the order of $(M_n^2-1)^3$ (see (1.2.31) and (1.2.32)). Thus in case of transonic flows with weak shocks, the irrotational flow assumption is a good approximation.

Derivation of the TFP equation.
We assume that the flow is steady, isenthalpic, irrotational and thus isentropic. A potential $\phi=\phi(x,y)$ can be introduced:

$$u=\frac{\partial\phi}{\partial x}\ ;\ v=\frac{\partial\phi}{\partial y}\ . \tag{1.4.2}$$

The continuity equation becomes

$$\frac{\partial}{\partial x}(\rho\frac{\partial\phi}{\partial x})+\frac{\partial}{\partial y}(\rho\frac{\partial\phi}{\partial y})=0\ . \tag{1.4.3}$$

Because the flow is isenthalpic and isentropic, we have

$$\frac{c^2}{\gamma-1}+\frac{1}{2}(u^2+v^2)=\frac{c_\infty^2}{\gamma-1}+\frac{1}{2}u_\infty^2 \tag{1.4.4}$$

$$\frac{p}{\rho^\gamma}=\frac{p_\infty}{p_\infty^\gamma}\ . \tag{1.4.5}$$

(With ∞ we denote a freestream value at infinity). Using $c^2=\gamma p/\rho$, we can eliminate p and derive the following expression for ρ:

$$\rho=\rho_\infty\{1+\frac{\gamma-1}{2}M_\infty^2(1-\frac{u^2+v^2}{u_\infty^2})\}^{\frac{1}{\gamma-1}} \tag{1.4.6}$$

where $M_\infty=u_\infty/c_\infty$ is the Mach number at infinity. Equation (1.4.3) with ρ given by (1.4.6) is the TFP equation in conservative form. The TFP equation in non-conserative form is also well known and will be derived for completeness. Define

$$q^2=u^2+v^2=(\frac{\partial\phi}{\partial x})^2+(\frac{\partial\phi}{\partial y})^2 \tag{1.4.7}$$

then

$$\frac{c^2}{c_\infty^2}=1+\frac{\gamma-1}{2}M_\infty^2(1-(\frac{q}{u_\infty})^2) \tag{1.4.8}$$

and

$$\rho=\rho_\infty\{\frac{c^2}{c_\infty^2}\}^{\frac{1}{\gamma-1}}\ . \tag{1.4.9}$$

With these expressions, it is easily seen from (1.4.6) that

$$d\rho=-\frac{\rho}{2c^2}dq^2\ . \tag{1.4.10}$$

Using (1.4.3) we have

$$\rho\Delta\phi+\nabla\rho\ .\ \nabla\phi=0$$

thus

$$\Delta\phi+\frac{1}{\rho}\frac{d\rho}{dq^2}\nabla q^2\ .\ \nabla\phi=0$$

or

$$\Delta\phi - \frac{1}{2c^2}\nabla q^2 \,.\, \nabla\phi = 0$$

i.e.

$$(1-\frac{u^2}{c^2})\phi_{xx} + (1-\frac{v^2}{c^2})\phi_{yy} - \frac{2uv}{c^2}\phi_{xy} = 0\,. \tag{1.4.11}$$

This equation with c^2 given by (1.4.8) is the TFP equation in nonconservative form. The TFP equation is a second-order non-linear partial differential equation of mixed elliptic-hyperbolic type.
Besides the wrong modelling of strong shocks, another disadvantage of the TFP equation is that contact discontinuities or slip lines cannot be modelled. It can be easily seen that no contact discontinuity can appear in an isenthalpic and isentropic flow.
Examples of discrepancies between potential flow solutions and solutions of the Euler equations can be found in [1,6]. Even at quite moderate Mach numbers, such as the NACA0012 airfoil at Mach 0.8 and an angle of attack of 1.25° large discrepancies were observed.

II. The Transonic Small Perturbation (TSP) equation.
The TSP equation is a simplification of the TFP equation and is therefore even more restrictive for general applications. The TSP equation is derived in the following way. Write the TFP equation as follows

$$\begin{aligned} &(c^2-u^2)u_x + (c^2-v^2)u_y - uv(u_y+v_x) = 0 \\ &c^2 = c_\infty^2 + \frac{\gamma-1}{2}(u_\infty^2 - (u^2+v^2)) \end{aligned} \tag{1.4.12}$$

Define

$$\begin{aligned} u &:= u_\infty(1+\tilde{u}) \\ v &= u_\infty \tilde{v} \end{aligned} \tag{1.4.13}$$

$\tilde{u},\tilde{v}$ are called "perturbation" velocity components. We assume that $|\tilde{u}|, |\tilde{v}| \ll 1$. Substituting (1.4.13) in (1.4.12) and neglecting terms containing squares of the perturbation velocities, in comparison to those containing first powers, we obtain the simpler equation

$$(1-M_\infty^2)\tilde{u}_x + \tilde{v}_y = (\gamma+1)M_\infty^2\tilde{u}\tilde{u}_x + (\gamma-1)M_\infty^2\tilde{u}\tilde{v}_y + M_\infty^2\tilde{v}(\tilde{u}_y+\tilde{v}_x) \tag{1.4.14}$$

where $M_\infty = u_\infty/c_\infty$, $\tilde{u}_x = \dfrac{\partial\tilde{u}}{\partial x}$ etc. The TSP equation is obtained from (1.4.14) by neglecting the last two terms in the right-hand side. The first term of the right-hand side of (1.4.14) cannot be neglected in general. For instance, in transonic flow, where $M_\infty \mapsto 1$, the coefficient of $\tilde{u}_x$, on the left-hand side, becomes very small. Then it is not possible to neglect the first term on the right-hand side of (1.4.14). Thus the TSP-equation becomes ($\tilde{u} = \phi_x, \tilde{v} = \phi_q$):

$$(1-M_\infty^2)\phi_{xx}+\phi_{yy}=(\gamma+1)M_\infty^2\,\phi_x\phi_{xx} \tag{1.4.15}$$

or

$$[(1-M_\infty^2)\phi_x-\tfrac{1}{2}(\gamma+1)M_\infty^2\,\phi_x^2]_x+[\phi_y]_y=0\,. \tag{1.4.16}$$

Finally, if the term with ϕ_x^2 is neglected in the TSP equation (1.4.16), we obtain the linear equation

$$(1-M_\infty^2)\phi_{xx}+\phi_{yy}=0\,. \tag{1.4.17}$$

This equation furnishes a useful approximation only for flows in which M is not close to 1 (subsonic or supersonic flows).

REFERENCES.

[1] T.J. BAKER (1983). *The Computation of Transonic Potential Flow.* In: *Computational Methods for Turbulent, Transonic and Viscous Flows,* J.A. ESSERS (ed.), Hemisphere Publishing Corporation, New York.

[2] S.K. GODUNOV (1959). *Finite Difference Method for Numerical Computation of Discontinuous Solutions of the Equations of Fluid Dynamics,* Math. Sbornik 47, 271-306; also: Cornell Aeronautical Lab., (Calspan Translation).

[3] P.D. LAX (1973). *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves,* Regional Conference Series in Applied Mathematics II, Siam, Philadelphia.

[4] H.W. LIEPMANN and A. ROSHKO (1966). *Elements of Gasdynamics,* Wiley, New York.

[5] S. OSHER and F. SOLOMON (1982). *Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws,* Math. Comp. 38, 339-374.

[6] A. RIZZI and H. VIVIAND (eds.) (1981). *Numerical Mehods for the Computation of Inviscid Transonic Flows with Shock Waves.* Notes on Numerical Fluid Mechanics, Vol. 3., Vieweg Verlag, Braunschweig.

[7] J. SMOLLER (1983). *Shock Waves and Reaction-Diffusion Equations.* Grundlehren der Mathematischen Wissenschaften 258, Springer Verlag, Berlin.

Chapter II

Finite-Volume Upwind Discretization of the Steady Euler Equations

2.1. Introduction

The subject of this work is the numerical solution of the steady Euler equations in 2D. The numerical solution of the steady Euler equations consists of two separate parts; the discretization and the solution of the system of discretized equations. In this chapter we consider the discretization while a solution method for the system of discretized equations is developed in the next two chapters.

We have seen in chapter I that in general solutions of the steady Euler equations contain discontinuities (shock waves, contact discontinuities). At discontinuities the differential form of the Euler equations is not valid. On the other hand, the integral form is valid both at discontinuities and in the smooth part of the flow field. This observation suggests that it will be better to base the discretization on the integral form instead of the differential form. Let $\Omega \subset \mathbb{R}^2$ be the physical domain in which we wish to solve the steady Euler equations numerically. The differential form gives relations at each point $(x,y) \in \Omega$, while the integral form gives relations for each control volume $\Omega^* \subset \Omega$. The Euler equations in integral form are

$$\frac{d}{dt}\int_{\Omega^*} q\,dv + \int_{\partial\Omega^*} \{\cos\phi f(q) + \sin\phi g(q)\}\,d\sigma = 0 \quad \forall \Omega^* \subset \Omega \tag{2.1.1a}$$

where

$$\begin{aligned} q &= (\rho, \rho u, \rho v, E)^T \\ f(q) &= (\rho u, \rho u^2 + p, \rho uv, (E+p)u)^T \\ g(q) &= (\rho v, \rho uv, \rho v^2 + p, (E+p)v)^T \end{aligned} \tag{2.1.1b}$$

and Ω^* is an arbitrary simply connected region in Ω, $(\cos\phi, \sin\phi) = n$ is the outward unit normal on the boundary $\partial\Omega^*$. Equation (2.1.1) is a direct consequence of equations (1.1.1), (1.1.3) and (1.1.8) derived in chapter I. Using the rotational invariance of the Euler equations (see (1.3.2.4)) equation (2.1.1a) is found to be equivalent with

$$\frac{d}{dt}\int_{\Omega^*} q\,dv + \int_{\partial\Omega^*} T(\phi)^{-1} f(T(\phi)q)\,d\sigma = 0 \quad \forall \Omega^* \subset \Omega \tag{2.1.2}$$

where $T(\phi)$ is the rotation matrix

$$T(\phi)=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \cos\phi & \sin\phi & 0\\ 0 & -\sin\phi & \cos\phi & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \tag{2.1.3}$$

To discretize the integral form, we subdivide the domain Ω into a finite number of disjunct control volumes (or finite volumes). Just for practical reasons (namely simple implementation) we will use only finite volumes that are quadrilateral, and use only structured grids. A structured grid is characterized by the fact that each interior finite volume has a common boundary with precisely four neighbours.

Different possibilities exist for the shape of the finite volumes. Triangular volumes are a reasonable choice as well. The use of quadrilateral finite volumes has the advantage that, on a smooth 2D grid, discretizations of the Euler equations in 1D can be generalized in a straightforward manner. As a consequence of the choice of a structured grid with finite volumes we can order the finite volumes such that the neighbouring volumes of $\Omega_{i,j}$ are $\Omega_{i+1,j}$, $\Omega_{i,j+1}$, $\Omega_{i-1,j}$ and $\Omega_{i,j-1}$. An example of a subdivision of a physical domain Ω in disjunct quadrilateral finite volumes is given in fig. 2.1a where Ω is a windtunnel section [12].

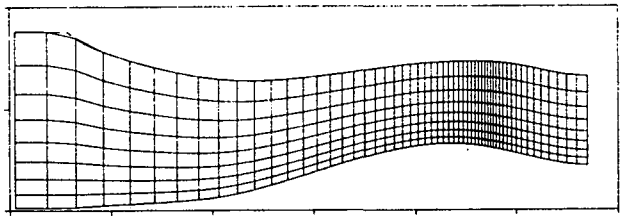


FIGURE 2.1a. Subdivision of a windtunnel section in disjunct quadrilateral finite volumes.

Once the domain Ω has been subdivided, we approximate the integral form in each volume $\Omega_{i,j}$. Assume that at time t, the mean values of a solution $q=q(x,y,t)$ are known in each control volume i.e. the set $\{q_{i,j}(t)\}$ is known where

$$q_{i,j}(t)=\frac{1}{V_{i,j}}\int_{\Omega_{i,j}} q(x,y,t)dv \tag{2.1.4}$$

here $V_{i,j}$ is the area of $\Omega_{i,j}$. From equation (2.1.2) we see that

$$V_{i,j}\frac{d}{dt}q_{i,j}(t)+\int_{\partial\Omega_{i,j}} T(\phi)^{-1}f(T(\phi)q(x,y,t))d\sigma=0 \quad \forall(i,j)\,. \tag{2.1.5}$$

Hence, the space discretization is determined by the way the total flux

$$\int_{\partial\Omega_{i,j}} T(\phi)^{-1}f(T(\phi)q(x,y,t))d\sigma$$

is approximated. Due to the fact that the control volumes have a quadrilateral shape, the total flux consists of four parts:

$$\int_{\partial\Omega_{i,j}} T(\phi)^{-1} f(T(\phi)q(x,y,t))d\sigma =$$
$$= \int_{\partial\Omega_{i+\frac{1}{2},j}} T(\phi_{i+\frac{1}{2},j})^{-1} f(T(\phi_{i+\frac{1}{2},j})q(x,y,t))d\sigma +$$
$$+ \int_{\partial\Omega_{i,j+\frac{1}{2}}} T(\phi_{i,j+\frac{1}{2}})^{-1} f(T(\phi_{i,j+\frac{1}{2}})q(x,y,t))d\sigma +$$
$$- \int_{\partial\Omega_{i-\frac{1}{2},j}} T(\phi_{i-\frac{1}{2},j})^{-1} f(T(\phi_{i-\frac{1}{2},j})q(x,y,t))d\sigma +$$
$$- \int_{\partial\Omega_{i,j-\frac{1}{2}}} T(\phi_{i,j-\frac{1}{2}})^{-1} f(T(\phi_{i,j-\frac{1}{2}})q(x,y,t))d\sigma \qquad (2.1.6)$$

where $\partial\Omega_{i+\frac{1}{2},j} = \partial\Omega_{i,j} \cap \partial\Omega_{i+i,j}$ and $\phi_{i+\frac{1}{2},j}$ is the angle between the outward unit normal on the boundary $\partial\Omega_{i+\frac{1}{2},j}$ and the x-axis (see fig. 2.1b), and similarly for the other three boundaries.

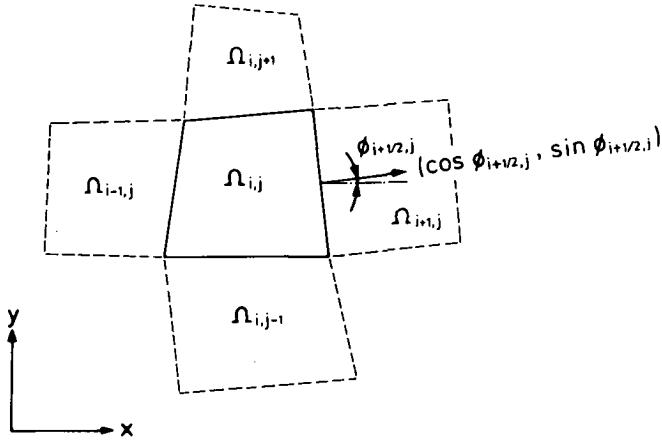


FIGURE 2.1b. Geometry of a control volume $\Omega_{i,j}$.

A very simple way to approximate the flux through $\partial\Omega_{i+\frac{1}{2},j}$ is

$$\int_{\partial\Omega_{i+\frac{1}{2},j}} T(\phi_{i+\frac{1}{2},j})^{-1} f(T(\phi_{i+\frac{1}{2},j})q(x,y,t))d\sigma \approx$$
$$l_{i+\frac{1}{2},j} T(\phi_{i+\frac{1}{2},j})^{-1} f(T(\phi_{i+\frac{1}{2},j}) \cdot \tfrac{1}{2}(q_{i,j}(t) + q_{i+1,j}(t))) \qquad (2.1.7)$$

where $l_{i+\frac{1}{2},j}$ is the length of boundary $\partial\Omega_{i+\frac{1}{2},j}$. Formula (2.1.7) leads to a central difference scheme on a Cartesian grid and is second-order accurate if the mesh is sufficiently smooth. This scheme is not resistent to high frequency oscillations between odd and even mesh points and dissipative terms must be added to suppress spurious oscillations (wiggles) of this type. Moreover, dissipative terms are also necessary to prevent wiggles in the neighbourhood of

shock waves. The numerical solution of the (steady) Euler equations by a central difference scheme with additional dissipative terms is advocated by Jameson [9,10]. An important drawback of Jameson's scheme is that the dissipative terms must be tuned, i.e. the amount of dissipation (or artificial viscosity) depends on the problem considered. But certainly, at the moment, Jameson's scheme is the most widely used scheme for solving practical aerodynamical problems. Another approach, which becomes more and more popular, is given by upwind schemes. Upwind schemes are based on the Riemann-problem and can be interpreted in the following way. First, assume that each state $q_{i,j}(t)$ is constant in $\Omega_{i,j}$. Then, at the boundary $\partial\Omega_{i+\frac{1}{2},j}$, the states $q_{i,j}(t)$ and $q_{i+1,j}(t)$ meet in a discontinuity. Fix the states $q_{i,j}(t)$ and $q_{i+1,j}(t)$ at time t: $q_{i,j}=q_{i,j}(t)$ and $q_{i+1,j}=q_{i+1,j}(t)$. Notice that

$$T(\phi_{i+\frac{1}{2},j})q_{i,j}\equiv\begin{bmatrix}\rho\\ \rho\tilde{u}\\ \rho\tilde{v}\\ E\end{bmatrix}_{i,j}=\tilde{q}_{i,j}\ ;\ T(\phi_{i+\frac{1}{2},j})q_{i+1,j}\equiv\begin{bmatrix}\rho\\ \rho\tilde{u}\\ \rho\tilde{v}\\ E\end{bmatrix}_{i+1,j}=\tilde{q}_{i+1,j} \tag{2.1.8}$$

where $\tilde{u}$ denotes the velocity component normal to $\partial\Omega_{i+\frac{1}{2},j}$ and $\tilde{v}$ denotes the velocity component tangential to $\partial\Omega_{i+\frac{1}{2},j}$ (see fig. 2.1c).

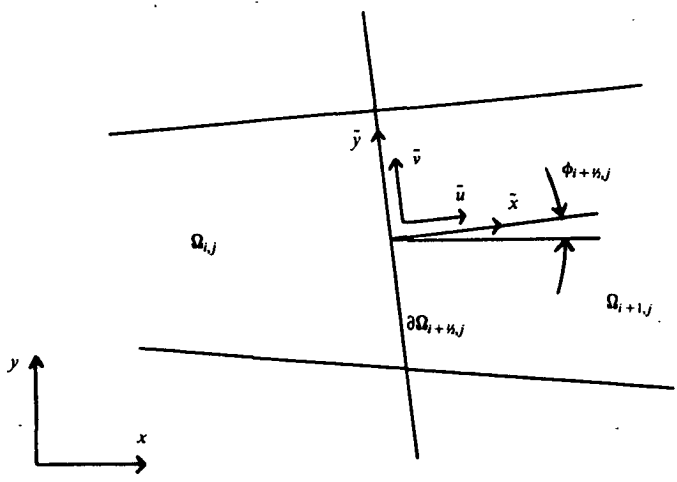


FIGURE 2.1c. The boundary $\partial\Omega_{i+\frac{1}{2},j}$ with local cartesian frame $(\tilde{x},\tilde{y})$.

With respect to a new Cartesian frame $(\tilde{x},\tilde{y})$ (see fig 2.1c) we consider the following Riemann-problem:

$$\begin{cases}\dfrac{\partial}{\partial t}\tilde{q}+\dfrac{\partial}{\partial\tilde{x}}f(\tilde{q})=0\\ \tilde{q}(\tilde{x},0)=\begin{cases}\tilde{q}_{i,j}=T(\phi_{i+\frac{1}{2},j})q_{i,j} & \tilde{x}<0\\ \tilde{q}_{i+1,j}=T(\phi_{i+\frac{1}{2},j})q_{i+1,j} & \tilde{x}>0\end{cases}\end{cases} \tag{2.1.9}$$

As we have seen in section 1.3.1, the exact solution of (2.1.9) is a similarity solution; $\tilde{q}(\tilde{x},t)=\tilde{q}_R(\tilde{x}/t)$. Thus the state at $\tilde{x}=0$, given by $\tilde{q}_R(0)$, is constant for all $t>0$. Then the flux at $\tilde{x}=0$ becomes $f(\tilde{q}_R(0))$. Notice that $f(\tilde{q}_R(0))$ represents the amount of mass, $\tilde{x}$-momentum, $\tilde{y}$-momentum and energy transported per unit of length and time across $\partial\Omega_{i+\frac{1}{2},j}$ from $\Omega_{i,j}$ to $\Omega_{i+1,j}$. Therefore, the amount of mass, x-momentum, y-momentum and energy transported per unit of length and time across $\partial\Omega_{i+\frac{1}{2},j}$ from $\Omega_{i,j}$ to $\Omega_{i+1,j}$ is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi_{i+\frac{1}{2},j} & -\sin\phi_{i+\frac{1}{2},j} & 0 \\ 0 & \sin\phi_{i+\frac{1}{2},j} & \cos\phi_{i+\frac{1}{2},j} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot f(\tilde{q}_R(0)) = T(\phi_{i+\frac{1}{2},j})^{-1} f(\tilde{q}_R(0)) . \tag{2.1.10}$$

This motivates the following approximation, as an alternative for (2.1.7):

$$\int_{\partial\Omega_{i+\frac{1}{2},j}} T(\phi_{i+\frac{1}{2},j})^{-1} f(T(\phi_{i+\frac{1}{2},j}) q(x,y,t)) d\sigma \approx$$

$$l_{i+\frac{1}{2},j} T(\phi_{i+\frac{1}{2},j})^{-1} f(\tilde{q}_R(0)) \tag{2.1.11}$$

where $\tilde{q}_R(\tilde{x}/t)$ is the exact solution of (2.1.9).

Formula (2.1.11) leads to the upwind scheme of Godunov [3,28] and is first-order accurate, provided that the mesh is smooth enough. Godunov's scheme achieves high resolution of stationary discontinuities if the discontinuity is aligned with the grid. Then, the resolution is perfect in the sense that a discontinuity has only one interior grid point (i.e. a finite volume). But Godunov's scheme has some severe disadvantages. The flux calculation (2.1.11) requires too much computational effort. The computation of the state $\tilde{q}_R(0)$ requires the numerical solution of a nonlinear algebraic equation [19]. Furthermore, the flux across $\partial\Omega_{i+\frac{1}{2},j}$ is not continuously differentiable with respect to the states $q_{i,j}$ and $q_{i+1,j}$. Differentiability of the flux is very desirable for the relaxation method for solving the system of discretized equations, as we shall see in chapter III. Due to these drawbacks of Godunov's scheme, new upwind schemes have been developed which are all based on an approximate solution of the Riemann problem (2.1.9). To introduce these schemes, we again consider the Riemann problem

$$\begin{cases} \dfrac{\partial}{\partial t} q + \dfrac{\partial}{\partial x} f(q) = 0 \\ q(x,0) = \begin{cases} q_L & x<0 \\ q_R & x>0 \end{cases} \end{cases} \tag{2.1.12}$$

Let $f_R(q_L,q_R)$ approximate $f(q_R(0))$ where $q_R(x/t)$ is the exact solution of (2.1.12). The function $f_R:\mathbb{R}^4\times\mathbb{R}^4\mapsto\mathbb{R}^4$ is called an approximate Riemann solver or numerical flux function. Several approximate Riemann solvers have been proposed.

Well known are the approximate Riemann solvers proposed by Steger and Warming [22], Van Leer [27], Roe [18] and Osher [16,17]. With a given approximate Riemann solver f_R the flux across the cell boundary $\partial\Omega_{i+\frac{1}{2},j}$ is approximated by

$$\int_{\partial\Omega_{i+\frac{1}{2},j}} T(\phi_{i+\frac{1}{2},j})^{-1} f(T(\phi_{i+\frac{1}{2},j})q(x,y,t))d\sigma \approx$$

$$l_{i+\frac{1}{2},j} T(\phi_{i+\frac{1}{2},j})^{-1} f_R(T(\phi_{i+\frac{1}{2},j})q_{i,j}, T(\phi_{i+\frac{1}{2},j})q_{i+1,j}) . \qquad (2.1.13)$$

Notice that if $f_R(q_L,q_R)=f(q_R(0))$ with $q_R(x/t)$ is the exact solution of (2.1.12). then (2.1.13) is equivalent with (2.1.11) i.e. Godunov's scheme.
Using (2.1.5), (2.1.6) and (2.1.13), we arrive at the following semidiscretization:

$$V_{i,j}\frac{d}{dt}q_{i,j}(t) + l_{i+\frac{1}{2},j} T(\phi_{i+\frac{1}{2},j})^{-1} f_R(T(\phi_{i+\frac{1}{2},j})q_{i,j}(t), T(\phi_{i+\frac{1}{2},j})q_{i+1,j}(t)) +$$

$$+ l_{i,j+\frac{1}{2}} T(\phi_{i,j+\frac{1}{2}})^{-1} f_R(T(\phi_{i,j+\frac{1}{2}})q_{i,j}(t), T(\phi_{i,j+\frac{1}{2}})q_{i,j+1}(t)) +$$

$$- l_{i-\frac{1}{2},j} T(\phi_{i-\frac{1}{2},j})^{-1} f_R(T(\phi_{i-\frac{1}{2},j})q_{i-1,j}(t), T(\phi_{i-\frac{1}{2},j})q_{i,j}(t)) +$$

$$- l_{i,j-\frac{1}{2}} T(\phi_{i,j-\frac{1}{2}})^{-1} f_R(T(\phi_{i,j-\frac{1}{2}})q_{i,j-1}(t), T(\phi_{i,j-\frac{1}{2}})q_{i,j}(t)) = 0. \qquad (2.1.14)$$

The term $\frac{d}{dt}q_{i,j}(t)$ must be approximated by a time integrator to solve the time dependent Euler equations. But in this work we restrict ourselves to the numerical solution of the steady Euler equations. Hence, the term $\frac{d}{dt}q_{i,j}(t)=0$, and we obtain the following nonlinear system of discretized equations

$$\begin{aligned} & l_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j}^{-1} f_R(T_{i+\frac{1}{2},j}q_{i,j}, T_{i+\frac{1}{2},j}q_{i+1,j}) \\ & + l_{i,j+\frac{1}{2}} T_{i,j+\frac{1}{2}}^{-1} f_R(T_{i,j+\frac{1}{2}}q_{i,j}, T_{i,j+\frac{1}{2}}q_{i,j+1}) \\ & - l_{i-\frac{1}{2},j} T_{i-\frac{1}{2},j}^{-1} f_R(T_{i-\frac{1}{2},j}q_{i-1,j}, T_{i-\frac{1}{2},j}q_{i,j}) \\ & - l_{i,j-\frac{1}{2}} T_{i,j-\frac{1}{2}}^{-1} f_R(T_{i,j-\frac{1}{2}}q_{i,j-1}, T_{i,j-\frac{1}{2}}q_{i,j}) = 0 \end{aligned} \qquad (2.1.15)$$

where

$$\begin{aligned} T_{i+\frac{1}{2},j} &= T(\phi_{i+\frac{1}{2},j}) \\ T_{i,j+\frac{1}{2}} &= T(\phi_{i,j+\frac{1}{2}}) . \end{aligned} \qquad (2.1.16)$$

Throughout the remainder of this chapter, we shall use this abbreviated notation for $T(\phi_{i+\frac{1}{2},j})$ etc. The system of discretized equations in the interior of the grid is determined completely after the construction of an approximate Riemann solver $f_R: \mathbb{R}^4 \times \mathbb{R}^4 \mapsto \mathbb{R}^4$.
The construction of an approximate Riemannn solver is the main topic of the next section. It will appear that, for our purposes, Osher's approximate Riemann solver is the best.
In section 2.3 we consider the treatment of boundary conditions. Just as the Riemann problem is the underlying physical model for the flux computation at

interior finite volume boundaries, the Riemann boundary problem is the underlying physical model for the flux computation at finite volume boundaries which are part of the boundary of the domain Ω.
In section 2.4 we consider the linearization of the discretized equations. Local linearization is used in the nonlinear relaxation method for solving the system of discretized equations.
In section 2.5 we consider the accuracy of the space discretization. It will appear that the space discretization (2.1.15) is only first-order accurate. First-order accuracy is too low in regions where the flow is smooth, and for the resolution of oblique (with respect to the grid) shocks or contact discontinuities. Therefore, the scheme is extended to second-order accuracy. In general, solutions of second-order schemes suffer from spurious oscillations in the neighbourhood of discontinuities. To prevent these oscillations a monotonicity concept is introduced and it is shown that it is possible to construct monotone second-order accurate schemes. Solutions of monotone second-order accurate schemes are second-order accurate in smooth parts of the flow field and admit steep oblique discontinuities without showing under- or overshoot.

In the publications [7,8,20,21] a large part of the contents of this chapter can be found.

2.2. Approximate solution of the Riemann problem

2.2.1. General observations

In this section we concentrate on the approximate solution of the Riemann problem. First, we consider the Riemann problem for a scalar hyperbolic equation and then for a hyperbolic system. Furthermore, we distinguish linear and nonlinear equations. In the linear case, an approximate solution of the Riemann problem is not necessary; an exact Riemann solver is easily obtained. In the nonlinear case, we generalize the exact Riemann solver of the linear case. The generalization can be performed in several ways and leads to different approximate Riemann solvers. The most simple generalization is the flux-splitting method; examples for the Euler equations are the method of Steger & Warming [22] and Van Leer [27]. A more refined approach is the flux-difference-splitting method; examples for the Euler equations are the method of Roe [18] and Osher [16]. We shall prefer Osher's approximate Riemann solver, which is described in detail in section 2.2.2.

I. Approximate solution of the Riemann problem for a scalar hyperbolic equation.
Consider the Riemann problem

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)=0 \qquad (2.2.1.1)$$

with

$$q(x,0)=\begin{cases} q_L & x<0 \\ q_R & x>0 \end{cases}$$

and $q:\mathbb{R}\times\mathbb{R}^+\mapsto\mathbb{R}, f\in C^1:\mathbb{R}\mapsto\mathbb{R}$. First, we consider the linear case i.e. $f(q)=aq$ where $a\in\mathbb{R}$ is constant. Then q is constant in the characteristic direction $\frac{dx}{dt}=a$. The exact solution of (2.2.1.1) is

$$q(x,t)=\begin{cases} q_L & \text{if } x-at<0 \\ q_R & \text{if } x-at>0 \end{cases} \tag{2.2.1.2}$$

Thus

$$q(0,t)=\begin{cases} q_L & \text{if } a>0 \\ q_R & \text{if } a<0 \end{cases} \tag{2.2.1.3}$$

and $f(q(0,t))=a^+q_L+a^-q_R$ where $a^+=\max(a,0)$, $a^-=\min(a,0)$. Hence, the exact Riemann solver is

$$f_R(q_L,q_R)=a^+q_L+a^-q_R. \tag{2.2.1.4}$$

In the nonlinear case, the exact Riemann solver is also very simple when $f(q)$ has the property that $\frac{d}{dq}f(q)\geq 0\ \forall q\in\mathbb{R}$ or $\frac{d}{dq}f(q)\leq 0\ \forall q\in\mathbb{R}$. From (2.2.1.1) we see that in the characteristic direction $\frac{dx}{dt}=\frac{d}{dq}f(q)$, q is constant. Hence, if $\frac{d}{dq}f(q)\geq 0\ \forall q\in\mathbb{R}$ then the exact Riemann solver is $f_R(q_L,q_R)=f(q_L)$ and if $\frac{d}{dq}f(q)\leq 0\ \forall q\in\mathbb{R}$ the exact Riemann solver becomes simply $f_R(q_L,q_R)=f(q_R)$. This result motivates the flux splitting method where the function $f(q)$ is split in a forward and a backward flux:

$$f(q)=f^+(q)+f^-(q) \tag{2.2.1.5}$$

where

$$\frac{d}{dq}f^+(q)\geq 0;\frac{d}{dq}f^-(q)\leq 0\ \ \forall q\in\mathbb{R}. \tag{2.2.1.6}$$

The approximate Riemann solver $f_R(q_L,q_R)$ is taken to be

$$f_R(q_L,q_R)=f^+(q_L)+f^-(q_R). \tag{2.2.1.7}$$

In the following example, we show how to split the function $f(q)$. Furthermore, we show the difference between the exact Riemann solver and the approximate Riemann solver (2.2.1.7), see also [28].

EXAMPLE (2.2.1a). The inviscid Burgers' equation.
We consider (2.2.1.1) with $f(q)=\frac{1}{2}q^2$. With this choice for f, eq. (2.2.1.1) is the inviscid Burgers' equation. The function $f(q)=\frac{1}{2}q^2$ can be split in a forward and backward flux as follows

$$f(q)=\tfrac{1}{2}q^2=f^+(q)+f^-(q)$$

where

$$f^+(q)=\tfrac{1}{2}\{q^+\}^2\ ;\ f^-(q)=\tfrac{1}{2}\{q^-\}^2$$
$$q^+=\max(q,0)\ ;\ q^-=\min(q,0).$$

Then the approximate Riemann solver becomes

$$f_R(q_L,q_R)=f^+(q_L)+f^-(q_R)=\tfrac{1}{2}\{q_L^+\}^2+\tfrac{1}{2}\{q_R^-\}^2.$$

It is easily seen that $f_R(q_L,q_R)$ is continuously differentiable.
In this case the exact solution of (2.2.1.1) is not difficult to obtain. Notice that in the characteristic direction $\frac{dx}{dt}=q$, q is constant. Let us therefore distinguish the cases $q_L\leqslant q_R$ and $q_L>q_R$. When $q_L\leqslant q_R$ the exact solution of the inviscid Burgers' equation is the simple wave (or rarefraction wave) solution

$$q(x,t)=\begin{cases}q_L & \text{if } x/t\leqslant q_L\\ x/t & \text{if } q_L<x/t<q_R\\ q_R & \text{if } x/t\geqslant q_R\ .\end{cases}$$

When $q_L>q_R$ the exact solution is a shock wave. From the Rankine-Hugoniot relation (1.3.1.16) it follows that the shock speed $s=\frac{\Delta x}{\Delta t}$ is

$$s=\frac{\Delta x}{\Delta t}=\frac{\tfrac{1}{2}(q_L^2-q_R^2)}{q_L-q_R}=\tfrac{1}{2}(q_L+q_R)$$

Hence, the exact solution is

$$q(x,t)=\begin{cases}q_L & \text{if } x/t<\tfrac{1}{2}(q_L+q_R)\\ q_R & \text{if } x/t>\tfrac{1}{2}(q_L+q_R)\ .\end{cases}$$

From these exact solutions it can be derived that the exact Riemann solver $f_R^{ex}(q_L,q_R)$ becomes

$$f_R^{ex}(q_L,q_R)=\max\{\tfrac{1}{2}(q_L^+)^2,\tfrac{1}{2}(q_R^-)^2\}.$$

Of course, $f_R^{ex}(q_L,q_R)$ corresponds with the Godunov scheme for the inviscid Burger's equation. The function $f_R^{ex}(q_L,q_R)$ is not continuously differentiable.

Now, we show that a steady shock has only one interior grid point with Godunov's scheme and two interior grid points with the flux splitting scheme. Consider a sequence $\{q_i\}_{i\in\mathbf{Z}}$. This sequence is a steady solution of Godunov's scheme when

$$f_R^{ex}(q_i,q_{i+1})=f_R^{ex}(q_{i-1},q_i)\quad \forall i\in\mathbf{Z}.$$

A sequence $\{q_i\}$ of the form

$$q_i=\begin{cases}q_L & i\leqslant -1\\ q_M & i=0\\ q_R & i\geqslant 1\end{cases}\qquad (2.2.1.8)$$

is a solution of Godunov's scheme when

$$f(q_L)=f_R^{ex}(q_L,q_M)=f_R^{ex}(q_M,q_R)=f(q_R)$$

i.e.

$$\begin{aligned}\tfrac{1}{2}q_L^2&=\max(\tfrac{1}{2}\{q_L^+\}^2,\tfrac{1}{2}\{q_M^-\}^2)=\\&=\max(\tfrac{1}{2}\{q_M^+\}^2,\tfrac{1}{2}\{q_R^-\}^2)=\tfrac{1}{2}q_R^2.\end{aligned}\qquad(2.2.1.9)$$

From this equation we see that

$$\begin{aligned}q_L<0 &\Rightarrow q_M=q_L,\ q_R=q_L\\q_R>0 &\Rightarrow q_M=q_R,\ q_L=q_R\end{aligned}$$

Thus, a shock structure is only possible when $q_L>0, q_R<0, q_R=-q_L$. Equation (2.2.1.9) is fulfilled for all $q_R<q_M<q_L$. Hence, we have a shock with one interior grid point.

A shock structure with two interior grid points is not possible. The derivation is analogous with the derivation that a three point shock structure is not possible for the flux-splitting scheme, as is shown further on.

A sequence of the form (2.2.1.8) is a solution of the flux splitting scheme when

$$\tfrac{1}{2}q_L^2=\tfrac{1}{2}\{q_L^+\}^2+\tfrac{1}{2}q_M^-\}^2=\tfrac{1}{2}\{q_M^+\}^2+\tfrac{1}{2}\{q_R^-\}^2=\tfrac{1}{2}q_R^2\qquad(2.2.1.10)$$

From this equation we see again that

$$\begin{aligned}q_L<0 &\Rightarrow q_M=q_L,\ q_R=q_L\\q_R>0 &\Rightarrow q_M=q_R,\ q_L=q_R\end{aligned}$$

Thus, a shock structure is only possible when $q_L>0, q_R<0, q_R=-q_L$: Equation (2.2.2.10) is fulfilled only if $q_M=0$. This is a special case of the general shock structure with two interior grid points. A sequence of the form

$$q_i=\begin{cases}q_L & i\leqslant -2\\ q_A & i=-1\\ q_B & i=0\\ q_R & i\geqslant 1\end{cases}$$

is a solution of the flux-splitting scheme when

$$\begin{aligned}\tfrac{1}{2}q_L^2&=\tfrac{1}{2}(q_L^+)^2+\tfrac{1}{2}(q_A^-)^2=\tfrac{1}{2}(q_A^+)^2+\tfrac{1}{2}(q_B^-)^2=\\&=\tfrac{1}{2}(q_B^+)^2+\tfrac{1}{2}(q_R^-)^2=\tfrac{1}{2}q_R^2\end{aligned}\qquad(2.2.1.11)$$

From this equation we see that

$$\begin{aligned}q_L<0 &\Rightarrow q_A=q_L,\ q_B=q_L\quad q_R=q_L\\q_R>0 &\Rightarrow q_B=q_R,\ q_A=q_R,\ q_L=q_R\end{aligned}$$

thus, a shock structure is only possible when $q_L>0, q_R<0, q_R=-q_L$. Equation

(2.2.1.11) is fulfilled when $q_A>0, q_B<0, \frac{1}{2}q_A^2+\frac{1}{2}q_B^2=\frac{1}{2}q_L^2$. Then, we have a shock with two interior grid points. We shall show that a shock structure with three interior grid points is not possible. For such a structure we must have

$$\frac{1}{2}q_L^2=\frac{1}{2}(q_L^+)^2+\frac{1}{2}(q_A^-)^2=\frac{1}{2}(q_A^+)^2+\frac{1}{2}(q_B^-)^2=$$
$$=\frac{1}{2}(q_B^+)^2+\frac{1}{2}(q_C^-)^2=\frac{1}{2}(q_C^+)^2+\frac{1}{2}(q_R^-)^2=\frac{1}{2}q_R^2. \qquad (2.2.1.12)$$

When $q_L<0$, or $q_R>0$ then $q_A=q_B=q_C=q_L=q_R$. Thus, suppose $q_L>0, q_R=-q_L<0$. Then (2.2.1.12) is fulfilled when $q_A>0, q_C<0$ and

$$\frac{1}{2}q_L^2=\frac{1}{2}q_A^2+\frac{1}{2}(q_B^-)^2=\frac{1}{2}(q_B^+)^2+\frac{1}{2}q_C^2=\frac{1}{2}q_R^2$$

Thus $q_B>0\Rightarrow q_A=q_L$ and $q_B<0\Rightarrow q_C=q_R$. Hence, no shock structure with three interior grid points is possible.

II. Approximate solution of the Riemann problem for a hyperbolic system.
Consider the Riemann problem (1.3.1.7) for a general hyperbolic system. First, suppose $f(q)$ is a linear function $f(q)=Aq$ where A is a constant $n\times n$ matrix. Then the exact solution is (see example 1.3.1a,b):

$$q(x,t)=\sum_{i=1}^{n}\{\beta_i H(x-\lambda_i t)+\alpha_i(1-H(x-\lambda_i t))\}R_i$$

where $q_L=\sum_{i=1}^{n}\alpha_i R_i, q_R=\sum_{i=1}^{n}\beta_i R_i$. Suppose $\lambda_1\leqslant\cdots\leqslant\lambda_k\leqslant 0<\lambda_{k+1}\leqslant\cdots\leqslant\lambda_n$. Then

$$q(0,t)=\sum_{i=1}^{k}\beta_i R_i+\sum_{i=k+1}^{n}\alpha_i R_i. \qquad (2.2.1.13)$$

and

$$f(q(0,t))=Aq(0,t)$$
$$=\sum_{i=1}^{k}\beta_i\lambda_i R_i+\sum_{i=k+1}^{n}\alpha_i\lambda_i R_i. \qquad (2.2.1.14)$$

Define the nonsingular matrix $R=(R_1\cdots R_n)$ and the diagonal matrix D by $D_{ii}=\lambda_i, i=1,\ldots,n$. Thus $A=RDR^{-1}$. Define the diagonal matrices D^+, D^- and $|D|$ by

$$D_{ii}^+=\lambda_i^+; D_{ii}^-=\lambda_i^-, |D|_{ii}=|\lambda_i| \quad i=1,\ldots,n$$

where $\lambda_i^+=\max(\lambda_i,0), \lambda_i=\min(\lambda_i,0)$. Let

$$A^+=RD^+R^{-1}, A^-=RD^-R^{-1} \text{ and } |A|=R|D|R^{-1}.$$

Hence $A=A^++A^-, |A|=A^+-A^-$. We find

$$A^+q_L=A^+(\sum_{i=1}^{n}\alpha_i R_i)=\sum_{i=1}^{n}\alpha_i\lambda_i^+R_i=\sum_{i=k+1}^{n}\alpha_i\lambda_i R_i \qquad (2.2.1.15a)$$

and

$$A^- q_R = A^-(\sum_{i=1}^{n} \beta_i R_i) = \sum_{i=1}^{n} \beta_i \lambda_i^- R_i = \sum_{i=1}^{k} \beta_i \lambda_i R_i \qquad (2.2.1.15b)$$

Combining (2.2.1.14) and (2.2.1.15) we find that

$$f(q(0,t)) = A^+ q_L + A^- q_R \qquad (2.2.1.16)$$

Thus, the exact Riemann solver is

$$f_R(q_L, q_R) = A^+ q_L + A^- q_R = \tfrac{1}{2}\{Aq_L + Aq_R - |A|(q_R - q_L)\} \qquad (2.2.1.17)$$

This expression is a generalization of (2.2.1.4). Thus, even for systems, the exact Riemann solver is easily obtained when the equation is linear.

Finally, we have to consider the Riemann problem for a general nonlinear hyperbolic system. The most simple approach is the flux splitting method where $f:\mathbb{R}^n \mapsto \mathbb{R}^n$ is split in a forward flux $f^+:\mathbb{R}^n \mapsto \mathbb{R}^n$ and a backward flux $f^-:\mathbb{R}^n \mapsto \mathbb{R}^n$ such that

$$f(q) = f^+(q) + f^-(q) \quad \forall q \in \mathbb{R}^n \qquad (2.2.1.18a)$$

and

$$\begin{aligned} &\frac{df^+}{dq} \text{ has all eigenvalues } \geqslant 0 \\ &\frac{df^-}{dq} \text{ has all eigenvalues } \leqslant 0 \end{aligned} \qquad (2.2.18b)$$

Then, the approximate Riemann solver $f_R(q_L, q_R)$ becomes

$$f_R(q_L, q_R) = f^+(q_L) + f^-(q_R). \qquad (2.2.1.19)$$

This expression can be considered as a generalization of the scalar case (2.2.1.7) and the linear case (2.2.1.17).

The splitting of $f(q)$ in a forward- and backward flux is not unique. Flux-splitting methods for the Euler equations have been proposed by Steger & Warming [22] and Van Leer [27]. In contrast with the Steger & Warming flux-splitting, Van Leer's flux-splitting leads to a continuously differentiable forward and backward flux. As noted before, differentiability of the approximate Riemann solver $f_R(q_L, q_R)$ is desirable in the relaxation method for solving the system of discretized equations. (Newton's method is applied in the relaxation method). Therefore, Van Leer's flux splitting method is preferable to Steger & Warming's method. The requirement that $f_R(q_L, q_R)$ is continuously differentiable is very restrictive. Only Van Leer's method and Osher's method result in a continuous differentiable approximate Riemann solver, the other well known approximate Riemann solvers of Steger & Warming and Roe do not.

Osher's method can be seen as a refinement of the flux splitting method and can be understood in the following way. Because (1.3.1.7) is a hyperbolic system, the matrix $A(q) = \frac{df}{dq}(q)$ has a linearly independent set of eigenvectors

$R_1(q),\ldots,R_n(q)$ with corresponding eigenvalues $\lambda_1(q),\ldots,\lambda_n(q)$ labeled in increasing order $\lambda_1(q)\leqslant\cdots\leqslant\lambda_n(q)$. Define the nonsingular matrix $R(q)=(R_1(q),\ldots,R_n(q))$ and the diagonal matrix $D(q)$ by $(D(q))_{i,i}=\lambda_i(q)$. Just as in the linear case, introduce the diagonal matrices $D^+(q),D^-(q)$ and $|D|(q)$:

$$\begin{aligned}&(D^+(q))_{i,i}=\lambda_i^+(q)\ ;\ (D^-(q))_{i,i}=\lambda_i^-(q)\\&(|D|(q))_{i,i}=|\lambda_i(q)|\quad ;\ i=1,\ldots,n\end{aligned}\tag{2.2.1.20}$$

where $\lambda_i^+(q)=\max(\lambda_i(q),0);\lambda_i^-(q)=\min(\lambda_i(q),0)$.
Introduce

$$\begin{aligned}&A^+(q)=R(q)D^+(q)R^{-1}(q)\\&A^-(q)=R(q)D^-(q)R^{-1}(q)\\&|A|(q)=R(q)|D|(q)R^{-1}(q)=A^+(q)-A^-(q)\end{aligned}\tag{2.2.1.21}$$

Suppose there exist functions $f^+(q)$ and $f^-(q)$ such that

$$f(q)=f^+(q)+f^-(q)\tag{2.2.1.22a}$$

and

$$\frac{d}{dq}f^+(q)=A^+(q);\frac{d}{dq}f^-(q)=A^-(q).\tag{2.2.1.22b}$$

Then, a natural approximate Riemann solver is

$$f_R(q_L,q_R)=f^+(q_L)+f^-(q_R)\tag{2.2.1.23}$$

which can also be written as

$$f_R(q_L,q_R)=f^+(q_L)+f^-(q_R)\tag{2.2.1.24a}$$

$$=f(q_L)-f^-(q_L)+f^-(q_R)=f(q_L)+\int_{q_L}^{q_R}A^-(q)dq\tag{2.2.1.24b}$$

$$=f(q_R)-f^+(q_R)+f^+(q_L)=f(q_R)-\int_{q_L}^{q_R}A^+(q)dq\tag{2.2.1.24c}$$

$$=\tfrac{1}{2}\{f(q_L)+f(q_R)-\int_{q_L}^{q_R}|A(q)|dq\}\tag{2.2.1.24d}$$

and the integrals in (2.2.1.24b,c,d) are independent of the integration path. Notice that the integrals are evaluated in the state space i.e. in $\mathbb{R}^n$. Unfortunately, in general no functions $f^+(q)$ and $f^-(q)$ exist such that (2.2.1.22) is valid. This is equivalent with the observations that the integrals

$$\int_{q_L}^{q_R}A^-(q)dq\ \text{and}\ \int_{q_L}^{q_R}A^+(q)dq\tag{2.2.1.25}$$

depend on their integration path. Now, Osher's scheme is defined as

$$f_R(q_L,q_R)=f(q_L)+\int_{q_L}^{q_R} A^-(q)dq \tag{2.2.1.26a}$$

$$=f(q_R)-\int_{q_L}^{q_R} A^+(q)dq \tag{2.2.1.26b}$$

$$=\tfrac{1}{2}\{f(q_L)+f(q_R)-\int_{q_L}^{q_R}|A|(q)dq\} \tag{2.2.1.26c}$$

where the integration path is chosen in such a way that the evaluation of (2.2.1.25) is easy.

Suppose that the states q_L and q_R can be connected with each other by an integration path Γ_k which is tangential to the eigenvector R_k i.e.

$$\begin{cases} \dfrac{dq}{d\xi}(\xi)=R_k(q(\xi)) \\ \\ q(0)=q_L; \qquad q(\xi_R)=q_R \end{cases} \tag{2.2.1.27}$$

Then, we see that

$$\begin{aligned} \int_{q_L}^{q_R} A^-(q)dq &= \int_0^{\xi_R} A^-(q(\xi))\frac{dq}{d\xi}d\xi \\ &= \int_0^{\xi_R} A^-(q(\xi))R_k(q(\xi))d\xi \\ &= \int_0^{\xi_R} \lambda_k^-(q(\xi))R_k(q(\xi))d\xi \end{aligned} \tag{2.2.1.28}$$

Let us distinguish the following possibilities:

A $\lambda_k(q(\xi))$ does not change sign along the integration path.

If $\lambda_k(q(\xi))\geq 0 \quad \forall\xi\in(0,\xi_R)$ then

$$\int_{q_L}^{q_R} A^-(q)dq=0. \tag{2.2.1.29}$$

If $\lambda_k(q(\xi))\leq 0 \quad \forall\xi\in(0,\xi_R)$ then

$$\begin{aligned} \int_{q_L}^{q_R} A^-(q)dq &= \int_0^{\xi_R} \lambda_k^-(q(\xi))R_k(q(\xi))d\xi \\ &= \int_0^{\xi_R} \lambda_k(q(\xi))R_k(q(\xi))d\xi \\ &= \int_0^{\xi_R} A(q(\xi))R_k(q(\xi))d\xi \end{aligned}$$

$$= \int_0^{\xi_R} \frac{df}{dq}(q(\xi))\frac{dq}{d\xi}(\xi)d\xi = f(q_R) - f(q_L) \qquad (2.2.1.30)$$

B $\lambda_k(q(\xi))$ changes sign along the integration path.
Suppose $\lambda_k(q(\xi))$ changes sign only once at $\xi=\xi_S$ $0<\xi_S<\xi_R$. Define $q_S = q(\xi_S)$.
If $\lambda_k(q(\xi)) \geq 0$ $\forall \xi \in (0,\xi_S)$ and $\lambda_k(q(\xi)) \leq 0$ $\forall \xi \in (\xi_S,\xi_R)$ then

$$\int_{q_L}^{q_R} A^-(q)dq = \int_0^{\xi_R} \lambda_k^-(q(\xi))R_k(q(\xi))d\xi$$

$$= \int_{\xi_S}^{\xi_R} \lambda_k(q(\xi))R_k(q(\xi))d\xi = f(q_R) - f(q_S) \qquad (2.2.1.31)$$

If $\lambda_k(q(\xi)) \leq 0$ $\forall \xi \in (0,\xi_S)$ and $\lambda_k(q(\xi)) \geq 0$ $\forall \xi \in (\xi_S,\xi_R)$ then

$$\int_{q_L}^{q_R} A^-(q)dq = \int_0^{\xi_R} \lambda_k^-(q(\xi))R_k(q(\xi))d\xi$$

$$= \int_0^{\xi_S} \lambda_k(q(\xi))R_k(q(\xi))d\xi = f(q_S) - f(q_L) \qquad (2.2.1.32)$$

Thus, when the states q_L and q_R can be connected with each other by an integration path which is tangential to the kth eigenvector $R_k(q)$ then Osher's approximate Riemann solver is

$$f_R(q_L,q_R) = \begin{cases} f(q_L) & \text{if } \lambda_k \geq 0 \text{ along } \Gamma_k \\ f(q_R) & \text{if } \lambda_k \leq 0 \text{ along } \Gamma_k \\ f(q_R) - f(q_S) + f(q_L) & \text{if } \lambda_k(q_L)>0,\ \lambda_k(q_R)<0,\ \lambda_k(q_S)=0 \\ f(q_S) & \text{if } \lambda_k(q_L)<0,\ \lambda_k(q_R)>0,\ \lambda_k(q_S)=0, \end{cases} \qquad (2.2.1.33)$$

where we have assumed that λ_k changes sign along Γ_k at most once. The point q_S is called a sonic point. If the eigenvector $R_k(q)$ is genuinely nonlinear (see definition 1.3.1.c) then

$$\frac{d}{d\xi}\lambda_k(q(\xi)) = \nabla\lambda_k(q(\xi))\frac{dq}{d\xi}(\xi)$$

$$= \nabla\lambda_k(q(\xi))R_k(q(\xi)) \neq 0 \qquad (2.2.1.34)$$

and this implies that λ_k is monotone along Γ_k. Thus if $R_k(q)$ is genuinely nonlinear then indeed λ_k changes sign along Γ_k at most once. If the eigenvector $R_k(q)$ is linearly degenerate then λ_k is constant along Γ_k. Then, we find

$$\int_{q_L}^{q_R} A^-(q)dq = \begin{cases} 0 & \text{if } \lambda_k > 0 \\ f(q_R) - f(q_L) & \text{if } \lambda_k < 0 \end{cases} \qquad (2.2.1.35)$$

and

$$f_R(q_L,q_R)=\begin{cases} f(q_L) & \text{if } \lambda_k>0 \\ f(q_R) & \text{if } \lambda_k<0 \end{cases} \tag{2.2.1.36}$$

A general pair (q_L,q_R) can be connected by a continuous integral path Γ which is decomposed into n subcurves Γ_k:

$$\Gamma=\bigcup_{k=1}^{n}\Gamma_k \tag{2.2.1.37}$$

where each subcurve Γ_k is tangential to the eigenvector $R_k(q)$. The subcurve Γ_1 starts in $q_L\equiv q_0$ and the subcurve Γ_n ends in $q_R\equiv q_1$. Define the $n-1$ points of intersection $q_{k/n}$, $k=1,\ldots,n-1$ by

$$q_{k/n}=\Gamma_k\cap\Gamma_{k+1}. \tag{2.2.1.38}$$

The intersection points are easily found with the use of Riemann invariants (see definition 1.3.1d); along the subcurve Γ_k the Riemann invariants $\psi_k^1,\ldots,\psi_k^{n-1}$ ($\psi_k^l:\mathbb{R}^n\mapsto\mathbb{R}$) are constant, then

$$\begin{aligned}&\psi_k^1(q_{k-1/n})=\psi_k^1(q_{k/n});\psi_k^2(q_{k-1/n})=\psi_k^2(q_{k/n});\cdots\cdots\\&\psi_k^{n-1}(q_{k-1/n})=\psi_k^{n-1}(q_{k/n});\ k=1,\ldots,n\end{aligned} \tag{2.2.1.39}$$

In this way, we obtain $n(n-1)$ equations for the $n(n-1)$ unknowns $q_{1/n},\ldots,q_{n-1/n}$. Once the points of intersection are known, the integrals in (2.2.1.26) along each subcurve Γ_k are evaluated in the manner described by (2.2.1.27-32).

This approximate Riemann solver $f_R(q_L,q_R)$ differs somewhat from the approximate Riemann solver proposed by Osher. Osher proposed a reverse ordering of the subcurves Γ_k, i.e. the subcurve Γ_n starts in q_0 and the subcurve Γ_1 ends in q_1. Then $n-1$ points of intersection $q_{k/n},k=1,\ldots,n-1$ are defined by

$$q_{k/n}=\Gamma_{n-k+1}\cap\Gamma_{n-k}\quad k=1,\ldots,n-1 \tag{2.2.1.40}$$

and are found by

$$\begin{aligned}&\psi_{n-k+1}^1(q_{k-1/n})=\psi_{n-k+1}^1(q_{k/n});\psi_{n-k+1}^2(q_{k-1/n})=\psi_{n-k+1}^2(q_{k/n})\cdots\\&\psi_{n-k+1}^{n-1}(q_{k-1/n})=\psi_{n-k+1}^{n-1}(q_{k/n});\quad k=1,\ldots,n\end{aligned} \tag{2.2.1.41}$$

The computation of the integrals in (2.2.1.26) along each subcurve Γ_k remains the same. The only difference is the reverse ordening of the subcurves.

We call the ordening corresponding with (2.2.1.38,39) the *P*-(Physical) variant and the ordening corresponding with (2.2.1.40,41) the *O*-(Osher) variant. The *P*-variant is more natural in the sense that when the states (q_L,q_R) are such that the exact solution of the Riemann initial value problem contains no shock waves then the flux $f_R(q_L,q_R)$ corresponding with the *P*-variant is exact. On the other hand, Osher claims that (in the case of the Euler equations) his ordening rules out overshoot in the two point transition region between the

constant states of a steady discrete shock [16]. However, it is our experience that both for the *O*- and *P*-variant a steady shock is monotone (no overshoot) and has two interior grid points. The construction of the integral path Γ corresponding with the *O*- and *P*-variant are depicted in fig. 2.2.1a for $n=3$.

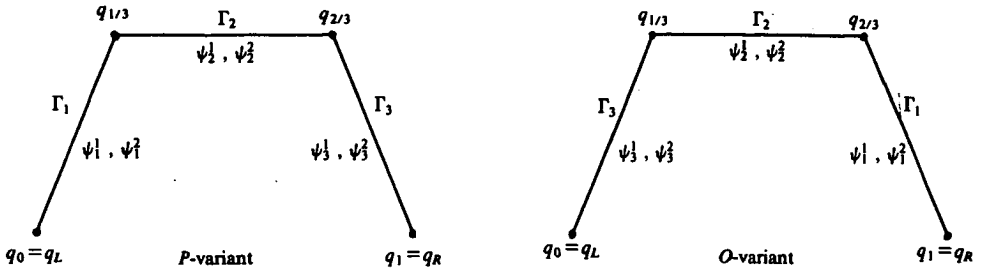


FIGURE 2.2.1a. The integral path Γ corresponding with the *P*- and *O*-variant for $n=3$. The Riemann invariants $\psi_k^1, \psi_k^2 : \mathbf{R}^3 \mapsto \mathbf{R}$ are constant along the subcurve $\Gamma_k, k=1,2,3$.

Finally, we have to explain why Osher's approximate Riemann solver (with the *O*- or *P*-variant) is the most attractive for our purposes. In the following table some important properties are listed for several well known approximate Riemann solvers for the Euler equations.

		Godunov	Osher	Van Leer	Roe
1	$f_R(q_L,q_R)$ is C^1.	no	yes	yes	no
2	Number of interior grid points in a steady shock.	one	two	two	one
3	Good resolution of a steady contact discontinuity.	yes	yes	no	yes
4	Physical inadmissible expansion shock is excluded.	yes	yes	yes	(no, but can be repaired)
5	Computational cost of $f_R(q_L,q_R)$.	high	low	low	low

As noted before, we need differentiability of $f_R(q_L,q_R)$. Furthermore, we need the property that a shock has at least two interior grid points. If a shock has only one interior grid point the discretized equation becomes singular at a shock (see equation (2.2.1.9); if $q_L>0, q_R=-q_L$ the equation is satisfied for all q_M with $q_R<q_M<q_L$). This is disastrous for a (local) relaxation method for solving the system of discretized equations. Therefore, Roe's and Godunov's method do not suit our purposes. A choice must by made between Van Leer's and Osher's approximate Riemann solver. We prefer Osher's method because it has the ability to resolve steady contact discontinuities. Furthermore, as we shall see in section 2.3, with Osher's scheme, the flux computation at the

boundary of the domain can be performed in a way fully consistent with the flux calculation at interior control volume boundaries. The only disadvantage of the Osher scheme is its complexity (from a computational point of view). In the next section we shall show that the computational complexity can be reduced significantly by choosing suitable independent variables. Both the *O*- and *P*-variant are considered.

2.2.2. Osher's approximate Riemann solver for the Euler equations.

Consider the Riemann problem (1.3.2.14) for the Euler equations. In the preceding subsection we have seen that Osher's approximate Riemann solver is given by

$$\begin{aligned} f_R(q_L,q_R) &= f(q_L) + \int_{q_L}^{q_R} A^-(q)dq \\ &= f(q_R) - \int_{q_L}^{q_R} A^+(q)dq \\ &= \tfrac{1}{2}\{f(q_L)+f(q_R) - \int_{q_L}^{q_R} |A(q)|dq\} \end{aligned} \tag{2.2.2.1}$$

The integrals are evaluated in the state space $\mathbb{R}^4$. Let the state q be represented as $q=(c,u,v,z)$ where z is the unscaled entropy:

$$z = \ln\left(\frac{p}{\rho^\gamma}\right) \tag{2.2.2.2}$$

First, consider the *P*-variant of the Osher scheme.

P-Variant.
Define $q_0 \equiv q_L, q_1 \equiv q_R$ and the intersection points

$$q_{1/4} = \Gamma_1 \cap \Gamma_2;\ q_{2/4} = \Gamma_2 \cap \Gamma_3;\ q_{3/4} = \Gamma_3 \cap \Gamma_4 .$$

Due to the fact that $R_2(q)$ and $R_3(q)$ are linearly degenerate (theorem 1.3.2.c) and $\lambda_2(q)=\lambda_3(q)=u$, we can omit the intersection point $q_{2/4}$ because

$$\int_{\Gamma_2} A^-(q)dq = \begin{cases} f(q_{2/4}) - f(q_{1/4}) & \text{if } u<0 \\ 0 & \text{if } u>0 \end{cases}$$

$$\int_{\Gamma_3} A^-(q)dq = \begin{cases} f(q_{3/4}) - f(q_{2/4}) & \text{if } u<0 \\ 0 & \text{if } u>0 \end{cases}$$

Thus

$$\int_{\Gamma_2 \cup \Gamma_3} A^-(q)dq = \int_{\Gamma_2} A^-(q)dq + \int_{\Gamma_3} A^-(q)dq$$

$$= \begin{cases} f(q_{3/4}) - f(q_{1/4}) & \text{if } u<0 \\ 0 & \text{if } u>0 \end{cases}$$

Hence, $f_R(q_L,q_R)$ does not depend on $q_{2/4}$. Therefore, we redefine the intersection points

$$q_{1/3}=q_{1/4},\quad q_{2/3}=q_{3/4}.$$

Thus $q_{1/3}$ and q_0 are connected by Γ_1, $q_{2/3}$ and q_1 are connected by Γ_4 and $q_{1/3}$ and $q_{2/3}$ are connected by an integral path that is composed of $\Gamma_2 \cup \Gamma_3$. Using the Riemann invariants mentioned in theorem (1.3.2.d), we find that

$$\begin{aligned} u_0+\frac{2}{\gamma-1}c_0 &= u_{1/3}+\frac{2}{\gamma-1}c_{1/3} \equiv \Psi_0 \\ v_0 &= v_{1/3} \\ z_0 &= z_{1/3} \\ u_{1/3} &= u_{2/3} \equiv u_H \\ p_{1/3} &= p_{2/3} \\ u_{2/3}-\frac{2}{\gamma-1}c_{2/3} &= u_1-\frac{2}{\gamma-1}c_1 \equiv \Psi_1 \\ v_{2/3} &= v_1 \\ z_{2/3} &= z_1 \end{aligned} \tag{2.2.2.3}$$

We have 8 equations for the 8 unknowns $q_{1/3}$ and $q_{2/3}$. We obtain directly $z_{1/3}=z_0$, $z_{2/3}=z_1$, $v_{1/3}=v_0$, $v_{2/3}=v_1$. Because

$$p=p(c,z)=\left(\gamma^{\gamma}\frac{e^z}{c^{2\gamma}}\right)^{\frac{1}{1-\gamma}},$$

$p_{1/3}=p_{2/3}$ leads to

$$\frac{c_{2/3}}{c_{1/3}}=\exp\left(\frac{z_{2/3}-z_{1/3}}{2\gamma}\right)=\exp\left(\frac{z_1-z_0}{2\gamma}\right)\equiv\alpha \tag{2.2.2.4}$$

We arrive at the linear system

$$\begin{aligned} u_H+\frac{2}{\gamma-1}c_{1/3} &= \Psi_0 \\ u_H-\frac{2}{\gamma-1}c_{2/3} &= \Psi_1 \\ c_{2/3} &= \alpha c_{1/3} \end{aligned} \tag{2.2.2.5}$$

This system is easily solved:

$$\begin{aligned} c_{1/3} &= \frac{\gamma-1}{2}\left(\frac{\Psi_0-\Psi_1}{1+\alpha}\right) \\ c_{2/3} &= \alpha c_{1/3} \end{aligned}$$

$$u_H = \frac{\Psi_1 + \alpha\Psi_0}{1+\alpha} \tag{2.2.2.6}$$

A meaningful solution does not exist in the unlikely case that $\Psi_0 - \Psi_1 < 0$.

Thus the evaluation of just one exponential and some arithmetic operations are sufficient for the computation of $q_{1/3}$ and $q_{2/3}$. This is made possible by the representation of the state q as $q=(c,u,v,z)$.

For the evaluation of the integrals in (2.2.2.1) we may need the sonic points on Γ_1 and Γ_4.

A sonic point q_S^0 exists along Γ_1 when

$$\lambda_1(q_0)\,\lambda_1(q_{1/3}) = (u_0 - c_0)\,(u_H - c_{1/3}) < 0 \tag{2.2.2.7}$$

and is found by

$$\begin{aligned} u_0 + \frac{2}{\gamma-1}c_0 &= u_S^0 + \frac{2}{\gamma-1}c_S^0 \\ v_0 &= v_S^0 \\ z_0 &= z_S^0 \\ u_S^0 - c_S^0 &= 0 \end{aligned} \tag{2.2.2.8}$$

Thus $v_S^0 = v_0, z_S^0 = z_0$ and

$$u_S^0 = c_S^0 = \frac{\gamma-1}{\gamma+1}\left(u_0 + \frac{2}{\gamma-1}c_0\right). \tag{2.2.2.9}$$

A sonic point q_S^1 exists along Γ_4 when

$$\lambda_4(q_{2/3})\,\lambda_4(q_1) = (u_H + c_{2/3})(u_1 + c_1) < 0 \tag{2.2.2.10}$$

and is found by

$$\begin{aligned} u_S^1 - \frac{2}{\gamma-1}c_S^1 &= u_1 - \frac{2}{\gamma-1}c_1 \\ v_S^1 &= v_1 \\ z_S^1 &= z^1 \\ u_S^1 + c_S^1 &= 0 \end{aligned} \tag{2.2.2.11}$$

Thus $v_S^1 = v_1, z_S^1 = z_1$ and

$$u_S^1 = -c_S^1 = \frac{\gamma-1}{\gamma+1}\left(u_1 - \frac{2}{\gamma-1}c_1\right) \tag{2.2.2.12}$$

With these results the evaluation of the integrals in (2.2.2.1) becomes straightforward. The result is summarized in table 2.2.2a. The verification is left to the reader.

	$u_0<c_0,u_1>-c_1$	$u_0>c_0,u_1>-c_1$	$u_0<c_0,u_1<-c_1$	$u_0>c_0,u_1<-c_1$
$c_{1/3}<u_H$	$f(q_S^0)$	$f(q_0)$	$f(q_S^0)-f(q_S^1)+f(q_1)$	$f(q_0)-f(q_S^1)+f(q_1)$
$0<u_H<c_{1/3}$	$f(q_{1/3})$	$f(q_0)-f(q_S^0)+f(q_{1/3})$	$f(q_{1/3})-f(q_S^1)+f(q_1)$	$f(q_0)-f(q_S^0)+f(q_{1/3})$ $-f(q_S^1)+f(q_1)$
$-c_{2/3}<u_H<0$	$f(q_{2/3})$	$f(q_0)-f(q_S^0)+f(q_{2/3})$	$f(q_{2/3})-f(q_S^1)+f(q_1)$	$f(q_0)-f(q_S^0)+f(q_{2/3})$ $-f(q_S^1)+f(q_1)$
$u_H<-c_{2/3}$	$f(q_S^1)$	$f(q_0)-f(q_S^0)+f(q_S^1)$	$f(q_1)$	$f(q_0)-f(q_S^0)+f(q_1)$

Table 2.2.2a. Osher's approximate Riemann solver $f_R(q_0,q_1)$ for the Euler equations: *P*-variant.

Table 2.2.2a is to be read in the following way: if, for instance, $(u_1>-c_1,u_0>c_0$ and $0<u_H<c_{1/3})$ then $f_R(q_0,q_1)=f(q_0)-f(q_S^0)+f(q_{1/3})$.

The case $u_0>c_0,u_1<-c_1$ is very unlikely while the case $u_0<c_0,u_1>-c_1$ is the common subsonic situation. Then $f_R(q_0,q_1)$ requires only one flux calculation. The case $(u_0>c_0,u_1>-c_1$ and $u_H<0)$ is unlikely too, just as the situation that $(u_0<c_0,u_1<-c_1$ and $u_H>0)$. The situations $(u_0>c_0,u_1>-c_1$ and $u_H>c_{1/3})$ and $(u_0<c_0,u_1<-c_1$ and $u_H<-c_{2/3})$ correspond with supersonic flow. The situations $(u_0>c_0,u_1>-c_1$ and $0<u_H<c_{1/3})$ and $(u_0<c_0,u_1<-c_1$ and $-c_{2/3}<u_H<0)$ correspond with a shock wave.

From table 2.2.2a it is easily seen that $f_R(q_0,q_1)$ is continuous.

The flux

$$f(q)=(\rho u,\rho u^2+p,\rho uv,(E+p)u)$$

is computed from the state $q=(c,\ u,\ v,\ z)$ as follows:

$$a=\frac{c^2}{\gamma};\rho=\exp\left(\frac{\ln(a)-z}{\gamma-1}\right);$$

$$p=a\rho;E=\tfrac{1}{2}\rho(u^2+v^2)+p/(\gamma-1). \qquad (2.2.2.13)$$

Thus, in the common situation where the computation of $f_R(q_0,q_1)$ requires only one flux computation, the computation takes two exponentials, one logarithm and some elementary operations and Boolean evaluations. This is not true for the *O*-variant, as we shall see.

O-Variant

Define $q_0=q_L,q_1=q_R$. Introduce the intersection points $q_{1/3}$ and $q_{2/3};q_0$ and $q_{1/3}$ are connected by $\Gamma_4,q_{2/3}$ and q_1 are connected by Γ_1, and $q_{1/3}$ and $q_{2/3}$ are connected by an integration path that is composed by $\Gamma_2\cup\Gamma_3$. Using the Riemann invariants we find

$$u_0-\frac{2}{\gamma-1}c_0=u_{1/3}-\frac{2}{\gamma-1}c_{1/3}\equiv\Psi_0$$

$$v_0=v_{1/3}$$

$$z_0=z_{1/3}$$

$$u_{1/3}=u_{2/3}\equiv u_H$$

$$p_{1/3}=p_{2/3}$$
$$u_{2/3}+\frac{2}{\gamma-1}c_{2/3}=u_1+\frac{2}{\gamma-1}c_1\equiv\Psi_1$$
$$v_{2/3}=v_1$$
$$z_{2/3}=z_1 \qquad (2.2.2.14)$$

Hence, $z_{1/3}=z_0, z_{2/3}=z_1, v_{1/3}=v_0$, $c_{2/3}=\alpha c_{1/3}$ where α is given by (2.2.2.4). We arrive at the linear system

$$u_H-\frac{2}{\gamma-1}c_{1/3}=\Psi_0$$
$$u_H+\frac{2}{\gamma-1}c_{2/3}=\Psi_1$$
$$c_{2/3}=\alpha c_{1/3} \qquad (2.2.2.15)$$

and this system is easily solved by

$$c_{1/3}=\frac{\gamma-1}{2}\frac{\Psi_1-\Psi_0}{1+\alpha}$$
$$c_{2/3}=\alpha c_{1/3}$$
$$u_H=\frac{\Psi_1+\alpha\Psi_0}{1+\alpha} \qquad (2.2.2.16)$$

A meaningful solution does not exist in the unlikely case that $\Psi_1-\Psi_0<0$. A sonic point q_S^0 exists along Γ_4 when

$$\lambda_4(q_0)\cdot\lambda_4(q_{1/3})=(u_0+c_0)\,(u_H+c_{1/3})<0 \qquad (2.2.2.17)$$

and is found by

$$u_0-\frac{2}{\gamma-1}c_0=u_S^0-\frac{2}{\gamma-1}c_S^0$$
$$v_0=v_S^0$$
$$z_0=z_S^0$$
$$u_S^0+c_S^0=0 \qquad (2.2.2.18)$$

Thus $v_S^0=v_0, z_S^0=z_0$ and

$$u_S^0=-c_S^0=\frac{\gamma-1}{\gamma+1}\left(u_0-\frac{2}{\gamma-1}c_0\right). \qquad (2.2.2.19)$$

A sonic point q_S^1 exists along Γ_1 when

$$\lambda_1(q_{2/3})\lambda_1(q_1)=(u_H-c_{2/3})(u_1-c_1)<0 \qquad (2.2.2.20)$$

and is found by

$$u_S^1+\frac{2}{\gamma-1}c_S^1=u_1+\frac{2}{\gamma-1}c_1$$

$$v_S^1 = v_1$$
$$z_S^1 = z_1$$
$$u_S^1 - c_S^1 = 0. \tag{2.2.2.21}$$

Thus $v_S^1 = v_1, z_S^1 = z_1$ and

$$u_S^1 = c_S^1 = \frac{\gamma-1}{\gamma+1}(u_1 + \frac{2}{\gamma-1}c_1). \tag{2.2.2.22}$$

With these results, the evaluation of the integrals in (2.2.2.1) becomes straightforward. The result in summarized in table 2.2.2b and the verification is left to the reader.

	$u_0 > -c_0, u_1 < c_1$	$u_0 > -c_0, u_1 > c_1$	$u_0 < -c_0, u_1 < c_1$	$u_0 < -c_0, u_1 > c_1$
$c_{2/3} < u_H$	$f(q_0) - f(q_S^1) + f(q_1)$	$f(q_0)$	$f(q_S^1) - f(q_S^1) + f(q_1)$	$f(q_S^1)$
$0 < u_H < c_{2/3}$	$f(q_0) - f(q_{2/3}) + f(q_1)$	$f(q_S^1) - f(q_{2/3}) + f(q_S^1)$	$f(q_S^1) - f(q_{2/3}) + f(q_1)$	$f(q_S^1) - f(q_{2/3}) + f(q_S^1)$
$-c_{1/3} < u_H < 0$	$f(q_0) - f(q_{1/3}) + f(q_1)$	$f(q_0) - f(q_{1/3}) + f(q_S^1)$	$f(q_S^1) - f(q_{1/3}) + f(q_1)$	$f(q_S^1) - f(q_{1/3}) + f(q_S^1)$
$u_H < -c_{1/3}$	$f(q_0) - f(q_S^1) + f(q_1)$	$f(q_0) - f(q_S^1) + f(q_S^1)$	$f(q_1)$	$f(q_S^1)$

Table 2.2.2b. Osher's approximate Riemann solver $f(q_0, q_1)$ for the Euler equations: *O*-variant.

Thus, for the *O*-variant, the computation of $f_R(q_0, q_1)$ is found to be a sum of three terms $f(q)$ in general. Therefore, the computation of $f_R(q_0, q_1)$ takes 7 exponentials, 6 logarithms and some elementary operations and Boolean evaluations. From the point of view of efficiency the *P*-variant is preferable to the *O*-variant. From table 2.2.2b it is easily seen that $f_R(q_0, q_1)$ is continuous. Finally, we mention the following theorem:

THEOREM (2.2.2a).
Osher's approximate Riemann solver $f_R(q_0, q_1)$ (P-variant or O-variant) has the following properties:

$$(i) \quad f_R(q,q) = f(q) \tag{2.2.2.23}$$

for all admissible state q,

$$(ii) \quad f_R(q_0, q_1) + E f_R(Eq_1, Eq_0) = 0 \tag{2.2.2.24}$$

for all admissible states q_0, q_1, where E is the reflection matrix

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.2.2.25}$$

PROOF.
These relations are evident from a physical point of view. The mathematical verification of these relations requires straightforward but tedious calculation.

□

REMARK.
The relations (2.2.2.23,24) are not typical for Osher's approximate Riemann solver. Other approximate Riemann solvers should obey these relations too, in order to be consistent with the differential equations.

2.3. APPROXIMATE SOLUTION OF THE RIEMANN BOUNDARY PROBLEM.

2.3.1. OSHER'S METHOD.

In sections 2.1 and 2.2 we have discussed the space discretization of the steady Euler equations in the interior of the physical domain Ω according to Osher's method.

In this subsection we consider the computation of the flux at finite-volume boundaries which are part of the boundary of Ω. One of the merits of Osher's method is that this can be performed in a fully consistent way with the interior flux computation.

Suppose $\Omega_{i,j}$ is a control volume and $\partial\Omega_{i+\frac{1}{2},j}$ is part of $\partial\Omega$ (see fig. 2.3.1a).

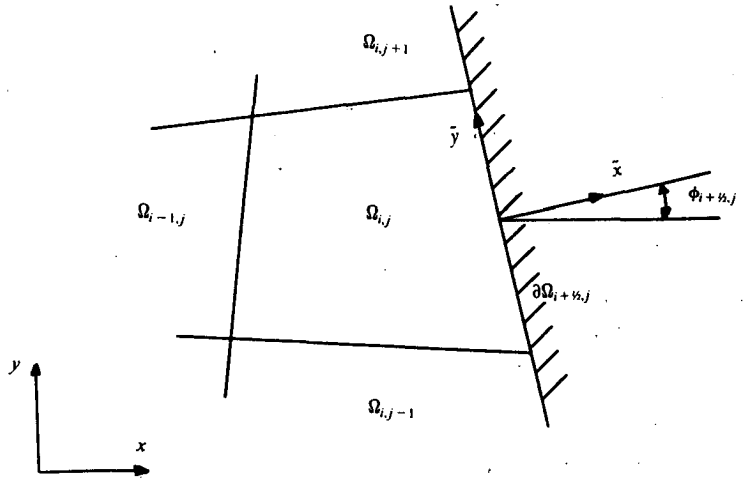


FIGURE 2.3.1a. The boundary $\partial\Omega_{i+\frac{1}{2},j} \subset \partial\Omega$ with local Cartesian frame $(\tilde{x},\tilde{y})$.

In the same manner as for interior control volume boundaries (see fig. 2.1c and (2.1.9)) we choose a local Cartesian frame $(\tilde{x},\tilde{y})$ and consider the Riemann boundary problem

$$\begin{cases} \dfrac{\partial \tilde{q}}{\partial t} + \dfrac{\partial}{\partial \tilde{x}} f(\tilde{q}) = 0 \\ \tilde{q}(\tilde{x},0) = \tilde{q}_{i,j} = T_{i+\frac{1}{2},j} q_{ij} \quad \tilde{x} < 0 \end{cases} \tag{2.3.1.1}$$

and boundary conditions $B(\tilde{q})|_{\tilde{x}=0} = 0$, with $B: \mathbb{R}^4 \mapsto \mathbb{R}^l$, where $l \in [0, \ldots, 4]$

denotes the number of boundary conditions. Note the use of the abbreviated notation $T_{i+\frac{1}{2},j}=T(\phi_{i+\frac{1}{2},j})$. Suppose that the boundary conditions are such that there is a unique solution $\tilde{q}=\tilde{q}(\tilde{x},t)$, $\tilde{x}<0, t>0$. Then the solution is a similarity solution $\tilde{q}(x,t)=\tilde{q}_R(x/t)$.

Godunov's scheme uses the exact solution and the flux $f_{i+\frac{1}{2},j}$ at $\partial\Omega_{i+\frac{1}{2},j}$ becomes

$$f_{i+\frac{1}{2},j}=l_{i+\frac{1}{2},j}\ T^{-1}_{i+\frac{1}{2},j} f(\tilde{q}_R(0)) \tag{2.3.1.2}$$

(Compare this formula with (2.1.11)). In Osher's scheme, an approximation $\tilde{q}_{i+\frac{1}{2},j}$ of $\tilde{q}_R(0)$ is constructed and the flux $f_{i+\frac{1}{2},j}$ at $\partial\Omega_{i+\frac{1}{2},j}$ is taken as

$$f_{i+\frac{1}{2},j}=l_{i+\frac{1}{2},j}\ T^{-1}_{i+\frac{1}{2},j} f(\tilde{q}_{i+\frac{1}{2},j}) \tag{2.3.1.3}$$

The construction of $\tilde{q}_{i+\frac{1}{2},j}$ is the main topic of this subsection. In this subsection we consider the approximate solution of the Riemann boundary problem for a general hyperbolic system. In the next subsection (2.3.2) the results are applied to the Euler equations. Consider the Riemann boundary problem for a general hyperbolic system:

$$\begin{cases} \dfrac{\partial q}{\partial t}+\dfrac{\partial}{\partial x} f(q)=0 & x<0,\ t>0 \\ q(x,0)=q_L & x<0 \\ B(q)|_{x=0}=0\ ;\ B:\mathbb{R}^n\mapsto\mathbb{R}^l & \end{cases} \tag{2.3.1.4}$$

where $q=(q_1,\ldots,q_n)^T\in\mathbb{R}^n$, $f:\mathbb{R}^n\mapsto\mathbb{R}^n$, $f\in C^1$.

As an example we first consider the linear case $f(q)=Aq$, where A is a constant $n\times n$ matrix (see examples (1.3.1a,b)). Assume that the eigenvalues $\{\lambda_i\}_{i=1,\ldots,n}$ are such that $\lambda_1\leqslant\cdots\leqslant\lambda_k<0<\lambda_{k+1}\leqslant\cdots\leqslant\lambda_n$. The exact solution of (2.3.1.4) is a similarity solution $q(x,t)=\tilde{q}(x/t)$. Represent q_L and $q_B=\tilde{q}(0)$ with respect to the base of eigenvectors $\{R_1,\ldots,R_n\}$;

$$q_L=\sum_{i=1}^{n}\alpha_i R_i\ ,\quad q_B=\sum_{i=1}^{n}\beta_i R_i\ . \tag{2.3.1.5}$$

From the exact solution of the pure initial value problem (1.3.1.4,5) it is clear that

$$\beta_i=\alpha_i\quad i=k+1,\ldots,n. \tag{2.3.1.6}$$

Hence, we need k boundary conditions to specify the state q_B, thus $B:\mathbb{R}^n\mapsto\mathbb{R}^k$. The state q_B is the intersection point of the $n-k$ dimensional manifold $B(q)=0$ and the k dimensional plane through q_L spanned by $\{R_1,\ldots,R_k\}$. In other words, the state q_B lies in the $n-k$ dimensional manifold $B(q)=0$ and the state q_L and q_B can be connected by a continuous (integration) path Γ which is decomposed into k subcurves Γ_i:

$$\Gamma=\bigcup_{i=1}^{k}\Gamma_i \tag{2.3.1.7}$$

where each subcurve Γ_i is tangential to the eigenvector R_i. The subcurve Γ_1 starts in q_L and the subcurve Γ_k ends in q_B (P-variant) or the subcurve Γ_k starts in q_L and the subcurve Γ_1 ends in q_B (O-variant). In the linear case, q_B is the same for the P- and O-variant.

These considerations lead to a straightforward generalization of the computation of q_B for a general hyperbolic system. Suppose $f(q)$ is nonlinear and the eigenvalues of $A(q)=\frac{df}{dq}(q)$ are such that $\lambda_1(q)\leqslant\cdots\leqslant\lambda_n(q)$. Assume that

$$\lambda_1(q_L)\leqslant\cdots\leqslant\lambda_k(q_L)<0<\lambda_{k+1}(q_L)\leqslant\cdots\leqslant\lambda_n(q_L) \tag{2.3.1.8}$$

Then we need k boundary conditions: $B:\mathbb{R}^n\mapsto\mathbb{R}^k$. The boundary state q_B lies in the $n-k$ dimensional manifold $B(q)=0$ and can be connected with q_L by a continuous (integration) path Γ which is decomposed in k subcurves Γ_i: $\Gamma=\bigcup_{i=1}^{k}\Gamma_i$ where each subcurve Γ_i is tangential to the eigenvector $R_i(q)$.
In the P-variant, the subcurve Γ_1 starts in q_L and the subcurve Γ_k ends in q_B. In the O-variant the subcurve Γ_k starts in q_L and the subcurve Γ_1 ends in q_B. For both variants, there are $k-1$ intersection points between q_L and q_B, thus there are k unknown states and nk unknowns. Using the Riemann invariants and the k boundary conditions we find $(n-1)k+k$ equations.

REMARK (2.3.1a).
We call (2.3.1.4) a left Riemann boundary problem ($x<0$). A right Riemann boundary problem is defined as

$$\begin{cases}\dfrac{\partial q}{\partial t}+\dfrac{\partial}{\partial x}f(q)=0 & x>0,t>0\\ q(x,0)=q_R & x>0\\ B(q)|_{x=0}=0\ ;\ B:\mathbb{R}^n\mapsto\mathbb{R}^l\end{cases} \tag{2.3.1.9}$$

It is sufficient to consider only left Riemann boundary problems by stipulating that at control volume boundaries which are part of $\partial\Omega$, the boundary state is computed by using a local Cartesian frame $(\tilde{x},\tilde{y})$ such that the positive $\tilde{x}$-axis is directed outward.

REMARK (2.3.1b).
Due to the fact that the integral path $\Gamma=\bigcup_{i=1}^{k}\Gamma_i$ corresponds with negative eigenvalues (see 2.3.1.8) we will usually have that

$$\int_{\Gamma}A^{-1}(q)dq=f(q_B)-f(q_L) \tag{2.3.1.10}$$

Hence, (see 2.2.1.26)

$$f_R(q_L,q_B)=f(q_B). \tag{2.3.1.11}$$

From this observation we see that an alternative for (2.3.1.3) is

$$f_{i+\frac{1}{2},j} = l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(\tilde{q}_{i,j}, \tilde{q}_{i+\frac{1}{2},j})$$
$$= l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(T_{i+\frac{1}{2},j} q_{i,j}, T_{i+\frac{1}{2},j} q_{i+\frac{1}{2},j}) \qquad (2.3.1.12)$$

where

$$q_{i+\frac{1}{2},j} = T^{-1}_{i+\frac{1}{2},j} \tilde{q}_{i+\frac{1}{2},j} \qquad (2.3.1.13)$$

is the boundary state with respect to the (x,y) frame. This expression of the flux agrees with the expression of the flux at interior control volume boundaries (see 2.1.13). For implementation purposes, we prefer (2.3.1.12) to (2.3.1.3).

2.3.2. Application to the Euler equations; boundary condition treatment at inflow, outflow and solid wall.

Consider the Riemann boundary problem (2.3.1.4) for the Euler equations. We consider 5 different cases (k is the number of boundary conditions: see the preceding subsection).

1. *Supersonic Outflow:* ($k=0$)

$$\lambda_1(q_L)>0\ ,\ \lambda_2(q_2)=\lambda_3(q_L)>0\ ,\ \lambda_4(q_L)>0$$

2. *Supersonic Inflow:* ($k=4$)

$$\lambda_1(q_L)<0\ ,\ \lambda_2(q_L)=\lambda_3(q_L)<0\ ,\ \lambda_4(q_L)<0$$

3. *Subsonic Outflow:* ($k=1$)

$$\lambda_1(q_L)<0\ ,\ \lambda_2(q_L)=\lambda_3(q_L)>0\ ,\ \lambda_4(q_L)>0$$

4. *Subsonic Inflow:* ($k=3$)

$$\lambda_1(q_L)<0\ ,\ \lambda_2(q_L)=\lambda_3(q_L)<0\ ,\ \lambda_4(q_L)>0$$

5. *Solid Wall:* ($k=1$ *or* $k=3$)

$$\lambda_1(q_L)<0\ ,\ \lambda_4(q_L)>0.$$

Only in case of subsonic inflow the boundary state q_B is different for the P- and O-variant. Each case is treated as follows.

1. *Supersonic Outflow*: $u_L>c_L$. No boundary condition is to be specified: $q_B=q_L$.
2. *Supersonic Inflow*: $u_L<-c_L$. A full set of four boundary conditions is necessary; $B:\mathbb{R}^4\mapsto\mathbb{R}^4$ and $B(q_B)=0$ has to specify q_B completely.
3. *Subsonic Outflow*: $0<u_L<c_L$. One boundary condition is necessary; $B:\mathbb{R}^4\mapsto\mathbb{R}$. The states q_L and q_B are connected by an integral path Γ_1 which is tangential to $R_1(q)$. Using the Riemann invariant we find

$$u_B+\frac{2}{\gamma-1}c_B=u_L+\frac{2}{\gamma-1}c_L$$

$$v_B = v_L$$
$$z_B = z_L \tag{2.3.2.1}$$

The single boundary condition $B(q_B)=0$ and the 3 relations (2.3.2.1) determine q_B.

EXAMPLE (2.3.2a).
Assume that the pressure p_B is given. From (2.3.2.1) we see that $z_B = z_L$, $v_B = v_L$. Thus

$$\rho_B = (p_B e^{-z_L})^{\frac{1}{\gamma}}$$
$$c_B = \sqrt{\gamma p_B / \rho_B}$$
$$u_B = u_L + \frac{2}{\gamma - 1}(c_L - c_B) \tag{2.3.2.2}$$

4. *Subsonic Inflow*: $-c_L < u_L < 0$. Three boundary conditions are necessary; $B: \mathbb{R}^4 \mapsto \mathbb{R}^3$. There is one intersection point q_I.
In the *P*-variant the states q_L and q_I are connected by Γ_1 and the states q_I and q_B are connected by an integration path that is composed of $\Gamma_2 \cup \Gamma_3$. Using the Riemann invariants we find:

$$u_L + \frac{2}{\gamma - 1} c_L = u_I + \frac{2}{\gamma - 1} c_I$$
$$v_L = v_I$$
$$z_L = z_I$$
$$u_I = u_B$$
$$p_I = p_B \tag{2.3.2.3}$$

Together with the three boundary conditions we have 8 relations and 8 unknowns (the components of q_I and q_B).

In the *O*-variant the states q_L and q_I are connected by an integration path that is composed of $\Gamma_2 \cup \Gamma_3$ and the states q_I and q_B are connected by Γ_1. Using the Riemann invariants we find

$$u_L = u_I$$
$$p_L = p_I$$
$$u_I + \frac{2}{\gamma - 1} c_I = u_B + \frac{2}{\gamma - 1} c_B$$
$$v_I = v_B$$
$$z_I = z_B. \tag{2.3.2.4}$$

EXAMPLE (2.3.2b).
Assume that u_B, v_B and z_B are prescribed.
P-Variant.
Using the relations (2.3.2.3) we find

$$v_I = v_L \ , \ z_I = z_L \ , \ u_I = u_B$$

$$c_I = c_L + \frac{\gamma - 1}{2}(u_L - u_B)$$

$$\rho_I = \left[\frac{c_I^2}{\gamma} e^{-z_L}\right]^{\frac{1}{\gamma - 1}}$$

$$p_B = p_I = \frac{\rho_I c_I^2}{\gamma}$$

$$\rho_B = \left[p_B e^{-z_B}\right]^{\frac{1}{\gamma}}$$

$$c_B = \sqrt{\gamma p_B / \rho_B}. \qquad (2.3.2.5)$$

O-Variant
Using the relations (2.3.2.4) we find

$$u_I = u_L \ , \ z_I = z_B \ , \ v_I = v_B \ , \ p_I = p_L$$

$$\rho_I = \left[p_L e^{-z_B}\right]^{\frac{1}{\gamma}} \ , \ c_I = \sqrt{\gamma p_L / \rho_I}$$

$$c_B = c_I + \frac{\gamma - 1}{2}(u_L - u_B). \qquad (2.3.2.6)$$

EXAMPLE (2.3.2c).
Assume that u_B, v_B and c_B are prescribed.
P-Variant
Using the relations (2.3.2.3) we find that v_I, z_I, u_I, c_I, ρ_I, $p_B = p_I$ are the same as in (2.3.2.5) and

$$\rho_B = \frac{\gamma p_B}{c_B^2}$$

$$z_B = \ln\left[\frac{p_B}{\rho_B^{\gamma}}\right] \qquad (2.3.2.7)$$

O-Variant
Using the relations (2.3.2.4) we find

$$u_I = u_L \ , \ v_I = v_B \ , \ p_I = p_L$$

$$c_I = c_B + \frac{\gamma - 1}{2}(u_B - u_L)$$

$$\rho_I = \frac{\gamma p_L}{c_I^2}$$

$$z_B = z_I = \ln\left[\frac{p_L}{\rho_I^\gamma}\right] \tag{2.3.2.8}$$

5. *Solid Wall*: At a solid wall, one boundary condition is prescribed, namely $u_B = 0$. The state q_B is computed as in the case of subsonic outflow. We find (see 2.3.2.1):

$$\begin{aligned} u_B + \frac{2}{\gamma-1} c_B &= u_L + \frac{2}{\gamma-1} c_L \\ v_B &= v_L \\ z_B &= z_L \\ u_B &= 0 \end{aligned} \tag{2.3.2.9}$$

Hence,

$$\begin{aligned} c_B &= c_L + \frac{\gamma-1}{2} u_L \\ \rho_B &= \left[\frac{c_B^2}{\gamma} e^{-z_L}\right]^{\frac{1}{\gamma-1}} \\ p_B &= \frac{\rho_B c_B^2}{\gamma} . \end{aligned} \tag{2.3.2.10}$$

Notice that the flux $f(q_B) = (0, p_B, 0, 0)$. Hence, the pressure p_B determines the flux completely.
It is also possible to compute the flux $f(q_B)$ as in the case of subsonic inflow. Then q_B is not uniquely determined, but, in case of the *P*-variant, the pressure p_B is uniquely determined and given by (2.3.2.10). This is not true for the *O*-variant.

2.4. Linearization of Osher's scheme

2.4.1. Introduction

The space discretization of the steady Euler equations in the interior and at the boundary of a physical domain Ω is described in the preceding sections of this chapter. For an interior control volume $\Omega_{i,j}$, we have (see (2.1.15)):

$$F_{i,j} \equiv f_{i+\frac{1}{2},j} + f_{i,j+\frac{1}{2}} - f_{i-\frac{1}{2},j} - f_{i,j-\frac{1}{2}} = 0 \tag{2.4.1.1}$$

with

$$f_{i+\frac{1}{2},j} = l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(T_{i+\frac{1}{2},j} q_{i,j} \ , \ T_{i+\frac{1}{2},j} q_{i+1,j}) \tag{2.4.1.2a}$$

$$f_{i,j+\frac{1}{2}} = l_{i,j+\frac{1}{2}} T^{-1}_{i,j+\frac{1}{2}} f_R(T_{i,j+\frac{1}{2}} q_{i,j} \ , \ T_{i,j+\frac{1}{2}} q_{i,j+1}) \qquad (2.4.1.2b)$$

where $l_{i+\frac{1}{2},j}$ is the length of $\partial\Omega_{i+\frac{1}{2},j}$, $T_{i+\frac{1}{2},j} = T(\phi_{i+\frac{1}{2},j})$ and $(\cos\phi_{i+\frac{1}{2},j}, \sin\phi_{i+\frac{1}{2},j})$ is the unit normal on $\partial\Omega_{i+\frac{1}{2},j}$ directed from $\Omega_{i,j}$ to $\Omega_{i+1,j}$ (see fig. 2.1b). Similarly, $l_{i,j+\frac{1}{2}}$ is the length of $\partial\Omega_{i,j+\frac{1}{2}}$, $T_{i,j+\frac{1}{2}} = T(\phi_{i,j+\frac{1}{2}})$ and $(\cos\phi_{i,j+\frac{1}{2}}, \sin\phi_{i,j+\frac{1}{2}})$ is the unit normal on $\partial\Omega_{i,j+\frac{1}{2}}$ directed from $\Omega_{i,j}$ to $\Omega_{i,j+1}$. Notice that

$$F_{i,j} = F_{i,j}(q_{i,j}, q_{i+1,j}, q_{i,j+1}, q_{i-1,j}, q_{i,j-1}) \qquad (2.4.1.3)$$

A nonlinear relaxation method is used in the solution method for solving the system of discretized equations (see chapter 3). For our nonlinear relaxation method we need $\frac{\partial F_{i,j}}{\partial q_{i,j}}$ (a 4×4 matrix). Because

$$\begin{aligned}
\frac{\partial F_{i,j}}{\partial q_{i,j}} &= l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} \frac{\partial f_R}{\partial q_L}(T_{i+\frac{1}{2},j} q_{i,j} \ , \ T_{i+\frac{1}{2},j} q_{i+1,j}) T_{i+\frac{1}{2},j} \\
&+ l_{i,j+\frac{1}{2}} T^{-1}_{i,j+\frac{1}{2}} \frac{\partial f_R}{\partial q_L}(T_{i,j+\frac{1}{2},j} q_{i,j} \ , \ T_{i,j+\frac{1}{2}} q_{i,j+1}) T_{i,j+\frac{1}{2}} \\
&- l_{i-\frac{1}{2},j} T^{-1}_{i-\frac{1}{2},j} \frac{\partial f_R}{\partial q_R}(T_{i-\frac{1}{2},j} q_{i-1,j} \ , \ T_{i-\frac{1}{2},j} q_{i,j}) T_{i-\frac{1}{2},j} \\
&- l_{i,j-\frac{1}{2}} T^{-1}_{i,j-\frac{1}{2}} \frac{\partial f_R}{\partial q_R}(T_{i,j-\frac{1}{2}} q_{i,j-1} \ , \ T_{i,j-\frac{1}{2}} q_{i,j}) T_{i,j-\frac{1}{2}}
\end{aligned} \qquad (2.4.1.4)$$

we need expressions for $\frac{\partial f_R}{\partial q_R}(q_L, q_R)$ and $\frac{\partial f_R}{\partial q_L}(q_L, q_R)$. These expressions are given in subsection 2.4.2.

For a control volume $\Omega_{i,j}$ which has a boundary (say $\partial\Omega_{i+\frac{1}{2},j}$ which is part of $\partial\Omega$ we also have (2.4.1.1) but (see (2.3.1.12, 13)):

$$f_{i+\frac{1}{2},j} = l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(T_{i+\frac{1}{2},j} q_{i,j} \ , \ T_{i+\frac{1}{2},j} q_{i+\frac{1}{2},j}) \qquad (2.4.1.5)$$

where $q_{i+\frac{1}{2},j}$ is the boundary state at $\partial\Omega_{i+\frac{1}{2},j}$: $q_{i+\frac{1}{2},j}$ is determined by $q_{i,j}$ and the boundary conditions. Suppose that $q_B = q_B(q_L)$ is the solution of the Riemann boundary problem corresponding with the boundary conditions at $\partial\Omega_{i+\frac{1}{2},j}$. Then we have (see (2.3.1.1) and (2.3.1.13),

$$q_{i+\frac{1}{2},j} = T^{-1}_{i+\frac{1}{2},j} q_B(T_{i+\frac{1}{2},j} q_{i,j}) \qquad (2.4.1.6)$$

and

$$\frac{dq_{i+\frac{1}{2},j}}{dq_{i,j}} = T^{-1}_{i+\frac{1}{2},j} \frac{dq_B}{dq_L}(T_{i+\frac{1}{2},j} q_{i,j}) T_{i+\frac{1}{2},j} \ . \qquad (2.4.1.7)$$

The linearization of $f_{i+\frac{1}{2},j}$ with respect to $q_{i,j}$ now becomes

$$\begin{aligned}
\frac{\partial f_{i+\frac{1}{2},j}}{\partial q_{i,j}} &= l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} \frac{\partial f_R}{\partial q_L}(T_{i+\frac{1}{2},j} q_{i,j} \ , \ T_{i+\frac{1}{2},j} q_{i+\frac{1}{2},j}) T_{i+\frac{1}{2},j} \\
&+ l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} \frac{\partial f_R}{\partial q_R}(T_{i+\frac{1}{2},j} q_{i,j} \ , \ T_{i+\frac{1}{2},j} q_{i+\frac{1}{2},j}) T_{i+\frac{1}{2},j} \frac{dq_{i+\frac{1}{2},j}}{dq_{i,j}}
\end{aligned} \qquad (2.4.1.8)$$

where $\frac{dq_{i+\frac{1}{2},j}}{dq_{i,j}}$ is given by (2.4.1.7).

Subsection 2.4.3 gives $\frac{dq_B}{dq}(q_L)$ for the boundary conditions as described in subsection 2.3.2.

Notice that if $\partial\Omega_{i+\frac{1}{2},j} \subset \partial\Omega$ then $(\cos\phi_{i+\frac{1}{2},j}, \sin\phi_{i+\frac{1}{2},j})$ is directed outward from Ω, but if $\partial\Omega_{i-\frac{1}{2},j} \subset \partial\Omega$ then $(\cos\phi_{i-\frac{1}{2},j}, \sin\phi_{i-\frac{1}{2},j})$ is directed inward to Ω. (Similar observations hold for the case $\partial\Omega_{i,j+\frac{1}{2}} \subset \partial\Omega$ or $\partial\Omega_{i,j-\frac{1}{2}} \subset \partial\Omega$). Then, due to the conventions introduced in remark 2.3.1a we have

$$q_{i-\frac{1}{2},j} = T(\phi_{i-\frac{1}{2},j}+\pi)^{-1} q_B(T(\phi_{i-\frac{1}{2},j}+\pi)q_{i,j})$$

where $q_B = q_B(q_L)$ is the solution of the Riemannn boundary problem corresponding with the boundary conditions at $\partial\Omega_{i-\frac{1}{2},j}$. Furthermore,

$$f_{i-\frac{1}{2},j} = T^{-1}_{i-\frac{1}{2},j} f_R(T_{i-\frac{1}{2},j}q_{i-\frac{1}{2},j}\ ,\ T_{i-\frac{1}{2},j}q_{i,j})$$

with $T_{i-\frac{1}{2},j} = T(\phi_{i-\frac{1}{2},j})$. The linearization of $q_{i-\frac{1}{2},j}, f_{i-\frac{1}{2},j}$ with respect to $q_{i,j}$ is similar to (2.4.1.7, 8).

2.4.2. Linearization of Osher's approximate Riemann solver.

The topic of this subsection is the computation of

$$\frac{\partial}{\partial q_L} f_R(q_L,q_R)\ ,\ \frac{\partial}{\partial q_R} f_R(q_L,q_R) \tag{2.4.2.1}$$

where $f_R : \mathbb{R}^4 \mapsto \mathbb{R}^4$ is Osher's approximate Riemann solver. For the computation of (2.4.2.1) we need the Jacobian of the flux

$$f'(q) = \frac{df}{dq} = \frac{\partial(\rho u, \rho u^2+p, \rho uv, (E+p)u)}{\partial(c,u,v,z)}\ . \tag{2.4.2.2}$$

The flux-vector $f = (\rho u, \rho u^2+p, \rho uv, (E+p)u)^T$ is found as a function of q by noting that (see 2.2.2.13):

$$\rho = \left[\frac{c^2 e^{-z}}{\gamma}\right]^{\frac{1}{\gamma-1}}$$

$$p = \frac{\rho c^2}{\gamma}$$

$$E = \tfrac{1}{2}\rho(u^2+v^2) + \frac{p}{\gamma-1} \tag{2.4.2.3}$$

From these relations is easily seen that

$$\begin{aligned}
&\frac{\partial\rho}{\partial c} = \frac{2\rho}{(\gamma-1)c}\ ;\ \frac{\partial\rho}{\partial z} = -\frac{\rho}{\gamma-1}\\
&\frac{\partial p}{\partial c} = \frac{2\rho c}{\gamma-1}\ ;\ \frac{\partial p}{\partial z} = -\frac{p}{\gamma-1}\\
&\frac{\partial E}{\partial c} = \frac{2(E+p)}{(\gamma-1)c}\ ;\ \frac{\partial E}{\partial z} = -\frac{E}{\gamma-1}
\end{aligned} \tag{2.4.2.4}$$

Hence,

$$f'(q)=\begin{bmatrix} \frac{2\rho u}{(\gamma-1)c} & \rho & 0 & -\frac{\rho u}{\gamma-1} \\ \frac{2\rho(u^2+c^2)}{(\gamma-1)c} & 2\rho u & 0 & -\frac{\rho u^2+p}{\gamma-1} \\ \frac{2\rho uv}{(\gamma-1)c} & \rho v & \rho u & -\frac{\rho uv}{\gamma-1} \\ \frac{2u(E+(\gamma+1)p)}{(\gamma-1)c} & E+p+\rho u^2 & \rho uv & -\frac{(E+p)u}{\gamma-1} \end{bmatrix} . \quad (2.4.2.5)$$

For the computation of (2.4.2.1) we consider both the *P*- and *O-Variant*.

P-Variant.
From table 2.2.2a we deduce the following tables ($q_0 \equiv q_L$; $q_1 \equiv q_R$):

	$u_0<c_0$	$u_0>c_0$
$c_{1/3}<u_H$	$f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}$	$f'(q_0)$
$0<u_H<c_{1/3}$	$f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_0}$	$f'(q_0)-f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}+f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_0}$
$-c_{2/3}<u_H<0$	$f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_0}$	$f'(q_0)-f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}+f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_0}$
$u_H<-c_{2/3}$	0	$f'(q_0)-f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}$

TABLE 2.4.2a: $\frac{\partial f_R}{\partial q_0}(q_0,q_1)$ for the *P*-Variant.

	$u_1>-c_1$	$u_1<-c_1$
$c_{1/3}<u_H$	0	$-f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}+f'(q_1)$
$0<u_H<c_{1/3}$	$f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_1}$	$f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_1}-f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}+f'(q_1)$
$-c_{2/3}<u_H<0$	$f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_1}$	$f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_1}-f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}+f'(q_1)$
$u_H<-c_{2/3}$	$f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}$	$f'(q_1)$

TABLE 2.4.2b: $\frac{\partial f_R}{\partial q_1}(q_0,q_1)$ for the *P*-Variant.

Thus we need $\frac{\partial q_S^0}{\partial q_0}$, $\frac{\partial q_{1/3}}{\partial q_0}$ and $\frac{\partial q_{2/3}}{\partial q_0}$. These 4 × 4 matrices are easily

obtained from the relations (2.2.2.3) to (2.2.2.12) which yield

$$\begin{aligned}
\partial\alpha &= -\frac{\alpha}{2\gamma}\partial z_0 \\
\partial\Psi_0 &= \partial u_0 + \frac{2}{\gamma-1}\partial c_0 \\
\partial\Psi_1 &= 0 \\
\partial u_H &= \frac{\alpha}{1+\alpha}\left[\partial\Psi_0 - \frac{1}{\gamma(\gamma-1)}c_{1/3}\partial z_0\right] \\
\partial c_{1/3} &= \frac{\gamma-1}{2}(\partial\Psi_0 - \partial u_H) \\
\partial c_{2/3} &= \frac{\gamma-1}{2}\partial u_H \\
\partial c_S^0 &= \frac{\gamma-1}{\gamma+1}\left[\partial u_0 + \frac{2}{\gamma-1}\partial c_0\right]
\end{aligned} \tag{2.4.2.6}$$

and

$$\begin{aligned}
\partial q_S^0 &= (\partial c_S^0, \partial c_S^0, \partial v_0, \partial z_0)^T \\
\partial q_{1/3} &= (\partial c_{1/3}, \partial u_H, \partial v_0, \partial z_0)^T \\
\partial q_{2/3} &= (\partial c_{2/3}, \partial u_H, 0, 0)^T
\end{aligned} \tag{2.4.2.7}$$

Hence,

$$\frac{\partial q_S^0}{\partial q_0} = \begin{bmatrix} \frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ \frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4.2.8}$$

$$\frac{\partial q_{1/3}}{\partial q_0} = \begin{bmatrix} \frac{1}{1+\alpha} & \frac{\gamma-1}{2(1+\alpha)} & 0 & \frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ \frac{2}{\gamma-1}\cdot\frac{\alpha}{1+\alpha} & \frac{\alpha}{1+\alpha} & 0 & -\frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4.2.9}$$

$$\frac{\partial q_{2/3}}{\partial q_0} = \begin{bmatrix} \frac{\alpha}{1+\alpha} & \frac{\gamma-1}{2}\frac{\alpha}{1+\alpha} & 0 & -\frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ \frac{2}{\gamma-1}\cdot\frac{\alpha}{1+\alpha} & \frac{\alpha}{1+\alpha} & 0 & -\frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (2.4.2.10)$$

The term $\frac{\partial}{\partial q_1} f_R(q_0,q_1)$ follows from table 2.4.2b after the computation of $\frac{\partial q_{1/3}}{\partial q_1}$, $\frac{\partial q_{2/3}}{\partial q_1}$ and $\frac{\partial q_S^1}{\partial q_1}$. These 4 × 4 matrices are also easily obtained from the relations (2.2.2.3) to (2.2.2.12) which yield

$$\begin{aligned} \partial\alpha &= \frac{\alpha}{2\gamma}\partial z_1 \\ \partial\Psi_0 &= 0 \\ \partial\Psi_1 &= \partial u_1 - \frac{2}{\gamma-1}\partial c_1 \\ \partial u_H &= \frac{1}{1+\alpha}\left[\partial\Psi_1 + \frac{1}{\gamma(\gamma-1)}c_{2/3}\partial z_1\right] \\ \partial c_{1/3} &= -\frac{\gamma-1}{2}\partial u_H \\ \partial c_{2/3} &= \frac{\gamma-1}{2}(\partial u_H - \partial\Psi_1) \\ \partial c_S^1 &= \frac{\gamma-1}{\gamma+1}\left[\frac{2}{\gamma-1}\partial c_1 - \partial u_1\right] \end{aligned} \qquad (2.4.2.11)$$

and

$$\begin{aligned} \partial q_{1/3} &= (\partial c_{1/3}, \partial u_H, 0, 0)^T \\ \partial q_{2/3} &= (\partial c_{2/3}, \partial u_H, \partial v_1, \partial z_1)^T \\ \partial q_s^1 &= (\partial c_s^1, -\partial c_s^1, \partial v_1, \partial z_1)^T \end{aligned} \qquad (2.4.2.12)$$

Hence,

$$\frac{\partial q_S^1}{\partial q_1} = \begin{bmatrix} \frac{2}{\gamma+1} & -\frac{\gamma-1}{\gamma+1} & 0 & 0 \\ -\frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (2.4.2.13)$$

$$\frac{\partial q_{2/3}}{\partial q_1} = \begin{bmatrix} \frac{\alpha}{1+\alpha} & -\frac{\gamma-1}{2}\cdot\frac{\alpha}{1+\alpha} & 0 & \frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ -\frac{2}{\gamma-1}\cdot\frac{1}{1+\alpha} & \frac{1}{1+\alpha} & 0 & \frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4.2.14}$$

$$\frac{\partial q_{1/3}}{\partial q_1} = \begin{bmatrix} \frac{1}{1+\alpha} & -\frac{\gamma-1}{2}\cdot\frac{1}{1+\alpha} & 0 & -\frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ -\frac{2}{\gamma-1}\cdot\frac{1}{1+\alpha} & \frac{1}{1+\alpha} & 0 & \frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.4.2.15}$$

It is not difficult to verify that both $\frac{\partial}{\partial q_0} f_R(q_0,q_1)$ and $\frac{\partial}{\partial q_1} f_R(q_0,q_1)$ are continuous functions of q_0 and q_1 as long as $u_H \neq 0$.

O-Variant.

The computation of $\frac{\partial}{\partial q_0} f_R(q_0,q_1)$ and $\frac{\partial}{\partial q_1} f_R(q_0,q_1)$ for the *O*-variant is completely analogous to the computation of these terms for the *P*-variant. We only give the results.

	$u_0 > -c_0$	$u_0 < -c_0$
$c_{2/3} < u_H$	$f'(q_0)$	$f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}$
$0 < u_H < c_{2/3}$	$f'(q_0) - f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_0}$	$f'(q_S^0)\frac{\partial q_S^0}{\partial q_0} - f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_0}$
$-c_{1/3} < u_H < 0$	$f'(q_0) - f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_0}$	$f'(q_S^0)\frac{\partial q_S^0}{\partial q_0} - f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_0}$
$u_H < -c_{1/3}$	$f'(q_0) - f'(q_S^0)\frac{\partial q_S^0}{\partial q_0}$	0

TABLE 2.2.2c: $\frac{\partial f_R}{\partial q_0}(q_0,q_1)$ for the *O*-Variant.

	$u_1<c_1$	$u_1>c_1$
$c_{2/3}<u_H$	$-f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}+f'(q_1)$	0
$0<u_H<c_{2/3}$	$-f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_1}+f'(q_1)$	$-f'(q_{2/3})\frac{\partial q_{2/3}}{\partial q_1}+f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}$
$-c_{1/3}<u_H<0$	$-f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_1}+f'(q_1)$	$-f'(q_{1/3})\frac{\partial q_{1/3}}{\partial q_1}+f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}$
$u_H<-c_{1/3}$	$f'(q_1)$	$+f'(q_S^1)\frac{\partial q_S^1}{\partial q_1}$

TABLE 2.2.2d: $\frac{\partial f_R}{\partial q_1}(q_0,q_1)$ for the *O*-Variant.

Furthermore,

$$\frac{\partial q_S^0}{\partial q_0}=\begin{pmatrix} \frac{2}{\gamma+1} & -\frac{\gamma-1}{\gamma+1} & 0 & 0 \\ -\frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{\partial q_{1/3}}{\partial q_0}=\begin{pmatrix} \frac{1}{1+\alpha} & -\frac{\gamma-1}{2}\cdot\frac{1}{1+\alpha} & 0 & \frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ -\frac{2}{\gamma-1}\cdot\frac{\alpha}{1+\alpha} & \frac{\alpha}{1+\alpha} & 0 & \frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{\partial q_{2/3}}{\partial q_0}=\begin{pmatrix} \frac{\alpha}{1+\alpha} & -\frac{\gamma-1}{2}\cdot\frac{\alpha}{1+\alpha} & 0 & -\frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ -\frac{2}{\gamma-1}\cdot\frac{\alpha}{1+\alpha} & \frac{\alpha}{1+\alpha} & 0 & \frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\frac{\partial q_S^1}{\partial q_1} = \begin{pmatrix} \frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ \frac{2}{\gamma+1} & \frac{\gamma-1}{\gamma+1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{\partial q_{2/3}}{\partial q_1} = \begin{pmatrix} \frac{\alpha}{1+\alpha} & \frac{\gamma-1}{2}\frac{\alpha}{1+\alpha} & 0 & \frac{1}{2\gamma}\frac{c_{2/3}}{1+\alpha} \\ \frac{2}{\gamma-1}\frac{1}{1+\alpha} & \frac{1}{1+\alpha} & 0 & -\frac{1}{\gamma(\gamma-1)}\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{\partial q_{1/3}}{\partial q_1} = \begin{pmatrix} \frac{1}{1+\alpha} & \frac{\gamma-1}{2}\cdot\frac{1}{1+\alpha} & 0 & -\frac{1}{2\gamma}\cdot\frac{c_{2/3}}{1+\alpha} \\ \frac{2}{\gamma-1}\cdot\frac{1}{1+\alpha} & \frac{1}{1+\alpha} & 0 & -\frac{1}{\gamma(\gamma-1)}\cdot\frac{c_{2/3}}{1+\alpha} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Just as for the P-Variant, $\frac{\partial}{\partial q_0} f_R(q_0,q_1)$ and $\frac{\partial}{\partial q_1} f_R(q_0,q_1)$ are continuous functions of q_0, q_1 as long as $u_H \neq 0$.

2.4.3. Linearization of boundary conditions

In this subsection we consider the computation of $\frac{dq_B}{dq_L}$ where $q_B = q_B(q_L)$ is an approximate solution of a Riemann boundary problem. The computation is done for the boundary conditions described in subsection 2.3.2. The relations given in (2.4.2.4) are useful in the computation of $\frac{dq_B}{dq_L}$. We consider:

1. *Supersonic Outflow.*

$$q_B = q_L \text{ and } \frac{dq_B}{dq_L} = I\,.$$

2. *Supersonic Inflow.*
q_B is completely determined by the boundary conditions and independent of q_L, $\frac{dq_B}{dq_L}=0$.

3. *Subsonic Outflow.*
Assume that the pressure p_B is prescribed. From (2.3.2.1,2) it follows that

$$\frac{dq_B}{dq_L}=\begin{pmatrix} 0 & 0 & 0 & \frac{c_B}{2\gamma} \\ \frac{2}{\gamma-1} & 1 & 0 & -\frac{c_B}{\gamma(\gamma-1)} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.4.3.1}$$

4. *Subsonic Inflow.*
We consider two cases:

4a. u_B, v_B and z_B are prescribed.
P-Variant.
From (2.3.2.3) and (2.3.2.5) it follows that

$$\frac{dq_B}{dq_L}=\begin{pmatrix} \frac{c_I\rho_I}{c_B\rho_B} & \frac{\gamma-1}{2}\cdot\frac{c_I\rho_I}{c_B\rho_B} & 0 & -\frac{c_B}{2\gamma} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.4.3.2}$$

O-Variant.
From (2.3.2.4) and (2.3.2.6) it follows that

$$\frac{dq_B}{dq_L}=\begin{pmatrix} \frac{c_L\rho_L}{c_I\rho_I} & \frac{\gamma-1}{2} & 0 & -\frac{c_I}{2\gamma} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.4.3.3}$$

4b. u_B, v_B and c_B are prescribed.
P-Variant.
From (2.3.2.3) and (2.3.2.7) it follows that

$$\frac{dq_B}{dq_L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{2\gamma}{c_I} & -\frac{\gamma(\gamma-1)}{c_I} & 0 & 1 \end{bmatrix} \tag{2.4.3.4}$$

O-Variant.
From (2.3.2.4) and (2.3.2.8) it follows that

$$\frac{dq_B}{dq_L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{2\gamma}{c_L} & -\frac{\gamma(\gamma-1)}{c_I} & 0 & 1 \end{bmatrix} \tag{2.4.3.5}$$

5. *Solid Wall.*
$u_B = 0$. From (2.3.2.9) it follows that

$$\frac{dq_B}{dq_L} = \begin{bmatrix} 1 & \frac{(\gamma-1)}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4.3.6}$$

2.5. Second-order discretizations

2.5.1. Introduction

The space discretization of the steady Euler equations described in sections 2.1 and 2.3 is only first-order accurate, as we shall see in the next subsection. It is highly desirable to improve the order of accuracy. In the smooth part of the flow field, first-order accuracy is too low for practical purposes. Furthermore, oblique (with respect to the mesh) shocks and contact discontinuities are smeared out disastrously because of the viscosity hidden in the first-order scheme. Therefore, we wish to improve the order of accuracy and to steepen oblique discontinuities without introducing over- or under-shoots.

The order of accuracy can be improved in a very simple way. The space discretization in the interior of the domain is completely determined by the flux computation at the control volume boundaries. For the first-order scheme, the flux $f_{i+\frac{1}{2},j}$ at the interior control volume boundary $\partial\Omega_{i+\frac{1}{2},j}$ is (see 2.4.1.1,2):

$$f_{i+\frac{1}{2},j}=l_{i+\frac{1}{2},j}T^{-1}_{i+\frac{1}{2},j}f_R(T_{i+\frac{1}{2},j}q_{i,j},\ T_{i+\frac{1}{2},j}q_{i+\frac{1}{2},j}) \tag{2.5.1.1}$$

The order of accuracy is improved by taking

$$f_{i+\frac{1}{2},j}=l_{i+\frac{1}{2},j}T^{-1}_{i+\frac{1}{2},j}f_R(T_{i+\frac{1}{2},j}q^L_{i+\frac{1}{2},j},T_{i+\frac{1}{2},j}q^R_{i+\frac{1}{2},j} \tag{2.5.1.2}$$

where $q^L_{i+\frac{1}{2},j}$ and $q^R_{i+\frac{1}{2},j}$ are obtained by a more accurate interpolation. The states $q^L_{i+\frac{1}{2},j}$ and $q^R_{i+\frac{1}{2},j}$ are located at the left and right side of the volume boundary $\partial\Omega_{i+\frac{1}{2},j}$ (figure 2.5.1a).

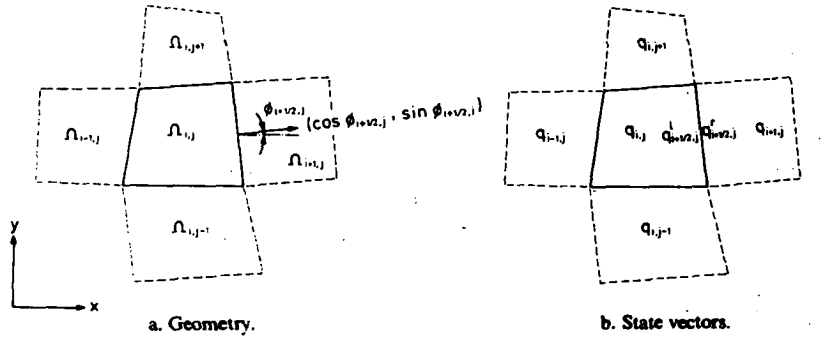


FIGURE 2.5.1a: Finite volume $\Omega_{i,j}$.

Second-order accuracy can be obtained by for example the κ-schemes introduced by Van Leer [29]:

$$q^L_{i+\frac{1}{2},j}=q_{i,j}+\frac{1+\kappa}{4}(q_{i+1,j}-q_{i,j})+\frac{1-\kappa}{4}(q_{i,j}-q_{i-1,j})$$

$$q^R_{i+\frac{1}{2},j}=q_{i+1,j}+\frac{1+\kappa}{4}(q_{i,j}-q_{i+1,j})+\frac{1-\kappa}{4}(q_{i+1,j}-q_{i+2,j}) \tag{2.5.1.3}$$

with $\kappa\in[-1,1]$. For $\kappa=-1$, $\kappa=0$, $\kappa=\frac{1}{3}$ and $\kappa=1$ we find respectively: the fully one-sided upwind scheme, the Fromm scheme, the upwind biased scheme (third-order accurate for 1D problems as we shall see in the next subsection) and the central scheme. A disadvantage of these κ-schemes is that near discontinuities spurious non-monotonicity (wiggles or over- and undershoots) apppears [13].

The space discretization corresponding with (2.5.1.2) is sometimes called the projection-evolution approach [29]. The projection stage is the computation of the states $q^L_{i+\frac{1}{2},j}$ and $q^R_{i+\frac{1}{2},j}$, while the evolution stage is the computation of the flux by an approximate Riemann solver $f_R:\mathbb{R}^4\times\mathbb{R}^4\mapsto\mathbb{R}^4$.

In subsection 2.5.2 the accuracy of the space discretization corresponding with (2.5.1.2) is considered under the assumption that the mesh is sufficiently

smooth.
In subsection 2.5.3 a monotonicity concept is introduced and it is shown that it is possible to construct a monotone second-order accurate scheme. A solution of a monotone second-order scheme has the desired properties: second-order accuracy in the smooth part of the flow field and steepening of oblique discontinuities without introducing spurious non-monotonicity.
The analysis concerns a general nonlinear scalar hyperbolic conservation law. Without the complexity of hyperbolic systems, the analysis is more complete and more transparant. It appears that it is essential to compute the states $q^L_{i+\frac{1}{2},j}$, $q^R_{i+\frac{1}{2},j}$ by nonlinear interpolation. The nonlinear part of the interpolation is called a limiter. The results of the scalar analysis is generalized in a straightforward manner to the Euler equations.

2.5.2. Accuracy on a smooth mesh

Consider the Euler equations

$$\frac{\partial}{\partial t}q+\frac{\partial}{\partial x}f(q)+\frac{\partial}{\partial y}g(q)=0 \tag{2.5.2.1}$$

on an open domain $\Omega\subset\mathbb{R}^2$, q, $f(q)$ and $g(q)$ are given in (2.1.1b). The physical domain Ω is subdivided into disjunct quadrilateral volumes $\Omega_{i,j}$, $(i,j)\in\{1,\ldots,M,1,\ldots,N\}$ in the way described in section 2.1 such that
i) $\Omega=\bigcup_{i,j}\Omega_{i,j}$
ii) $\Omega_{i,j}$, $\Omega_{i\pm1,j}$, $\Omega_{i,j\pm1}$ are neighbouring volumes,
iii) $(x_{i+\frac{1}{2},j+\frac{1}{2}}, y_{i+\frac{1}{2},j+\frac{1}{2}})=\bar{\Omega}_{i,j}\cap\bar{\Omega}_{i+1,j}\cap\bar{\Omega}_{i,j+1}\cap\bar{\Omega}_{i+1,j+1}$ is the common vertex of the volumes $\Omega_{i,j}$, $\Omega_{i+1,j}$, $\Omega_{i,j+1}$ and $\Omega_{i+1,j+1}$.
It is clear that the vertices $\{(x_{i+\frac{1}{2},j+\frac{1}{2}}, y_{i+\frac{1}{2},j+\frac{1}{2}})\}$ define the subdivision of Ω completely.

Let (ξ,η) and (x,y) denote Cartesian coordinates in respectively the computational and physical space. In the computational space we consider a rectangular domain Ω^* subdivided in square control volumes $\Omega^*_{i,j}$, $(i,j)\in\{1\cdots M, 1\cdots N\}$ of uniform size such that $(i\cdot h, j\cdot h)$ is the midpoint of $\Omega^*_{i,j}$; h denotes the length of the edges of the control volumes. Assume the existence of a sufficiently smooth 1-1 mapping between (ξ,η) and (x,y):

$$\begin{cases}\xi=\xi(x,y)\\ \eta=\eta(x,y)\end{cases}\Leftrightarrow\begin{cases}x=x(\xi,\eta)\\ y=y(\xi,\eta)\end{cases} \tag{2.5.2.2}$$

such that the vertices of the control volumes in the physical and computational space are related by this mapping: for all $i,j\in\{0\cdots M, 0\cdots N\}$

$$(x_{i+\frac{1}{2},j+\frac{1}{2}}, y_{i+\frac{1}{2},j+\frac{1}{2}})=(x(\xi_{i+\frac{1}{2}},\eta_{j+\frac{1}{2}}),y(\xi_{i+\frac{1}{2}},\eta_{j+\frac{1}{2}})) \tag{2.5.2.3}$$

where $\xi_{i+\frac{1}{2}}=(i+\frac{1}{2})h$ and $\eta_{j+\frac{1}{2}}=(j+\frac{1}{2})h$.

In the computational space (ξ,η) the Euler equations become

$$\frac{\partial}{\partial t}(Jq)+\frac{\partial}{\partial\xi}(y_\eta f(q)-x_\eta g(q))+\frac{\partial}{\partial\eta}(x_\xi g(q)-y_\xi f(q))=0 \tag{2.5.2.4}$$

where J is the Jacobian of the mapping:

$$J = x_\xi y_\eta - y_\xi x_\eta \,. \tag{2.5.2.5}$$

The discretization of (2.5.2.1) on Ω is equivalent with the discretization of (2.5.2.4) on Ω^*. Therefore, we can study the order of accuracy of a discretization of (2.5.2.4) on Ω^* as well as of (2.5.2.1) on Ω.

Write (2.5.2.4) as

$$(Jq)_t + F(q) = 0 \tag{2.5.2.6}$$

and the steady Euler equations as

$$F(q) = 0 \,. \tag{2.5.2.7}$$

Here $F: X \mapsto Y$ is a nonlinear operator, $X \subset [L^2(\Omega^*)]^4$ is the space of possible fluid states and $Y = [L^2(\Omega^*)]^4$ is the space of rates of change (of states).

Define the finite dimensional vector spaces X_h and Y_h by

$$X_h = Y_h = \{q_{i,j} \in \mathbb{R}^4 \mid i = 1 \cdots M,\ j = 1 \cdots N\} \,.$$

The relation between the spaces X and X_h, Y and Y_h is obtained by introducing $R_h: X \mapsto X_h$ and $\bar{R}_h: Y \mapsto Y_h$:

$$(R_h q)_{i,j} = (\bar{R}_h q)_{i,j} = \frac{1}{h^2} \int\int_{\Omega^*_{i,j}} q(\xi,\eta)\, d\xi\, d\eta \tag{2.5.2.8}$$

for any $q \in [L^2(\Omega)]^4$. Thus $(R_h q)_{i,j}$ is the mean value of q in $\Omega^*_{i,j}$. We define the accuracy of a discretization of (2.5.2.7) as follows:

DEFINITION (2.5.2a).
A p-order accurate discretization of (2.5.2.7) is an associated problem:

$$F^p_h(q_h) = 0$$

where $F^p_h: X_h \mapsto Y_h$ has the property that for all sufficiently smooth $q \in X$

$$(F^p_h R_h q)_{i,j} - (\bar{R}_h F q)_{i,j} = O(h^p) \,. \tag{2.5.2.9}$$

The relation between the various spaces and mappings in the discretization is summarized in the following diagram:

$$\begin{array}{ccc} X & \xrightarrow{\quad F \quad} & Y \\ {\scriptstyle R_h}\downarrow & & \downarrow{\scriptstyle \bar{R}_h} \\ X_h & \xrightarrow{\quad F^p_h \quad} & Y_h \end{array}$$

Notice that

$$(\bar{R}_h F q)_{i,j} = \frac{1}{h^2} \left\{ \int_{\partial\Omega^*_{i+\frac{1}{2},j}} (y_\eta f - x_\eta g)\, (q((i+\tfrac{1}{2})h, \eta))\, d\eta \right.$$

$$
\begin{aligned}
&- \int_{\partial\Omega^*_{i-\frac{1}{2},j}} (y_\eta f - x_\eta g)\,(q((i-\tfrac{1}{2})h,\eta))d\eta \\
&+ \int_{\partial\Omega^*_{i,j+\frac{1}{2}}} (x_\xi g - y_\xi f)\,(q(\xi,(j+\tfrac{1}{2})h))d\xi \\
&- \int_{\partial\Omega^*_{i,j-\frac{1}{2}}} (x_\xi g - y_\xi f)\,(q(\xi,(j-\tfrac{1}{2})h))d\xi \Bigg\}
\end{aligned}
\tag{2.5.2.10}
$$

where $\partial\Omega^*_{i,j+\frac{1}{2}}$ etc, denote the boundaries of the control volume $\Omega^*_{i,j}$, defined by $\partial\Omega^*_{i,j+\frac{1}{2}} = \Omega^*_{i,j} \cap \Omega^*_{i+1,j}$ etc.
For analytical purposes, we introduce the operator $\tilde{F}_h : X_h \mapsto Y_h$ defined by

$$
\begin{aligned}
(\tilde{F}_h q)_{i,j} = \frac{1}{h}\Big\{ &(y_{\eta_{i+\frac{1}{2},j}} f - x_{\eta_{i+\frac{1}{2},j}} g)\,(\bar{q}_{i+\frac{1}{2},j}) - (y_{\eta_{i-\frac{1}{2},j}} f - x_{\eta_{i-\frac{1}{2},j}} g)\,(\bar{q}_{i-\frac{1}{2},j}) \\
&+ (x_{\xi_{i,j+\frac{1}{2}}} g - y_{\xi_{i,j+\frac{1}{2}}} f)\,(\bar{q}_{i,j+\frac{1}{2}}) - (x_{\xi_{i,j-\frac{1}{2}}} g - y_{\xi_{i,j-\frac{1}{2}}} f)\,(\bar{q}_{i,j-\frac{1}{2}})\Big\}
\end{aligned}
\tag{2.5.2.11}
$$

where

$$
\begin{aligned}
&y_{\eta_{i+\frac{1}{2},j}} = \frac{1}{h}(y_{i+\frac{1}{2},j+\frac{1}{2}} - y_{i+\frac{1}{2},j-\frac{1}{2}}) \,,\; x_{\eta_{i+\frac{1}{2},j}} = \frac{1}{h}(x_{i+\frac{1}{2},j+\frac{1}{2}} - x_{i+\frac{1}{2},j-\frac{1}{2}}) \\
&x_{\xi_{i,j+\frac{1}{2}}} = \frac{1}{h}(x_{i+\frac{1}{2},j+\frac{1}{2}} - x_{i-\frac{1}{2},j+\frac{1}{2}}) \,,\; y_{\xi_{i,j+\frac{1}{2}}} = \frac{1}{h}(y_{i+\frac{1}{2},j+\frac{1}{2}} - y_{i-\frac{1}{2},j+\frac{1}{2}})
\end{aligned}
\tag{2.5.2.12}
$$

and

$$
\bar{q}_{i+\frac{1}{2},j} = \frac{1}{h}\int_{\partial\Omega^*_{i+\frac{1}{2},j}} q((i+\tfrac{1}{2})h,\eta)d\eta \,. \tag{2.5.2.13}
$$

Thus $\bar{q}_{i+\frac{1}{2},j}$ is the mean value of $q(\xi,\eta)$ at the control volume boundary $\partial\Omega^*_{i+\frac{1}{2},j}$. In a similar way $\bar{q}_{i-\frac{1}{2},j}, \bar{q}_{i,j+\frac{1}{2}}$ and $\bar{q}_{i,j-\frac{1}{2}}$ are defined.
Assuming a sufficiently smooth mapping it can be shown by elementary interpolation theory that

$$
(\tilde{F}_h q)_{i,j} - (\bar{R}_h F q)_{i,j} = O(h^2) \tag{2.5.2.14}
$$

for all sufficiently smooth $q \in X$.
Even when $x(\xi,\eta)=\xi$, $y(\xi,\eta)=\eta$ the righthandside of (2.5.2.14) is not zero. This is due to the fact that at a cell boundary the mean flux differs from the flux computed in the mean state. But for one-dimensional problems we have $\tilde{F}_h = \bar{R}_h F$.
We only consider discretizations in which states are interpolated (see 2.5.1.2,3). From (2.5.2.14) we conclude that the order of accuracy of such discretizations is at most two. (For 1D problems the order of accuracy can be higher than two.)
From (2.5.2.14) we see that if $F^p_h : X_h \mapsto Y_h$, $p = 1, 2$ is such that

$$(F_h^R R_h q)_{i,j} - (\tilde{F}_h q)_{i,j} = O(h^p) \tag{2.5.2.15}$$

then also

$$(F_h^R R_h q)_{i,j} - (\bar{R}_h F q)_{i,j} = O(h^p)\,. \tag{2.5.2.16}$$

This means that we may approximate $\tilde{F}_h$ instead of $\bar{R}_h F$. Therefore, let us look more carefully at $\tilde{F}_h$. Define $l_{i+\frac{1}{2},j}$, $l_{i,j+\frac{1}{2}}$ by

$$l_{i+\frac{1}{2},j} = h(y^2_{\eta_{i+\frac{1}{2},j}} + x^2_{\eta_{i+\frac{1}{2},j}})^{\frac{1}{2}}\,,\quad l_{i,j+\frac{1}{2}} = h(x^2_{\xi_{i,j+\frac{1}{2}}} + y^2_{\xi_{i,j+\frac{1}{2}}})^{\frac{1}{2}} \tag{2.5.2.17}$$

and $\phi_{i+\frac{1}{2},j}$, $\phi_{i,j+\frac{1}{2}}$ by

$$\begin{aligned} l_{i+\frac{1}{2},j}\cos\phi_{i+\frac{1}{2},j} &= \;\; h y_{\eta_{i+\frac{1}{2},j}},\quad l_{i+\frac{1}{2},j}\sin\phi_{i+\frac{1}{2},j} = -h x_{\eta_{i+\frac{1}{2},j}} \\ l_{i,j+\frac{1}{2}}\cos\phi_{i,j+\frac{1}{2}} &= -h y_{\xi_{i,j+\frac{1}{2}}},\quad l_{i,j+\frac{1}{2}}\sin\phi_{i,j+\frac{1}{2}} = \;\; h x_{\xi_{i,j+\frac{1}{2}}} \end{aligned} \tag{2.5.2.18}$$

then using the rotational invariance (1.3.2.4,5) we see from (2.5.2.11) that

$$\begin{aligned} (\tilde{F}_h q)_{i,j} = \frac{1}{h^2}\Big\{ & l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f(T_{i+\frac{1}{2},j}\bar{q}_{i+\frac{1}{2},j}) - l_{i-\frac{1}{2},j} T^{-1}_{i-\frac{1}{2},j} f(T_{i-\frac{1}{2},j}\bar{q}_{i-\frac{1}{2},j}) \\ & + l_{i,j+\frac{1}{2}} T^{-1}_{i,j+\frac{1}{2}} f(T_{i,j+\frac{1}{2}}\bar{q}_{i,j+\frac{1}{2}}) - l_{i,j-\frac{1}{2}} T^{-1}_{i,j-\frac{1}{2}} f(T_{i,j-\frac{1}{2}}\bar{q}_{i,j-\frac{1}{2}})\Big\} \end{aligned} \tag{2.5.2.19}$$

where $T_{i+\frac{1}{2},j} = T(\phi_{i+\frac{1}{2},j})$ etc.
Notice that $l_{i+\frac{1}{2},j}$ is the length of the boundary $\partial\Omega_{i+\frac{1}{2},j}$ and $(\cos\phi_{i+\frac{1}{2},j}, \sin\phi_{i+\frac{1}{2},j})$ is the outward unit normal on $\partial\Omega_{i+\frac{1}{2},j}$ directed from $\Omega_{i,j}$ to $\Omega_{i+1,j}$ (under the assumption that the Jacobian of the mapping $J>0$).
We define $F_h^R: X_h^R \mapsto Y_h^R$ as follows

$$\begin{aligned} (F_h^R q_h)_{i,j} = \frac{1}{h^2}\Big\{ & l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(T_{i+\frac{1}{2},j} q^L_{i+\frac{1}{2},j},\; T_{i+\frac{1}{2},j} q^R_{i+\frac{1}{2},j}) \\ & - l_{i-\frac{1}{2},j} T^{-1}_{i-\frac{1}{2},j} f_R(T_{i-\frac{1}{2},j} q^L_{i-\frac{1}{2},j},\; T_{i-\frac{1}{2},j} q^R_{i-\frac{1}{2},j}) \\ & + l_{i,j+\frac{1}{2}} T^{-1}_{i,j+\frac{1}{2}} f_R(T_{i,j+\frac{1}{2}} q^L_{i,j+\frac{1}{2}},\; T_{i,j+\frac{1}{2}}) q^R_{i,j+\frac{1}{2}} \\ & - l_{i,j-\frac{1}{2}} T^{-1}_{i,j-\frac{1}{2}} f_R(T_{i,j-\frac{1}{2}} q^L_{i,j-\frac{1}{2}},\; T_{i,j-\frac{1}{2}} q^R_{i,j-\frac{1}{2}})\Big\} \end{aligned} \tag{2.5.2.20}$$

where $q^L_{i+\frac{1}{2},j}$, $q^R_{i+\frac{1}{2},j}$ etc. are obtained by interpolation of the states $\{q_{i,j}\} = q_h$ and f_R is an approximate Riemann solver.
Now we can establish the following theorem:

THEOREM (2.5.2a).
Let $q \in X$ be sufficiently smooth. Define $q_h \in X_h$ by $q_h = R_h q$ and

$$q_{i,j} = (R_h q)_{i,j} = \frac{1}{h^2}\int\!\!\int_{\Omega_{i,j}} q(\xi,\eta)\,d\xi\,d\eta\,. \tag{2.5.2.21}$$

Define the mean values of q at the control volume boundaries by

$$\bar{q}_{i+\frac{1}{2},j} = \frac{1}{h} \int_{\partial\Omega_{i+\frac{1}{2},j}} q((i+\tfrac{1}{2})h, \eta) d\eta \qquad (2.5.2.22)$$

and a similar formula for $\bar{q}_{i,j+\frac{1}{2}}$.
Define the states $q^L_{i+\frac{1}{2},j}$, $q^R_{i+\frac{1}{2},j}$ etc. by interpolation of the states $\{q_{i,j}\}$ such that

$$\begin{aligned}
q^R_{i+\frac{1}{2},j} - q^R_{i-\frac{1}{2},j} &= \bar{q}_{i+\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j} + O(h^{p+1}) \\
q^L_{i+\frac{1}{2},j} - q^L_{i-\frac{1}{2},j} &= \bar{q}_{i+\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j} + O(h^{p+1}) \\
q^R_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j} &= O(h^p) \\
q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j} &= O(h^p) \qquad (2.5.2.23)
\end{aligned}$$

with $p=1$ or $p=2$, and let similar formula hold for the j-direction. Then F^p_h defined by (2.5.2.20) is a pth-order accurate discretization.

PROOF.
From (2.5.2.14,15,16,19,20) is easily seen that it suffices to show that

$$\begin{aligned}
& l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} \{ f_R(T_{i+\frac{1}{2},j} q^L_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} q^R_{i+\frac{1}{2},j}) - f(T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j}) \} \qquad (2.5.2.24) \\
& - l_{i-\frac{1}{2},j} T^{-1}_{i-\frac{1}{2},j} \{ f_R(T_{i-\frac{1}{2},j} q^L_{i-\frac{1}{2},j},\ T_{i-\frac{1}{2},j} q^L_{i-\frac{1}{2},j}) - f(T_{i-\frac{1}{2},j} \bar{q}_{i-\frac{1}{2},j}) \} = O(h^{p+2}) .
\end{aligned}$$

With the notation

$$\begin{aligned}
\frac{\partial f_R}{\partial q_0} \Big|_{i+\frac{1}{2},j} &= \frac{\partial f_R}{\partial q_0} (T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j}) \\
\frac{\partial f_R}{\partial q_1} \Big|_{i+\frac{1}{2},j} &= \frac{\partial f_R}{\partial q_1} (T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j})
\end{aligned}$$

and assuming that the approximate Riemann solver f_R is sufficiently smooth, it follows by Taylor expansion that

$$\begin{aligned}
& f_R(T_{i+\frac{1}{2},j} q^L_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} q^R_{i+\frac{1}{2},j}) = f_R(T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j}) + \\
& \frac{\partial f_R}{\partial q_0} \Big|_{i+\frac{1}{2},j} \cdot T_{i+\frac{1}{2},j} (q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + \frac{\partial f_R}{\partial q_1} \Big|_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j} (q^R_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + \\
& O(|q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}|^2,\ |q^R_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}|^2) = f(T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j}) + \\
& \frac{\partial f_R}{\partial q_0} \Big|_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j} (q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + \frac{\partial f_R}{\partial q_1} \Big|_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j} (q^R_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + O(h^{2p})
\end{aligned} \qquad (2.5.2.25)$$

where we have used the consistency of f_R: $f_R(q,q) = f(q)$ (see (2.2.2.23)).
With the notation

$$\begin{aligned}
X_{i+\frac{1}{2},j} &= T^{-1}_{i+\frac{1}{2},j} \frac{\partial f_R}{\partial q_0} \Big|_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j} \\
Y_{i+\frac{1}{2},j} &= T^{-1}_{i+\frac{1}{2},j} \frac{\partial f_R}{\partial q_1} \Big|_{i+\frac{1}{2},j} T_{i+\frac{1}{2},j}
\end{aligned}$$

we find

$$
\begin{aligned}
& l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} \{ f_R(T_{i+\frac{1}{2},j} q^L_{i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j} q^R_{i+\frac{1}{2},j}) - f(T_{i+\frac{1}{2},j} \bar{q}_{i+\frac{1}{2},j}) \} \\
&= l_{i+\frac{1}{2},j} X_{i+\frac{1}{2},j} (q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + l_{i+\frac{1}{2},j} Y_{i+\frac{1}{2},j} (q^R_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j}) + O(h^{2p+1}) \\
&= (l_{i-\frac{1}{2},j} + O(h^2))\ (X_{i-\frac{1}{2},j} + O(h))\ (q^L_{i-\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j} + O(h^{p+1})) + \\
&\quad (l_{i-\frac{1}{2},j} + O(h^2))\ (Y_{i-\frac{1}{2},j} + O(h))\ (q^R_{i-\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j} + O(h^{p+1})) + O(h^{2p+1})) \\
&= l_{i-\frac{1}{2},j} X_{i-\frac{1}{2},j} (q^L_{i-\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j}) + l_{i-\frac{1}{2},j} Y_{i-\frac{1}{2},j} (q^R_{i-\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j}) + O(h^{p+2}) + O(h^{2p+1}) \\
&= l_{i-\frac{1}{2},j} T^{-1}_{i-\frac{1}{2},j} \{ f_R(T_{i-\frac{1}{2},j} q^L_{i-\frac{1}{2},j}, T_{i-\frac{1}{2},j} q^R_{i-\frac{1}{2},j}) - f(T_{i-\frac{1}{2},j} \bar{q}_{i-\frac{1}{2},j}) \} + O(h^{p+2})
\end{aligned}
\tag{2.5.2.26}
$$

where we used $l_{i-\frac{1}{2},j} = O(h)$, see (2.5.2.17). From (2.5.2.26), eq. (2.5.2.24) follows directly. □

We consider two interpolations

$$
\begin{aligned}
&(I1): \quad q^L_{i+\frac{1}{2},j} = q_{i,j}\ ;\ q^R_{i+\frac{1}{2}} = q_{i+1,j}\ ; \\
&(I2): \quad q^L_{i+\frac{1}{2},j} = q_{i,j} + \frac{1+\kappa}{4}(q_{i+1,j} - q_{i,j}) + \frac{1-\kappa}{4}(q_{i,j} - q_{i-1,j})\ , \\
&\qquad\quad q^R_{i+\frac{1}{2},j} = q_{i+1,j} + \frac{1+\kappa}{4}(q_{i,j} - q_{i+1,j}) + \frac{1-\kappa}{4}(q_{i+1,j} - q_{i+2,j})\ .
\end{aligned}
$$

We shall show that interpolation $(I1)$ is first-order accurate ($p=1$) and interpolation $(I2)$ is second-order accurate ($p=2$). For these interpolations we only show that

$$
\begin{aligned}
& q^L_{i+\frac{1}{2},j} - q^L_{i-\frac{1}{2},j} = \bar{q}_{i+\frac{1}{2},j} - \bar{q}_{i-\frac{1}{2},j} + O(h^{p+1}) \\
& q^L_{i+\frac{1}{2},j} - \bar{q}_{i+\frac{1}{2},j} = O(h^p)\ ,
\end{aligned}
\tag{2.5.2.27}
$$

the other relations in (2.5.2.23) are derived in a similar way. To verify (2.5.2.27), assume that the midpoint of $\Omega^*_{i,j}$ is (0,0) and

$$
\begin{aligned}
q(\xi,\eta) = {} & q_0 + q_1\xi + q_2\eta + q_3\xi^2 + q_4\xi\eta \\
& + q_5\eta^2 + q_6\xi^3 + q_7\xi^2\eta + q_8\xi\eta^2 + q_9\eta^3 + O(h^4)\ .
\end{aligned}
\tag{2.5.2.28}
$$

Then it is easily seen that

$$
q_{i,j} = \frac{1}{h^2} \int_{-\frac{h}{2}}^{\frac{h}{2}} \int_{-\frac{h}{2}}^{\frac{h}{2}} q(\xi,\eta)\,d\xi\,d\eta = q_0 + q_3\frac{h^2}{12} + q_5\frac{h^2}{12} + O(h^4)
$$

$$
q_{i+1,j} = \frac{1}{h^2} \int_{\frac{h}{2}}^{\frac{3h}{2}} \int_{-\frac{h}{2}}^{\frac{h}{2}} q(\xi,\eta)\,d\xi\,d\eta = q_0 + q_1 h + q_3\frac{13}{12}h^2 + q_5\frac{h^2}{12} + q_6\frac{5}{4}h^3 + q_8\frac{h^3}{12} + O(h^4)
$$

$$
q_{i-1,j} = \frac{1}{h^2} \int_{-\frac{3h}{2}}^{-\frac{h}{2}} \int_{-\frac{h}{2}}^{\frac{h}{2}} q(\xi,\eta)\,d\xi\,d\eta = q_0 - q_1 h + q_3\frac{13}{12}h^2 + q_5\frac{h^2}{12} - q_6\frac{5}{4}h^3 - q_8\frac{h^3}{12} + O(h^4)
$$

$$q_{i-2,j}=\frac{1}{h^2}\int_{-\frac{5}{2}h}^{-\frac{3}{2}h}\int_{-\frac{h}{2}}^{\frac{h}{2}} q(\xi,\eta)d\xi d\eta=q_0-2q_1h+q_3\frac{49}{12}h^2+q_5\frac{h^2}{12}-q_6\frac{17}{2}h^3-q_8\frac{h^3}{6}+O(h^4)$$

$$\bar{q}_{i+\frac{1}{2},j}=\frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} q(\frac{h}{2},\eta)d\eta=q_0+q_1\frac{h}{2}+q_3\frac{h^2}{4}+q_5\frac{h^2}{12}+q_6\frac{h^3}{8}+q_8\frac{h^3}{24}+O(h^4)$$

$$\bar{q}_{i-\frac{1}{2},j}=\frac{1}{h}\int_{-\frac{h}{2}}^{\frac{h}{2}} q(-\frac{h}{2},\eta)d\eta=q_0-q_1\frac{h}{2}+q_3\frac{h^2}{4}+q_5\frac{h^2}{12}-q_6\frac{h^3}{8}-q_8\frac{h^3}{24}+O(h^4)$$

$$\bar{q}_{i+\frac{1}{2},j}-\bar{q}_{i-\frac{1}{2},j}=q_1h+q_6\frac{h^3}{4}+q_8\frac{h^3}{12}+O(h^4)\,. \qquad (2.5.2.29)$$

For interpolation $I1$ we find that

$$q^L_{i+\frac{1}{2},j}-q^L_{i-\frac{1}{2},j}=q_{i,j}-q_{i-1,j}=q_1h-q_3h^2+O(h^3)=\bar{q}_{i+\frac{1}{2},j}-\bar{q}_{i-\frac{1}{2},j}+O(h^2)$$

$$q^L_{i+\frac{1}{2},j}-\bar{q}_{i+\frac{1}{2},j}=-q_1\frac{h}{2}+O(h^2)=O(h)\,. \qquad (2.5.2.30)$$

Thus the interpolation $I1$ leads to a first-order accurate scheme.
For interpolation $I2$ we find that

$$\begin{aligned}q^L_{i+\frac{1}{2},j}-q^L_{i-\frac{1}{2},j}&=\frac{1+\kappa}{4}(q_{i+1,j}-q_{i,j})+(1-\frac{\kappa}{2})\,(q_{i,j}-q_{i-1,j})-(\frac{1-\kappa}{4})\,(q_{i-1,j}-q_{i-2,j})\\&=q_1h+q_6\frac{h^3}{4}(6\kappa-1)+q_8\frac{h^3}{12}+O(h^4)\\&=\bar{q}_{i+\frac{1}{2},j}-\bar{q}_{i-\frac{1}{2},j}+q_6\frac{h^3}{2}(3\kappa-1)+O(h^4) \qquad (2.5.2.31)\end{aligned}$$

$$\begin{aligned}q^L_{i+\frac{1}{2},j}&=q_{i,j}+\frac{1+\kappa}{4}(q_{i+1,j}-q_{i,j})+(\frac{1-\kappa}{4})\,(q_{i-1,j}-q_{i-1,j})\\&=q_0+q_1\frac{h}{2}+q_3\frac{h^2}{12}(1+6\kappa)+q_5\frac{h^2}{12}+O(h^3)\\&=\bar{q}_{i+\frac{1}{2},j}+q_3\frac{h^2}{6}(3\kappa-1)+O(h^3) \qquad (2.5.2.32)\end{aligned}$$

Hence, we have found the following result.

THEOREM (2.5.2b).
Scheme (2.5.2.20) is first-order accurate for interpolation $I1$ and second-order accurate for interpolation $I2$. Furthermore, interpolation $I2$ with $\kappa=\frac{1}{3}$ yields a third-order accurate scheme in 1D.

Schemes using interpolation $I2$ will be called κ-schemes.

2.5.3. MONOTONICITY AND SECOND-ORDER ACCURACY.

Solutions of the aforementioned second-order acccurate κ-schemes suffer from

spurious oscillations (wiggles) in the neighbourhood of discontinuities (shock waves, contact discontinuities). The purpose of this section is to study second-order schemes which do not exhibit spurious oscillations. Such a scheme is called monotone.

Several monotonicity concepts occur in the literature. They are all based on the following scalar conservation law:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial}{\partial x} f(u) = 0 \\ u(x, 0) = u_0(x) \end{cases} \tag{2.5.3.1}$$

which is assumed to be discretized in conservation form as follows:

$$U_i^{n+1} = U_i^n - \lambda(h_{i+\frac{1}{2}} - h_{i-\frac{1}{2}}) \equiv H(U_{i-l}^n, U_{i-l+1}^n, \ldots, U_{i+m}^n) \tag{2.5.3.2}$$

where $\lambda = \dfrac{\Delta t}{\Delta x}$ and

$$h_{i+\frac{1}{2}} = h(U_{i-l+1}^n, \ldots, U_{i+m}^n) \tag{2.5.3.3}$$

is the so-called numerical flux function, satisfying the consistency condition

$$h(U, U, \ldots, U) = f(U) . \tag{2.5.3.4}$$

The main reason for considering the scalar conservation law (2.5.3.1) is the following monotonicity property:

For any weak solution of (2.5.3.1) (see Lax [14], Harten [6]), we have

(M1) No new local maximum or minimum can appear for $t > 0$.

(M2) The value of a local maximum is nonincreasing, that of a local minimum is nondecreasing.

and therefore

(M3) The total variation

$$TV[u(t)] := \sup \sum_i |u(x_{i+1}, t) - u(x_i, t)|$$

is a nonincreasing function of time t.

The common, well known definition of a monotone scheme is due to Harten, Hyman and Lax [5]. They call the finite difference scheme (2.5.3.2) monotone if the function H is a monotone nondecreasing function of its $(l+m+1)$ arguments.

They were able to prove the following theorem:

THEOREM (2.5.3a) (cf. [5]).

Assume that the finite difference scheme (2.5.3.2) is monotone in the sense of Harten, Hyman and Lax. Assume that, as Δt and Δx tend to zero, $\lambda = \Delta t / \Delta x = $ const., U_i^n converges boundedly almost everywhere to some function $u(x,t)$. Then according to the theorem of Lax and Wendroff [15] $u(x,t)$ is a weak solution of (2.5.3.1), moreover an entropy condition is satisfied for all discontinuites of u, i.e. $u(x,t)$ is the unique physicaly significant solution.

For a review of several, more or less equivalent entropy conditions for (2.5.3.1) we refer to [14].
Unfortunately, a scheme which is monotone in the sense of Harten, Hyman and Lax is only first-order accurate [5]. To allow higher order of accuracy, Harten [6] introduced a weaker concept of monotonicity: scheme (2.5.3.2) is called TVD (Total Variation Diminishing) when

$$TV(U^{n+1}) \leqslant TV(U^n) \tag{2.5.3.5}$$

where

$$TV(U^n) = TV(\{U_i^n\}) = \sum_{i=-\infty}^{\infty} |U_i^n - U_{i-1}^n| . \tag{2.5.3.6}$$

A grid function U is called monotone if for all i

$$\min(U_{i-1}, U_{i+1}) \leqslant U_i \leqslant \max(U_{i-1}, U_{i+1}) . \tag{2.5.3.7}$$

Following Harten [6], scheme (2.5.3.2) is called monotonicity preserving if monotonicity of U^{n+1} follows from monotonicity of U^n.

The relation between the above three properties is given by the following theorem.

THEOREM (2.5.3b) (cf. Harten [6]).
(i) *A scheme which is monotone in the sense of Harten, Hyman and Lax is TVD.*
(ii) *A TVD scheme is monotonicity preserving.*

It is well known (see [3,6]) that a linear scheme

$$U_i^{n+1} = \sum_{k=-m}^{m} c_k U_{i+k}^n$$

is monotonicity preserving if and only if

$$c_k \geqslant 0 \quad -m \leqslant k \leqslant m$$

(PROOF.
(a) If $c_k \geqslant 0$ for every k and $U_i^n - U_{i-1}^n \geqslant 0$ for all i, then

$$U_i^{n+1} - U_{i-1}^{n+1} = \sum_{k=-m}^{m} c_k U_{i+k}^n - \sum_{k=-m}^{m} c_k U_{i-1+k}^n$$

$$= \sum_{k=-m}^{m} c_k (U_{i+k}^n - U_{i+k-1}^n) \geqslant 0 .$$

The case of nonincreasing U^n is handled similarly.
(b) Conversely, supposing that $c_{k_0} < 0$, then for the particular function

$$U_i^n = \begin{cases} 1 & i \geqslant k_0 \\ 0 & i < k_0 \end{cases}$$

we obtain

$$U_0^{n+1} - U_{-1}^{n+1} = \sum_{k=-m}^{m} c_k (U_k^n - U_{k-1}^n) = c_{k_0}(U_{k_0}^n - U_{k_0-1}^n) = c_{k_0} < 0$$

and monotonicity is not preserved).

Hence, any *linear* monotonicity-preserving scheme is monotone in the sense of Harten, Hyman and Lax and consequently first-order accurate. By theorem (2.5.3b), any *linear* TVD scheme will also be first-order accurate.
Hence, only nonlinear schemes can be second-order and TVD. For the construction of such schemes, Harten's lemma [6] plays a fundamental role.

LEMMA (2.5.3c) (cf. Harten [6]).
Consider a discretization of (2.5.3.1) given by

$$U_i^{n+1} = U_i^n + \lambda A_{i+\frac{1}{2}}^n (U_{i+1}^n - U_i^n) + \lambda B_{i-\frac{1}{2}}^n (U_{i-1}^n - U_i^n) \qquad (2.5.3.8)$$

where $\lambda = \Delta t / \Delta x$ *and*

$$A_{i+\frac{1}{2}}^n = A(\ldots, U_{i-1}^n, U_i^n, U_{i+1}^n, \ldots)$$
$$B_{i-\frac{1}{2}}^n = B(\ldots, U_{i-1}^n, U_i^n, U_{i+1}^n, \ldots). \qquad (2.5.3.9)$$

If the coefficients $A_{i+\frac{1}{2}}^n$, $B_{i+\frac{1}{2}}^n$ *satisfy*

$$A_{i+\frac{1}{2}}^n \geqslant 0,\ B_{i+\frac{1}{2}}^n \geqslant 0,\ 1 - \lambda A_{i+\frac{1}{2}}^n - \lambda B_{i+\frac{1}{2}}^n \geqslant 0 \qquad (2.5.3.10)$$

then scheme (2.5.3.8) is TVD.

PROOF.
By rewriting (2.5.3.8) with i replaced by $i-1$ and subtracting from (2.5.3.8) one easily obtains

$$\begin{aligned} |U_i^{n+1} - U_{i-1}^{n+1}| \leqslant\ & (1 - \lambda B_{i-\frac{1}{2}}^n - \lambda A_{i-\frac{1}{2}}^n) |U_i^n - U_{i-1}^n| \\ & + \lambda A_{i+\frac{1}{2}}^n |U_{i+1}^n - U_i^n| + \lambda B_{i-3/2}^n |U_{i-1}^n - U_{i-2}^n| \end{aligned}$$

so that

$$\begin{aligned} TV(U^{n+1}) &= \sum_{i=-\infty}^{\infty} |U_i^{n+1} - U_{i-1}^{n+1}| \\ &\leqslant \sum_{i=-\infty}^{\infty} (1 - \lambda B_{i-\frac{1}{2}}^n - \lambda A_{i-\frac{1}{2}}^n) \cdot |U_i^n - U_{i-1}^n| \\ &\quad + \sum_{i=-\infty}^{\infty} \lambda A_{i-\frac{1}{2}}^n |U_i^n - U_{i-1}^n| + \sum_{i=-\infty}^{\infty} \lambda B_{i-\frac{1}{2}}^n |U_i^n - U_{i-1}^n| \\ &= \sum_{i=-\infty}^{\infty} |U_i^n - U_{i-1}^n| = TV(U^n) \end{aligned}$$

□

The usefulness of this lemma for the construction of higher order TVD schemes is shown in [23].

Recently, however, Goodman and Le Veque have obtained the following negative result: every conservative TVD scheme for scalar hyperbolic conservation laws in two space dimensions, is at most first-order accurate [4].
This result forces us to use a monotonicity concept weaker than TVD.
Consider the following nonlinear scalar conservation law in two space dimensions:

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}f(u)+\frac{\partial}{\partial y}g(u)=0 \tag{2.5.3.11}$$

Assume that (2.5.3.11) is discretized on an equidistant mesh with mesh size h.
Consider a discretization of (2.5.3.11) given by

$$\begin{aligned}U_{i,j}^{n+1}=U_{i,j}^{n}&+\lambda A_{i+\frac{1}{2},j}^{n}(U_{i+1,j}^{n}-U_{i,j}^{n})+\lambda B_{i,j+\frac{1}{2}}^{n}(U_{i,j+1}^{n}-U_{i,j}^{n})\\&+\lambda C_{i-\frac{1}{2},j}^{n}(U_{i-1,j}^{n}-U_{i,j}^{n})+\lambda D_{i,j-\frac{1}{2}}^{n}(U_{i,j-1}-U_{i,j}^{n})\end{aligned} \tag{2.5.3.12}$$

where $\lambda=\Delta t/h$ and

$$\begin{aligned}A_{i+\frac{1}{2},j}^{n}&=A\ (...,U_{i-1,j}^{n},U_{i,j}^{n},U_{i+1,j},...)\\B_{i,j+\frac{1}{2}}^{n}&=B\ (...,U_{i,j-1}^{n},U_{i,j}^{n},U_{i,j+1},...)\\C_{i-\frac{1}{2},j}^{n}&=C\ (...,U_{i-1,j}^{n},U_{i,j}^{n},U_{i+1,j},...)\\D_{i,j-\frac{1}{2}}^{n}&=D\ (...,U_{i,j-1}^{n},U_{i,j}^{n},U_{i,j+1},...)\ .\end{aligned} \tag{2.5.3.13}$$

We introduce the following monotonicity concept:

DEFINITION (2.5.3a).
Scheme (2.5.3.12) is called monotone if

$$A_{i+\frac{1}{2},j}^{n}\geqslant 0\ ;\ B_{i,j+\frac{1}{2}}^{n}\geqslant 0\ ;\ C_{i-\frac{1}{2},j}^{n}\geqslant 0\ ;\ D_{i,j-\frac{1}{2}}^{n}\geqslant 0 \tag{2.5.3.14a}$$

and if $A_{i+\frac{1}{2},j}^{n}$, $B_{i,j+\frac{1}{2}}^{n}$, $C_{i-\frac{1}{2},j}^{n}$, $D_{i,j-\frac{1}{2}}^{n}$ are uniformly bounded, i.e. there exists a $B>0$ such that for all (i,j)

$$A_{i+\frac{1}{2},j}^{n}\leqslant B\ ;\ B_{i,j+\frac{1}{2}}^{n}\leqslant B\ ;\ C_{i-\frac{1}{2},j}^{n}\leqslant B\ ;\ D_{i,j-\frac{1}{2}}^{n}\leqslant B \tag{2.5.3.14b}$$

This monotonicity concept is not to be confused with what we called monotonicity in the sense of Harten, Hyman and Lax. Monotonicity is weaker than TVD in more than one space dimension. In two dimensions we define the total variation of U as

$$TV(U)=\sum_{i,j}\{\,|\,U_{i,j}-U_{i-1,j}\,|\,+\,|\,U_{i,j}-U_{i,j-1}\,|\,\} \tag{2.5.3.15}$$

and U is called monotone if for all (i,j)

$$\min(U_{i-1,j}\,,\,U_{i+1,j}\,,\,U_{i,j-1}\,,\,U_{i,j+1})\leqslant U_{i,j}\leqslant\max(U_{i-1,j}\,,\,U_{i+1,j}\,,\,U_{i,j-1}\,,\,U_{i,j+1}) \tag{2.5.3.16}$$

THEOREM (2.5.3d).
(i) *A monotone scheme is TVD in one space dimension.*
(ii) *A monotone scheme is not necessarily TVD in two space dimensions.*

PROOF.
(i) For scheme (2.5.3.8) we have

$$0 \leqslant A^n_{i+\frac{1}{2},j} \leqslant B \; ; \; 0 \leqslant B^n_{i+\frac{1}{2},j} \leqslant B \; .$$

For λ sufficiently small ($\lambda \leqslant \frac{1}{2B}$) it is easily seen that (2.5.3.10) is fulfilled. Hence, (2.5.3.8) is TVD.

(ii) Consider the following grid function

$$U^n_{i,j} = \begin{cases} 1 & \text{for } (i,j)=(1,0) \\ 0 & \text{for } (i,j)\neq(1,0) \end{cases}$$

and let in (2.5.3.12)

$$A^n_{i+\frac{1}{2},j} = B^n_{i,j+\frac{1}{2}} = C^n_{i-\frac{1}{2},j} = D^n_{i,j-\frac{1}{2}} = 0 \quad \forall (i,j) \neq (0,0)$$

$$A_{\frac{1}{2},0} = 1 \; ; \; A_{0,\frac{1}{2}} = 0 \; ; \; C_{-\frac{1}{2},0} = 0 \; ; \; D_{0,-\frac{1}{2}} = 0 \; .$$

Then

$$U^{n+1}_{i,j} = U^n_{i,j} \quad \forall (i,j) \neq (0,0)$$

$$U^{n+1}_{0,0} = \lambda \; .$$

Hence, $TV(U^n)=4$, $TV(U^{n+1})=4+2\lambda$. Because $\lambda>0$ we find $TV(U^{n+1})>TV(U^n)$. Thus we have found a monotone scheme which is not TVD. □

Nevertheless, the monotonicity concept of definition (2.5.3a) has some use, as shown by the following theorem.

THEOREM (2.5.3e).
If scheme (2.5.3.12) is monotone then a steady state solution of (2.5.3.12) is monotone.

PROOF.
From (2.5.3.12) we see that for a steady state solution $\{U_{i,j}\}$ we have

$$U_{i,j} = \frac{A_{i+\frac{1}{2},j} U_{i+1,j} + B_{i,j+\frac{1}{2}} U_{i,j+1} + C_{i-\frac{1}{2},j} U_{i-1,j} + D_{i,j-\frac{1}{2}} U_{i,j-1}}{A_{i+\frac{1}{2},j} + B_{i,j+\frac{1}{2}} + C_{i-\frac{1}{2},j} + D_{i,j-\frac{1}{2}}}$$

which, due to the positivity of the coefficients, proves this theorem immediately. □

Thus a monotone scheme guarantees that in a steady solution spurious oscillations do not occur. There is no contradiction between monotonicity and

second-order accuracy (neither in one nor in more dimensions). This will be shown by constructing a second-order monotone scheme. Consider (2.5.3.11) and suppose that the flux functions $f(u)$ and $g(u)$ can be split in a forward and a backward flux (see (2.2.1.5)):

$$\begin{aligned} f(u) &= f^+(u) + f^-(u) \; ; \\ g(u) &= g^+(u) + g^-(u) \; ; \end{aligned} \tag{2.5.3.17a}$$

where

$$\begin{aligned} &\frac{d}{du} f^+(u) \geqslant 0 \; ; \; \frac{d}{du} f^-(u) \leqslant 0 \; ; \; \forall u \in \mathbb{R} \\ &\frac{d}{du} g^+(u) \geqslant 0 \; ; \; \frac{d}{du} g^-(u) \leqslant 0 \; ; \; \forall u \in \mathbb{R} \; . \end{aligned} \tag{2.5.3.17b}$$

A finite volume discretization is given by

$$\begin{aligned} U_{i,j}^{n+1} = U_{i,j}^n &+ \frac{\Delta t}{h}[\{f^+(U_{i-\frac{1}{2},j}^{L\,n}) + f^-(U_{i-\frac{1}{2},j}^{R\,n})\} - \{f^+(U_{i+\frac{1}{2},j}^{L\,n}) + f^-(U_{i+\frac{1}{2},j}^{R\,n})\}] \\ &+ \frac{\Delta t}{h}[\{g^+(U_{i,j-\frac{1}{2}}^{L\,n}) + g^-(U_{i,j-\frac{1}{2}}^{R\,n})\} - \{g^+(U_{i,j+\frac{1}{2}}^{L\,n}) + g^-(U_{i,j+\frac{1}{2}}^{R\,n})\}] \end{aligned} \tag{2.5.3.18}$$

where

$$\begin{aligned} U_{i+\frac{1}{2},j}^{L\,n} &= U_{i,j}^n + \tfrac{1}{2}\psi(R_{i,j}^n)(U_{i,j}^n - U_{i-1,j}^n) \\ U_{i-\frac{1}{2},j}^{R\,n} &= U_{i,j}^n + \tfrac{1}{2}\psi(\frac{1}{R_{i,j}^n})(U_{i,j}^n - U_{i+1,j}^n) \\ U_{i,j+\frac{1}{2}}^{L\,n} &= U_{i,j}^n + \tfrac{1}{2}\psi(S_{i,j}^n)(U_{i,j}^n - U_{i,j-1}^n) \\ U_{i,j-\frac{1}{2}}^{R\,n} &= U_{i,j}^n + \tfrac{1}{2}\psi(\frac{1}{S_{i,j}^n})(U_{i,j}^n - U_{i,j+1}^n) \end{aligned} \tag{2.5.3.19}$$

and

$$R_{i,j}^n = \frac{U_{i+1,j}^n - U_{i,j}^n}{U_{i,j}^n - U_{i-1,j}^n} \; ; \; S_{i,j}^n = \frac{U_{i,j+1}^n - U_{i,j}^n}{U_{i,j}^n - U_{i,j-1}^n} \; ; \tag{2.5.3.20}$$

and $\psi: \mathbb{R} \mapsto \mathbb{R}$ is a continuous function called the limiter. The value $U_{i,j}^n$ is a numerical approximation of the mean value of u in cell (i,j) at time $t = n\Delta t$, the values $U_{i+\frac{1}{2},j}^{L\,n}$, $U_{i+\frac{1}{2},j}^{R\,n}$ are approximations of $\frac{1}{h}\int_{(j-\frac{1}{2})h}^{(j+\frac{1}{2})h} u((i+\frac{1}{2})h, \eta, n\Delta t)d\eta$ at the left and right side of the cell wall $(i+\frac{1}{2}, j)$. See figure 2.5.3a.

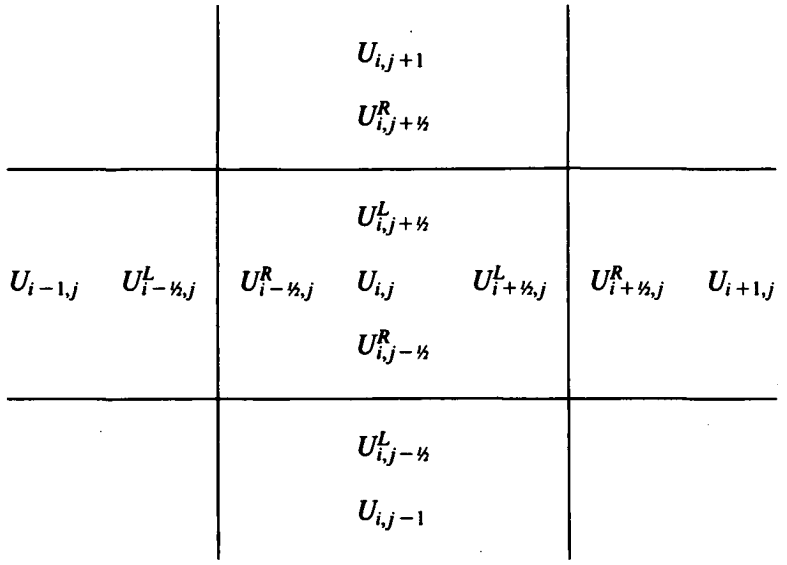


FIGURE 2.5.3a. Location of the various variables in the space discretization.

The limiter $\psi=\psi(R)$ is introduced in the discretization in order to construct a monotone, spatially second-order scheme. The limiter is a function of the consecutive gradients, a common practice in this field [23,25]. Notice that $\psi\equiv 0$ corresponds to the first-order upwind scheme, while $\psi\equiv 1$ yields the fully one-sided second-order upwind scheme ($\kappa=-1$: see (2.5.1.3)). We already know that $\psi(R)$ has to be a nonlinear function.

We wish to show under what conditions scheme (2.5.3.18) is monotone. It can be easily seen that scheme (2.5.3.18) can be written as (2.5.3.12) by taking

$$A^n_{i+½,j}=-\frac{f^-(U^{R}_{i+½,j}{}^n)-f^-(U^{R}_{i-½,j}{}^n)}{U^{R}_{i+½,j}{}^n-U^{R}_{i-½,j}{}^n}\cdot\frac{U^{R}_{i+½,j}{}^n-U^{R}_{i-½,j}{}^n}{U^n_{i+1,j}-U^n_{i,j}}$$

$$C^n_{i-½,j}=+\frac{f^+(U^{L}_{i+½,j}{}^n)-f^+(U^{L}_{i-½,j}{}^n)}{U^{L}_{i+½,j}{}^n-U^{L}_{i-½,j}{}^n}\cdot\frac{U^{L}_{i+½,j}{}^n-U^{L}_{i-½,j}{}^n}{U^n_{i,j}-U^n_{i-1,j}}$$

$$B^n_{i,j+½}=-\frac{g^-(U^{R}_{i,j+½}{}^n)-g^-(U^{R}_{i,j-½}{}^n)}{U^{R}_{i,j+½}{}^n-U^{R}_{i,j-½}{}^n}\cdot\frac{U^{R}_{i,j+½}{}^n-U^{R}_{i,j-½}{}^n}{U^n_{i,j+1}-U^n_{i,j}}$$

$$D^n_{i,j-½}=+\frac{g^+(U^{L}_{i,j+½}{}^n)-g^+(U^{L}_{i,j-½}{}^n)}{U^{L}_{i,j+½}{}^n-U^{L}_{i,j-½}{}^n}\cdot\frac{U^{L}_{i,j+½}{}^n-U^{L}_{i,j-½}{}^n}{U^n_{i,j}-U^n_{i,j-1}} \qquad (2.5.3.21)$$

To obtain positivity of the coefficients $A^n_{i+½,j}$, $B^n_{i,j+½}$ etc, it is sufficient (by the Mean Value theorem) that

$$\frac{U^{R}_{i+½,j}{}^n-U^{R}_{i-½,j}{}^n}{U^n_{i+1,j}-U^n_{i,j}}\geqslant 0\ ;\ \frac{U^{L}_{i+½,j}{}^n-U^{L}_{i-½,j}{}^n}{U^n_{i,j}-U^n_{i-1,j}}\geqslant 0\ ;$$

$$\frac{U^R_{i,j+\frac{1}{2}}{}^n - U^R_{i,j-\frac{1}{2}}{}^n}{U^n_{i,j+1} - U^n_{i,j}} \geqslant 0 \; ; \; \frac{U^L_{i,j+\frac{1}{2}}{}^n - U^L_{i,j-\frac{1}{2}}{}^n}{U^n_{i,j} - U^n_{i,j-1}} \geqslant 0 \, . \tag{2.5.3.22}$$

The coefficients $A^n_{i+\frac{1}{2},j}$, $B^n_{i,j+\frac{1}{2}}$ etc, are uniformly bounded by assuming uniform boundedness of the derivatives of $f^+(u)$, $f^-(u)$, $g^+(u)$, and $g^-(u)$, and taking care that the left handsides of inequalities (2.5.3.22) are also uniformly bounded.

By substitution of (2.5.3.19) in (2.5.3.22) it is easily seen that (2.5.3.22) is fulfilled if

$$1 + \tfrac{1}{2}\psi(R) - \tfrac{1}{2}\psi(S) \cdot \frac{1}{S} \geqslant 0 \quad \forall R,S \in \mathbb{R} \, . \tag{2.5.3.23}$$

Furthermore, uniform boundedness of the left-handside of the inequalities in (2.5.3.22) is obtained by requiring

$$\psi(R) - \psi(S) \cdot \frac{1}{S} \leqslant 2M, \quad \forall R,S \in \mathbb{R}, \; M \in (0,\infty) \, . \tag{2.5.3.24}$$

So, (2.5.3.18) is a monotone scheme if the limiter $\psi = \psi(R)$ satisfies

$$-2 \leqslant \psi(R) - \psi(S) \cdot \frac{1}{S} \leqslant 2M, \quad \forall R,S \in \mathbb{R} \, . \tag{2.5.3.25}$$

This inequality is satisfied if

$$\alpha \leqslant \psi(R) \leqslant M, \quad \forall R \in \mathbb{R} \tag{2.5.3.26a}$$

and

$$-M \leqslant \frac{\psi(R)}{R} \leqslant 2 + \alpha, \quad \forall R \in \mathbb{R} \, . \tag{2.5.3.26b}$$

The monotonicity region given by (2.5.3.26) is depicted in figure (2.5.3b). We assume $\alpha \in [-2,0]$.

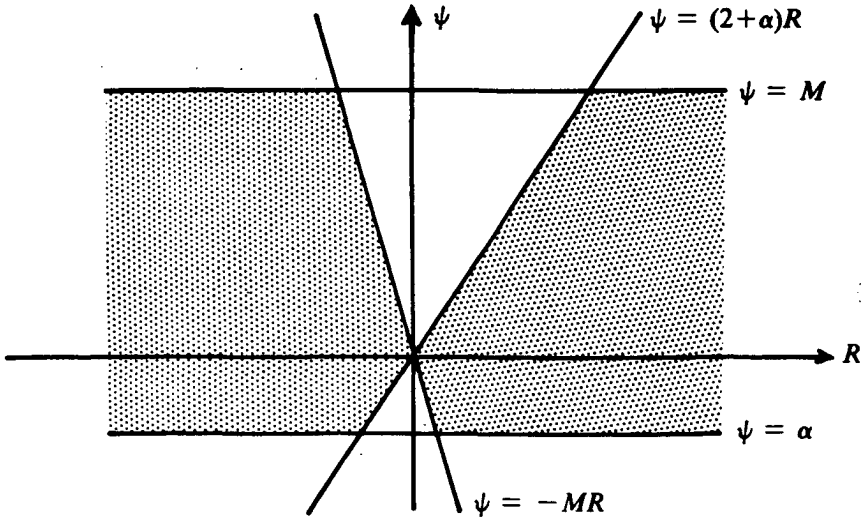


FIGURE (2.5.3b). Monotonicity region.

So, we have found the following theorem.

THEOREM (2.5.3f).
If the limiter $\psi=\psi(R)$ has the properties that there exist constants $M\in(0,\infty),\alpha\in[-2,0]$ such that $\alpha\leqslant\psi(R)\leqslant M$, $-M\leqslant\frac{\psi(R)}{R}\leqslant 2+\alpha,\forall R\in\mathbb{R}$, then (2.5.3.18) is a monotone scheme.

The requirements for this theorem imply that $\psi(0)=0$. Notice that $\psi\equiv 0$, which corresponds with the first-order upwind scheme, result in a monotone scheme, as is to be expected.
Now, we investigate under which conditions scheme (2.5.3.18) is second-order accurate with respect to the space discretization. Define

$$\begin{aligned}
\tilde{U}^L_{i+\frac{1}{2},j} &= U_{i,j}+\tfrac{1}{2}(U_{i,j}-U_{i-1,j})\\
U^L_{i+\frac{1}{2},j} &= U_{i,j}+\tfrac{1}{2}\psi(R_{i,j})\,(U_{i,j}-U_{i-1,j})\\
\tilde{U}^R_{i-\frac{1}{2},j} &= U_{i,j}+\tfrac{1}{2}(U_{i,j}-U_{i+1,j})\\
U^L_{i-\frac{1}{2},j} &= U_{i,j}+\tfrac{1}{2}\psi(\frac{1}{R_{i,j}})\,(U_{i,j}-U_{i-1,j})
\end{aligned}\qquad (2.5.3.27)$$

and similar formulae for $\tilde{U}^{L,R}_{i,j\pm\frac{1}{2}}$ and $U^{L,R}_{i,j\pm\frac{1}{2}}$.
Notice that $\tilde{U}$ corresponds with $\psi\equiv 1$, the fully one-sided second-order upwind scheme ($\kappa=-1$).

THEOREM (2.5.3g).
If the limiter $\psi=\psi(R)$ is constructed such that

$$U^R_{i+\frac{1}{2},j}-U^R_{i-\frac{1}{2},j}=\tilde{U}^R_{i+\frac{1}{2},j}-\tilde{U}^R_{i-\frac{1}{2},j}+O(h^{p+1}) \qquad (2.5.3.28a)$$

$$U^L_{i+\frac{1}{2},j}-U^L_{i-\frac{1}{2},j}=\tilde{U}^L_{i+\frac{1}{2},j}-\tilde{U}^L_{i-\frac{1}{2},j}+O(h^{p+1}) \qquad (2.5.3.28b)$$

$$U^R_{i+\frac{1}{2},j}-\tilde{U}^R_{i+\frac{1}{2},j}=O(h^p) \qquad (2.5.3.28c)$$

$$U^L_{i+\frac{1}{2},j}-\tilde{U}^L_{i+\frac{1}{2},j}=O(h^p) \qquad (2.5.3.28d)$$

with $p=1$ or $p=2$, and where $U^R_{i+\frac{1}{2},j}$, $U^L_{i+\frac{1}{2},j}$, $\tilde{U}^R_{i+\frac{1}{2},j}$, $\tilde{U}^L_{i+\frac{1}{2},j}$ etc, are given by (2.5.3.27), then scheme (2.5.3.18) is p-order accurate with respect to the space discretization.

PROOF.
This theorem is a direct consequence of theorems (2.5.2a,b). □

From (2.5.3.27) we see that

$$U^L_{i+\frac{1}{2},j}=\tilde{U}^L_{i+\frac{1}{2},j}+\tfrac{1}{2}(\psi(R_{i,j})-1)\,(U_{i,j}-U_{i-1,j})\,. \qquad (2.5.3.29)$$

Because

$$\psi(R_{i,j})-1=\psi(1)-1+\psi'(1)(R_{i,j}-1)+\tfrac{1}{2}\psi''(\rho_{i,j})(R_{i,j}-1)^2$$

with $\psi = d\psi/dR, \psi' = d^2\psi/dR^2$ and $\rho_{i,j}$ between 1 and $R_{i,j}$ and

$$(R_{i,j}-1)\,(U_{i,j}-U_{i-1,j}) = U_{i+1,j} - 2U_{i,j} + U_{i-1,j}$$

we find that the assumption

$$\psi(1) = 1 \tag{2.5.3.30}$$

leads to

$$U^L_{i+\frac{1}{2},j} = \tilde{U}^L_{i+\frac{1}{2},j} + \tfrac{1}{2}\psi'(1)(U_{i+1,j} - 2U_{i,j} + U_{i-1,j}) + \tfrac{1}{4}\psi''(\rho_{i,j})(R_{i,j}-1)(U_{i+1,j} - 2U_{i,j} + U_{i-1,j}) \tag{2.5.3.31}$$

Assume that $|\psi'|$ is uniformly bounded, then

$$U^L_{i+\frac{1}{2},j} - \tilde{U}^L_{i+\frac{1}{2},j} = O(h^2)$$

$$U^L_{i+\frac{1}{2},j} - U^L_{i-\frac{1}{2},j} = \tilde{U}^L_{i+\frac{1}{2},j} - \tilde{U}^L_{i-\frac{1}{2},j} + O(h^2)\,.$$

(Similar relations can be derived for $U^R_{i+\frac{1}{2},j} - \tilde{U}^R_{i+\frac{1}{2},j}$, $U^R_{i+\frac{1}{2},j} - \tilde{U}^R_{i-\frac{1}{2},j}$).

Hence, scheme (2.5.3.18) is first-order accurate with respect to the space discretization.

Furthermore, by assuming that $\dfrac{\partial u}{\partial x} \neq 0$, we see that

$$R_{i,j} - 1 = O(h)\,.$$

Then, assuming $|\psi''|$ is uniformly bounded, (2.5.3.31) leads to

$$U^L_{i+\frac{1}{2},j} - \tilde{U}^L_{i+\frac{1}{2},j} = O(h^2)$$

$$U^L_{i+\frac{1}{2},j} - U^L_{i-\frac{1}{2},j} = \tilde{U}^L_{i+\frac{1}{2},j} - \tilde{U}^L_{i-\frac{1}{2},j} + O(h^3)$$

and we see that the scheme is even second-order accurate in space.
Because in general the set

$$\Omega = \{(x,y) \mid \frac{\partial u}{\partial x} = 0 \text{ or } \frac{\partial u}{\partial y} = 0\}$$

has measure 0, we have the following theorem.

THEOREM (2.5.3h).

(a) *If* $\psi(1) = 1$ *and* $|\psi'|$ *is uniformly bounded, then scheme (2.5.3.18) is first-order accurate in space.*

(b) *Assume furthermore that* $|\psi''|$ *is uniformly bounded. Away from points where* $\dfrac{\partial u}{\partial x} = 0$ *or* $\dfrac{\partial u}{\partial y} = 0$ *the space discretization is second-order accurate (in the sense of definition 2.5.2a). Moreover,*

$$\sum_{i,j} h^2 \,|\, F_h R_h u - \bar{R}_h F u \,|_{i,j} = O(h^2) \tag{2.5.3.32}$$

where F_h *is the space discretization according to (2.5.3.18).*

Notice that scheme (2.5.3.18) is linear when $f(u)$ and $g(u)$ are linear and

$\psi(R)=a+bR$, $a, b\in\mathbb{R}$. Note that $\psi(R)=a+bR$ does not satisfy the requirements of theorems (2.5.3f,h), as we should expect.

Examples of limiters combining the property of second-order accuracy and monotonicity are:

EXAMPLE 1: The VAN LEER limiter [23,25,26].

$$\psi_{VL}(R)=\frac{R+|R|}{|R|+1}\,. \tag{2.5.3.33}$$

By taking $M=2$ and $\alpha=0$ it is easily seen that this limiter satisfies the monotonicity restriction (2.5.3.26). Because $\psi_{VL}(1)=1$ second-order accuracy is obtained.

EXAMPLE 2: The VAN ALBADA limiter [24].

$$\psi_{VA}(R)=\frac{R^2+R}{R^2+1} \tag{2.5.3.34}$$

By taking $M=2$ and $\alpha=-\frac{1}{2}$ it is easily seen that this limiter combines monotonicity with second-order accuracy. Notice that $\psi_{VA}\in C^\infty(\mathbb{R})$. Another advantage of this limiter is that $\lim_{R\mapsto\pm\infty}\psi_{VA}(R)=1$. This implies that at discontinuities the scheme renders the fully one-sided upwind scheme ($\kappa=-1$), which is a natural scheme at discontinuites.
For a review of other limiters see [23]. But notice that a limiter $\phi(r)$ of [23] is related to $\psi(R)$ by $R=1/r$, $\psi(R)=R\phi(r)$. A limiter $\phi(r)$ of [23] is only algebraically identical with $\psi(R)$ if $\psi(R)/R=\psi(1/R)$.

REMARK (2.5.3a).
It has been observed [26] that second-order accuracy can be achieved by assuming a linear distribution, rather than the uniform distribution, associated with first-order schemes. In a cell, a linear distribution in the x direction is achieved if

$$U^L_{i+\frac{1}{2},j}-U_{i,j}=U_{i,j}-U^R_{i-\frac{1}{2},j}\,,$$

and similarly in the y-direction. Using (2.5.3.27), this means

$$\psi(R_{i,j})(U_{i,j}-U_{i-1,j})=\psi(\frac{1}{R_{i,j}})(U_{i+1,j}-U_{i,j})$$

or, equivalently

$$\psi(\frac{1}{R_{i,j}})=\frac{\psi(R_{i,j})}{R_{i,j}}\,.$$

So, if a limiter satisfies

$$\psi(\frac{1}{R})=\frac{\psi(R)}{R}\quad\forall R\in\mathbb{R} \tag{2.5.3.35}$$

we can speak of linear distributions in each cell. It can be verified that both ψ_{VL} and ψ_{VA} possess this property. This is no coincidence: they were designed that way.
If a limiter satisfies (2.5.3.35) we have the following theorem:

THEOREM (2.5.3i).
If a limiter $\psi=\psi(R)$ has the property $\psi(\frac{1}{R})=\frac{\psi(R)}{R}$ and if $\psi_{\max}-\psi_{\min}\leqslant 2$, where

$$\psi_{\max}=\max_{R\in\mathbb{R}}(\psi(R)),\ \psi_{\min}=\min_{R\in\mathbb{R}}(\psi(R)),$$

then scheme (2.5.3.18) is monotone.

PROOF.
Because the limiter satisfies (2.5.3.35), the monotonicity conditions (2.5.3.26) are equivalent with

$$\alpha\leqslant\psi(R)\leqslant M\ ;\ -M\leqslant\psi(R)\leqslant 2+\alpha\ ;\quad \forall R\ . \tag{2.5.3.36}$$

Put

$$\alpha=\psi_{\min},\ M=2\ .$$

Formula (2.5.3.35) implies $\psi(0)=0$, hence $\alpha\leqslant 0$ and $\psi_{\max}\geqslant 0$.
Hence

$$\psi_{\max}\leqslant 2+\alpha\leqslant 2=M\ \Rightarrow\ \alpha\leqslant\psi(R)\leqslant M\quad\forall R$$

and

$$\psi_{\min}\geqslant\psi_{\max}-2\geqslant -2=-M\ \Rightarrow\ -M\leqslant\psi(R)\leqslant 2+\alpha\quad\forall R\ .$$

□

REMARK (2.5.3b).
The κ-schemes (2.5.1.3) can be written as (2.5.3.19) by taking

$$\psi(R)=\frac{1-\kappa}{2}+\frac{1+\kappa}{2}R\ . \tag{2.5.3.37}$$

As expected, the κ-schemes correspond with a linear limiter and are therefore not monotone. A class of limiters which combines monotonicity and second-order accuracy is derived from the κ-schemes by

$$\psi_\kappa(R)=(\frac{1-\kappa}{2}+\frac{1+\kappa}{2}R)\cdot\phi(R) \tag{2.5.3.38}$$

where

$$\phi(R)=\frac{2R}{R^2+1}\ . \tag{2.5.3.39}$$

Notice that $\phi(1)=1$, $\phi'(1)=0$ and therefore the ψ_κ limiters resemble the κ-schemes in the smooth part of the flow field where $R\approx 1$. Concerning accuracy, the $\psi_{1/3}$ limiter is the best. Notice that $\psi_0(R)$ is the Van Albada limiter.

The ψ_κ limiters are depicted in figure (2.5.3c) for $\kappa=-1,-\frac{2}{3},-\frac{1}{3},0,\frac{1}{3},\frac{2}{3},1$. Not all $\psi_\kappa(R)$ limiters lie in the monotonicity region depicted in fig (2.5.3b). Notice that $\psi'_\kappa(0)=1-\kappa$. Thus $\psi_{-1}(R)$ does not satisfy the monotonicity conditions (2.5.3.26) because $\psi'_{-1}(0)=2$ and $\psi_{-1}(-1)=-1$. From figure (2.5.3c) it is easily seen that $\psi_\kappa(R)$, $\kappa\in[-\frac{1}{3},1]$ satisfies the monotonicity conditions for $\alpha=-\frac{1}{2}$, $M=2$.

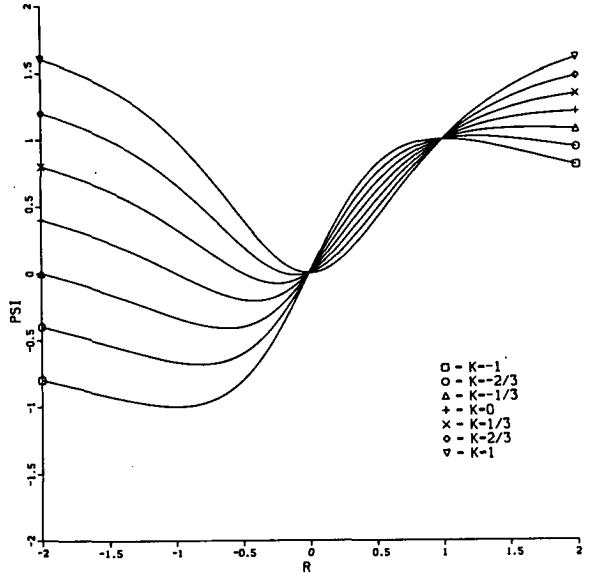


FIGURE (2.5.3c). The $\psi_\kappa(R)$ limiters for $\kappa=-1,-\frac{2}{3},-\frac{1}{3},0,\frac{1}{3},\frac{2}{3},1$.

A disadvantage of the $\psi_{\frac{1}{3}}(R)$ limiter is that $\lim_{R\mapsto\pm\infty}\psi_{\frac{1}{3}}(R)=\frac{4}{3}$. An improvement of the $\psi_{\frac{1}{3}}(R)$ limiter is

$$\tilde{\psi}_{\frac{1}{3}}(R)=(\tfrac{1}{3}+\tfrac{2}{3}R)\tilde{\phi}(R) \tag{2.5.3.40a}$$

with

$$\tilde{\phi}(R)=\frac{3R^3-2R^2+3R}{2(R^4+1)}\,. \tag{2.5.3.40b}$$

Notice that $\tilde{\phi}(1)=1$, $\tilde{\phi}'(1)=0$ and $\tilde{\phi}(R)\sim\frac{3}{2R}$ for $R\mapsto\pm\infty$. Thus $\tilde{\psi}_{\frac{1}{3}}(R)$ resembles the $\kappa=\frac{1}{3}$ scheme and $\lim_{R\mapsto\pm\infty}\psi_{\frac{1}{3}}(R)=1$.

The $\tilde{\psi}_{\frac{1}{3}}(R)$ limiter is depicted in figure (2.5.3d). It is easily seen that $\tilde{\psi}_{\frac{1}{3}}(R)$ satisfies the monotonicity conditions for $\alpha=-\frac{1}{2}$, $M=2$.

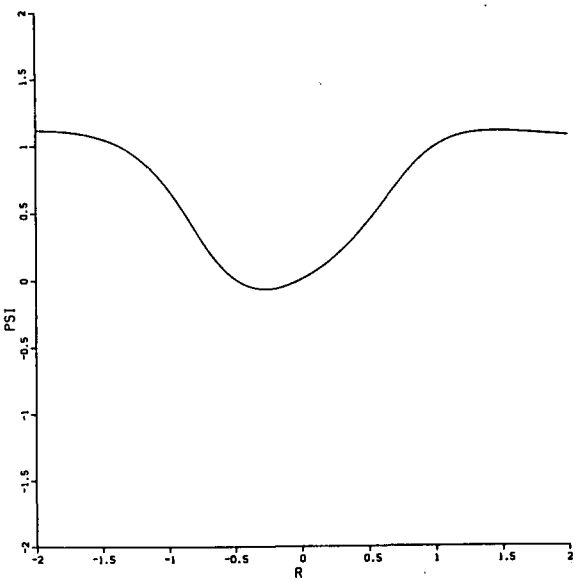


FIGURE (2.5.3d). The special limiter $\tilde{\psi}_{1/3}(R)$.

The use of the $\psi_{1/3}(R)$ or $\tilde{\psi}_{1/3}(R)$ limiter instead of the $\psi_0(R)$ limiter is especially important for the computation of boundary layer flow modelled by the Navier-Stokes equations, see [2].

REMARK (2.5.3c).
The use of a limiter in the interpolation is in fact equivalent with a nonlinear monotone interpolation. This becomes clear in the following example where a limiter is constructed by the interpolation of three states by an exponential.
Consider three states U_{-1}, U_0, U_1 which are the mean values of a function $u(x)$ on the intervals $(-\frac{3}{2}h,-\frac{1}{2}h)$, $(-\frac{1}{2}h,\frac{1}{2}h)$, $(\frac{1}{2}h,\frac{3}{2}h)$ respectively. Suppose

$$u(x)=A+Be^{\alpha x}\,. \tag{2.5.3.41}$$

The unknowns A,B and α are derived from

$$U_k=\frac{1}{h}\int_{(k-1/2)h}^{(k+1/2)h}\left\{A+Be^{\alpha x}\right\}dx=A+\frac{B}{\alpha h}e^{\alpha hk}\left\{e^{1/2\alpha h}-e^{-1/2\alpha h}\right\}\quad k=-1,0,1\,.$$

Define

$$R\equiv\frac{U_1-U_0}{U_0-U_{-1}}=e^{\alpha h}\,.$$

We require

$$U_0+\tfrac{1}{2}\psi(R)\,(U_0-U_{-1})=u(\tfrac{1}{2}h)=A+Be^{1/2\alpha h} \tag{2.5.3.42}$$

It can be easily seen that (2.5.3.42) implies

$$\psi(R)=\frac{2R}{(R-1)^2}\{1-R+R\log R\}\,. \tag{2.5.3.43}$$

By defining $\psi(1)=1$ it is easily seen that $\psi\in C^\infty(0,\infty)$ and $\psi'(1)=\frac{2}{3}$. Thus, this limiter corresponds with the $\kappa=1/3$ scheme. Because $\lim_{R\downarrow 0}\psi'(R)=2$ we have to take $\alpha=0$ and $M=+\infty$. By defining $\psi(R)=0$, $R<0$ it is easily seen that this limiter lies in the monotonicity region depicted in figure (2.5.3b) for $\alpha=0$ and $M=+\infty$. This means that steady state solutions obtained with this limiter are monotone.

The use of limiters in the second-order space discretization of the steady Euler equations is straightforward. Let $q^{L}_{i+1/2,j}{}^{(k)}$ and $q^{R}_{i+1/2,j}{}^{(k)}$ be the k-th component ($k=1,2,3,4$) of $q^{L}_{i+1/2,j}$ and $q^{L}_{i+1/2,j}$ (see (2.5.1.2)). Then we take

$$\begin{aligned} q^{L}_{i+1/2,j}{}^{(k)} &= q^{(k)}_{i,j}+\tfrac{1}{2}\psi_\kappa(R^{(k)}_{i,j})\,(q^{(k)}_{i,j}-q^{(k)}_{i-1,j}) \\ q^{R}_{i+1/2,j}{}^{(k)} &= q^{(k)}_{i+1,j}+\tfrac{1}{2}\psi_\kappa\Big(\frac{1}{R^{(k)}_{i+1,j}}\Big)\,(q^{(k)}_{i+1,j}-q^{(k)}_{i+2,j}) \end{aligned} \tag{2.5.3.44}$$

where

$$R^{(k)}_{i,j}=\frac{q^{(k)}_{i+1,j}-q^{(k)}_{i,j}}{q^{(k)}_{i,j}-q^{(k)}_{i-1,j}} \tag{2.5.3.45}$$

Thus the limiter $\psi_\kappa(R)$ is applied on each component c,u,v or z of the states $\{q_{i,j}\}$ separately.
In case $\Omega_{i,j}$ is a boundary volume, so that, for example $\partial\Omega_{i+1/2,j}$ is part of the domain boundary, no limiter can be used to compute $q^{L}_{i+1/2,j}$ and $q^{R}_{i-1/2,j}$. In this case we use a simple linear interpolation, i.e.

$$\begin{aligned} q^{L}_{i+1/2,j} &= q_{i,j}+\tfrac{1}{2}(q_{i,j}-q_{i-1,j}) \\ q^{R}_{i-1/2,j} &= q_{i,j}-\tfrac{1}{2}(q_{i,j}-q_{i-1,j})\,. \end{aligned} \tag{2.5.3.46}$$

The boundary conditions, together with the state $q^{L}_{i+1/2,j}$, are used to compute the boundary state $q^{R}_{i+1/2,j}$, by considering the Riemann boundary problem. The flux $f_{i+1/2,j}$ at $\partial\Omega_{i+1/2,j}$ is computed by (2.5.1.2).

REFERENCES.

1. S.R. CHAKRAVARTHY and S. OSHER (1985). *Computing with High-Resolution Upwind Schemes for Hyperbolic Equations.* In: Lectures in Applied Mathematics, (B. ENGQUIST, S. OSHER, R. SOMERVILLE eds.), Volume 22, Part I, 57-86, AMS, Providence, R.I.
2. S.R. CHAKRAVARTHY, K. SZEMA, U.C. GOLDBERG, J.J. GORSKI S. OSHER (1985). *Application of a New Class of High Accuracy TVD Schemes to the Navier-Stokes Equations.* AIAA-85-0165. AIAA-23r Aerospace Sciences Meeting, Reno/Nevada.
3. S.K. GODUNOV (1959). *Finite Difference Method for Numerical*

Computation of Discontinuous Solutions of the Equations of Fluid Dynamics. Math. Sbornik 47, 271-306. Also: Cornell Aeronautical Lab. (Calspan Translation).

4. J.B. GOODMAN and R.J. LE VEGUE (1985). *On the Accuracy of Stable Schemes for 2D Scalar Conservation Laws.* Math. Comp. 45, 156-21.
5. A. HARTEN, J.M. HYMAN, P.D. LAX (1976). *On Finite-Difference Approximations and Entropy Conditions for Shocks.* Comm. Pure Appl. Math. 29, 297-322.
6. A. HARTEN (1983). *High Resolution Schemes for Hyperbolic Conservation Laws.* J. Comp. Phys. 49, 357-393.
7. P.W. HEMKER, S.P. SPEKREIJSE (1985). *Multigrid Solutions of the Steady Euler Equations.* In: Advances in Multi-Grid Methods Proceedings of the Conference held in Oberwolfach, December 8-13-1984. (D. BRAESS, W. HACKBUSH, U. TROTTENBERG eds.). *Notes on Numerical Fluid Mechanics,* Volume 11, 33-44. Vieweg, Braunschweig.
8. P.W. HEMKER, S.P. SPEKREIJSE (1986). *Multiple Grid and Osher's Scheme for the Efficient Solution of the Steady Euler Equations.* Appl. Num. Math.2, 475-493.
9. A. JAMESON, W. SCHMIDT and E. TURKEL (1981). *Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes.* AIAA-81-1259 AIAA-14th Fluid and Plasma Dynamics Conference, Palo Alto, California.
10. A. JAMESON (1983). *Solution of the Euler equations for Two-Dimensional Flow by a Multigrid Method.* Appl. Math. Comput 13, 327-355.
11. A. JAMESON and D. MAVRIPLIS (1985). *Finite Volume Solution of the Two-Dimensional Euler Equations on a Regular Triangular Mesh.* AIAA-85-0435. AIAA-23rd Aerospace Sciences Meeting, Reno, Nevada.
12. B. KOREN (1986). *Euler Flow Solutions for a Transonic Windtunnel Section.* Report NM-R8601, Centre for Mathematics and Computer Science, Amsterdam.
13. B. KOREN (1986). *Evaluation of Second-Order Schemes and Defect Correction for the Multigrid Computation of Airfoil Flows with the Steady Euler Equations.* Report NM-R8616, Centre for Mathematics and Computer Science, Amsterdam. To appear in J. Comput. Phys.
14. P.D. LAX (1973). *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves.* Regional Conference Series in Appl. Math. 11, Siam, Philadelphia.
15. P.D. LAX and B. WENDROFF (1960). *Systems of Conservation Laws.* Comm. Pure Appl. Math., Volume 13, 217-237.
16. S. OSHER and F. SOLOMON (1982). *Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws.* Math. Comp. 38, 339-374.
17. S. OSHER and S. CHAKRAVARTHY (1983). *Upwind Schemes and Boundary Conditions with Applications to Euler Equations in General Geometries.* J. Comp. Phys., 50, 447-481.
18. P.L. ROE (1981). *Approximate Riemann Solvers, Parameter Vectors and Difference Schemes.* J. Comp. Phys. 43, 357-372.

19. J. SMOLLER (1983). *Shock Waves and Reaction-Diffusion Equations.* Grundlehren der Mathematischen Wissenschaften 258. Springer Verlag, Berlin.
20. S.P. SPEKREIJSE (1986). *Second-Order Accurate Upwind Solutions of the 2D Steady Euler Equations by the Use of a Defect Correction Method.* In: Proceedings of the 2nd European Conference on Multigrid Methods, Cologne, October 1-4-1985. (W. HACKBUSH, U. TROTTENBERG eds.), Lecture Notes in Mathematics 1228, 285-300, Springer Verlag, Berlin.
21. S.P. SPEKREIJSE (1987). *Multigrid Solution of Monotone Second-Order Discretizations of Hyperbolic Conservation Laws.* Math. Comp. 49, 135-155.
22. J.L. STEGER, R.F. WARMING (1981). *Flux Vector Splitting of the Inviscid Gas Dynamics Equations with Applications to Finite Difference Methods.* J. Comp. Phys. 40, 263-293.
23. P.K. SWEBY (1984). *High Resolution Schemes Using Flux Limiters for Hyperbolic Conservation Laws.* Siam J. Num. Anal. 21, 995-1011.
24. G.D. VAN ALBADA, B. VAN LEER and W.W. ROBERTS (1982). *A Comparative Study of Computational Methods in Cosmic Gas Dynamics.* Astron. Astrophys. 108, 76-84.
25. B. VAN LEER (1974). *Towards the Ultimate Conservative Difference Scheme II.* Monotonicity and Conservation Combined in a Second-Order Scheme. J. Comp. Phys. 14, 361-370.
26. B. VAN LEER (1977). *Towards the Ultimate Conservative Difference Scheme IV.* A New Approach to Numerical Convection. J. Comp. Phys. 23, 276-299.
27. B. VAN LEER (1982). *Flux-Vector Splitting for the Euler Equations.* In: Procs. 8th Intern. Conf. on Numerical Methods in Fluid Dynamics, (E. KRAUSE ed.), Aachen 1982. Lecture Notes in Physics 170, 507-512, Springer Verlag, Berlin.
28. B. VAN LEER (1984). *On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist-Osher and Roe.* Siam J. Sci. Stat. Comput 5, 1-20.
29. B. VAN LEER (1985). *Upwind-Difference Methods for Aerodynamic Problems Governed by the Euler Equations.* In: Lectures in Applied Mathematics, (B. ENGQUIST, S. OSHER, R. SOMERVILLE eds.), Volume 22, Part II, 327-336, AMS, Providence, R.I.

Chapter III

Multigrid Solution of the First-Order Discretization

3.1. Introduction

The multigrid method has become a well-established technique for the acceleration of relaxation-iterations to solve the sparse systems that arise from discretization of elliptic partial differential equations. The advantage of multigrid over other acceleration techniques is the fact that - under suitable, but quite general circumstances - the rate of convergence is independent of the size of the system to be solved. For other methods the rate slows down rapidly for finer discretizations when the systems get larger. This makes the multiple grid method superior to other methods - at least for very large elliptic systems. For readers unfamiliar with multigrid techniques we refer to [1,2].
The multigrid technique has also been applied with success to other types of equations, such as parabolic partial differential equations and integral equations. Based on the pioneering work of Brandt [1] it is expected that with the multigrid method, for many equations, a sufficiently accurate approximate solution can be computed in an amount of work that is equivalent to a small number of evaluations of the (nonlinear) operator.

Also for the steady solution of hyperbolic equations, such as the Euler equations, the multigrid technique has been used for the acceleration of the solution process. For a survey of multigrid approaches to the Euler equations, see [3].

In this chapter, we study a multigrid method for the solution of first-order discretizations only. As noted before, first-order accuracy is too low for practical problems. Therefore, the ultimate goal is an efficient and robust solution method for systems obtained by second-order discretization. Such a method, based on the defect correction principle, is developed in chapter IV. The basic tool of this method is the multigrid solver for first-order discretization which is the topic of this chapter.

3.2. Nested iteration and nonlinear multigrid

Let

$$F_h^1(q_h)=r_h \tag{3.2.1}$$

be the first-order accurate discretization of the 2D steady Euler equations with source term r. Hence, (see (2.1.15))

$$(F_h^l(q_h))_{i,j} = f_{i+\frac{1}{2},j} + f_{i,j+\frac{1}{2}} - f_{i-\frac{1}{2},j} - f_{i,j-\frac{1}{2}} \tag{3.2.2}$$

with

$$\begin{aligned} f_{i+\frac{1}{2},j} &= l_{i+\frac{1}{2},j} T^{-1}_{i+\frac{1}{2},j} f_R(T_{i+\frac{1}{2},j} q_{h,i,j}, T_{i+\frac{1}{2},j} q_{h,i+1,j}) \\ f_{i,j+\frac{1}{2}} &= l_{i,j+\frac{1}{2}} T^{-1}_{i,j+\frac{1}{2}} f_R(T_{i,j+\frac{1}{2}} q_{h,i,j}, T_{i,j+\frac{1}{2}} q_{h,i,j+1}) \end{aligned} \tag{3.2.3}$$

where $l_{i+\frac{1}{2},j}$ is the length of $\partial\Omega_{i+\frac{1}{2},j}$, $T_{i+\frac{1}{2},j} = T(\phi_{i+\frac{1}{2},j})$ and $(\cos\phi_{i+\frac{1}{2},j}, \sin\phi_{i+\frac{1}{2},j})$ is the unit normal on $\partial\Omega_{i+\frac{1}{2},j}$ directed from $\Omega_{i,j}$ to $\Omega_{i+1,j}$ (see fig. 2.1.b). Similarly, $l_{i,j+\frac{1}{2}}$ is the length of $\partial\Omega_{i,j+\frac{1}{2}}$, $T_{i,j+\frac{1}{2}} = T(\phi_{i,j+\frac{1}{2}})$ and $(\cos\phi_{i,j+\frac{1}{2}}, \sin\phi_{i,j+\frac{1}{2}})$ is the unit normal on $\partial\Omega_{i,j+\frac{1}{2}}$ directed from $\Omega_{i,j}$ to $\Omega_{i,j+1}$. The subscript h denotes the meshwidth and f_R is Osher's approximate Riemann-solver. Although in general $r=0$, we prefer to describe the solution method for first-order systems with an arbitrary (but small) right-hand side.
We will develop an efficient multigrid solution method for (3.2.1). A nested sequence of finite volume grids is constructed, such that each finite volume in a given grid is the union of four finite volumes in the next finer grid, as indicated in fig. 3.2a. The grids are denoted by Ω^k, $k=1,\ldots,l$; their mesh-size is $h_1 > h_2 > \cdots > h_l = h$. Hence, Ω^1 is the coarsest grid, Ω^l the finest grid.

The solution method consists of two successive stages: nested iteration (or full multigrid: FMG) and nonlinear multigrid (NMG) (or full approximation scheme: FAS).

Stage I: Nested iteration
Let

$$F_k^l(q_k) = r_k \tag{3.2.4}$$

be the first-order discretization on Ω^k, $k=1,\ldots,l$. Denote with q_k^* the solution (3.2.4), $k=1,\ldots,l$. Nested iteration starts with some initial estimate of q_1^* and proceeds recursively. Given an approximation of q_k^*, an approximation of q_{k+1}^* is obtained as follows. The approximation of q_k^* is improved by a single NMG-iteration (see stage II) and this improved approximation is interpolated to the finer grid Ω^{k+1}. These steps are repeated until an approximation of q_l^* has been obtained.

The interpolation used to obtain the approximation on a finer grid is a piecewise constant interpolation (assign the coarse volume value $(q_k)_{i,j}$ to the corresponding 4 finer volumes of Ω^{k+1}).

Stage II: The nonlinear multigrid (NMG) method
To converge rapidly to the solution of (3.2.1), NMG-iterations are applied on the finest grid Ω^l. One NMG-iteration on a general grid Ω^k is defined recursively by the following steps:

(0) Start with an approximation q_k of q_k^*.
(1) Improve q_k by application of p (pre-)relaxation iterations to $F_k^l(q_k) = r_k$.
(2) Compute the defect (or residual) $d_k := r_k - F_k^l(q_k)$.

(3) Find an approximation q_{k-1} of q_{k-1}^* on the next coarser grid Ω^{k-1}. One possibility is to take the last obtained approximation to q_{k-1}^*. Another possibility, to be used here, is $q_{k-1} := \hat{I}_k^{k-1} q_k$ where $\hat{I}_k^{k-1}$ is a restriction operator.
(4) Compute $r_{k-1} := F_{k-1}^1(q_{k-1}) + I_k^{k-1} d_k$ where I_k^{k-1} is another restriction operator.
(5) Approximate the solution of $F_{k-1}^1(q_{k-1}) = r_{k-1}$ by σ NMG-iterations on Ω^{k-1}. The result is called $\tilde{q}_{k-1}$. ($\sigma=1$ results in a so-called V-cycle, $\sigma=2$ in a W-cycle).
(6) Correct the current approximation by $q_k := q_k + I_{k-1}^k(\tilde{q}_{k-1} - q_{k-1})$, where I_{k-1}^k is a prolongation operator.
(7) Improve q_k by application of q (post-)relaxation iterations to $F_k^1(q_k) = r_k$.

The steps (2)-(6) are called the coarse-grid correction. These steps are skipped on the coarsest grid.

In order to complete the description of NMG we have to discuss: (i) the choice of the transfer operators I_{k-1}^k, I_k^{k-1}, $\hat{I}_k^{k-1}$, (ii) the relaxation method, and (iii) the multigrid schedule i.e. the numbers p,q and σ.

(i) Choice of the transfer operators
The coarse grid cell $(\Omega^{k-1})_{i,j}$ is the union of the fine grid cells $(\Omega^k)_{2i,2j}, (\Omega^k)_{2i-1,2j}, (\Omega^k)_{2i,2j-1}, (\Omega^k)_{2i-1,2j-1}$ (see fig. 3.2a).

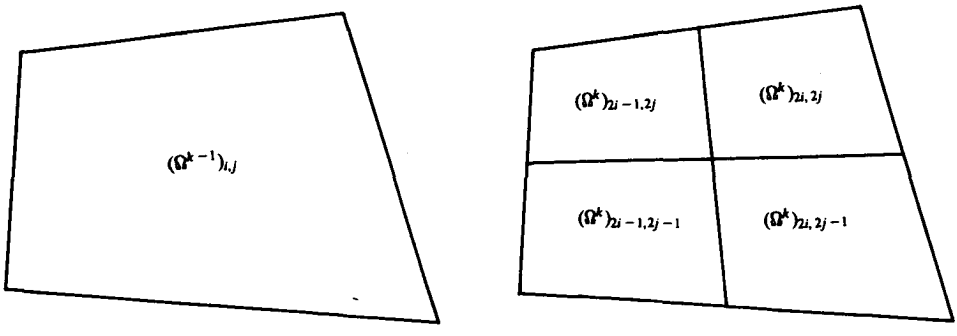


FIGURE 3.2a. The subdivision of a coarse grid cell in four fine grid cells.

The restriction operators $\hat{I}_k^{k-1}$ and I_k^{k-1} are defined by:

$$(q_{k-1})_{i,j} = (\hat{I}_k^{k-1} q_k)_{i,j}$$
$$:= \frac{1}{4}\{(q_k)_{2i,2j} + (q_k)_{2i-1,2j} + (q_k)_{2i,2j-1} + (q_k)_{2i-1,1j-1}\} \tag{3.2.5}$$

$$(d_{k-1})_{i,j} = (I_k^{k-1} d_k)_{i,j}$$
$$:= (d_k)_{2i,2j} + (d_k)_{2i-1,2j} + (d_k)_{2i,2j-1} + (d_k)_{2i-1,2j-1} . \quad (3.2.6)$$

Notice that a state q is represented as $q=(c,u,v,z)$ (see section 2.2.2).
The prolongation operator I_{k-1}^k is piecewise constant interpolation defined by

$$(I_{k-1}^k q_{k-1})_{2i,2j} = (I_{k-1}^k q_{k-1})_{2i-1,2j} = (I_{k-1}^k q_{k-1})_{2i,2j-1}$$
$$= (I_{k-1}^k q_{k-1})_{2i-1,2j-1} := (q_{k-1})_{i,j} . \quad (3.2.7)$$

By defining the transfer operators in this way, we have the following theorem.

THEOREM (3.2a).
The first-order coarse grid discretizations of the steady Euler equations are Galerkin approximations of the fine grid discretization i.e.

$$F_{k-1}^l = I_k^{k-1} F_k^l I_{k-1}^k \quad k=l,\ldots,2 . \quad (3.2.8)$$

PROOF.
From the definitions of the transfer operators I_k^{k-1}, I_{k-1}^k it follows that

$$(I_k^{k-1} F_k^l I_{k-1}^k q_{k-1})_{i,j} = (F_k^l q_k)_{2i,2j} + (F_k^l q_k)_{2i-1,2j}$$
$$+ (F_k^l q_k)_{2i,2j-1} + (F_k^l q_k)_{2i-1,2j-1} \quad (3.2.9)$$

where $q_k = I_{k-1}^k q_{k-1}$.
Because

$$(q_k)_{2i,2j} = (q_k)_{2i-1,2j} = (q_k)_{2i,2j-1} = (q_k)_{2i-1,2j-1} = (q_{k-1})_{i,j}$$

it is easily seen that the righthandside of (3.2.9) equals $(F_{k-1}^l q_{k-1})_{i,j}$.
□

From (3.2.8) it follows that we have a nested sequence of discretizations, i.e. the following scheme (fig. 3.2b) of operators and spaces commutes (X_k is the vector space of states $\{q_k\}$ at Ω^k, Y_k is the vector space of total fluxes at Ω^k).

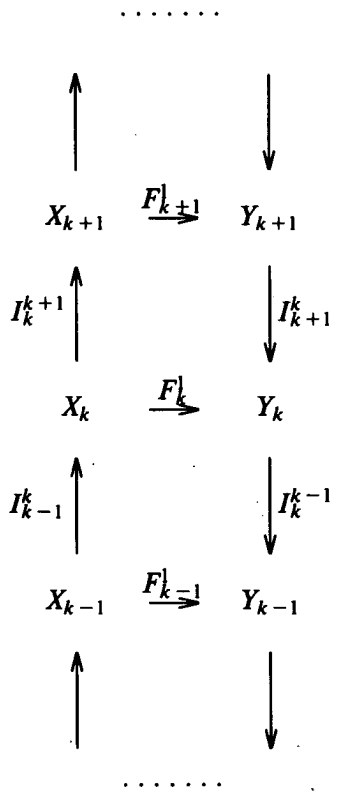


FIGURE 3.2b. The nested sequence of discretizations.

Another property which follows directly from the definitions (3.2.5)-(3.2.7) is

$$\hat{I}^{k-1}_k I^k_{k-1} = I_{k-1} \tag{3.2.10}$$

where I_{k-1} is the identity operator on X_{k-1}.

The effect of the Galerkin approximation (3.2.8) on the FAS-iteration process is the following. Let q_k be an approximation of q^*_k and $\tilde{q}_k$ the improvement of q_k after a coarse grid correction which is assumed to be solved exactly for the moment. Thus

$$\tilde{q}_k = q_k + I^k_{k-1}(\tilde{q}_{k-1} - \hat{I}^{k-1}_k q_k) \tag{3.2.11}$$

where $\tilde{q}_{k-1}$ is the exact solution of

$$F^1_{k-1}(\tilde{q}_{k-1}) = F^1_{k-1}(\hat{I}^{k-1}_k q_k) + I^{k-1}_k(r_k - F^1_k(q_k)) . \tag{3.2.12}$$

From (3.2.10,11) it follows that

$$\hat{I}_k^{k-1}\tilde{q}_k = \tilde{q}_{k-1}\,. \tag{3.2.13}$$

Using the relations (3.2.8, 10, 11, 12, 13) we find that

$$\begin{aligned}
&I_k^{k-1}(r_k - F_k^l I_{k-1}^k \hat{I}_k^{k-1}\tilde{q}_k)\\
&= I_k^{k-1} r_k - F_{k-1}^l \hat{I}_k^{k-1}\tilde{q}^k\\
&= I_k^{k-1} r_k - F_{k-1}^l \tilde{q}_{k-1}\\
&= I_k^{k-1} r_k - F_{k-1}^l \hat{I}_k^{k-1} q_k - I_k^{k-1}(r_k - F_k^l q_k)\\
&= I_k^{k-1} F_k^l q_k - I_k^{k-1} F_k^l I_{k-1}^k \hat{I}_k^{k-1} q_k\\
&= I_k^{k-1}(F_k^l q_k - F_k^l I_{k-1}^k \hat{I}_k^{k-1} q_k)\,.
\end{aligned} \tag{3.2.14}$$

For the restriction of the residual we obtain

$$\begin{aligned}
&I_k^{k-1}(r_k - F_k^l \tilde{q}_k)\\
&= I_k^{k-1}(r_k - F_k^l I_{k-1}^k \hat{I}_k^{k-1}\tilde{q}_k + F_k^l I_{k-1}^k \hat{I}_k^{k-1}\tilde{q}_k - F_k^l \tilde{q}_k)\\
&= I_k^{k-1}(\{F_k^l q_k - F_k^l I_{k-1}^k \hat{I}_k^{k-1} q_k\} - \{F_k^l \tilde{q}_k - F_k^l I_{k-1}^k \hat{I}_k^{k-1}\tilde{q}_k\})
\end{aligned} \tag{3.2.15}$$

In two particular cases the restriction of the residual vanishes for a Galerkin approximation. First, suppose $q_k \in \text{Range}\,(I_{k-1}^k)$. From (3.2.11) it follows that $\tilde{q}_k \in \text{Range}\,(I_{k-1}^k)$ and using (3.2.10) we see from (3.2.15) that

$$I_k^{k-1}(r_k - F_k^l \tilde{q}_k) = 0\,.$$

That is, after coarse grid correction the residual belongs to Ker (I_k^{k-1}).
When F_k^l is linear, this remains true when $q_k \notin \text{Range}\,(I_{k-1}^k)$. This is a well known result. It follows from (3.2.15) in the following way:

$$\begin{aligned}
&I_k^{k-1}(r_k - F_k^l \tilde{q}_k)\\
&= I_k^{k-1}(F_k^l - F_k^l I_{k-1}^k \hat{I}_k^{k-1})\,(q_k - \tilde{q}_k)\\
&= I_k^{k-1}(F_k^l - F_k^l I_{k-1}^k \hat{I}_k^{k-1}) I_{k-1}^k (\hat{I}_k^{k-1} q_k - \tilde{q}_{k-1})\\
&= I_k^{k-1}(F_k^l I_{k-1}^k - F_k^l I_{k-1}^k)\,(\hat{I}_k^{k-1} q_k - \tilde{q}_{k-1}) = 0\,.
\end{aligned}$$

In the neighbourhood of a solution, the difference between $\tilde{q}_k - q_k$ will be small and F_k^l will approximately behave as a linear operator: the restriction of its residual will be very small viz. $O(\|\tilde{q}_k - q_k\|^2)$.
Since I_k^{k-1} corresponds to taking local averages with positive weights, grid functions in Ker(I_k^{k-1}) have many sign-changes and hence are non-smooth. Local relaxation methods exist that reduce non-smooth residuals efficiently.

(ii) The relaxation method
We use Collective Symmetric Gauss-Seidel (CSGS-) relaxation. This means that, at a particular level, all cells are scanned one by one in some prescribed order and at each volume visited, the 4 unknowns (c,u,v,z) are changed simultaneously ('Collectively') by solving the 4 nonlinear equations by Newton's

method (local linearization).
One Newton iteration applied to the equation

$$(F_k^l(q_k^l))_{i,j} = (r_k)_{i,j} \tag{3.2.16}$$

is defined by

$$\frac{\partial}{\partial(q_k)_{i,j}}(F_k^l(q_k))_{i,j} \cdot (\delta q_k)_{i,j} = (r_k)_{i,j} - (F_k^l(q_k))_{i,j} \tag{3.2.17a}$$

$$(q_k)_{i,j}^{new} := (q_k)_{i,j}^{old} + (\delta q_k)_{i,j} \,. \tag{3.2.17b}$$

The computation of the term $\frac{\partial}{\partial(q_k)_{i,j}}(F_k^l(q_k))_{i,j}$ is described in section 2.4. The solution of the linear 4×4 system (3.2.17a) is obtained by Gauss-elimination with column pivoting. One or more steps of Newton-iteration are used until the local residual is reduced below a specified magnitude. It appeared most efficient to take this tolerance so crude that no more than one iteration step per cell is performed.

Several relaxations, all being of the Collective Gauss-Seidel type, were considered in [4,5]. It appeared that the following choice is very robust and efficient [4,5]: use CSGS from North-West to South-East and vice versa for the pré-relaxation, use CSGS from North-East to South-West and vice versa for the post-relaxation. For all computations presented in this work we use this particular relaxation method.

For an analysis of relaxations of Collective Gauss-Seidel type for the Euler equations and related model equations see [6].

(iii) The multigrid schedule

A multigrid schedule is a rule for the order in which the grids are visited. We use a fixed schedule for all computations presented in the next section. We take $\sigma = 1$ and $p = q = 1$, i.e. we use V-cycles with one pré- and one post relaxation, both in the nested iteration and the NMG-stage. Experiments show that different (p,q,σ)-strategies are not much different in efficiency [4,5]. Usually a smaller convergence factor is compensated by a corresponding amount of additional work.

Figure 3.2c gives an illustration of the multigrid schedule. Suppose there are 5 nested grids ($l=5$). Between two succeeding points A,B we have one NMG-iteration (V-cycle). Between two succeeding points B,A we have a piecewise constant interpolation in the nested iteration stage.

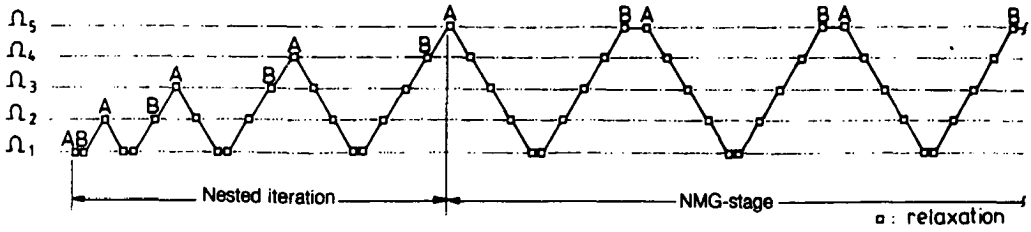


FIGURE 3.2c. Schematic representation of the multigrid schedule (5 levels).

3.3. NUMERICAL RESULTS

In this section numerical results are presented for four testproblems. The first two testproblems concern the computation of channel flows. Testproblem three concerns the resolution of a contact discontinuity and in testproblem four we consider a cylinder in a supersonic free stream. For each testproblem we give the convergence history of the multigrid method described in the previous section. The first-order solutions obtained are presented graphically in order to compare their quality with the second-order accurate solutions presented in chapter IV.

PROBLEM 1. *Flow in channel over a circular arc bump with thickness 4.2%.*
The geometry of the channel is given by the following mapping from the (ξ,η)-computational space to the (x,y)-physical space.

$$\begin{aligned} -2 \leqslant \xi \leqslant -\tfrac{10}{7} &\Rightarrow \tilde{\xi}=(77\xi+90)/32 \\ -\tfrac{10}{7} \leqslant \xi \leqslant \tfrac{15}{7} &\Rightarrow \tilde{\xi}=(14\xi-5)/40 \\ \tfrac{15}{7} \leqslant \xi \leqslant 3 &\Rightarrow \tilde{\xi}=(133\xi-255)/48 \end{aligned} \tag{3.3.1a}$$

$$\begin{aligned} -2 \leqslant \tilde{\xi} \leqslant -0.625 &\Rightarrow x=-2+1.375\cdot\left[\frac{e^{-\beta_1(\tilde{\xi}+2)}-1}{e^{-\beta_1\cdot 1.375}-1}\right] \\ -0.625 \leqslant \tilde{\xi} \leqslant 0.625 &\Rightarrow x=\tilde{\xi} \\ 0.625 \leqslant \tilde{\xi} \leqslant 3 &\Rightarrow x=3-2.375\left[\frac{e^{\beta_2(\tilde{\xi}-3)}-1}{e^{\beta_2\cdot -2.375}-1}\right] \end{aligned} \tag{3.3.1b}$$

$$\tilde{\eta}=2\frac{e^{\beta_3\eta}-1}{e^{2\beta_3}-1} \tag{3.3.1c}$$

$$\begin{aligned} |x| \leqslant 0.5 &\Rightarrow y=\tilde{\eta}+(1-\frac{\tilde{\eta}}{2})(\sqrt{9-x^2}-2.958) \\ |x| > 0.5 &\Rightarrow y=\tilde{\eta} \end{aligned} \tag{3.3.1d}$$

with $\beta_1=2.26$, $\beta_2=1.39$, $\beta_3=1.25$.
At level l, $l=1,2,3,4,5$ the vertices of the quadrilateral volumes $\Omega_{i,j}$ in the (x,y)-space correspond to a regular square mesh consisting of $5.2^{(l-1)}\times 2.2^{(l-1)}$ volumes covering $[-2,3]\times[0,2]$ in the (ξ,η)-plane.
The nested sequence of grids is given in fig. 3.3.1.

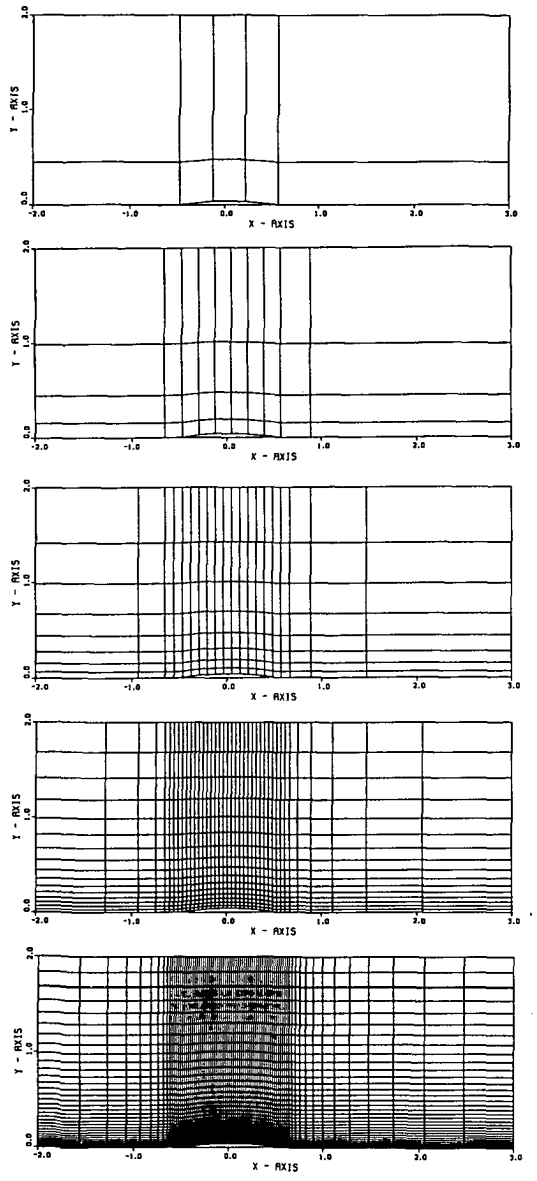


FIGURE 3.3.1. The nested sequence of grids for testproblem 1.

We consider the convergence history of the NMG-iteration proces for three different parallel flows given by the following boundary conditions

Problem 1a: Subsonic flow: $M_{\text{inlet}}=0.3$, $p_{\text{inlet}}=p_{\text{outlet}}$.

Problem 1b: Transonic flow: $M_{\text{inlet}}=0.85$, $p_{\text{inlet}}=p_{\text{outlet}}$.

Problem 1c: Supersonic flow: $M_{\text{inlet}}=3.0$.

M_{inlet} is the entrance Mach number at $x=-2$. We take $p_{\text{inlet}}=p_{\text{outlet}}$; this

makes the numerical solution essentially independent of the value of $p_{\text{inlet}} = p_{\text{outlet}}$.

The solid boundaries are treated as described in section 2.3.2. For problem 1a,b at the outflow boundary ($x=3$), p_{outlet} is prescribed and the boundary condition is treated as described in example (2.3.2a). At the inflow boudary ($x=-2$) we prescribe $v=0$, $c=1$, $u=M_{\text{inlet}}$ and, in case of problem 1a,b, z such that $p_{\text{inlet}}=p_{\text{outlet}}$ (overspecification), in case of problem 1c, $z=1$.

In fig. 3.3.2 we present the convergence history of the NMG-iteration process at different grids for the subsonic testproblem 1a. At the ordinate the logarithm (base 10) of a norm of the residual is depicted. The norm used is the sum of the four L_1-norms of the components in the residual, i.e.

$$||F_h^1(q_h)|| = \sum_{k=1}^{4} \left[\sum_{(i,j)} |(F_h^1(q_h))_{i,j}^k| \right] \tag{3.3.2}$$

In fig. 3.3.3,4 we present the convergence history at different grids for the transonic testproblem 1b. In fig. 3.3.3 we have used the P-variant of the Osher-scheme, in fig. 3.3.4 the O-variant has been used. No difference is observed. The P-variant of the Osher scheme is always used unless mentioned otherwise.

In fig. 3.3.5 we present the convergence history at different levels for the supersonic testproblem 1b.

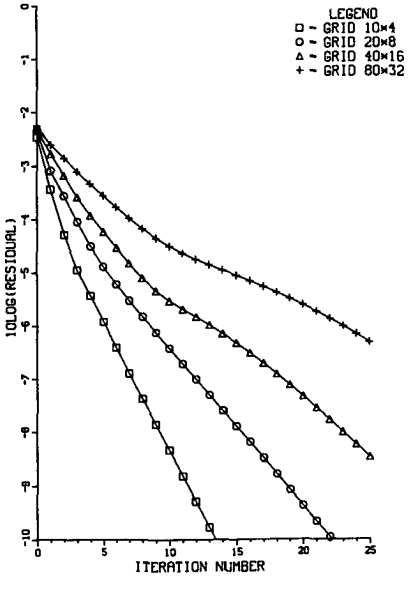


FIGURE 3.3.2. Convergence history of the NMG-iteration process at different grids for testproblem 1a.

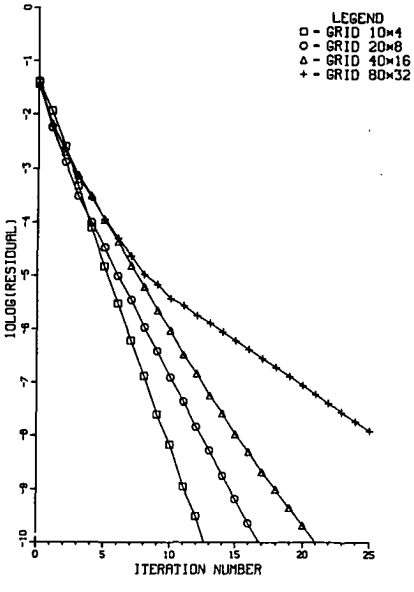


FIGURE 3.3.3. As figure 3.3.2 but for testproblem 1b.

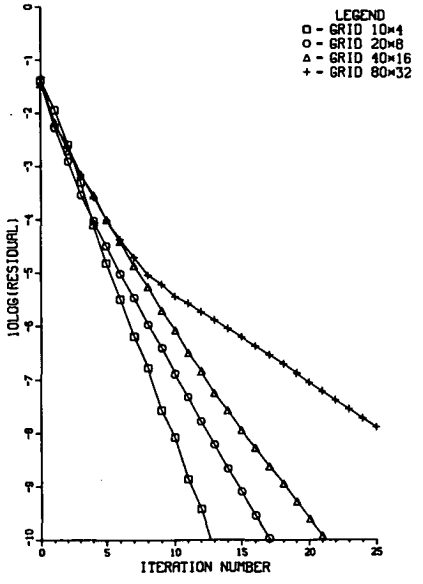


FIGURE 3.3.4. As figure 3.3.3 but with the *O*-variant.

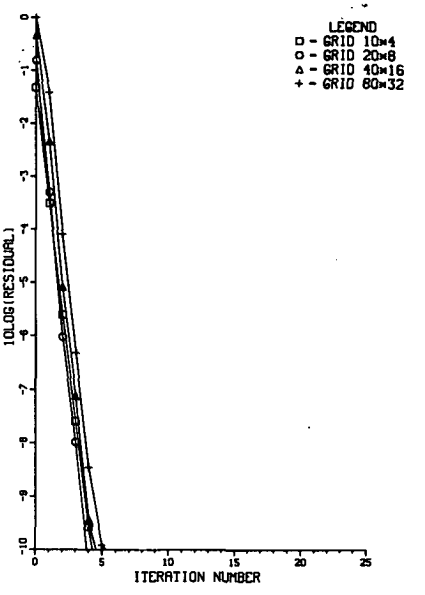


FIGURE 3.3.5. As figure 3.3.2 but for testproblem 1c.

From the experiments we conclude that for supersonic and transonic flow the rate of convergence of the NMG-iteration process is (for practical purposes, where one only wants to get below truncation error) independent of the meshwidth. Nested iteration alone already brings us close to truncation error. Convergence is slower and dependent on the meshwidth for small Mach numbers.

The transonic testproblem 1b is a standard problem, used to compare many different methods [9]. For this problem only we give results of the first-order solution obtained. The results are presented in fig. 3.3.6-12 and are obtained at the 40 × 16 grid. In fig. 3.3.6,7 we give the Mach number distribution. Fig. 3.3.6 is obtained with the *P*-variant, fig. 3.3.7 is obtained with the *O*-variant. Again, no difference between the *P*- and *O*-variant is observed. In fig. 3.3.8 we give the c_p distribution. The pressure coefficient is defined as

$$c_p = \frac{p - p_\infty}{\frac{1}{2}\rho_\infty u_\infty^2} \tag{3.3.3}$$

where the values at infinity are obtained by averaging the outflow values. In fig. 3.3.9 we give the entropy distribution of the flow field. The entropy is presented as $(s - s_{-\infty})/s_{-\infty}$ with $s = p\rho^{-\gamma}$.

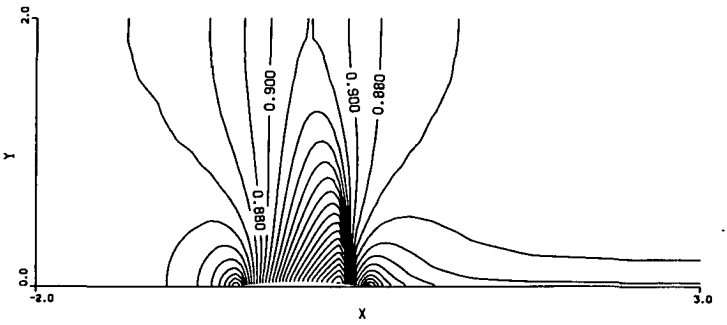


FIGURE 3.3.6. Iso-Mach lines for testproblem 1b (40 × 16 grid).

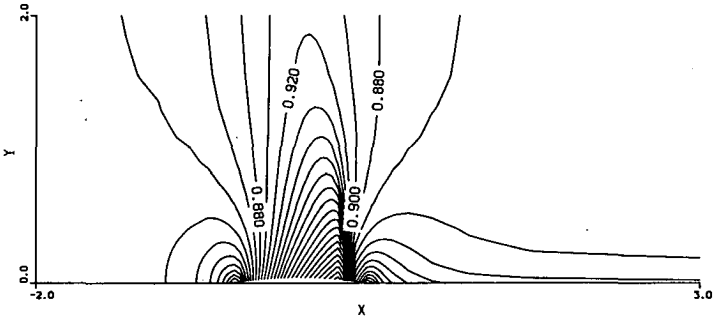


FIGURE 3.3.7. As figure 3.3.6 but the solution is obtained with the *O*-variant.

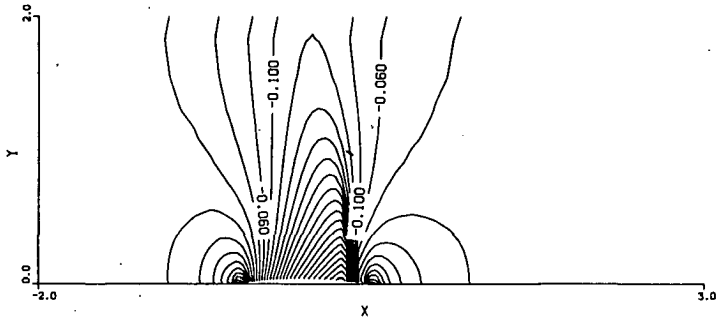


FIGURE 3.3.8. Pressure contours for testproblem 1b (40 × 16 grid).

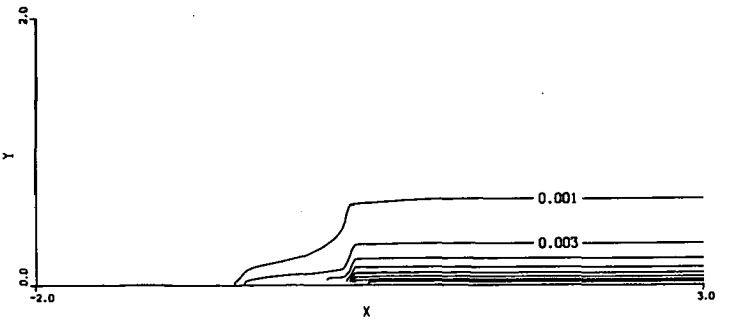


FIGURE 3.3.9. Entropy contours for testproblem 1b (40 × 16 grid).

In fig. 3.3.10-12 we give the Mach number, $-c_p$ and entropy distribution along the lower surface of the channel. The shock is well captured but notice the spurious entropy generation at both corners of the circular bump.

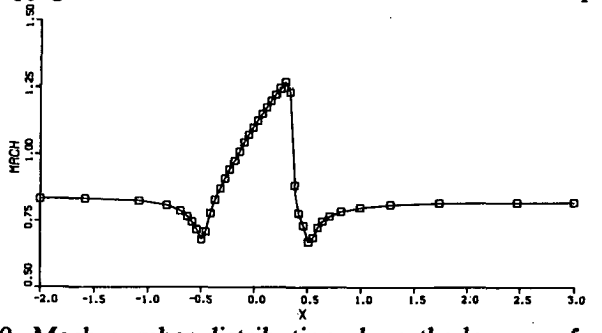


FIGURE 3.3.10. Mach number distribution along the lower surface of the channel for testproblem 1b (40 × 16 grid).

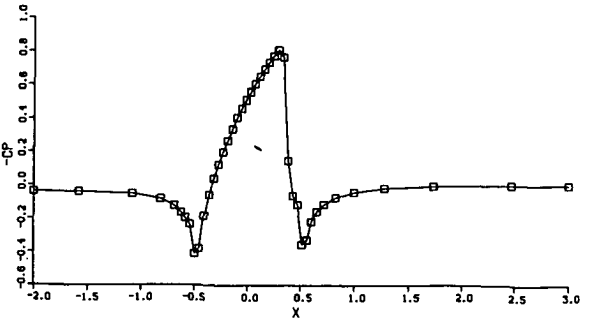


FIGURE 3.3.11. Pressure distribution along the lower surface of the channel for testproblem 1b (40 × 16 grid).

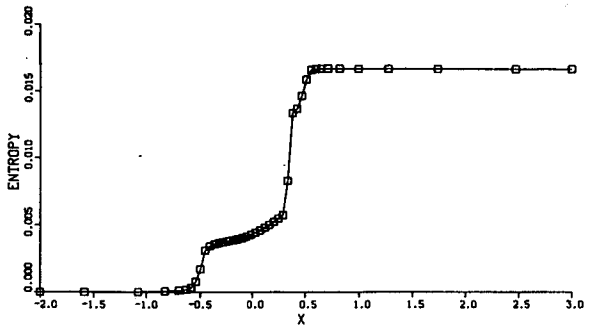


FIGURE 3.3.12. Entropy distribution along the lower surface of the channel for testproblem 1b (40 × 16 grid).

PROBLEM 2. *Supersonic flow in a channel with a 4% thick circular arc bump.*
This is a standard test problem, considered in [7,10].
The geometry of the channel is given by the following mapping from the (ξ,η)-computational space to the (x,y)-physical space with $(\xi,\eta)\in[-1,2]\times[0,1]$. The mapping is given by

$$\begin{aligned}
-1 \leqslant \xi \leqslant -\tfrac{1}{4} &\Rightarrow \tilde{\xi}=(4\xi+1)/3 \\
-\tfrac{1}{4} \leqslant \xi \leqslant \tfrac{5}{4} &\Rightarrow \tilde{\xi}=(4\xi+1)/6 \\
\tfrac{5}{4} \leqslant \xi \leqslant 2 &\Rightarrow \tilde{\xi}=(4\xi-2)/3
\end{aligned} \tag{3.3.4a}$$

$$\begin{aligned}
-1 \leqslant \tilde{\xi} \leqslant 0 &\Rightarrow x=-1+\frac{e^{-\beta_1(\tilde{\xi}+1)}-1}{e^{-\beta_1}-1} \\
0 \leqslant \tilde{\xi} \leqslant 1 &\Rightarrow x=\tilde{\xi} \\
1 \leqslant \tilde{\xi} \leqslant 2 &\Rightarrow x=2-\frac{e^{\beta_1(\tilde{\xi}-2)}-1}{e^{-\beta_1}-1}
\end{aligned} \tag{3.3.4b}$$

$$\eta=\frac{e^{\beta_2\eta}-1}{e^{\beta_2}-1} \tag{3.3.4c}$$

$$\begin{aligned}
0\leqslant x\leqslant 1 &\Rightarrow y=\tilde{\eta}+(1-\tilde{\eta})\left(\sqrt{9.89105-(x-\tfrac{1}{2})^2}-3.105\right) \\
x\leqslant 0 \text{ or } x>1 &\Rightarrow y=\tilde{\eta}
\end{aligned} \tag{3.3.4d}$$

with $\beta_1=1.26$ and $\beta_2=1.01$.
At level l, $l=1,2,3,4,5$ the vertices of the quadrilateral volumes $\Omega_{i,j}$ in the (x,y)-space correspond to a regular square mesh over $6.2^{l-1}\times 2.2^{l-1}$ volumes on $[-1,2]\times[0,1]$ in the (ξ,η)-plane.
The nested sequence of grids that was employed is given in fig. 3.3.13.

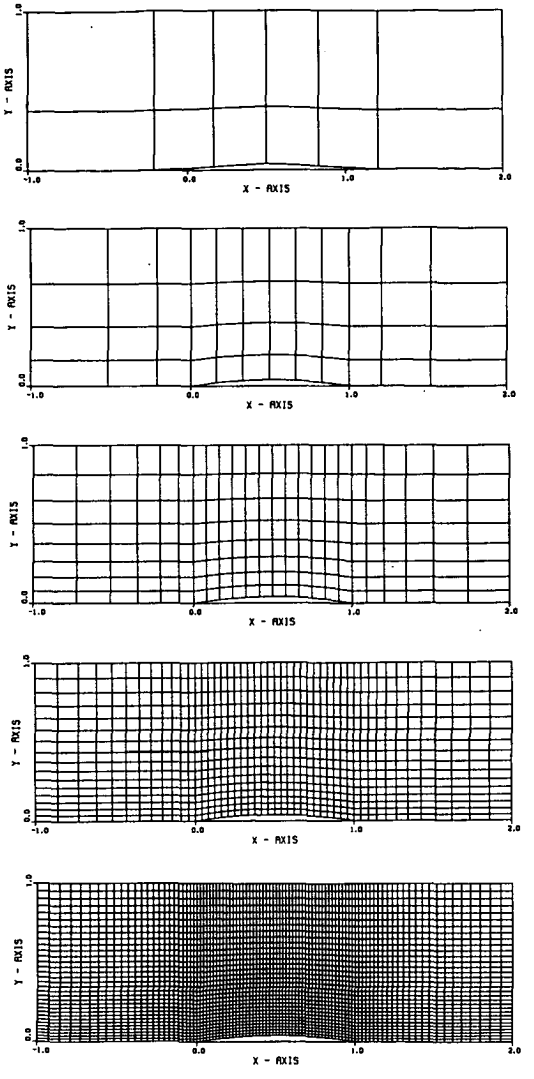


FIGURE 3.3.13. The nested sequence of grids for testproblem 2.

At the inflow boundary ($x=-1$) we prescribe $M_{\text{inlet}}=1.4$. We take $u_{\text{inlet}}=M_{\text{inlet}}$, $v_{\text{inlet}}=0$, $c_{\text{inlet}}=1$, $z_{\text{inlet}}=-\gamma\ln(\gamma)$.
In fig. 3.3.14,15, we present the convergence histories of the NMG-iteration process on grid 48×16 and grid 96×32, respectively. As expected, the convergence rate is excellent. For later comparison (see chapter IV), the first-order solutions on grids 48×16 and 96×32 are presented.

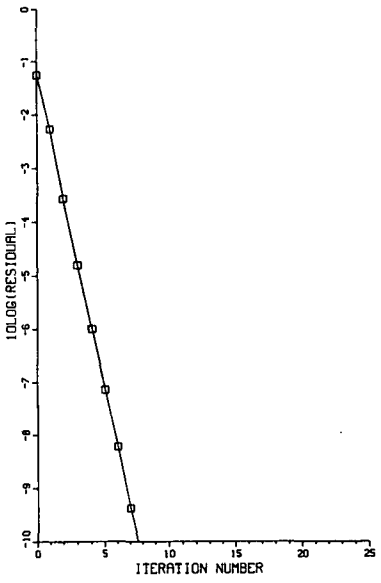


FIGURE 3.3.14. Convergence history of the NMG-iteration process for testproblem 2 (48 × 16 grid).

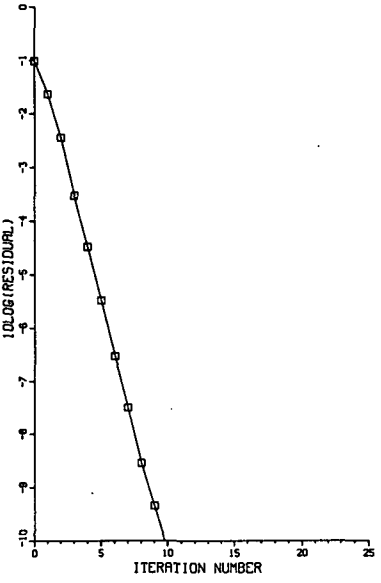


FIGURE 3.3.15. As figure 3.3.14 but on grid 96 × 32.

In fig. 3.3.16,17 the Mach number distributions obtained are shown. Two oblique shocks are formed at both corners of the bump. Due to discretization errors, the shocks loose sharpness as one moves out from the lower wall; the reflection of the leading-edge shock by the upper wall is hardly visible (cf. the second-order solutions presented in chapter IV). The leading-edge shock is spread numerically over 6 volumes halfway the channel ($y = 0.5$).

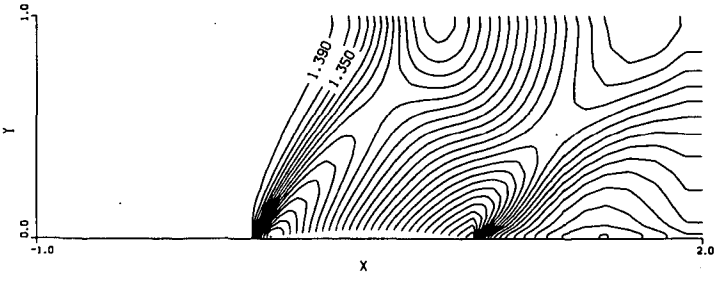


FIGURE 3.3.16. Iso-Mach lines for testproblem 2 (48 × 16 grid).

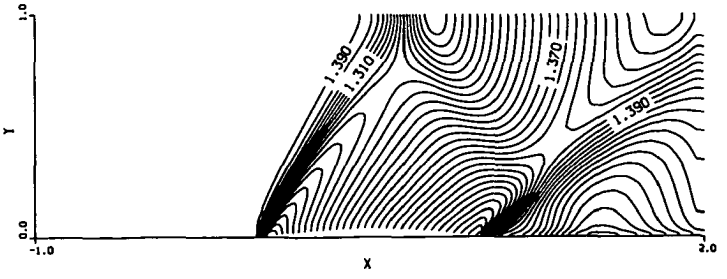


FIGURE 3.3.17. As figure 3.3.16 but on grid 96 × 32.

In fig. 3.3.18-21 we give the Mach number and the entropy distributions along the lower surface of the channel.

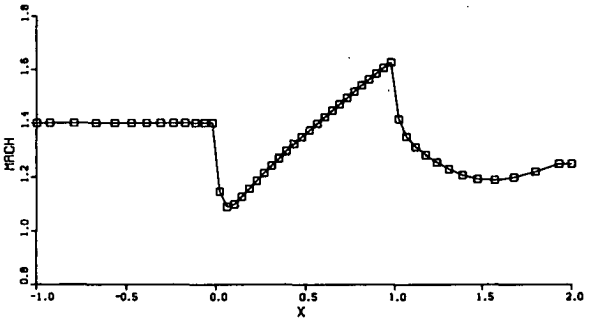


FIGURE 3.3.18. Mach number distribution along the lower surface of the channel for testproblem 2 (48 × 16 grid).

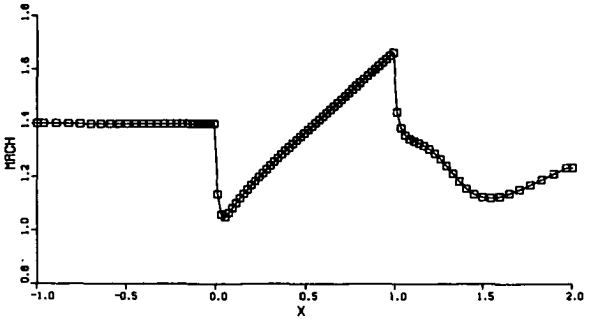


FIGURE 3.3.19. As figure 3.3.18 but on grid 96 × 32.

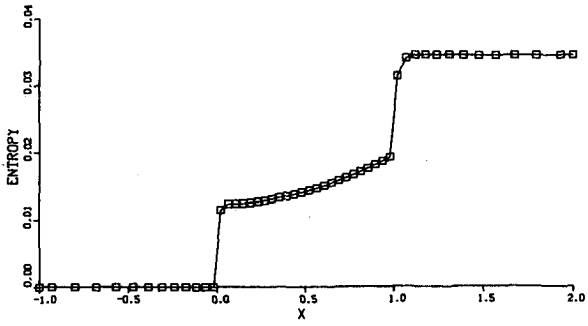


FIGURE 3.3.20. Entropy distribution along the lower surface of the channel for testproblem 2 (48 × 16 grid).

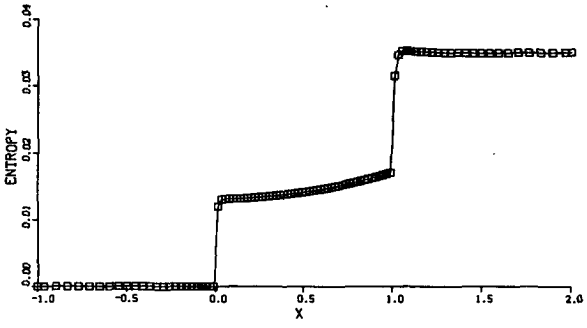


FIGURE 3.3.21. As figure 3.3.20 but on grid 96 × 32.

PROBLEM 3. *Resolution of contact discontinuities.*
Here the physical and computational domain are the same, i.e. the mapping is the identity: $x=\xi$, $y=\eta$. We take $\Omega=[0,1]\times[0,1]$, the coarsest grid is 2×2 volumes, the finest grid is 32×32 volumes (level 5).
Two problems are considered:

PROBLEM 3a. *Contact discontinuity aligned to the grid.*
The boundary conditions are ($s=p/\rho^\gamma$, $z=\ln(s)$).

$$x=0,\ 0\leqslant y\leqslant 0.5 : u=0.25,\ v=0,\ s=0.5$$

$$x=0,\ 0.5\leqslant y\leqslant 1 : u=0.75,\ v=0,\ s=1$$

$$\textit{other boundaries} : p=1\,.$$

(see fig. 3.3.22).
At the north, east and south boundary of the domain Ω, the boundary condition is treated as described in example 2.3.2a (subsonic outflow). At the west boundary, the boundary condition treatment is as described in example 2.3.2b. The exact solution of this problem has a contact discontinuity at $y=0.5$. In both parts of the domain the solution has a uniform state: for $y<0.5$ we have $u=0.25$, $v=0$, $s=0.5$, $p=1$; for $y>0.5$ we have $u=0.75$, $v=0$, $s=1$, $p=1$.

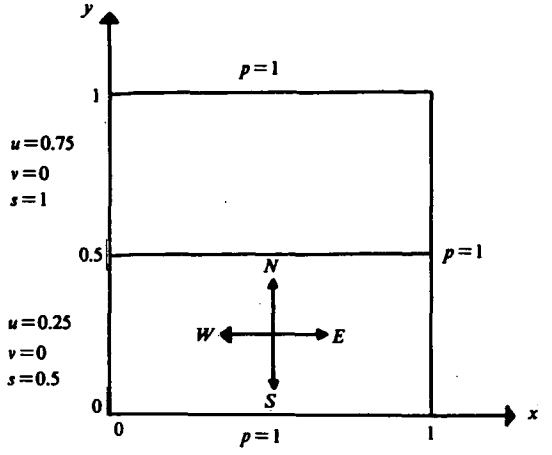


FIGURE 3.3.22. The domain Ω with boundary conditions corresponding to testproblem 3a.

Because the contact discontinuity coincides with a grid line there is no discretization error, and the numerical solution should be exact. The convergence history of the NMG-iteration process is shown in fig. 3.3.23 and the entropy distribution along the line $x=0.5$ is shown in fig. 3.3.24. From this last figure we see that the solution obtained is indeed exact.

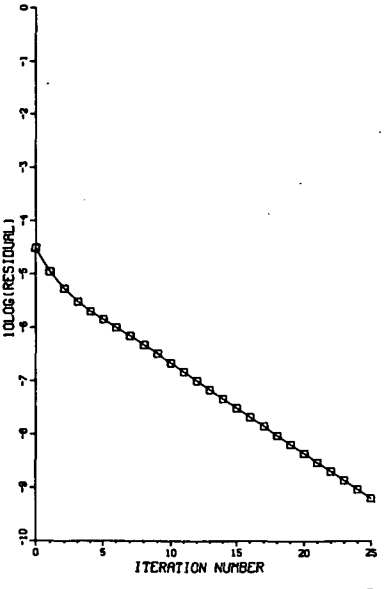


FIGURE 3.3.23. Convergence history of the NMG-iteration process of testproblem 3a (32 × 32 grid).

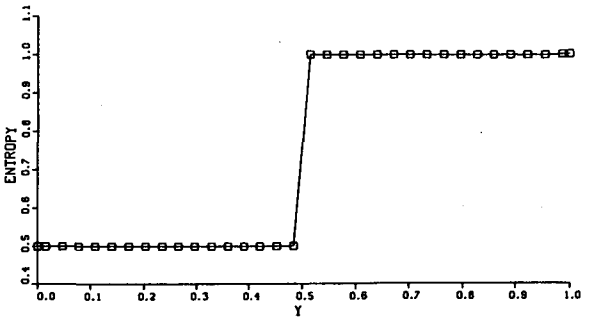


FIGURE 3.3.24. The entropy distribution along the line $x=0.5$ for testproblem 3a (32×32 grid).

PROBLEM 3b. *Oblique contact discontinuity.*
The boundary conditions are

West boundary $x=0 : u=\sqrt{2}/8,\ v=-\sqrt{2}/8,\ s=1/2$.

North boundary $y=1 : u=3\sqrt{2}/8,\ v=-3\sqrt{2}/8,\ s=1$.

East boundary $x=1 : p=1$.

South boundary $y=0 : p=1$.

The exact solution of this problem has a contact discontinuity at $x+y=1$. In both parts of the domain the solution is uniform. For $x+y<1$ we have $u=\sqrt{2}/8,\ v=-\sqrt{2}/8,\ s=1/2,\ p=1$; for $x+y>1$ we have $u=3\sqrt{2}/8$, $v=-3\sqrt{2}/8,\ s=1,\ p=1$.
The outflow boundaries ($p=1$) are treated as in example 2.3.2a, the inflow boundaries are treated as in example 2.3.2b.

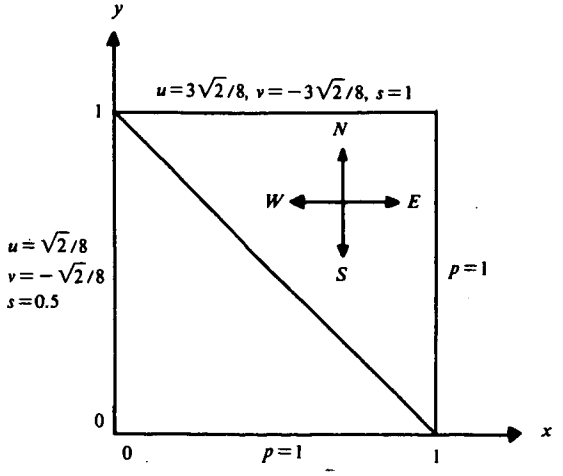


FIGURE 3.3.25. The domain Ω with boundary conditions for testproblem 3b.

In fig. 3.3.26 we present the convergence history of the NMG-iteration process. In fig. 3.3.27 we present the entropy distribution along the line $x=0.5$ and entropy contours are shown in fig. 3.3.28. We observe considerable smearing out of the discontinuity.

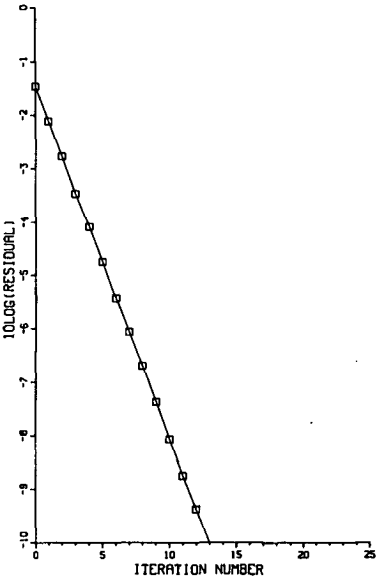


FIGURE 3.3.26. Convergence history of the NMG-iteration process for testproblem 3b (32×32 grid).

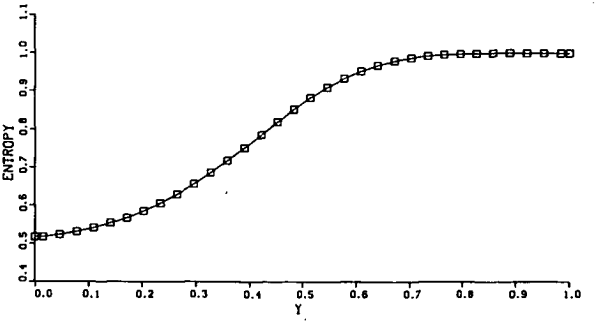


FIGURE 3.3.27. The entropy distribution along the line $x=0.5$ for testproblem 3b (32×32 grid).

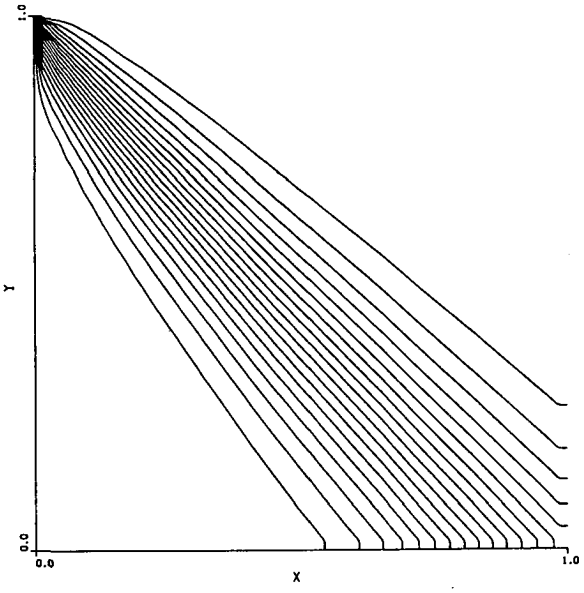
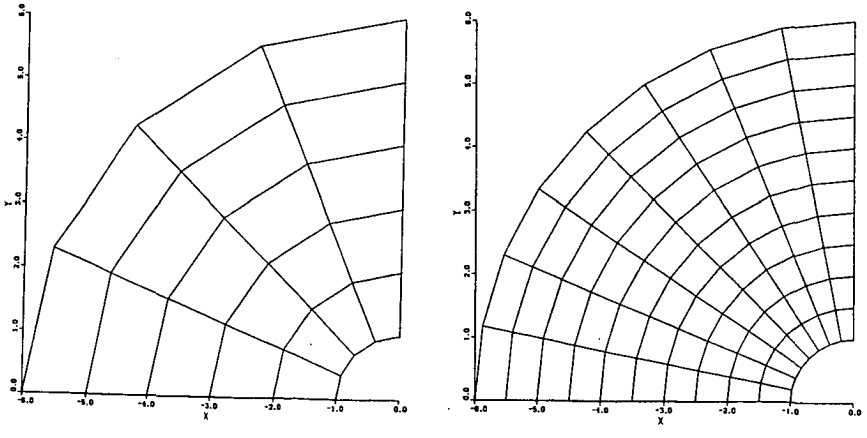


FIGURE 3.3.28. Entropy contours for testproblem 3b (32×32 grid).

PROBLEM 4. *Cylinder in a supersonic flow*
The grids used for the multigrid computation are shown in fig. 3.3.29. The mapping from the (ξ,η)-computational space to the (x,y)-physical space with $(\xi,\eta)\in[1,6]\times[\pi/2,\pi]$ is

$$x=\xi\cos\eta,\ y=\xi\sin\eta\ . \tag{3.3.5}$$

At level l, $l=1,2,3,4$ the computational space is subdivided in $5.2^{l-1}\times4.2^{l-1}$ rectangular volumes.



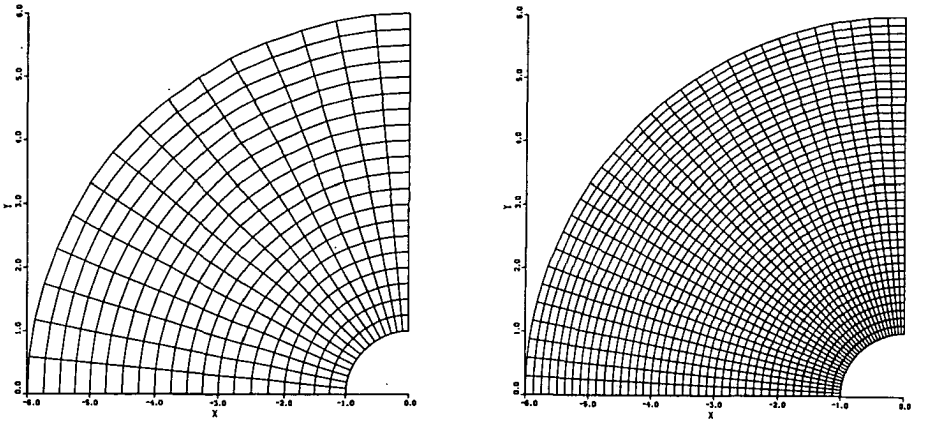


FIGURE 3.3.29. The nested sequence of grids for testproblem 4.

The free-stream Mach number is $M_{in}=2.0$. At inflow the supersonic boundary condition is used (u_{in}, v_{in}, c_{in} , z_{in} are prescribed), at outflow the supersonic outflow boundary condition is used (no boundary values are prescribed) and the other two boundaries are treated as a solid wall (see section 2.3.2).

The standard multigrid solution method starts, in the nested iteration stage with a uniform constant flow field given by u_{in}, v_{in}, c_{in} and z_{in} on the coarsest mesh. Unfortunately, divergence was observed for the local Newton iteration process in the CSGS-relaxation method on the coarsest grid. A simple remedy for this problem is the following continuation method. At the coarsest grid, we start with $M_{in}=0.1$ and with a corresponding uniform flow field. Then 2 CSGS-relaxations are performed to improve the solution. After the improvement, the inflow boundary condition is changed such that $M_{in}^{new}:=M_{in}^{old}+\Delta M$ and 2 CSGS-ralaxations are performed with this new inflow boundary condition. We take $\Delta M=0.1$. This process is repeated until $M_{in}=2.0$. This inexpensive continuation process results in a good initial approximation on the coarsest mesh. The standard multigrid method starts, in the nested iteration stage, with this initial approximation and no further problems were encountered.

In figure 3.3.30, the convergence history of the NMG-iteration process is shown.

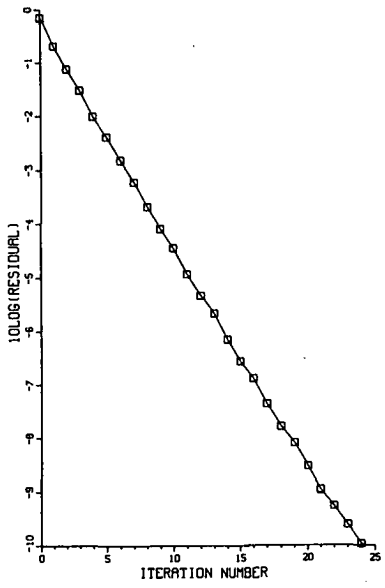


FIGURE 3.3.30. Convergence history of the NMG-iteration process for testproblem 4 (40 × 32 grid).

In fig. 3.3.31,32 the Mach number and pressure destributions are shown. The bow shock is clearly visible. The bow shock starts at $x = -2.5$ and this result agrees well with the first-order results published in [8]. In fig. 3.3.33 the surface pressure distribution along the cylinder is shown.

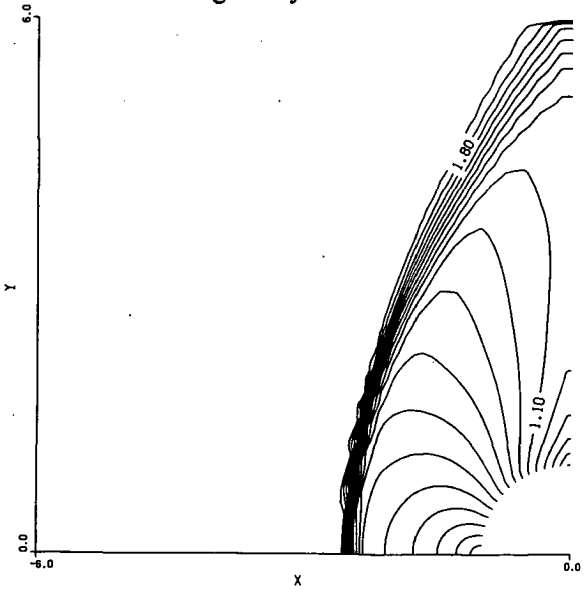


FIGURE 3.3.31. Iso-Mach lines for testproblem 4 (40×32 grid).

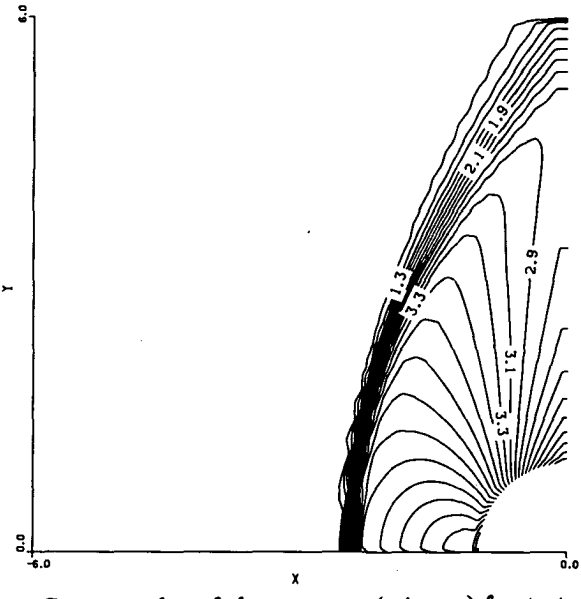


FIGURE 3.3.32. Contour plot of the pressure ($p/p_{-\infty}$) for testproblem 4 (40 × 32 grid).

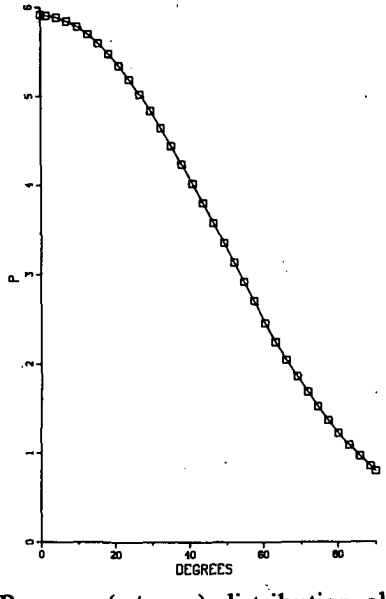


FIGURE 3.3.33. Pressure ($p/p_{-\infty}$) distribution along the surface of the cylinder for testproblem 4 (40 × 32 grid).

For these four testproblems we can conclude that the multigrid method has the following features:

- *Robustness.* All problems are solved with a fixed multigrid schedule.

- *Efficiency.* The convergence histories are excellent. For practical purposes, where one only wants to get below truncation error, after a start with nested iterations, a few (two or three) NMG-iterations are sufficient. This means that first-order solutions are obtained in an amount of work that is equivalent with about $3 \times \frac{4}{3} \times 2$ CSGS-relaxation on the finest grid.

- *Grid independence of the convergence rate.* This is observed for the transonic and supersonic test-problems 1a,b. However, for flows with smaller Mach numbers the convergence rate slightly decreases (see testproblem 1c).

However, the quality of the solutions obtained is not satisfactory. Especially oblique discontinuities are captured badly. The accuracy of the solutions can be improved. This becomes clear by comparing the first-order solutions with the second-order solutions presented in chapter IV.

References

[1] A. Brandt (1982). *Guide to Multigrid Development.* In: Multigrid-Methods (W. Hackbusch and U. Trottenberg, eds.), Lecture Notes in Mathematics 960, pp. 220-312, Springer Verlag, Berlin.

[2] W. Hackbusch (1985). *Multi-Grid Methods and Applications.* Springer Series in Computational Mathematics 4. Springer-Verlag, Berlin.

[3] P.W. Hemker and G.M. Johnson (1987). *Approaches to the Euler Equations.* In: SIAM 'Frontiers in Appl. Math'. Vol. 5, (S. McCormick, ed.) To appear.

[4] P.W. Hemker and S.P. Spekreijse (1985). *Multigrid Solution of the Steady Euler Equations.* In: Advances in Multigrid Methods (D. Braess, W. Hackbusch, U. Trottenberg, eds.). Notes on Numerical Fluid Mechanics, Vol. 11, pp. 33-44, Vieweg, Braunschweig.

[5] P.W. Hemker and S.P. Spekreijse (1986). *Multiple Grid and Osher's Scheme for the Efficient Solution of the Steady Euler Equations.* Appl. Num. Math. 2, pp. 475-493.

[6] D.C. Jespersen (1983). *Design and Implementation of a Multigrid Code for the Euler Equations.* Appl. Math. and Computat. 13, pp. 357-374.

[7] R.H. Ni (1982). *A Multiple-Grid Scheme for Solving the Euler Equations.* AIAA J. 20, pp. 1565-1571.

[8] M.H. Ray (1986). *A Relaxation Approach to Patched-Grid Calculations with the Euler Equations.* J. Comp. Phys. 66, pp. 99-131.

[9] A. Rizzi and H. Viviand (1981). *Numerical Methods for the Computation of Inviscid Transonic Flows with Shock Waves.* Notes on numerical fluid mechanics, Vol. 3. Vieweg, Braunschweig.

[10] J.Y. Yang (1985). *Numerical Solution of the Two-Dimensional Euler Equations by Second-Order Upwind Difference Schemes.* AIAA-85-0292. AIAA 23rd Aerospace Sciences Meeting, Reno, Nevada.

Chapter IV

Defect Correction for Second-Order Accuracy

4.1. Introduction

In chapter II we have derived a first-order and a monotone second-order upwind scheme for the steady Euler equations. Both discretizations are conservative and can be written as

$$F_h(q_h)=r_h \tag{4.1.1}$$

where

$$F_h(q_h)_{i,j}=f_{i+\frac{1}{2},j}+f_{i,j+\frac{1}{2}}-f_{i-\frac{1}{2},j}-f_{i,j-\frac{1}{2}} \tag{4.1.2}$$

and

$$\begin{aligned} f_{i+\frac{1}{2},j}&=l_{i+\frac{1}{2},j}T^{-1}_{i+\frac{1}{2},j}f_R(T_{i+\frac{1}{2},j}q^L_{h,i+\frac{1}{2},j},\ T_{i+\frac{1}{2},j}q^R_{h,i+\frac{1}{2},j})\\ f_{i,j+\frac{1}{2}}&=l_{i,j+\frac{1}{2}}T^{-1}_{i,j+\frac{1}{2}}f_R(T_{i,j+\frac{1}{2}}q^L_{h,i,j+\frac{1}{2}},\ T_{i,j+\frac{1}{2}}q^R_{h,i,j+\frac{1}{2}}) \end{aligned} \tag{4.1.3}$$

(see 2.1.15, 2.5.1.1,2). The discretization (4.1.1) is first-order accurate with (omitting the subscript h)

$$q^L_{i+\frac{1}{2},j}=q_{i,j}\ \ ;\ \ q^R_{i+\frac{1}{2},j}=q_{i+1,j} \tag{4.1.4}$$

and analogous expressions for $q^{L,R}_{i,j+\frac{1}{2}}$.
Then $F_h(q_h)=F^1_h(q_h)$, the first-order space discretization operator. The discretization (4.1.1) is monotone and second-order accurate with (again omitting the subscript h, and denoting the kth component ($k=1,2,3,4$) with superscript k)

$$\begin{aligned} q^{L\,(k)}_{i+\frac{1}{2},j}&=q^{(k)}_{i,j}+\frac{1}{2}\psi_0(R^{(k)}_{i,j})\,(q^{(k)}_{i,j}-q^{(k)}_{i-1,j})\\ q^{R\,(k)}_{i+\frac{1}{2},j}&=q^{(k)}_{i+1,j}+\frac{1}{2}\psi_0\Big(\frac{1}{R^{(k)}_{i+1,j}}\Big)\,(q^{(k)}_{i+1,j}-q^{(k)}_{i+2,j}) \end{aligned} \tag{4.1.5}$$

where

$$R^{(k)}_{i,j}=\frac{q^{(k)}_{i+1,j}-q^{(k)}_{i,j}}{q^{(k)}_{i,j}-q^{(k)}_{i-1,j}}\,. \tag{4.1.6}$$

and where $\psi_0 : \mathbb{R}\mapsto\mathbb{R}$ is the Van Albada limiter defined as

$$\psi_0(R)=\frac{R^2+R}{R^2+1} \tag{4.1.7}$$

(see 2.5.3.44,45). Then $F_h(q_h)=F_h^2(q_h)$, the monotone second-order space discretization operator.

In chapter III we have developed a robust and efficient nonlinear multigrid method for solving

$$F_h^1(q_h)=r_h \ . \tag{4.1.8}$$

We have already mentioned that first-order accuracy is too low for practical problems. Therefore, the accuracy has to be improved. There are roughly two ways to improve the accuracy. A first possibility is to construct an efficient and robust multigrid solver for

$$F_h^2(q_h)=r_h \ . \tag{4.1.9}$$

Unfortunately, for this system of equations there are no relaxation methods available with good smoothing properties for damping short wavelength error components. The smoothing properties of point-relaxations are insufficient [6]. A good alternative seems the block relaxation method proposed in [12]. But experiments with a multigrid method which uses this block relaxation as a smoothing operator show a rather disappointing convergence behaviour in case of solutions with shocks [unpublished results].

A second possilibity is to solve (4.1.9) in an indirect way, making use of the excellent multigrid solver for (4.1.8). This can be done by the following Defect Correction (DeC-) iteration process:

$$\begin{cases} F_h^1(q_h^1)=r_h \\ F_h^1(q_h^{i+1})=F_h^1(q_h^i)+(r_h-F_h^2(q_h^i)) \quad i=1,2,.... \end{cases} \tag{4.1.10}$$

It is clear that the fixed point of this iteration process is the solution of (4.1.9). But what is the convergence rate of this DeC-iteration process? In section 4.3, numerical results show that the convergence rate is in general rather slow. Therefore, the DeC-iteration process is not an efficient process to obtain the exact solution of (4.1.9). Fortunately, to obtain second-order accurate solutions it is not necessary at all to iterate until convergence. For problems with smooth solutions, a single DeC-iteration is sufficient to obtain second-order accuracy [4]. In case of the Euler equations, where solutions are in general discontinuous, experiments show that a few (5-10) DeC-iterations significantly improve the accuracy of the solution [7]. From these considerations it follows that (4.1.10) must be taken as a finite process (i.e. a small number of iterations are performed) only used to improve the accuracy of the first-order solution q_h^1.

In section 4.2 we derive some general theoretical results for the DeC-iteration process. A description is given of the complete solution process to obtain second-order accurate solutions of the steady Euler equations.

In section 4.3 numerical results are given for the testproblems of section 3.3. It is shown that the first-order solutions presented in section 3.3 are improved considerably by a few DeC-iterations.

Finally, in section 4.4 we consider the special and interesting case of the steady

Euler equations with a source term. In that case, second-order accurate solutions are obtained again very easily by the defect correction method.

4.2. The defect correction method

In this section, the defect correction method is considered in a general context. The main result is theorem 4.2b. From this theorem it follows that a small number of DeC-iterations is sufficient to improve the accuracy. An alternative proof is also given. Finally, the results of the analysis are used to develop a complete solution process for obtaining second-order accurate solutions of the steady Euler equations. A large part of this section has been published elsewhere [1, 2, 3, 11].

Consider the problem

$$Fq = r^* \tag{4.2.1}$$

where $F : X \mapsto Y$ and $r^* \in Y$ are given and X and Y are normed vector spaces. We may think of X and Y as being infinite dimensional function spaces. A discretization of (4.2.1) is an associated problem

$$F_h q_h = r_h^* \tag{4.2.2}$$

where $F_h : X_h \mapsto Y_h$ and $r_h^* \in Y_h$ is given and X_h and Y_h are normed vector spaces. The relation between the problem and its discretization is obtained by introducing surjections $R_h : X \mapsto X_h$, $\bar{R}_h : Y \mapsto Y_h$. The relation between the various spaces and mappings in the discretization is summarized in the following diagram (see also subsection 2.5.2):

$$\begin{array}{ccc} X & \xrightarrow{F} & Y \\ \Big\downarrow{\scriptstyle R_h} & & \Big\downarrow{\scriptstyle \bar{R}_h} \\ X_h & \xrightarrow{F_h} & Y_h \end{array}$$

By assuming $h \in (0,H)$, $H>0$, a sequence of discretizations of (4.2.1) is obtained.
Suppose

$$r_h^* = \bar{R}_h r^* \tag{4.2.3}$$

and let q^* denote the solution of (4.2.1) and q_h^* the solution of (4.2.2). The truncation error τ_h^* is defined as

$$\tau_h^* = r_h^* - F_h R_h q^* = (\bar{R}_h F - F_h R_h) q^* \tag{4.2.4}$$

and the sequence of discretizations (4.2.2) with $h \in (0,H)$ is called *consistent* (of order p) *with the problem* (4.2.1) if

$$||\tau_h^*||_{Y_h} = O(h^p) . \tag{4.2.5}$$

Define the truncation error operator $\tau_h : X \mapsto Y_h$ by

$$\tau_h = \overline{R}_h F - F_h R_h . \tag{4.2.6}$$

DEFINITION 4.2a. (*Consistency*).
The sequence of discretizations is *consistent* of order p if

$$||\tau_h(q)||_{Y_h} \leqslant C_1(q) h^p \quad \forall q \in X,\ h \in (0,H) . \tag{4.2.7}$$

The discretization error $\epsilon_h^* \in X_h$ is defined as

$$\epsilon_h^* = R_h q^* - q_h^* = (R_h - F_h^{-1} \overline{R}_h F) q^* \tag{4.2.8}$$

where we have assumed that F_h is a bijection. The sequence of discretizations (4.2.2) with $h \in (0,H)$ is called discrete *convergent* (of order p) *to the solution of* (4.2.1) if

$$||\epsilon_h^*||_{X_h} = O(h^p) . \tag{4.2.9}$$

Define the discretization error operator $\epsilon_h : X \mapsto X_h$ by

$$\epsilon_h = R_h - F_h^{-1} \overline{R}_h F . \tag{4.2.10}$$

DEFINITION 4.2b. (*Convergence*).
The sequence of discretizations is *convergent* of order p if

$$||\epsilon_h(q)||_{X_h} \leqslant C_2(q) h^p \quad \forall q \in X,\ h \in (0,H) . \tag{4.2.11}$$

Another important concept is the *stability* of the discretizations:

DEFINITION 4.2c. (*Stability*).
The sequence of discretizations is *stable* if there exist a $C_3 > 0$ (independent of h) such that

$$||F_h^{-1} r_h - F_h^{-1} \tilde{r}_h||_{X_h} \leqslant C_3 ||r_h - \tilde{r}_h||_{Y_h} \quad \forall r_h,\ \tilde{r}_h \in Y_h,\ h \in (0,H) . \tag{4.2.12}$$

The following well known theorem is proved easily.

THEOREM 4.2a. (*Equivalence theorem*).
If a sequence of discretizations is stable and consistent of order p then it is convergent of order p.

PROOF.

$$\begin{aligned} ||\epsilon_h(q)||_{X_h} &= ||R_h q - F_h^{-1} \overline{R}_h F q||_{X_h} \\ &= ||F_h^{-1} F_h R_h q - F_h^{-1} \overline{R}_h F q||_{X_h} \end{aligned}$$

$$\leq C_3 \| F_h R_h q - \bar{R}_h F q \|_{Y_h}$$
$$\leq C_3 C_1(q) \cdot h^p \qquad \square$$

We will now formulate the main theorem about the defect correction method. Consider two different discretizations F_h, $\tilde{F}_h : X_h \mapsto Y_h$. Assume that F_h is stable and consistent of order p, $\tilde{F}_h$ is consistent of order $\tilde{p} > p$. Consider the DeC-iteration process:

$$\begin{cases} F_h q_h^1 = r_h^* \\ F_h q_h^{i+1} = F_h q_h^i + (r_h^* - \tilde{F}_h q_h^i) \quad i = 1,2,... \end{cases} \tag{4.2.13}$$

First we need the concept of relative consistency.

DEFINITION 4.2d. (*Relative consistency*).
Two sequences of discretizations F_h, $\tilde{F}_h$ are *relatively consistent* of order p if there exist a $C_4 > 0$ such that

$$\| (F_h - \tilde{F}_h) q_h - (F_h - \tilde{F}_h) \tilde{q}_h \|_{Y_h} \leq C_4 h^p \| q_h - \tilde{q}_h \|_{X_h} \quad \forall q_h, \tilde{q}_h \in X_h,\ h \in (0,H) . \tag{4.2.14}$$

THEOREM 4.2b. (*DeC-iteration*).
Let F_h, $\tilde{F}_h : X_h \mapsto Y_h$ be two different discretizations. Assume:

- *F_h is stable and consistent of order p.*
- *$\tilde{F}_h$ is consistent of order $\tilde{p} > p$.*
- *F_h and $\tilde{F}_h$ are relatively consistent of order p.*

Then the i^{th} iterate q_h^i of the DeC-iteration process (4.2.13) satisfies

$$\| q_h^i - R_h q^* \|_{X_h} \leq C h^{\min(\tilde{p},\, ip)} . \tag{4.2.15}$$

PROOF.
The theorem is proven by induction; (4.2.15) is true for $i = 1$. Assume that (4.2.15) is true for i. Then

$$\| q_h^{i+1} - R_h q^* \|_{X_h} = \| F_h^{-1} (F_h q_h^i + r_h^* - \tilde{F}_h q_h^i) - R_h q^* \|_{X_h}$$
$$\leq C_3 \| F_h q_h^i + \bar{R}_h F q^* - \tilde{F}_h q_h^i - \tilde{F}_h R_h q^* + \tilde{F}_h R_h q^* - F_h R_h q^* \|_{Y_h}$$
$$\leq C_3 \left\{ \| (\bar{R}_h F - \tilde{F}_h R_h) q^* \|_{Y_h} + \| (F_h - \tilde{F}_h) q_h^i - (F_h - \tilde{F}_h) R_h q^* \|_{Y_h} \right\}$$
$$\leq C_3 \left\{ C_1(q^*) h^{\tilde{p}} + C_4 h^p \| q_h^i - R_h q^* \|_{Y_h} \right\}$$

$$\leqslant \tilde{C}h^{\tilde{p}} + \tilde{C} \cdot h^{p} \cdot h^{\min(\tilde{p}, ip)}$$

$$\leqslant Ch^{\min(\tilde{p}, (i+1)p)}. \qquad \square$$

REMARK (4.2a).
In general, the relative consistency of the discretizations F_h and $\tilde{F}_h$ can be established only for $q_h = R_h q$, $\tilde{q}_h = R_h \tilde{q}$ where q, $\tilde{q} \in X$ are sufficiently smooth. Therefore theorem 4.2b is only applicable to problems with sufficiently smooth solutions.

In case of the Euler equations we have $F_h = F_h^1$ (the first-order space discretiation operator) and $\tilde{F}_h = F_h^2$ (the monotone second-order space discretization operator). So $p = 1$ and $\tilde{p} = 2$ and it follows from theorem 4.2b that a single DeC-iteration is sufficient to obtain second-order accuracy, at least for smooth problems. This has been confirmed by experiments [4].
An alternative proof of theorem 4.2a can be given when F, F_h and $\tilde{F}_h$ are linear scalar differential operators with constant coefficients. Then the symbols $F(\omega)$, $F_h(\omega)$ and $\tilde{F}_h(\omega)$ of respectively F, F_h and $\tilde{F}_h$ are defined as

$$F(e^{i\omega x}) = F(\omega)e^{i\omega x};\ F_h(e^{i\omega x}) = F_h(\omega)e^{i\omega x};\ \tilde{F}_h(e^{i\omega x}) = \tilde{F}_h(\omega)e^{i\omega x} \tag{4.2.16}$$

and the consistency of F_h and $\tilde{F}_h$ can be expressed as

$$\begin{aligned} F(\omega) - F_h(\omega) &= O(h^p) \\ F(\omega) - \tilde{F}_h(\omega) &= O(h^{\tilde{p}}) \end{aligned} \tag{4.2.17}$$

with ω fixed.

EXAMPLE 4.2a.
Suppose

$$F = aq_x + bq_y, \quad a,\ b > 0 \tag{4.2.18a}$$

$$(F_h q)_{i,j} = \frac{a}{h}(q_{i,j} - q_{i-1,j}) + \frac{b}{h}(q_{i,j} - q_{i,j-1}) \tag{4.2.18b}$$

$$\begin{aligned} (\tilde{F}_h q)_{i,j} = \frac{a}{h}\{ & q_{i,j} + \frac{1+\kappa}{4}(q_{i+1,j} - q_{i,j}) + \frac{1-\kappa}{4}(q_{i,j} - q_{i-1,j}) \\ & - q_{i-1,j} - \frac{1+\kappa}{4}(q_{i,j} - q_{i-1,j}) - \frac{1-\kappa}{4}(q_{i-1,j} - q_{i-2,j})\} \\ + \frac{b}{h}\{ & q_{i,j} + \frac{1+\kappa}{4}(q_{i,j+1} - q_{i,j}) + \frac{1-\kappa}{4}(q_{i,j} - q_{i,j-1}) \\ & - q_{i,j-1} - \frac{1+\kappa}{4}(q_{i,j} - q_{i,j-1}) - \frac{1-\kappa}{4}(q_{i,j-1} - q_{i,j-2})\} \end{aligned} \tag{4.2.18c}$$

Notice that F_h is the first-order upwind discretization of F, $\tilde{F}_h$ is the second-order κ-scheme (see 2.5.1.3).
It is easily verified that

$$F(\omega)=i(\omega_1 a+\omega_2 b) \tag{4.2.19a}$$

$$F_h(\omega)=\frac{a}{h}(1-e^{-i\omega_1 h})+\frac{b}{h}(1-e^{-i\omega_2 h}) \tag{4.2.19b}$$

$$\tilde{F}_h(\omega)=\frac{a}{h}(1-e^{-i\omega_1 h})\left(1-\frac{\kappa}{2}+\frac{1+\kappa}{4}e^{i\omega_1 h}-\frac{1-\kappa}{4}e^{-i\omega_1 h}\right)$$
$$+\frac{b}{h}(1-e^{-i\omega_2 h})\left(1-\frac{\kappa}{2}+\frac{1+\kappa}{4}e^{i\omega_2 h}-\frac{1-\kappa}{4}e^{-i\omega_2 h}\right) \tag{4.2.19c}$$

and

$$F(\omega)-F_h(\omega)=-\frac{1}{2}(a\omega_1^2+b\omega_2^2)h+O(h^2) \tag{4.2.20a}$$

$$F(\omega)-\tilde{F}_h(\omega)=\frac{3\kappa-1}{12}i(a\omega_1^3+b\omega_2^3)h^2+O(h^3)\,. \tag{4.2.20b}$$

Consider the DeC-iteration process (4.2.13) and assume that F, F_h and $\tilde{F}_h$ are linear scalar operators satisfying (4.2.17). Suppose that the ith iterand of the DeC-iteration process satisfies the linear equation

$$A_h^i q_h^i=r_h^* \tag{4.2.21}$$

and denote with $A_h^i(\omega)$ the symbol of the linear operator A_h^i. Notice that $A_h^1=F_h$, $A_h^1(\omega)=F_h(\omega)$. Because

$$F_h q_h^{i+1}=F_h q_h^i+A_h^i q_h^i-\tilde{F}_h q_h^i \tag{4.2.22}$$

we find

$$(A_h^i+F_h-\tilde{F}_h)^{-1}F_h q_h^{i+1}=q_h^i=(A_h^i)^{-1}r_h^*\,. \tag{4.2.23}$$

Hence,

$$A_h^{i+1}=A_h^i(A_h^i+F_h-\tilde{F}_h)^{-1}F_h \tag{4.2.24}$$

and

$$A_h^{i+1}(\omega)=\frac{A_h^i(\omega)F_h(\omega)}{A_h^i(\omega)+F_h(\omega)-\tilde{F}_h(\omega)}\,. \tag{4.2.25}$$

It is easily seen that

$$F(\omega)-A_h^{i+1}(\omega)=\frac{F(\omega)(F(\omega)-\tilde{F}_h(\omega))-(F(\omega)-A_h^i(\omega))(F(\omega)-F_h(\omega))}{A_h^i(\omega)+F_h(\omega)-\tilde{F}_h(\omega)}\,. \tag{4.2.26}$$

Assume

$$F(\omega)-A_h^i(\omega)=O(h^{\min(\tilde{p},ip)}) \tag{4.2.27}$$

then it follows from (4.2.26) that

$$F(\omega)-A_h^{i+1}(\omega)=O(h^{\tilde{p}})+O(h^{\min(\tilde{p},ip)})\cdot O(h^p)$$
$$=O(h^{\min(\tilde{p},(i+1)p)})\,. \tag{4.2.28}$$

Because (4.2.27) is true for $i=1$, it follows by induction that (4.2.28) is satisfied for all i. Hence, the ith iterand of the DeC-iteration satisfies (4.2.21)

where A_h^i is a consistent discretization of order $\min(\tilde{p},ip)$.

In case of the Euler equations, we see from (4.2.13) that for each DeC-iteration we have to solve a first-order system with an appropriate right-hand side. It was found that it is inefficient to solve this system very accurately. Application of a single NMG-iteration to approximate q_h^i in (4.2.13) usually is the most efficient strategy [7]. We have to solve the first-order system $F_h^1(q_h^1)=r_h^*$ with $r_h^*=0$ to obtain the first iterand q_h^1 of the DeC-iteration process. It is also inefficient to solve this system very accurately. Therefore q_h^1 is obtained by the nested iteration-NMG method with only a single NMG-iteration (see chapter III). So, the complete multigrid solution process to obtain second-order accurate solutions consists of three successive parts: nested iteration, the NMG-stage and the DeC-stage. In fig. 4.2a we give an illustration of the complete process. Suppose there are 5 nested grids. Between two succeeding points A, B we have one NMG-iteration (V-cycle). Between two succeeding points B, A we have piecewise constant interpolation in the nested iteration stage, and an appropriate right-hand side computation in the DeC-stage.

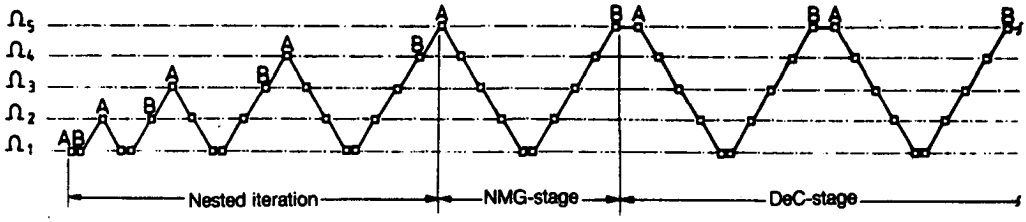


FIGURE 4.2a. Schematic representation of the complete multigrid process to obtain second-order accurate solutions.

4.3. NUMERICAL RESULTS

In this section numerical results are presented for the same testproblems as in section 3.3. For each testproblem we consider the convergence history of the DeC-iteration process, i.e. after each DeC-iteration we compute the L_1-norm of the residual

$$||F_h^2(q_h^i)|| = \sum_{k=1}^{4} \left[\sum_{(i,j)} \left| (F_h^2(q_h^i)_{i,j}^{(k)} \right| \right] \qquad (4.3.1)$$

where q_h^i is the current approximate solution. A fixed number (25) of DeC-iterations is performed for each testproblem. This rather large number (10 DeC-iterations would be sufficient) is only chosen to give a good impression of the convergence behaviour of the DeC-iteration process.
The improvement of the solutions obtained can be seen by comparison with the first-order solutions presented in section 3.3.

PROBLEM 1. *Transonic flow in a channel with a 4.2% circular bump.*
The geometry of the channel and the grids have been given in section 3.3, problem 1. The transonic testproblem is specified by $M_{\text{inlet}}=0.85$, $p_{\text{inlet}}=p_{\text{outlet}}$; (see section 3.3, problem 1b). The result has been obtained on the 40×16 grid. In fig. 4.3.1 we show the convergence history of the DeC-iteration process. Although the convergence is rather slow we may expect that it is possible to drive the residual to machine-zero.
Fig. 4.3.2-4 show respectively the iso-Mach lines, pressure contours and entropy contours of the second-order solution obtained after 25 DeC-iterations. The pressure coefficient c_p is defined in (3.3.3) and the entropy is defined as $(s-s_{-\infty})/s_{-\infty}$ with $s=p\rho^{-\gamma}$. The improved capturing and sharpness of the shock is clearly observed (compare with fig. 3.3.6, 8, 9).
In fig. 4.3.5-7 we give the Mach number, $-c_p$ and entropy distribution along the lower surface of the channel. Especially the entropy distribution shows a clear improvement (compare with fig. 3.3.12). The spurious entropy generation at both corners of the bump is reduced and the spurious entropy rise along the entire bump has disappeared completely. For reference results we refer to [10].

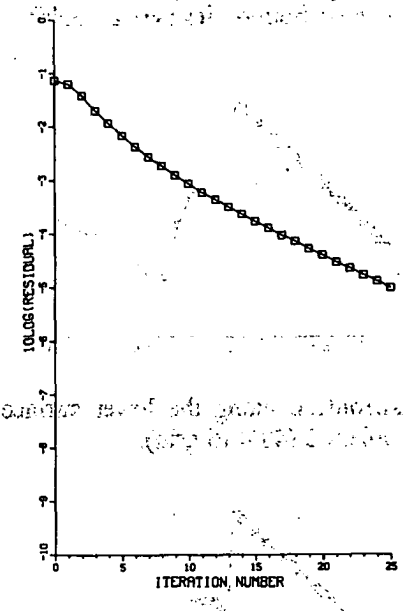


FIGURE 4.3.1. Convergence history of the DeC-iteration process for test-problem 1 (40×16 grid).

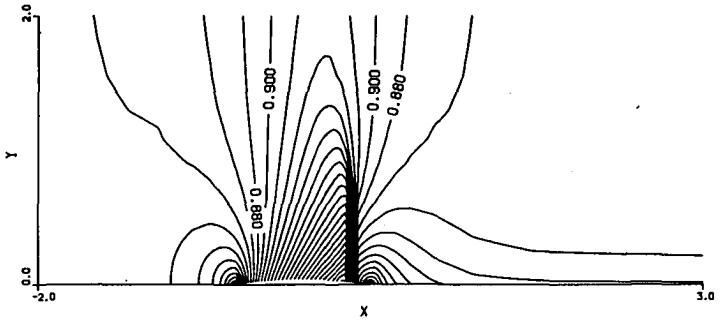


FIGURE 4.3.2. Iso-Mach lines for testproblem 1 (40 × 16 grid).

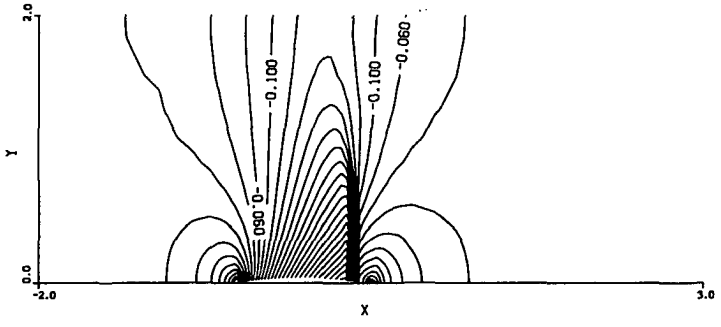


FIGURE 4.3.3. Pressure contours for testproblem 1 (40 × 16 grid).

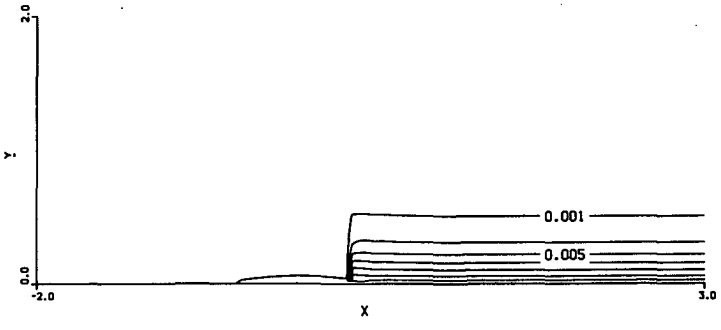


FIGURE 4.3.4. Entropy contours for testproblem 1 (40 × 16 grid).

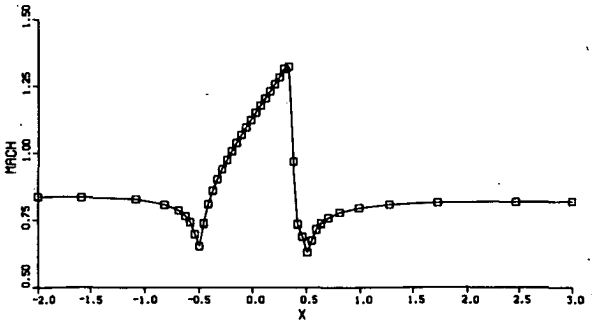


FIGURE 4.3.5. Mach number distribution along the lower channel wall for testproblem 1 (40 × 16 grid).

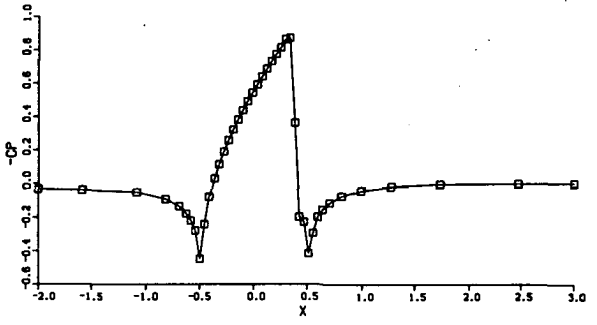


FIGURE 4.3.6. Pressure distribution along the lower channel wall for testproblem 1 (40 × 16 grid).

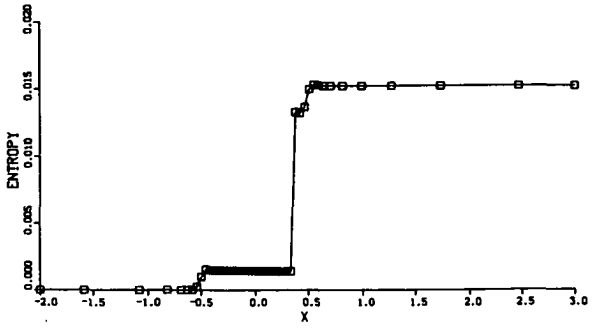


FIGURE 4.3.7. Entropy distribution along the lower channel wall for testproblem 1 (40 × 16 grid).

PROBLEM 2. *Supersonic flow in a channel with a 4% thick circular arc bump.* The geometry of the channel and the grids have been given in section 3.3, problem 2. At the inflow boundary ($x = -1$) the Mach number is prescribed: $M_{\text{inlet}} = 1.4$. We compare the second-order solutions obtained on two different

grids (viz. grid 48×16 and grid 96×32). In section 3.3, first-order solutions are presented on these grids. Therefore, a good comparison between first- and second-order solutions obtained on different grids is possible.
In fig. 4.3.8, 9, we show the convergence history of the DeC-iteration process on grid 48×16 and grid 96×32, respectively. A very slow convergence behaviour is observed for both cases and we may not expect that it is possible to drive the residual to machine-zero. In fig. 4.3.10, 11 we show the convergence history of the ω-DeC-iteration process on grid 48×16 and grid 96×32,

$$\begin{cases} F_h^1(q_h^i)=r_h \\ F_h^1(q_h^{i+1})=F_h^1(q_h^i)+\omega(r_h-F_h^2(q_h^i)) \quad i=1,2,... \end{cases} \tag{4.3.2}$$

where $\omega\in[0,1]$. We take $\omega=0.5$. By taking $\omega=0.5$ instead of $\omega=1$ (which corresponds to the standard DeC-iteration method), damping (under-relaxation) is introduced. In general, we may expect that the ω-DeC-iteration process with $\omega=0.5$ is more robust than with $\omega=1$ (see also problem 4 in this section). Here, we see that the convergence histories, are similar for $\omega=1$ and $\omega=0.5$. But we may expect that with $\omega=0.5$ it is possible to drive the residual to machine-zero. A disadvantage of the ω-DeC-iteration process with $\omega<1$ is that long wavelength error components (which determine the accuracy) are damped with a factor $1-\omega$ in each iteration. Therefore, even for very smooth problems, a single ω-DeC-iteration with $\omega=0.5$ is not sufficient to obtain second-order accuracy. On the other hand, 10 ω-DeC-iterations with $\omega=0.5$ reduce the long wavelength error components with a factor 0.001, which is sufficient in practice.

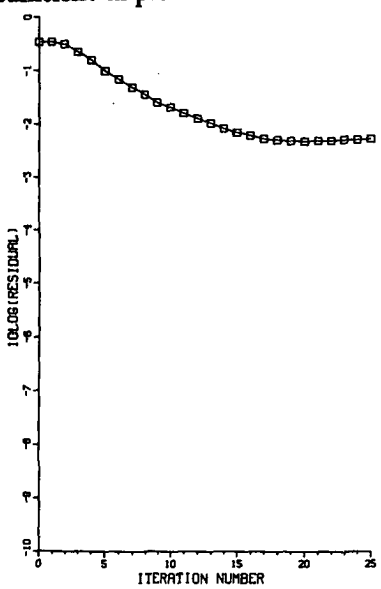


FIGURE 4.3.8. Convergence history of the DeC-iteration process on grid 48×16 for testproblem 2.

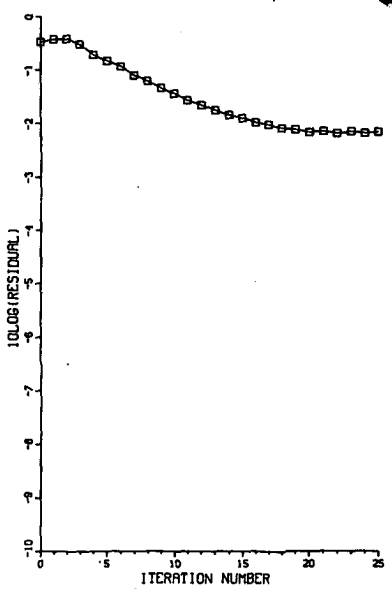


FIGURE 4.3.9. As figure 4.3.8 but on grid 96×32.

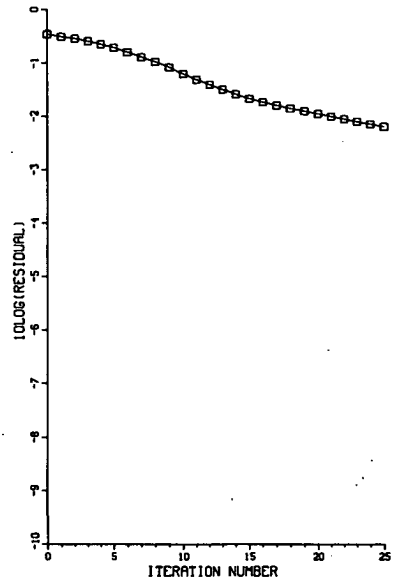


FIGURE 4.3.10. Convergence history of the ω-DeC-iteration process on gird 48×16 for testproblem 2.

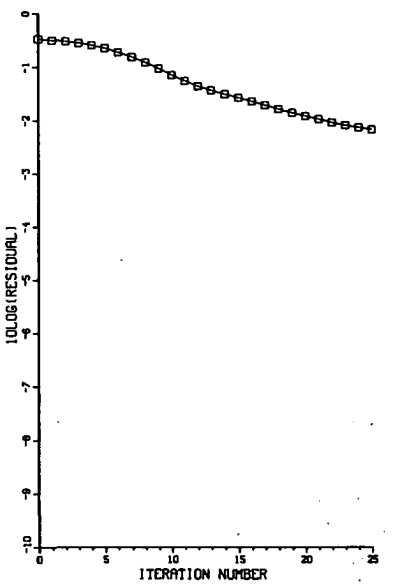


FIGURE 4.3.11. As figure 4.3.10 but on grid 96×32.

As mentioned before, the DeC-iteration process ($\omega=1$) is not used to obtain the exact solution of (4.1.9) but to improve the accuracy of the first-order solutions. The following figures, obtained by 25 DeC-iterations, show clearly that the solutions obtained after 25 DeC-iterations are much more accurate than the first-order solutions presented in section 3.3, problem 2. Fig. 4.3.12, 13 show the iso-Mach lines of the second-order solutions. The figures show very sharp shocks. The reflection of the leading edge shock at the upper wall, its intersection with the trailing edge shock, its further reflection at the lower wall and finally its merging with the trailing edge shock are all clearly visible. The leading-edge oblique shock is spread numerically over 2-3 volumes halfway the channel ($y=0.5$).

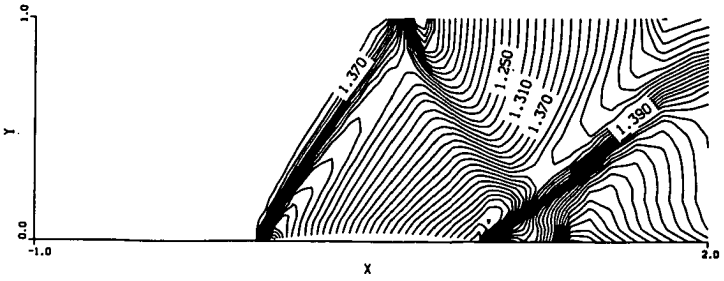


FIGURE 4.3.12. Iso-Mach lines of the second-order solution obtained on the 48×16 grid for testproblem 2.

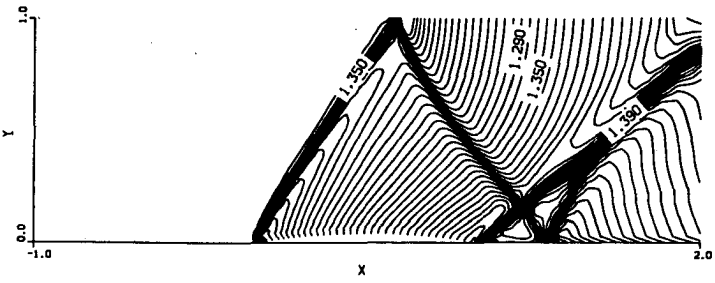


FIGURE 4.3.13. As figure 4.3.12 but on grid 96×32.

In fig. 4.3.14-17 we give the Mach number and the entropy distributions along the lower surface of the channel. Downstream of the bump, a large qualitative difference between the first- and second-order solutions is observed once more. The first-order solutions show spurious entropy generation along the entire bump (see fig. 3.3.20, 21). The second-order solution has no such entropy generation, but shows some spurious non-monotonicity. The latter is caused by the fact that no limiter can be used near boundaries (see 2.5.3.46).

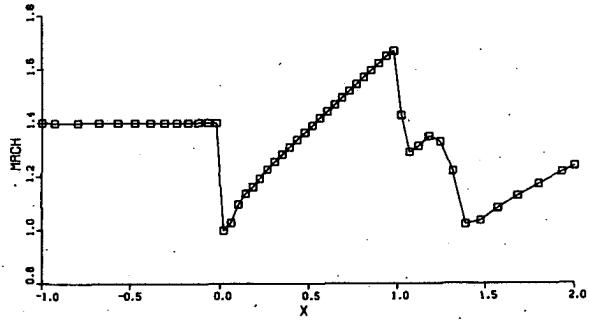


FIGURE 4.3.14. Mach number distribution along the lower surface of the channel for testproblem 2 (48×16 grid).

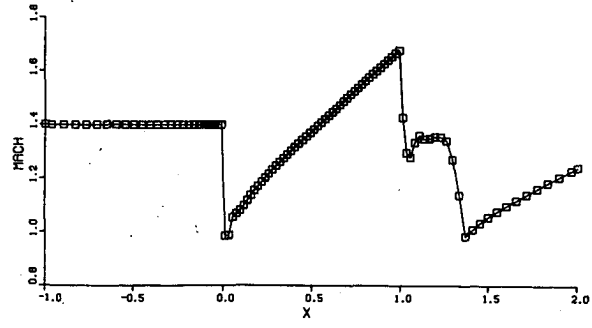


FIGURE 4.3.15. As figure 4.3.14 but on grid 96×32.

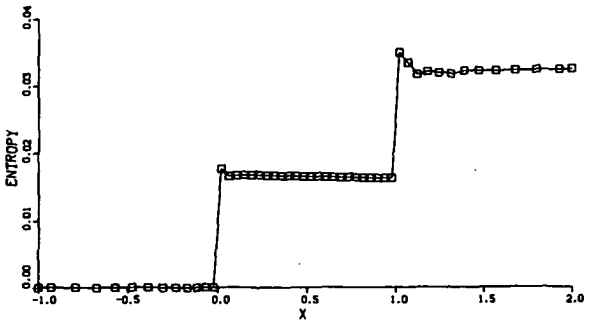


FIGURE 4.3.16. Entropy distribution along the lower surface of the channel for testproblem 2 (48×16 grid).

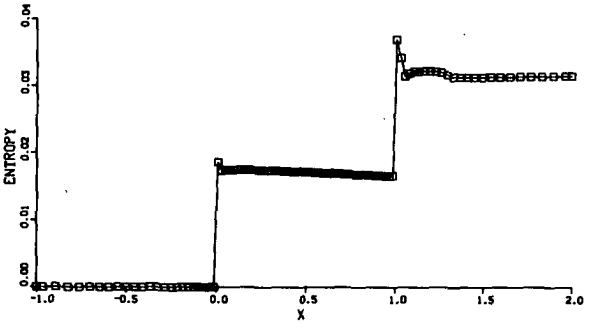


FIGURE 4.3.17. As figure 4.3.16 but on grid 96×32.

PROBLEM 3. *Resolution of contact discontinuities*

PROBLEM 3a. *Contact discontinuity aligned to the grid*

For a description of this problem see section 3.3, problem 3a. The first-order solution is exact and cannot be improved. Therefore, the solution obtained with the DeC-iteration process should be the same as the first-order solution presented in section 3.3, problem 3a.

The convergence history of the DeC-iteration process is shown in fig. 4.3.18 and the obtained entropy ($s=p\rho^{-\gamma}$) distribution along the line $x=0.5$ is shown in fig. 4.3.19. The entropy distribution is exact and the same as in fig. 3.3.24. The convergence history of the DeC-iteration process is rapid and similar to the convergence history of the NMG-iteration process (see fig. 3.3.23).

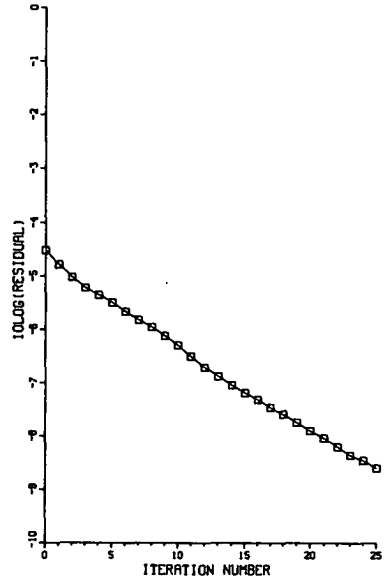


FIGURE 4.3.18. Convergence history of the DeC-iteration process for test-problem 3a (32 × 32 grid).

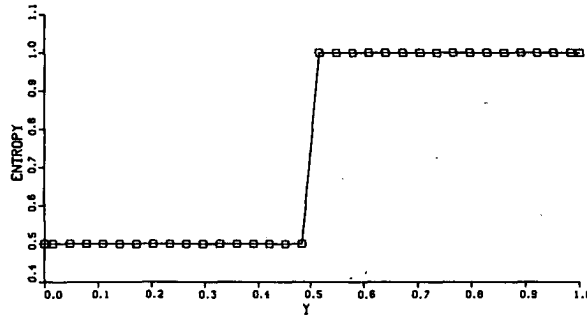


FIGURE 4.3.19. Entropy distribution along the line $x=0.5$ for test-problem 3a (32 × 32 grid).

PROBLEM 3b. *Oblique contact discontinuity*

For a description of this problem, see section 3.3, problem 3b. In fig. 4.3.20 we show the convergence history of the DeC-iteration process. The process converges but not very rapidly. In fig. 4.3.21 we show the entropy distribution along the line $x=0.5$ and entropy contours are shown in fig 4.3.22. In comparison with the first-order solution, the spreading of the contact discontinuity is reduced significantly. But it is clear that an oblique contact discontinuity is captured not so well as an oblique shock (the width of oblique shock and contact discontinuity is about 3 and 6 volumes, respectively).

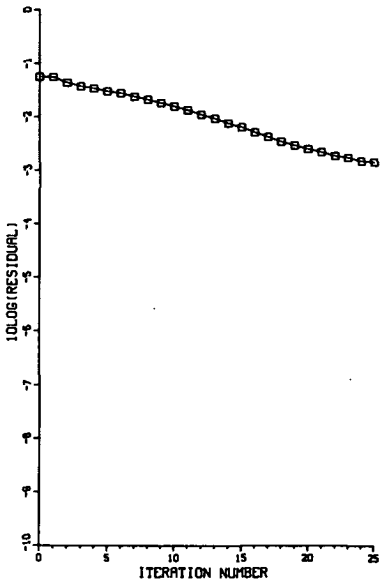


FIGURE 4.3.20. Convergence history of the DeC-iteration process for testproblem 3b (32 × 32 grid).

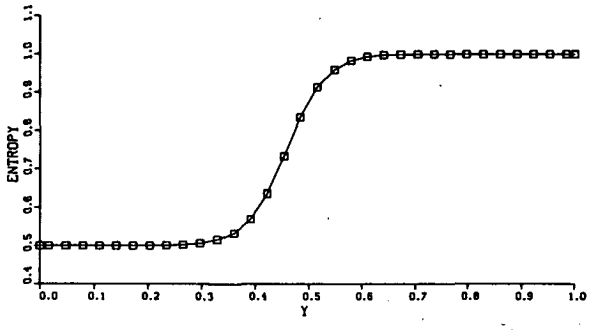


FIGURE 4.3.21. The entropy distribution along the line $x=0.5$ for testproblem 3b (32 × 32 grid).

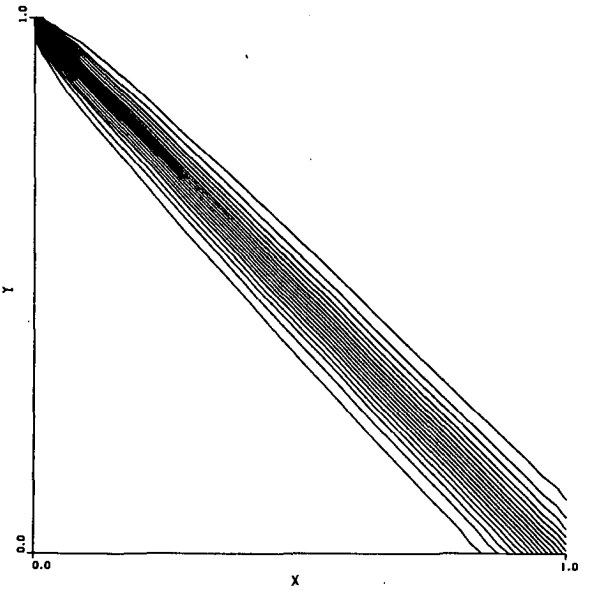


FIGURE 4.3.22. Entropy contours of the second-order solutions for testproblem 3b (32 × 32 grid).

PROBLEM 4. *Cylinder in a supersonic free-stream*

We refer to section 3.3, problem 4, for a description of this problem and the grids used.

For this problem the standard DeC-iteration process (4.1.10) does not work. The first iterand in the standard DeC-iteration process is an approximate first-order solution. Then, the second iterand is computed by solving a first-order system with an appropriate right-hand side. Unfortunately, in the solution process for the second-iterand, divergence was observed for the local Newton iteration process in the CSGS relaxation on the finest grid. Therefore, for this problem some damping is necessary in the DeC-iteration process. Damping is achieved by the ω-DeC-iteration process which is defined by (4.3.2). Again we take $\omega = 0.5$. The convergence history of the ω-DeC-iteration process is shown in fig. 4.3.23.

The pressure in the stagnation point in front of the cylinder can be computed analytically because of the fact that locally the shock is normal to the x-axis. Denote with q_0 the state ahead of the shock, with q_1 the state behind the shock and with q_2 the state in the stagnation point. The relation between q_1 and q_2 is given by

$$\frac{c_1^2}{\gamma-1} + \frac{1}{2}u_1^2 = \frac{c_2^2}{\gamma-1} \quad (\textit{isenthalpy}) \tag{4.3.3}$$

$$p_1\rho_1^{-\gamma} = p_2\rho_2^{-\gamma} \quad (\textit{isentropy}) \tag{4.3.4}$$

from which it is easily derived that

$$\frac{p_2}{p_1}=(1+\frac{\gamma-1}{2}M_1^2)^{\frac{\gamma}{\gamma-1}} \tag{4.3.5}$$

where M_1 is the Mach number of q_1.
Using the normal shock relations (1.2.27, 28) it can be derived that

$$M_1^2=\frac{1+\frac{\gamma-1}{2}M_0^2}{\gamma M_0^2-\frac{\gamma-1}{2}} \tag{4.3.6}$$

where M_0 is the Mach number of q_0.
Using (1.2.30) we also have

$$\frac{p_1}{p_0}=1+\frac{2\gamma}{\gamma+1}(M_0^2-1)\,. \tag{4.3.7}$$

Because $M_0=2$ we find the pressure ratio $p_2/p_0=5.64$. The first-order solution gives a pressure ratio $p_2/p_0=5.92$ (see fig. 3.3.33). The second-order solution (obtained after 25 ω-DeC-iteratios) gives a much better ratio: $p_2/p_0=5.65$. In fig. 4.3.24 we show the pressure ratio of the solutions obtained after each ω-DeC-iteration step. Fig. 4.3.25 shows the surface pressure distribution of the second-order solution along the surface of the cylinder. Finally, fig. 4.3.26, 27 show iso-Mach lines and pressure contours of the second-order solution obtained after 25 ω-DeC-iterations. There is a small change in the shock position: the bow shock starts in $x=-2.375$ while the bow shock position according to the first-order solution is $x=-2.5$. These results agree well with the results published in [9].

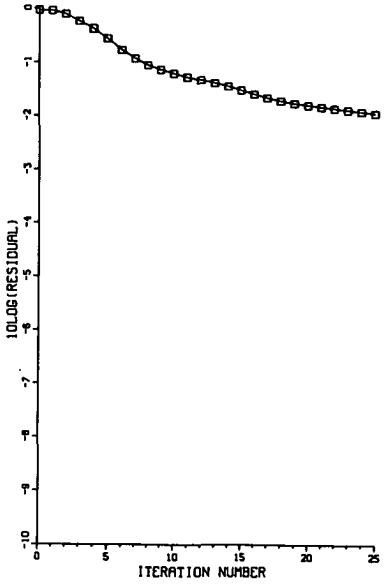


FIGURE 4.3.23. Convergence history of the ω-DeC-iteration process for testproblem 4 (40 × 32 grid).

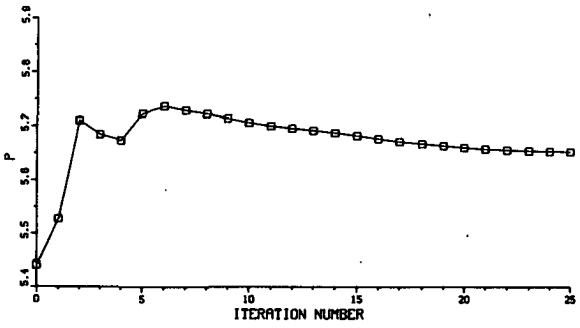


FIGURE 4.3.24. The stagnation pressure after each ω-DeC-iteration step for testproblem 4 (40 $\times$ 32 grid).

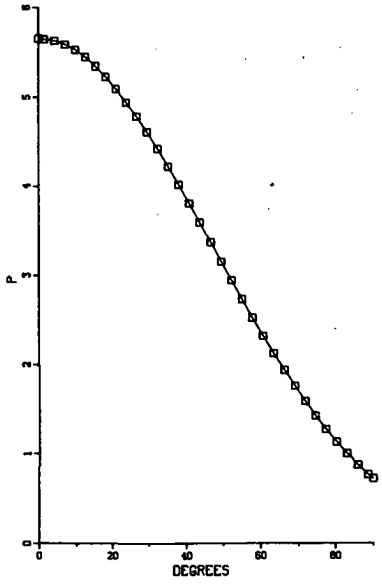


FIGURE 4.3.25. Pressure ($p/p_{-\infty}$) distribution along the surface of the cylinder for testproblem 4 (40 $\times$ 32 grid).

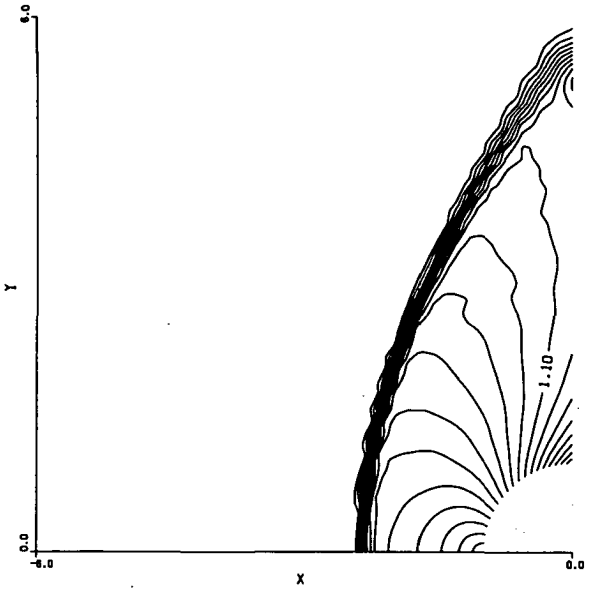


FIGURE 4.3.26. Iso-Mach lines for testproblem 4 (40 × 32 grid).

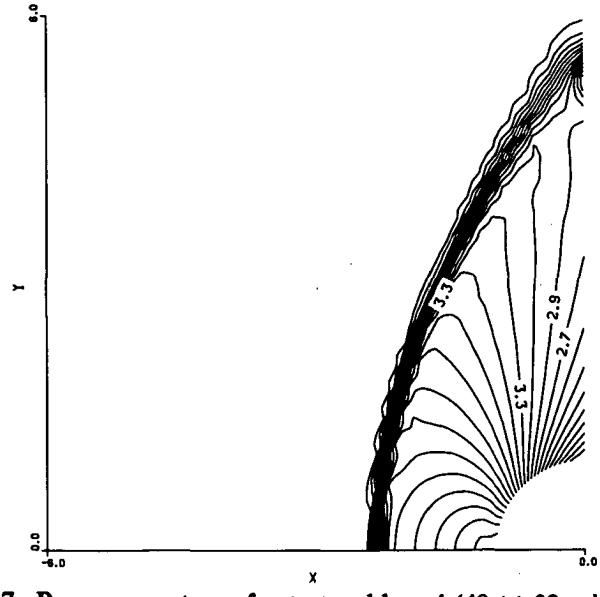


FIGURE 4.3.27. Pressure contours for testproblem 4 (40 × 32 grid).

For these four testproblems we can conclude that the DeC-iteration method is an effective way to improve the accuracy of the first-order solutions. In general, about 10 DeC-iterations are sufficient. Therefore, the amount of work to obtain second-order accuracy is equivalent with about $10\times\frac{4}{3}\times 2$ CSGS-relaxations on the finest grid.
Finally, we refer to the work of B. Koren showing the feasibility of the method

for airfoil flow computations [5,7].

4.4. Solution of the steady Euler equations with a source term

Consider the Euler equations with a source term:

$$\frac{\partial q}{\partial t}+\frac{\partial}{\partial x}f(q)+\frac{\partial}{\partial y}g(q)=r(q) \tag{4.4.1}$$

where q, $f(q)$, $g(q)$ are defined by (2.1.1b) and $r(q)$ is the source term. On a finite volume grid $\{\Omega_{i,j}\}$ the discretization of the steady Euler equations with source term $r(q)$ is given by

$$(F_h(q_h))_{i,j}=f_{i+\frac{1}{2},j}+f_{i,j+\frac{1}{2}}-f_{i-\frac{1}{2},j}-f_{i,j-\frac{1}{2}}=(r_h(q_h))_{i,j} \tag{4.4.2}$$

where

$$(r_h(q_h))_{i,j}=r((q_h)_{i,j})\cdot V_{i,j} \tag{4.4.3}$$

with $V_{i,j}$ the area of $\Omega_{i,j}$.
The operator $F_h(q_h)$ is first-order accurate when $f_{i+\frac{1}{2},j}$, $f_{i,j+\frac{1}{2}}$ are defined by (4.1.3,4), then we write $F_h(q_h)=F_h^1(q_h)$. The operator $F_h(q_h)$ is second-order accurate when $f_{i+\frac{1}{2},j}$, $f_{i,j+\frac{1}{2}}$ are defined by (4.1.3) and (4.1.5-7), then we write $F_h(q_h)=F_h^2(q_h)$.
The defect correction method

$$\begin{cases}F_h^1(q_h^1)=0\\ F_h^1(q_h^{i+1})=F_h^1(q_h^i)\ +(r_h(q_h^i)-F_h^2(q_h^i))\ \ i=1,\ 2,...\end{cases} \tag{4.4.4}$$

is a simple method to obtain a second-order accurate solution of the steady Euler equations with a source term.

A source term appears in the Euler equations when a body force $\underline{F}=(F_1,F_2)^T$ is present. Then the Euler equations become (see section 1.1):

Conservation of mass:

$$\frac{d}{dt}\int_\Omega \rho dv=-\int_{\partial\Omega}\rho(\underline{n}\cdot\underline{v})d\sigma \tag{4.4.5}$$

Conservation of momentum:

$$\frac{d}{dt}\int_\Omega \rho\underline{v}dv=-\int_{\partial\Omega}\rho\underline{v}(\underline{n}\cdot\underline{v})d\sigma-\int_{\partial\Omega}p\underline{n}d\sigma+\int_\Omega \underline{F}d\sigma \tag{4.4.6}$$

Conservation of energy:

$$\frac{d}{dt}\int_\Omega E dv=-\int_{\partial\Omega}E(\underline{n}\cdot\underline{v})d\sigma-\int_{\partial\Omega}p(\underline{n}\cdot\underline{v})d\sigma+\int_\Omega \underline{F}\cdot\underline{v}d\sigma\,. \tag{4.4.7}$$

From these equations we obtain (4.4.1) with

$$r=(0,F_1,F_2,uF_1+vF_2)^T\,. \tag{4.4.8}$$

Note that F is a force per unit of volume.
A special case of the presence of a body force is the actuator disk where F is a line distribution along a line segment l such that for an arbitrary control volume Ω we have

$$\iint_{\Omega} F dv = \int_{\Omega \cap l} \tilde{F} d\sigma . \tag{4.4.9}$$

Then $\tilde{F}$ is a force per unit of length. For an actuator disk the force $\tilde{F}$ is perpendicular to l and acts in such a way that there is a prescribed pressure jump at the disk. Assume that l coincides with the y-axis and let q_L and q_R be the left and right state at the disk in a steady flow.
First, we show that the normal velocity is discontinuous at the disk. Suppose there is no velocity jump. By taking a control volume with infinitesimal width but finite length across l (see fig. 1.2a) we find that

$$\begin{aligned}
&(\rho u)_R - (\rho u)_L = 0 \\
&(\rho u^2 + p)_R - (\rho u^2 + p)_L = F_1 \\
&(\rho uv)_R - (\rho uv)_L = 0 \\
&((E+p)u)_R - ((E+p)u)_L = F_1 u
\end{aligned} \tag{4.4.10}$$

where $u = u_L = u_R > 0$.
From the first and second equations it follows that $F_1 = p_R - p_L$. On the other hand, from the fourth equation we find that

$$\begin{aligned}
F_1 &= (E+p)_R - (E+p)_L = \rho(H_R - H_L) \\
&= \frac{\rho}{\gamma - 1}(c_R^2 - c_L^2) = \frac{\gamma}{\gamma - 1}(p_R - p_L)
\end{aligned}$$

and we have a contradiction. Hence, there must be a velocity and density jump at the disk. Thus, the normal velocity at the disk is not defined and it is not clear how to compute $F \cdot v$ at the disk.
A way to model the actuator disk is the following. Denote with δ_2 and δ_4 the x-momentum and energy source imposed by the actuator disk.
Thus

$$\begin{aligned}
&(\rho u)_R - (\rho u)_L = 0 \\
&(\rho u^2 + p)_R - (\rho u^2 + p)_L = \delta_2 \\
&(\rho uv)_R - (\rho uv)_L = 0 \\
&((E+p)u)_R - ((E+p)u)_L = \delta_4
\end{aligned} \tag{4.4.11}$$

The sources δ_2 and δ_4 are computed by assuming that the prescribed pressure jump is isentropic. The assumption that the flow is isentropic is based on the fact that the material derivative of the entropy is zero for a smooth flow, even when body forces are present. This result is derived in the following way (see also section 1.2, formulae 1.2.9-19).
When body forces are present, we have

$$\rho \frac{D}{Dt} \underline{v} = -\nabla p + \underline{F} \tag{4.4.12}$$

and

$$\rho \frac{D}{Dt}(e + \tfrac{1}{2}(\underline{v} \cdot \underline{v})) = -\operatorname{div}(p\underline{v}) + \underline{F} \cdot \underline{v} . \tag{4.4.13}$$

Combining the last two equations we find

$$\frac{De}{Dt} = -\frac{p}{\rho} \operatorname{div} \underline{v} \tag{4.4.14}$$

which is the same expression as in (1.2.12). Following the derivation (1.2.13-19) we find again

$$\frac{Ds}{Dt} = 0 . \tag{4.4.15}$$

Assume that $u > 0$ and $p_R = \alpha p_L$ with $\alpha > 1$. Then a force is acting on the disk in the negative x-direction. From the isentropy it follows that

$$p_R \rho_R^{-\gamma} = p_L \rho_L^{-\gamma} \tag{4.4.16}$$

Using (4.4.11), $p_R = \alpha p_L$ and (4.4.16) we find

$$\begin{aligned}
p_R &= \alpha p_L, \quad v_R = v_L \\
\rho_R &= \alpha^{\frac{1}{\gamma}} \rho_L \\
u_R &= \alpha^{-\frac{1}{\gamma}} u_L \\
c_R &= c_L \alpha^{\frac{\gamma-1}{2\gamma}} \\
H_R &= \frac{c_R^2}{\gamma - 1} + \tfrac{1}{2}(u_R^2 + v_R^2) \\
\delta_2 &= \rho_R u_R^2 + p_R - \rho_L u_L^2 - p_L \\
\delta_4 &= \rho_L u_L (H_R - H_L)
\end{aligned} \tag{4.4.17}$$

Hence, given a state q_L, from (4.4.17) the state q_R and the sources δ_2 and δ_4 can be computed for a given pressure jump.

To compute a flow with an actuator disk, the disk is modelled in the following way. Assume that $\partial\Omega_{i+\frac{1}{2},j}$ coincides with l (see fig. 4.4.1)

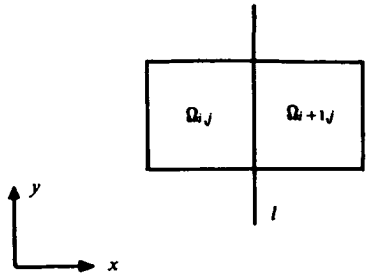


FIGURE 4.4.1. Modelling of an actuator disk.

Then we take

$$(r_h(q_h))_{i+1,j} = l_{i+\frac{1}{2},j}(0,(\delta_2)_{i,j},\ 0,(\delta_4)_{i,j})^T \tag{4.4.18}$$

where $l_{i+\frac{1}{2},j}$ is the length of $\partial\Omega_{i+\frac{1}{2},j}$ and $(\delta_2)_{i,j}$, $(\delta_4)_{i,j}$ are computed by (4.4.17) with $q_L = q_{i,j}$.

We present two flow computations with an actuator disk. See [8] for an airfoil computation with an actuator disk.

PROBLEM 1. *Actuator disk in a channel flow*
The first example is a channel flow with an actuator disk extending from the lower to the upper wall. The channel is straight. The physical and computational domain are the same (the mapping between the computational and physical domain is the identity: $x=\xi$, $y=\eta$). We take $\Omega=[0,5]\times[0,1]$, the coarsest grid consists of 5×1 volumes, the finest grid consists of 20×4 volumes. At inflow u,v,z are prescribed: $u=0{,}5$, $v=0$, $z=-\gamma\ln\gamma$, at outflow the pressure is prescribed: $p=1$. The other two boundaries are solid walls.
The actuator disk is located at $x=2.5$ and the prescribed pressure jump is $\alpha=1.2$. Hence, for the exact solution we have an upstream pressure $p_{\text{upstream}}=\frac{1}{1.2}\approx 0.833$ and a downstream pressure $p_{\text{downstream}}=1.0$. The entropy $s=\gamma^{-\gamma}\approx 0.6243$ is uniformly constant.
The numerical solution has been obtained by the defect correction iteration process (4.4.4) with $r_h(q_h)$ defined by (4.4.18). The iteration process has an average reduction factor 0.63. The pressure and entropy distribution obtained along the line $y=0.5$ are given in fig. 4.4.2,3. We see that the flow is indeed isentropic and the right pressure jump is obtained.

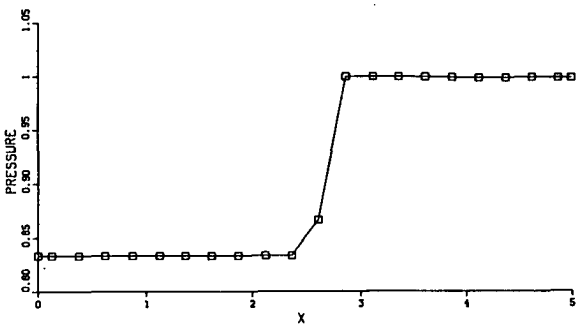


FIGURE 4.4.2. Pressure distribution along the line $y=0.5$. The actuator disk is located at $x=2.5$.

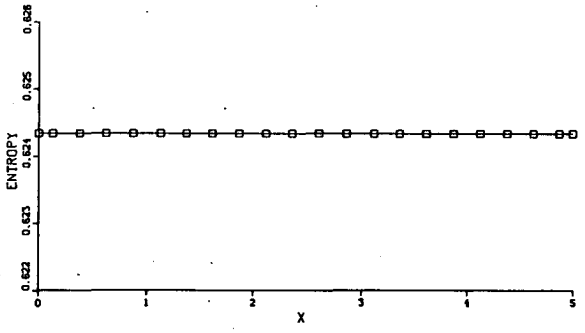


FIGURE 4.4.3. Entropy distribution along the line $y=0.5$. Isentropy is observed.

PROBLEM 2. *Actuator disk in a subsonic free stream*
An adaptive mesh has been used. The mapping from the computational space $(\xi,\eta)\in[-5,5]\times[-5,5]$ to the physical domain (x,y) is given by

$$
\begin{aligned}
-5.0 \leqslant \xi \leqslant -2.5 &\Rightarrow \tilde{\xi} = (8\xi+15)/5 \\
-2.5 \leqslant \xi \leqslant +2.5 &\Rightarrow \tilde{\xi} = 2\xi/5 \\
+2.5 \leqslant \xi \leqslant +5.0 &\Rightarrow \tilde{\xi} = (8\xi-15)/5 \\
-5.0 \leqslant \xi \leqslant -1.0 &\Rightarrow x = -5+4\left[\frac{e^{-\beta(\tilde{\xi}+5)}-1}{e^{-4\beta}-1}\right] \\
-1.0 \leqslant \xi \leqslant +1.0 &\Rightarrow x = \tilde{\xi} \\
+1.0 \leqslant \xi \leqslant +5.0 &\Rightarrow x = +5-4\left[\frac{e^{+\beta(\tilde{\xi}-5)}-1}{e^{-4\beta}-1}\right]
\end{aligned}
\tag{4.4.19}
$$

with $\beta=0.585$ and y depends on η in exactly the same way as x depends on ξ.

The finest grid has 32 × 32 volumes, the coarsest 2 × 2 volumes. Figure 4.4.4 shows the finest grid. The actuator disk is located at $x=0$, $y\in[-0.5,0.5]$. The prescribed isentropic pressure jump is again $\alpha=1.2$.

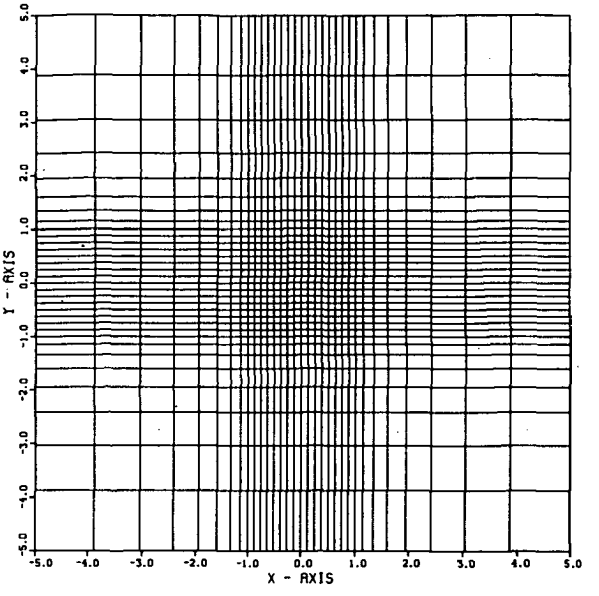


FIGURE 4.4.4. Finest grid (32 × 32) for problem 2.

Two cases are considered:

PROBLEM 2a. *Actuator disk perpendicular to the free stream flow direction.* The boundary conditions are as follows:

$x=-5$, $y\in[-5,5]$: subsonic inflow: $u=0.5$, $v=0$, $z=-\gamma\ln\gamma$
other three boundaries : subsonic outflow: $p=1.0$.

The solution has been obtained by a ω-DeC-iteration process with $\omega=0.5$, i.e.

$$\begin{cases} F_h^1(q_h^i)=0 \\ F_h^1(q_h^{i+1})=F_h^1(q_h^i) + \omega(r_h(q_h^i) - F_h^2(q_h^i)) \quad i=1,2,... \end{cases} \tag{4.4.20}$$

where $r_h(q_h)$ is computed by (4.4.18) (with $\omega=1$ similar difficulties occur as in case of problem 4, section 4.3). The ω-DeC-iteration process has an average reduction factor 0.8. Fig. 4.4.5,6 give a qualitative impression of the obtained Mach number and pressure distribution after 25 iterations. The largest observed entropy variation $\max_{(i,j)} |(s_{i,j}-s_\infty)/s_\infty|$ is less than 1%.

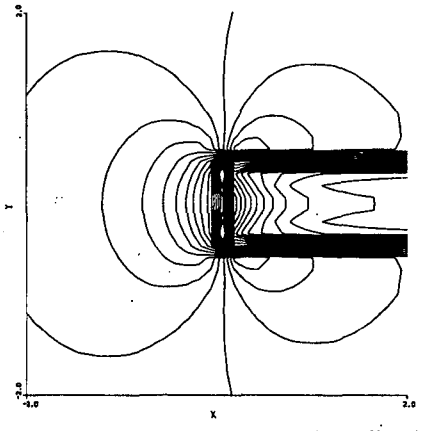


FIGURE 4.4.5. Mach number distribution for problem 2a.

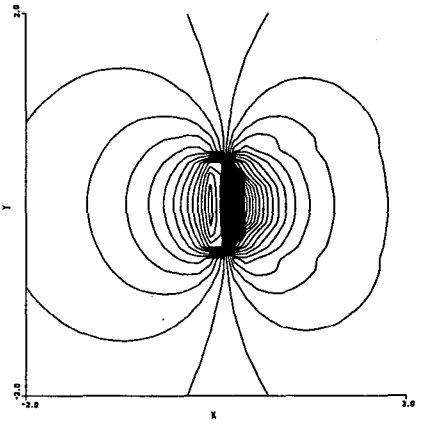


FIGURE 4.4.6. Pressure distribution for problem 2a.

PROBLEM 2b. *Inclined*
The angle of inclination is 45°.
The boundary conditions are:

$x=-5,\ y\in[-5,5]$: subsonic inflow: $u=\sqrt{2}/4,\ v=\sqrt{2}/4,\ z=-\gamma\ln\gamma$
$y=-5,\ x\in[-5,5]$: subsonic inflow: $u=\sqrt{2}/4,\ v=\sqrt{2}/4,\ z=-\gamma\ln\gamma$
other two boundaries : subsonic outflow: $p=1$.

The solution has been obtained by a ω-DeC-iteration process with $\omega=0.5$. The iteration process has an average reduction factor 0.85. The largest observed entropy variaton $\max_{(i,j)}|(s_{i,j}-s_\infty)/s_\infty|$ is less than 1.5%. Fig 4.4.7,8 give a qualitative impression of the Mach number and pressure distribution obtained after 25 iterations.

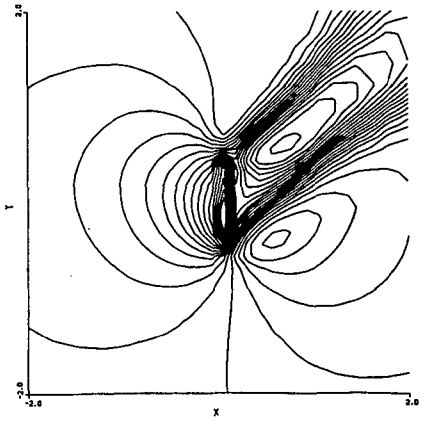


FIGURE 4.4.7. Mach number distribution for problem 2b.

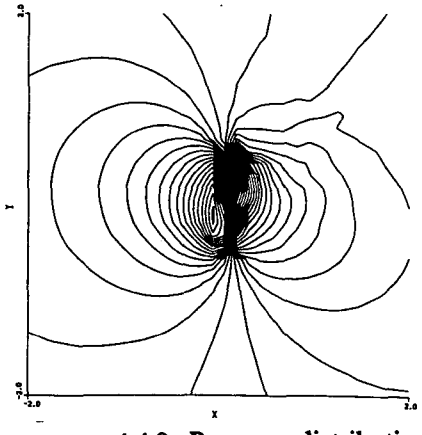


FIGURE 4.4.8. Pressure distribution for problem 2b.

Another example where source terms appear in a natural way is in case of axially symmetrical flow. In cylindrical polar co-ordinates R,z,θ (where θ is the azimuthal angle about the axis $R=0$) and suppressing all components and derivatives in the θ-direction, the Euler equations are

$$\frac{\partial}{\partial t}\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ E\end{bmatrix}+\frac{\partial}{\partial R}\begin{bmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ (E+p)u\end{bmatrix}+\frac{\partial}{\partial z}\begin{bmatrix}\rho u\\ \rho uv\\ \rho v^2+p\\ (E+p)v\end{bmatrix}=-\frac{1}{R}\begin{bmatrix}\rho u\\ \rho u^2\\ \rho uv\\ (E+p)u\end{bmatrix}$$

where u,v are the velocity components in the R-and z-direction respectively. Hence, Euler flow computations for steady axial symmetrical flow can be performed by merely adding source terms in an existing code for steady 2D planar flow computations.

References

1. K. Böhmer, P.W. Hemker and H.J. Stetter (1984). *The Defect Correction Approach.* In: Defect Correction Methods. (K. Böhmer, H.J. Stetter, eds.). Computing, Suppl. 5, 1-32, Springer - Verlag, Wien, New York.
2. W. Hackbush (1985). *Multi-Grid Methods and Applications.* Springer Series in Computational Mathematics 4, Springer Verlag, Berlin.
3. P.W. Hemker (1981). *Lecture Notes of a Seminar on Multiple Grid Methods.* Report NN-24/81, Centre for Mathematics and Computer Science, Amsterdam.
4. P.W. Hemker (1985). *Defect Correction and Higher-Order Schemes for the Multigrid Solution of the Steady Euler Equations.* In: Multigrid Methods II. (W. Hackbush, U. Trottenberg, eds.). Proceedings 2nd European Multigrid Conference, Cologne, 1985. Lecture Notes in Mathematics 1228, 149-165, Springer - Verlag, Berlin.
5. P.W. Hemker and B. Koren (1986). *A Non-Linear Multigrid Method for the Steady Euler Equations.* Report NM-R8621, Centre for Mathematics and Computer Science, Amsterdam. To appear in: Proceedings Gamm - Workshop on the Numerical Simulation of Compressible Euler Flows, Rocquencourt, 1986. Notes on Numerical Fluid Mechanics, Vieweg Verlag, Braunschweig.
6. D.C. Jespersen (1983). *Design and Implementation of a Multigrid Code for the Euler Equations.* App. Math. and Computat. 13, 357-374.
7. B. Koren (1986). *Evaluation of Second-Order Schemes and Defect Correction for the Multigrid Computation of Airfoil Flows with the Steady Euler Equations.* Report NM-R8616, Centre for Mathematics and Computer Science, Amsterdam. To appear in J. Comp. Phys.
8. B. Koren and S.P. Spekreijse, (1987). *Multigrid and Defect Correction for the Efficient Solution of the Steady Euler Equations.* In: Research in Numerical Fluid Dynamics. Proceedings of the 25th Meeting of the Ducth

Association for Numerical Fluid Dynamics (P. WESSELING, ed). Notes on Numerical fluid Mechanics 17, 87-100, Vieweg, Braunschweig.

9. M.H. RAY (1986). *A Relaxation Approach to Patched-Grid Calculations with the Euler Equations.* J. Comp. Phys. 66, 99-131.
10. A. RIZZI and H. VIVIAND (eds.) (1981). *Numerical Methods for the Computation of Inviscid Transonic Flows with Shock Waves.* Notes on Numerical Fluid Mechanics, Vol. 3. Vieweg Verlag, Braunschweig.
11. S.P. SPEKREIJSE (1986). *Second-Order Accurate Upwind Solutions of the 2D Steady Euler Equations by the Use of a Defect Correction Method.* In: Multigrid Methods II, (W. HACKBUSH and U. TROTTENBERG, eds.). Proceedings 2nd European Multigrid Conference, Cologne, 1985. Lecture Notes in Mathematics 1228, 285-300, Springer Verlag, Berlin.
12. S.P. SPEKREIJSE (1987). *Multigrid Solution of Monotone Second-Order Discretizations of Hyperbolic Conservation Laws.* Math. Comp. 49, 135-155.

Summary

Discretizations of the steady Euler equations are studied and robust and efficient solution methods are developed to solve the resulting highly nonlinear algebraic systems of equations.

The discretizations used are based on cell centered finite volume schemes, i.e. the physical domain, where the solution of the steady Euler equations is sought, is subdivided into a finite number of disjunct finite volumes (or cells) and the numerical approximations are stored inside the cells. The discretization is determined completely by the way in which the flux computations are performed at the cell boundaries. A flux at a cell boundary is the amount of mass, momentum and energy transported per unit of time across the cell boundary. The equations are obtained by demanding that the total flux is zero for each volume.

At each cell boundary, a flux is computed by solving approximately a local one dimensional Riemann problem. As a consequence, the schemes are conservative and characteristic-based or upwind. The approximate Riemann solver is the one proposed by Osher but the constituent parts of the integration path in the state space used in the Riemann solver are taken in an order opposite to that as originally proposed by Osher. In this way the implementation of Osher's scheme becomes rather simple, provided that the proper dependent variables are used.

In the first-order discretization, the numerical approximations are assumed to be uniformly constant in each volume. Second-order accuracy is obtained by using piecewise linear interpolation in each volume. In this approach the slopes are limited to prevent spurious oscillations in the neighbourhood of shocks or contact discontinuities. The limiting procedure must be nonlinear even when applied to linear problems. A novel, very simple and clear description of the limiting procedure is given for a general nonlinear scalar hyperbolic conservation law by considering the limiting procedure as a modification of the fully one-sided upwind scheme. It appears that limiting and flux-splitting are closely related. It is also shown that monotonicity and second-order accuracy can be achieved simultaneously, even in more than one dimension.

The second-part of this thesis (chapters III and IV) concerns the solutions of the first- and second-order discretizations. Due to the favorable properties of the first-order discretization (5-point stencil structure, upwind character, differentiability, consistency of flux computations at interior cell boundaries and at cell boundaries which are part of the boundary of the physical domain) a straightforward nonlinear multigrid solution method can be developed. In the multigrid method used, the coarse grid discretizations are Galerkin approximations of the fine grid discretization and a simple Collective Symmetric Gauss-Sidel (CSGS) relaxation method appears to be an excellent smoothing procedure. The numerical examples, covering channel flows, resolution of contact discontinuities and a blunt body in a supersonic flow, show that a degree of efficiency and robustness characteristic for successful multigrid methods is obtained. For practical purposes, where one only wants to get below

truncation error, a few (two or three) nonlinear multigrid iterations are sufficient. This means that first-order solutions are obtained in an amount of work equivalent with about $3\times\frac{4}{3}\times 2$ CSGS-relaxations on the finest grid.

A Defect Correction (DeC) iteration method is used to improve the accuracy of the first-order solutions. The DeC-iteration method makes effective use of the excellent multigrid solver for the first-order discretization. It is well known that for smooth problems only one DeC-iteration is sufficient to obtain a second-order accurate solution. For non-smooth problems, it appears that more (about 10) DeC-iterations are necessary. Because 10 DeC-iterations correspond with an amount of work equivalent with about $10\times\frac{4}{3}\times 2$ CSGS-relaxations on the finest grid, the method is still an efficient procedure to improve the accuracy. For the aforementioned numerical testproblems, the results obtained with the DeC-iteration method show an impressive improvement in accuracy compared with the first-order solutions. It should be mentioned that it is sometimes necessary (for flows where strong shocks are present) to use some damping in the DeC-iteration method.

Samenvatting

In dit proefschrift worden discretisaties van de stationaire Euler vergelijkingen bestudeerd en worden efficiente en robuuste oplossingsmethoden ontwikkeld voor de resulterende stelsels niet lineaire algebraïsche vergelijkingen.

De gebruikte discretisaties zijn gebaseerd op cel gecentreerde eindige volume schema's, d.w.z. het fysische gebied waar de oplossing van de stationaire Euler vergelijkingen wordt gezocht wordt opgedeeld in een eindig aantal disjuncte volumes (cellen) van eindige afmeting en de numerieke benaderingen worden gelocaliseerd in de volumes. De discretisatie wordt dan volledig bepaald door de wijze waarop de flux door de celwandjes wordt berekend. Een flux door een celwandje is de hoeveelheid massa, impuls en energie die het celwandje per tijdseenheid passeert. De vergelijkingen worden verkregen door te eisen dat voor ieder volume de netto flux gelijk is aan nul.

Een flux wordt berekend door locaal een één-dimensionaal Riemann probleem benaderend op te lossen. Een gevolg hiervan is dat de schema's conservatief zijn en gebaseerd zijn op de karakteristieken theorie, m.a.w. een upwind karakter hebben. De wijze waarop het Riemann probleem benaderend wordt opgelost is in essentie zoals voorgesteld door Osher. Echter, het integratie pad in de toestandsruimte, nodig bij het oplossingsprocédé van het Riemann probleem, heeft een volgorde omgekeerd aan die zoals oorspronkelijk door Osher is voorgesteld. Op deze wijze wordt het oplossingsprocédé voor het Riemann probleem aanzienlijk eenvoudiger (minder operaties), vooropgesteld dat de juiste afhankelijke variabelen worden gekozen.

De eerste-orde discretisatie wordt verkregen door aan te nemen dat de numerieke benaderingen uniform constant zijn in iedere cel afzonderlijk. Tweede-orde nauwkeurigheid wordt verkregen d.m.v. interpolatie zodanig dat de benaderende oplossing een lineaire verdeling heeft in iedere cel afzonderlijk. De hellingen van de lineaire verdelingen moeten op een bepaalde manier begrensd blijven om te voorkomen dat oneigenlijke oscillaties optreden in de oplossing nabij schokken of contact discontinuiteiten. De manier waarop de hellingen van de lineaire verdelingen begrensd moeten worden, wordt op een nieuwe, eenvoudige en duidelijke manier gepresenteerd voor een algemene niet lineaire scalaire hyperbolische behoudswet. De beschrijving is eenvoudig doordat is uitgegaan van het eenzijdig tweede-orde upwind schema. Aangetoond is dat een bepaalde zwakke vorm van monotoniciteit gelijktijdig gecombineerd kan worden met tweede-orde nauwkeurigheid. Dit geldt zowel in één dimensie als in meerdere dimensies.

In het tweede deel van dit proefschrift (hoofdstukken III en IV) worden oplossingsmethoden ontwikkeld voor de eerste- en tweede-orde discretisaties. Ten gevolge van de gunstige eigenschappen van de eerste-orde discretisatie (5-punts stencil struktuur, upwind karakter, differentieerbaarheid, consistentie van de berekening van de flux door inwendige celwandjes en door celwandjes die deel uit maken van de rand van het fysisch domein) kan een ongekunstelde niet lineaire multigrid oplossingsmethode ontwikkeld worden. In de toegepaste

multigrid methode zijn de grof net discretisaties Galerkin approximaties van de fijn net discretisatie en is een eenvoudige Collectieve Symmetrische Gauss-Seidel (CSGS) relaxatie methode een uitstekende smoothing procedure. De numerieke testproblemen hebben betrekking op kanaal stromingen, resolutie van contact discontinuiteiten en een stomp lichaam in een supersone stroming. De resultaten van de testproblemen tonen aan dat een mate van efficientie en robuustheid wordt verkregen welke karakteristiek is voor een goed werkende multigrid methode. Voor berekeningen aan praktische problemen, waarbij het alleen maar zin heeft te convergeren tot aan afbreekfout nauwkeurigheid, zijn slechts een paar (twee of drie) niet lineaire multigrid iteraties voldoende. Dit betekent dat eerste-orde oplossingen worden verkregen in een hoeveelheid werk equivalent met ongeveer $3\times\frac{4}{3}\times 2$ CSGS-relaxaties op het fijnste grid.

Een Defect Correctie (DeC) iteratie methode wordt gebruikt om de nauwkeurigheid van de eerste-orde oplossingen te verbeteren. De DeC-iteratie methode maakt op een effectieve manier gebruik van de multigrid methode voor het oplossen van de eerste-orde discretisaties. Het is bekend dat voor gladde problemen slechts één DeC-iteratie voldoende is om tweede-orde nauwkeurigheid te verkrijgen. Voor niet gladde problemen blijkt dat ongeveer 10 DeC-iteraties noodzakelijk zijn. Omdat 10 DeC-iteraties corresponderen met een hoeveelheid werk equivalent met ongeveer $10\times\frac{4}{3}\times 2$ CSGS-relaxaties op het fijnste net, mogen we concluderen dat de methode een redelijk efficiente manier is om de nauwkeurigheid van de oplossing te verbeteren. Voor de eerder genoemde testproblemen blijkt dat de resultaten verkregen na toepassing van de DeC-iteratie methode aanzienlijk nauwkeuriger zijn dan de eerste-orde oplossingen verkregen met de multigrid methode. Tenslotte merken we nog op dat het soms noodzakelijk is (voor stromingen waarin zeer sterke schokken aanwezig zijn) om enige demping in het DeC-iteratie proces toe te passen.

ACKNOWLEDGEMENTS

I wish to express my gratitude to:

DR. P.W. HEMKER, who put the train on the rails so that it was easy to progress;

PROF.DR.IR. P. WESSELING, for his careful reading of the manuscript and the valuable suggestions which led to many improvements;

DRS. P.M. DE ZEEUW, for his assistance in the computational work, especially for his implementation of user friendly plotting routines which was a pleasure to use;

PROF.DR. B. VAN LEER, for the many pleasant and stimulating discussions;

IR. B. KOREN, for his cooperation in this Euler research project;

MR. M. DELUSSU, who typed the manuscript so excellently.

Curriculum Vitae

De schrijver van dit proefschrift werd geboren te Borne, op 2 juli 1958.

Na het behalen van het VWO-diploma aan het Twickel College te Hengelo begon hij in 1976 met zijn studie Wiskunde aan de Rijksuniversiteit Groningen. In 1979 legde hij het kandidaatsexamen af en in 1983 het doctoraalexamen Technische Mechanica met bijvakken Wiskunde en Numerieke Wiskunde.

Vervolgens was hij van 1 oktober 1983 tot 1 oktober 1987 als wetenschappelijk medewerker verbonden aan de afdeling Numerieke Wiskunde van het Centrum voor Wiskunde en Informatica te Amsterdam. In deze periode werd het onderzoek verricht dat in dit proefschrift is beschreven.

Op 1 oktober 1987 trad hij in dienst van het Nationaal Lucht- en Ruimtevaartlaboratorium, vestiging Noordoostpolder.