Computational Fluid Dynamics

I + II

Numerische Strömungsmechanik

I + II

Prof. Dr.-Ing. Dieter Hänel
Matthias Meinke

Lennart Schneiders, Konrad Pausch, Marian Albers
Jannik Borgelt

Institute of Aerodynamics

RWTH Aachen University

Contents

Chapter 1

Computational Fluid Dynamics I

Introduction

Computational Fluid Dynamics (CFD) is a field of physics, which is concerned with the numerical solution of the governing equations describing a fluids motion. These equations are usually systems of partial differential equations for which analytical solutions are difficult or even impossible to find. With CFD methods, numerical solutions of fluid dynamical problems can be obtained, which allows in principle a prediction of any flow problem. Continuous improvements of the numerical solution methods and increasing computing power have made CFD become an important factor in science and engineering. Applications can be found wherever flow problems play an important role, e.g. the aerospace and automotive industry, but also in the fundamental research of flow phenomena, such as turbulence or combustion research.

CFD is based on the discrete approximation of the basic differential problems, e.g. the transformation of differentials into finite differences at discrete points of an integration domain. The approximation causes truncation errors which leads to a divergence of the numerical from the exact solution of the flow problem. Therefore, the most important effort of computational fluid dynamics is the formulation of discrete approximations such that the error remains limited and to guarantee that the numerical solution approaches the exact solution with increasing grid resolution, i.e. with decreasing step sizes (convergence of the numerical problem).

This course will explain the basics of the formulation of the governing equations of fluid mechanics with difference approximations. Section 1.1 introduces the governing equations in their most important formulations and approximations. A mathematical classification of these equations is performed in section 1.2. The basics of the dicrete formulation of partial differential equations are summarized in section 1.3. The derivation of consistent difference approximations and conditions for numerical stability and convergence of initial value problems are discussed. The iterative solution of elliptic partial differential equations, such as Poisson's equation and the Laplace equation are considered in section 1.4. The most important iteration schemes are discussed in this context.

In the second chapter of this course, important equqation systems in fluid mechanics are discussed. In the first section of chapter two, i.e., section 2.1, the numerical integration of Prandtl's boundary layer equations is discussed and a solution method is developed. In

the further sections of this course, the basics of the numerical solution of the Euler equations for compressible, unsteady flows are presented. Solution schemes are first introduced for hyperbolic scalar equations in section 2.2. Subsequently, different form of the Euler equations are introduced in the next section 2.3. The solution of this hyperbolic, nonlinear system is one of the most important tasks in computational fluid dynamics. This is not only relevant for the simulation of inviscid flows, but also is a prerequisite for the solution of the Navier-Stokes equations. The properties of the Euler equations are discussed including continuous as well as discontinuous solutions such as shock waves, see section 2.3.3. An appropriate numerical description of discontinuous solutions requires a so called conservative discretization. The most important schemes for such a discretization are presented and their properties are demonstrated for a one-dimensional model problem in section 2.4.4.

1.1 Governing equations of fluid dynamics

- The flow of a continuum, i.e., a gas or liquid, is described by the conservation laws of mass, momentum and energy.
- Additional relations are necessary for the solution of these conservation equations:

Caloric equation of state	e.g. $e = c_v\, T$
Thermal equation of state	e.g. $p = \rho R\, T$
Formulations for transport coefficients	e.g. $\eta = \eta(T),\ \lambda = \lambda(T)$

- The initial and boundary conditions for the conservation equations define the specific flow problem.

1.1.1 General formulation of the conservation equations

Integral formulation of the conservation equations

The integral form of the conservation equations is obtained from the fundamental conservation law of classical physics. This principle is applied to a control volume τ in a fluid flow with the surface A. To each surface element dA of A a normal vector $\vec{n}$ is assigned.

The conservation principle for the volume τ can be formulated in a general way for the conserved quantities mass, momentum and energy per volume unit:

Temporal change of the conservation quantities $\vec{U}$ in the volume τ + *Generalized flux $\vec{H}$ (fluxes, stresses) across the surface A* = *Effect of the volume forces $\vec{F}_{vol}$ exerted on τ*

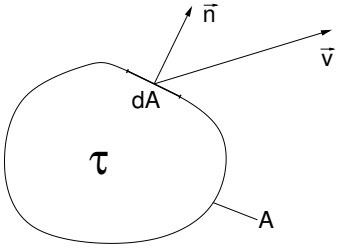


Figure 1.1.1: Control volume τ with the surface A, the normal vector $\vec{n}$ and the velocity vector $\vec{v}$.

The mathematical formulation of these items yields the integral form of the conservation equations:

$$\boxed{\int_\tau \frac{\partial \vec{U}}{\partial t} d\tau + \oint_A \vec{H} \cdot \vec{n} dA = \int_\tau \vec{F}_{vol}\, d\tau}$$

In the above equations $\vec{U}$ is the vector of conservation quantities with the elements mass/volume (ρ), momentum/volume ($\rho\,\vec{v}$) and energy/volume ($\rho\, E = \rho\,(e + \frac{1}{2}\,\vec{v}^{\,2})$).

$$\vec{U} = \begin{pmatrix} \rho \\ \rho\vec{v} \\ \rho E \end{pmatrix}$$

The generalized flux vector $\vec{H}$ summarizes the effects of the fluxes and the stresses. The first component of $\vec{H}$ describes the mass flow $\rho\,\vec{v}$, the second component includes the momentum flux $\rho\,\vec{v}\,\vec{v}$ and the stress tensor $\overline{\overline{\sigma}}$, the third component is composed from the energy flux $\rho\,\vec{v}\,E$ the effect of the stresses $\overline{\overline{\sigma}}\vec{v}$ and the heat flux $\vec{q}$:

$$\vec{H} = \begin{pmatrix} \rho\vec{v} \\ \rho\vec{v}\vec{v} + \overline{\overline{\sigma}} \\ \rho E\vec{v} + \overline{\overline{\sigma}}\vec{v} + \vec{q} \end{pmatrix}$$

The vector $\vec{F}_{vol}$ represents the volume forces. It includes e.g. gravity $\vec{f}_{vol} = \rho\,\vec{g}$. The effects of the volume forces $\vec{f}_{vol} \cdot \vec{v}$ includes :

$$\vec{F}_{vol} = \begin{pmatrix} 0 \\ \vec{f}_{vol} \\ \vec{f}_{vol} \cdot \vec{v} \end{pmatrix}$$

Component wise the conservation equations for mass, momentum and energy are:

$$\begin{aligned} \int_\tau \frac{\partial \rho}{\partial t} d\tau + \oint_A [\rho\vec{v}] \cdot \vec{n}\, dA &= 0 \\ \int_\tau \frac{\partial \rho\,\vec{v}}{\partial t} d\tau + \oint_A [\rho\vec{v}\vec{v} + \overline{\overline{\sigma}}] \cdot \vec{n}\, dA &= \int_\tau \vec{f}_{vol}\, d\tau \\ \int_\tau \frac{\partial \rho\, E}{\partial t} d\tau + \oint_A [\rho E\vec{v} + \overline{\overline{\sigma}}\vec{v} + \vec{q}] \cdot \vec{n}\, dA &= \int_\tau \vec{f}_{vol} \cdot \vec{v}\, d\tau \end{aligned}$$

The uniform appearance of the conservation equations for mass, momentum and energy and as a consequence of the direct applicability on grid defined control volumes, the integral form of the conservation equations is an important starting point for numerical discretization schemes (e.g. finite volume method).

Differential formulation of the conservation equations

The integral form of the conservation equations can be transferred into a system of partial differential equations by means of Gauss theorem. This directly delivers the conservative, respectively the divergence form of the conservation equations. This form is also an important starting point for the numerical discretization.

Further differential formulations of the conservation equations can be derived by introducing other dependent variables instead of the conserved variables $\vec{U}$. They are most often called non-conservative formulations. But apart from few exceptions (e.g. boundary layer equations), they play a minor role in the numerical discretization. Nevertheless, they are often better suited for the analysis of the solution behavior conservative formulations.

a) Conservative (Divergence-) form:

Under the assumption of functions that are continuous and can be derived to a suitable degree in space and time, the integral form:

$$\int_\tau \frac{\partial \vec{U}}{\partial t}\, d\tau \;+\; \oint_A \vec{H}\cdot\vec{n}\, d\,A \;=\; \int_\tau \vec{F}_{vol}\, d\,\tau$$

can be transferred into a differential form by application of the Gaussian Integral Theorem:

$$\oint_A \vec{H}\cdot\vec{n}\, d\,A \;=\; \int_\tau \nabla\cdot\vec{H}\, d\,\tau$$

and differentiation for the volume τ :

$$\boxed{\frac{\partial \vec{U}}{\partial t} \;+\; \nabla\cdot\vec{H} \;=\; \vec{F}_{vol}}$$

This system of partial differential equations is generally called the conservative form or divergence form (because of $\nabla\cdot\vec{H}$) of the conservations equations. The divergence form of the conservation equations is like the integral form an important starting point for the numerical solution.

Considering the components of the vector of the specific conservation quantities $\vec{U}$, the generalized flux vector $\vec{H} \;=\; \vec{H}\,(\vec{U})$ and the vector of volume forces $\vec{F}_{vol}\,(f_{vol})$ the system of conservation equations for mass, momentum and energy is obtained:

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla\cdot\rho\vec{v} &= 0 \\
\frac{\partial \rho\vec{v}}{\partial t} + \nabla\cdot(\rho\vec{v}\vec{v} + \overline{\overline{\sigma}}) &= \vec{f}_{vol} \\
\frac{\partial \rho E}{\partial t} + \nabla\cdot(\rho E\vec{v} + \overline{\overline{\sigma}}\vec{v} + \vec{q}) &= \vec{f}_{vol}\cdot\vec{v}
\end{aligned}$$

b) Non-conservative forms

If instead of the conservation quantities $\vec{U}$ another set of conservation quantities (e.g. $\rho, \vec{v}, E$) is chosen as dependent variables, the non-conservative form is obtained. In this case the divergence form can not be preserved and variable coefficients in front of the differentials occur, typically indicating the non-conservative form.

A non-conservative form is obtained for instance when a moving system with velocity $\vec{v}$ is chosen (Lagrangian formulation). Here, the temporal change in the moving system is obtained as the "substantial derivate":

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \nabla$$

With this the conservation equations can be written in the following form:

$$\begin{aligned} \frac{D\rho}{Dt} + \rho \nabla \cdot \vec{v} &= 0 \\ \frac{D\vec{v}}{Dt} + \frac{1}{\rho} \nabla \cdot \overline{\overline{\sigma}} &= \frac{1}{\rho} \vec{f}_{vol} \\ \frac{DE}{Dt} + \frac{1}{\rho} \nabla \cdot (\overline{\overline{\sigma}} \vec{v} + \vec{q}) &= \frac{1}{\rho} \vec{f}_{vol} \cdot \vec{v} \end{aligned}$$

This system is only one possible non-conservative formulation. Especially for the energy law, several different forms are possible, e.g.:

$$\rho \frac{De}{Dt} + \overline{\overline{\sigma}} \cdot \nabla \vec{v} + \nabla \cdot \vec{q} = 0$$

Thermal and caloric state equations

The description of compressible flows requires additional relations between the thermal and caloric state quantity for the closure of the conservation equations. The following equations are valid for an thermal and caloric perfect gas (e.g. air for temperatures up to 800 K):

- Thermal state equation:

$$p = \rho R T$$

- Caloric state equation:

$$e = c_v T \qquad h = e + p/\rho = c_p T \quad \text{with} \quad c_p = const. \quad \text{and} \quad c_v = const.$$

With this the relation between pressure p and inner energy e becomes:

$$p = (\gamma - 1) \rho e \qquad \text{with} \quad \gamma = \frac{c_p}{c_v}$$

Transport coefficients

For the calculation of shear stresses and the heat flow closure assumptions are necessary which couple these quantities with the flow variables. One decides between laminar flows on the one hand and turbulent flows on the other hand.

For most fluids in laminar flows the behavior of a Newtonian fluid can be assumed which states the linear dependence of the shear stress from the velocity gradient, e.g. $\tau = -\eta \frac{du}{dy}$. In analogy to this Fourier's law states the linear relationship between the heat flux and the temperature gradient, i.e. $\vec{q} = -\lambda \nabla T$.
The proportionality coefficients are the transport coefficients for the molecular momentum exchange (dynamic viscosity η) and for the energy exchange (heat conductivity λ). These are material constants and depend on the thermodynamic state of the fluid. For gases the viscosity is essentially a function of the temperature which can be approximated by a power law:

$$\frac{\eta}{\eta_0} = \left(\frac{T}{T_0}\right)^{\omega} \qquad \text{with} \quad .5 < \omega < 1 \qquad (\text{Left } \omega = .72\,)$$

Assuming a constant Prandtl number $Pr = \frac{c_p\,\eta}{\lambda}$ and constant c_p value the heat conductivity is proportional to the viscosity.

The mathematical formulation of the mechanisms in turbulent flows is much more challenging than for laminar flow and not completely solved until today. This is mainly due to the statistical fluctuations of the momentum and energy exchange. Most approaches for the solution of turbulent flows are based on Reynolds averaging theory (see e.g. course fluid dynamics II). In this theory the temporal flow quantity f is split into a time averaged mean value $\bar{f}$ and a fluctuating part f'. After introducing this approach into the conservation equations and temporal averaging one obtains the time averaged equations which differ from the original equations by additional stress and heat flux components (Reynolds stresses). Such additional components like the cross product of the velocity fluctuation $\overline{u'\,v'}$ are further unknowns that need to be coupled with the mean quantities by closure assumptions. A number of closure assumptions exist, starting from two-equation models (e.g $k-\epsilon$ model) until the simpler algebraic models. One of the most simple algebraic closure assumptions is given by Prandtls mixing length hypothesis

$$\overline{u'\,v'} = -l^2\,\left|\frac{\partial \bar{u}}{\partial y}\right|\,\frac{\partial \bar{u}}{\partial y} = -\eta_{turb}\,\frac{\partial \bar{u}}{\partial y}$$

Such an eddy viscosity approach allows to retain the assumption of a Newtonian fluid. This is achieved by replacing the viscosity η with the effective viscosity which consists of the sum of the laminar and the turbulent viscosity, i.e. $\eta_{eff} = \eta_{lam} + \eta_{turb}$. Therefore, the general structure of the conservation equations remains unchanged.

1.1.2 Conservation equations in Cartesian coordinates (x,y,t)

In this section the conservation equations are derived in Cartesian coordinates and the Navier-Stokes equations and their most important approximations for compressible flows are described.

The Navier-Stokes equations which describe the fluid flow, including friction and heat conduction represent the most complete description of continuum flows. The complexity of their solution requires a high computational effort. Therefore, approximations of these equations are used wherever this seem physically reasonable. One of the most important concepts in this respect is Prandtl's boundary layer approximation which is valid for high Reynolds numbers and attached flows. According to this theory the flow field around a body can be split into a thin viscous boundary layer on the contour line of the body and an inviscid outer flow. The pressure is constant normal to the contour line inside the boundary layer and is therefore determined by the inviscid outer flow. According to this separation of the flow, the Navier-Stokes equations can also divided in two more straightforward equation systems. For boundary layer flows the boundary layer equations are obtained, while the inviscid outer flow is determined by the Euler equations or their approximation, the potential equations.

In the following the effect of the volume forces is neglected ($\vec{F}_{vol} = 0$).

Definitions for Cartesian coordinates

Some definitions are necessary for the presentation in Cartesian coordinates:

Note : It is $\vec{f} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = (f_1\,,\, f_2)^T$

- Surface normal: $\vec{n}dA = (dy\,,\, -dx)^T$
- Volume: $\tau = \tau(x, y)$
- Nabla operator: $\nabla = (\partial/\partial x\,,\, \partial/\partial y)^T$
- Dyadic product of two vectors $\vec{a}\,\vec{b}$ e.g. for momentum flux $\rho\,\vec{v}\,\vec{v}$

  $$\vec{a}\,\vec{b} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} \begin{pmatrix} b_x \\ b_y \end{pmatrix} = \begin{pmatrix} a_x b_x & a_x b_y \\ a_y b_x & a_y b_y \end{pmatrix}$$

- inner vector product of tensor and vector $\overline{\overline{T}}\,\vec{a}$, e.g. for $\overline{\overline{\sigma}}\,\vec{v}$ or $\overline{\overline{\sigma}}\,\vec{n}$

  $$\overline{\overline{T}}\,\vec{a} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \begin{pmatrix} a_x \\ a_y \end{pmatrix} = \begin{pmatrix} t_{11}\,a_x + t_{12}\,a_y \\ t_{21}\,a_x + t_{22}\,a_y \end{pmatrix}$$

- velocity: $\vec{v} = (u\,,\, v)^T$
- Conservation quantities: $\vec{U} = (\rho, \rho u, \rho v, \rho E)^T$

  with $\quad E = c_v\,T + \frac{1}{2}(u^2 + v^2)$

- Stress tensor with pressure and friction components $\quad \overline{\overline{\sigma}} = p\,I + \sigma$

  with $\qquad \sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}$

$$\begin{array}{lrcl} & \sigma_{xx} &=& -\eta\,(2u_x - \frac{2}{3}\,(u_x + v_y)) \\ \text{and} & \sigma_{yy} &=& -\eta\,(2v_y - \frac{2}{3}\,(u_x + v_y)) \\ & \sigma_{xy} &=& -\eta\,(u_y + v_x) \end{array}$$

- Heat flux: $\vec{q} = (q_x\,,\, q_y)^T = -\lambda\,(T_x\,,\, T_y)^T$
- generalized flux vector with inviscid and viscous component $\vec{H} = \vec{H}_{inv} + \vec{H}_{vis}$ in Cartesian coordinates: $\vec{H}_{inv} = (E_{inv}\,,\, F_{inv})^T \quad \vec{H}_{vis} = (E_{vis}\,,\, F_{vis})^T$

  with

$$E_{inv} = (\rho u\,,\, \rho u^2 + p\,,\, \rho uv\,,\, \rho uE + up)^T \qquad E_{vis} = (0\,,\, \sigma_{xx}\,,\, \sigma_{xy}\,,\, u\sigma_{xx} + v\sigma_{xy} + q_x)^T$$
$$F_{inv} = (\rho v\,,\, \rho uv\,,\, \rho v^2 + p\,,\, \rho vE + vp)^T \qquad F_{vis} = (0\,,\, \sigma_{xy}\,,\, \sigma_{yy}\,,\, u\sigma_{xy} + v\sigma_{yy} + q_y)^T$$

Conservation equations for compressible flows

Navier-Stokes equations

The Navier-Stokes equations describe the viscous continuum flow and heat conduction. In this context the Navier-Stokes equations are understood as the complete set of conservation for mass, momentum and energy.

a) Integral form

With the above definitions the integral form can be obtained after performing the inner product $\vec{H} \cdot \vec{n}\, dA$.

$$\int_\tau \frac{\partial \vec{U}}{\partial t}\, d\tau \;+\; \oint_A (E_{inv} + E_{vis})\; dy - \oint_A (F_{inv} + F_{vis})\; dx \;=\; 0$$

The components $E\,dy - F\,dx$ correspond to the normal projection of the flux $\vec{H} = (E\,,\, F)^T$ on a surface element $dA = \sqrt{dx^2 + dy^2}$, multiplied by dA.

b) Conservative form (Divergence form)

With the Cartesian components of ∇ and $\vec{H}$ the divergence form is found to be:

$$\frac{\partial \vec{U}}{\partial t} \;+\; \frac{\partial}{\partial x}(E_{inv} + E_{vis}) \;+\; \frac{\partial}{\partial y}(F_{inv} + F_{vis}) \;=\; 0$$

The component wise presentation of the conservation equations for mass, momentum and energy yields:

$$\begin{array}{lcllcl} \frac{\partial}{\partial t}\,\rho & + & \frac{\partial}{\partial x}\,(\rho u) & + & \frac{\partial}{\partial y}\,(\rho v) & = 0 \\ \frac{\partial}{\partial t}\,(\rho u) & + & \frac{\partial}{\partial x}\,(\rho u^2 + p + \sigma_{xx}) & + & \frac{\partial}{\partial y}\,(\rho uv + \sigma_{xy}) & = 0 \\ \frac{\partial}{\partial t}\,(\rho v) & + & \frac{\partial}{\partial x}\,(\rho uv + \sigma_{xy}) & + & \frac{\partial}{\partial y}\,(\rho v^2 + p + \sigma_{yy}) & = 0 \\ \frac{\partial}{\partial t}\,(\rho E) & + & \frac{\partial}{\partial x}\,(\rho uE + up + u\sigma_{xx} + v\sigma_{xy} + q_x) & + & \frac{\partial}{\partial y}\,(\rho vE + vp + v\sigma_{yy} + u\sigma_{xy} + q_y) & = 0 \end{array}$$

c) Non-conservative form

For the above described non-conservative form with the variables $\vec{V} = (\rho\,,\, u\,,\, v\,,\, E)^T$ and the substantial derivative $\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$ the following system is obtained in Cartesian coordinates:

$$\begin{array}{lcllcll}
\frac{D}{Dt}\,\rho & + & \rho\frac{\partial}{\partial x}\,u & + & \rho\frac{\partial}{\partial y}\,v & = & 0 \\
\frac{D}{Dt}\,u & + & \frac{1}{\rho}\,\frac{\partial}{\partial x}(p+\sigma_{xx}) & + & \frac{1}{\rho}\,\frac{\partial}{\partial y}\,\sigma_{xy} & = & 0 \\
\frac{D}{Dt}\,v & + & \frac{1}{\rho}\,\frac{\partial}{\partial x}\,\sigma_{xy} & + & \frac{1}{\rho}\,\frac{\partial}{\partial y}(p+\sigma_{yy}) & = & 0 \\
\frac{D}{Dt}\,E & + & \frac{1}{\rho}\,\frac{\partial}{\partial x}(up+u\sigma_{xx}+v\sigma_{xy}+q_x) & + & \frac{1}{\rho}\,\frac{\partial}{\partial y}(vp+u\sigma_{xy}+v\sigma_{yy}+q_y) & = & 0
\end{array}$$

The Euler equations

One obtains the Euler equations from the Navier-Stokes equations, if the viscosity terms and the heat conduction terms are neglected: $\sigma = 0$ and $\vec{q} = 0$. This simplification causes another solution behavior of the conservation equations. The Euler equations are treated in detail in part II of this course.

In analogy to the Navier-Stokes equations the following forms can be presented:

a) Integral form

$$\int_\tau \frac{\partial \vec{U}}{\partial t}\,d\tau \;+\; \oint_A (E_{inv})\;dy - \oint_A (F_{inv})\;dx \;=\; 0$$

The components $E\,dy - F\,dx$ correspond to the normal projection of the flux $\vec{H} = (E\,,\, F)^T$ on a surface element $dA = \sqrt{dx^2 + dy^2}$, multiplied by dA.

b) Conservative Form (Divergence-Form)

With the Cartesian components of ∇ and $\vec{H}$ the divergence form is found to be:

$$\frac{\partial \vec{U}}{\partial t} \;+\; \frac{\partial}{\partial x}E_{inv} \;+\; \frac{\partial}{\partial y}F_{inv} \;=\; 0$$

The component wise presentation of the conservation equations for mass, momentum and energy yields:

$$\begin{array}{lcllcll}
\frac{\partial}{\partial t}\,\rho & + & \frac{\partial}{\partial x}\,(\rho u) & + & \frac{\partial}{\partial y}\,(\rho v) & = & 0 \\
\frac{\partial}{\partial t}\,(\rho u) & + & \frac{\partial}{\partial x}\,(\rho u^2+p) & + & \frac{\partial}{\partial y}\,(\rho uv) & = & 0 \\
\frac{\partial}{\partial t}\,(\rho v) & + & \frac{\partial}{\partial x}\,(\rho uv) & + & \frac{\partial}{\partial y}\,(\rho v^2+p) & = & 0 \\
\frac{\partial}{\partial t}\,(\rho E) & + & \frac{\partial}{\partial x}\,(\rho uE+up) & + & \frac{\partial}{\partial y}\,(\rho vE+vp) & = & 0
\end{array}$$

c) Non-conservative form

For the above described non-conservative form with the variables $\vec{V} = (\rho\,,\, u\,,\, v\,,\, E)^T$ and the substantial derivative $\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$ the following system is obtained in Cartesian coordinates:

$$\begin{array}{lcllcll} \frac{D}{Dt}\,\rho & + & \rho\frac{\partial}{\partial x}\,u & + & \rho\frac{\partial}{\partial y}\,v & = & 0 \\ \frac{D}{Dt}\,u & + & \frac{1}{\rho}\,\frac{\partial}{\partial x}\,p & + & 0 & = & 0 \\ \frac{D}{Dt}\,v & + & 0 & + & \frac{1}{\rho}\,\frac{\partial}{\partial y}\,p & = & 0 \\ \frac{D}{Dt}\,E & + & \frac{1}{\rho}\,\frac{\partial}{\partial x}\,(up) & + & \frac{1}{\rho}\,\frac{\partial}{\partial y}\,(vp) & = & 0 \end{array}$$

Potential equation

Potential flows require irrotational, isoenergetic flows. The following assumptions are made:

- irrotational flow $\quad \nabla \times \vec{v} = 0$, also given by the potential ϕ, if $\vec{v} = \nabla\phi \quad$ or $u = \phi_x \quad v = \phi_y$.
- steady flow $\quad \frac{\partial}{\partial t} = 0$
- isoenergetic flow $\; H_t = h + \frac{1}{2}(u^2 + v^2) = const.$
  this yields $\quad T/T_0 = 1/(1 + \frac{\gamma-1}{2}\,Ma^2)$
- isentropic flow follows from Croccos vorticity theorem $(\vec{v} \times (\nabla \times \vec{v}) = \nabla H_t - T\,\nabla S)$,
  this yields $\nabla S = 0$ and thus $\; T/T_0 = (\rho/\rho_0)^{\gamma-1} = (p/p_0)^{\frac{\gamma-1}{\gamma}}$

The density ρ and the speed of sound a are in this case functions which depend on the potential ϕ. They can be calculated with the isotropy relation and the energy law as function of $\vec{v} = \nabla\,\phi$.

a) Conservative form

This form is obtained from the continuity equation and the definition of the potential

$$(\rho\,\phi_x)_x + (\rho\,\phi_y)_y = 0$$

b) Non-conservative form

A non-conservative form is obtained from the Euler equations with $dp = a^2\,d\rho$ for $s = const.$

$$(u^2 - a^2)\,\frac{\partial^2\phi}{\partial x^2} + 2\,u\,v\,\frac{\partial^2\phi}{\partial x \partial y} + (v^2 - a^2)\frac{\partial^2\phi}{\partial y^2} = 0$$

Boundary layer equations

The boundary layer equations are derived from the Navier-Stokes equations by application of Prandtl's boundary layer approximation (see e.g.: H. Schlichting: "Grenzschichttheorie"). The most important assumptions for this are high Reynolds numbers $Re \gg 1$ and attached flow. For this case the viscosity is only influential in a thin layer of width δ close to the body with the properties $\frac{\delta}{L} \sim \frac{v}{U_\infty} \sim \frac{1}{\sqrt{Re_\infty}}$.

$$\begin{aligned}
\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} &= 0 \\
\rho u\frac{\partial u}{\partial x} + \rho v\frac{\partial u}{\partial y} + \frac{\partial p}{\partial x} &= \frac{\partial}{\partial y}\left(\eta\,\frac{\partial u}{\partial y}\right) \\
\frac{\partial p}{\partial y} &= 0 \\
\rho u\frac{\partial h}{\partial x} + \rho v\frac{\partial h}{\partial y} - u\frac{\partial p}{\partial x} &= \frac{\partial}{\partial y}\left(\lambda\frac{\partial T}{\partial y}\right) + \eta\left(\frac{\partial u}{\partial y}\right)^2
\end{aligned}$$

Conservation equations for incompressible flows

Many fluids, e.g. liquids, can be considered incompressible in most domains, i.e. the density ρ is constant. As a consequence the continuity equation is reduced to $div\,\vec{v} = 0$ which changes the solution behavior, since there's no time derivative in this equation.

Furthermore the pressure is no longer coupled with the density and the temperature by a state equation. This leads to the decoupling of the energy equation from the other equations (there still a weak coupling enforced by the transport coefficients, e.g. the viscosity $\eta(T)$, which is neglected in this case). Therefore, the continuity and momentum equations are sufficient for the solution of the flow.

As another result of the decoupling of the pressure, can no longer an explicit equation be derived for the pressure that also preserves the continuity (divergence free velocity field).

Therefore, solution schemes which are based on the equations with pressure and velocity as variables use an iteration process in which the pressure as a parameter is iterated such that for all times the continuity equation is fulfilled.

Navier-Stokes equations

a) $\vec{v}$, p formulation

For constant density and the velocity and pressure as dependent variables the following system of continuity and momentum equation is obtained:

$$\begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\ \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial x} &= \nu \nabla^2 u \\ \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial y} &= \nu \nabla^2 v \end{aligned}$$

with the Laplacian operator $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$

and the kinematic viscosity $\nu = \eta/\rho$

The pressure terms are eliminated by taking the curl of the momentum equations ($\nabla \times$ *momentumequation*). Then the vorticity transport equation is obtained with the z-component of the vorticity vector ζ as the variable.

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$$

The continuity equation is identically satisfied by the definition of the stream function.

$$\frac{\partial \psi}{\partial y} = u \quad , \qquad \frac{\partial \psi}{\partial x} = -v$$

Poisson's equation for the determination of the stream function ψ is obtained by insertion of the stream function in the definition of ζ.

With the above performed procedure the Navier-Stokes equations can be written as stream function and eddy transport equation:

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\zeta \qquad \| \text{ Poisson's equation for } \psi$$

$$\frac{\partial\zeta}{\partial t} + u\,\frac{\partial\zeta}{\partial x} + v\,\frac{\partial\zeta}{\partial y} = \nu\nabla^2\zeta \qquad \| \text{ Eddy transport equation}$$

The pressure p can be calculated afterwards from Poisson's equation for pressure which is derived from the divergence of the momentum equations:

$$\nabla^2 p = -\rho\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + 2\,\frac{\partial u}{\partial y}\,\frac{\partial v}{\partial x}\right]$$

Euler equations

The Euler equations for incompressible flows are obtained for diminishing viscosity ($\nu = 0$). In analogy to the Navier-Stokes equations there are two formulations:

a) $\vec{v}$, p formulation

$$\begin{aligned}
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial x} &= 0 \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial y} &= 0
\end{aligned}$$

b) $\psi - \zeta$ formulation (Stream function and Eddy transport equation)

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\zeta \qquad \| \text{ Poisson's equation for } \psi$$

$$\frac{\partial\zeta}{\partial t} + u\,\frac{\partial\zeta}{\partial x} + v\,\frac{\partial\zeta}{\partial y} = 0 \qquad \| \text{ Eddy transport equation}$$

The pressure p can be calculated afterwards from Poisson's equation for pressure.

Potential equation

The potential equations are obtained from the Euler equations for irrotational ($\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0$) and steady ($\frac{\partial}{\partial t} = 0$) flows. Depending on the dependent variable one distinguishes between:

- Velocity formulation (Cauchy-Riemann equations)

  The condition of irrotational flows replaces the momentum equations, while the continuity equation is retained.

$$\begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\ \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} &= 0 \end{aligned}$$

- Stream function formulation

  The definition of the stream function satisfies the continuity equation. Insertion in condition of irrotational flow yields Laplace's equation for the stream function:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0$$

- Potential formulation

  The definition of the potential $\vec{v} = \nabla \Phi$ yields Laplace's equation for the potential function, if inserted in the continuity equation:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0$$

The integration of the momentum equations in respect to the irrotational flow yields Bernoulli's equation for the pressure calculation:

$$p_0 = p + \frac{\rho}{2}\left(u^2 + v^2\right) = const.$$

Boundary layer equations

Prandtl's boundary layer equations for an incompressible flow are:

$$\begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\ u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{1}{\rho}\frac{\partial p}{\partial x} &= \frac{\partial}{\partial y}\left(\nu \frac{\partial u}{\partial y}\right) \\ \frac{\partial p}{\partial y} &= 0 \end{aligned}$$

1.1.3 Initial and boundary conditions

The initial and boundary conditions define the flow problem which is to be solved by solution of the conservation equations.

The number of necessary boundary conditions is given by the highest order derivative of an independent variable. The distinction between initial and boundary conditions is determined by the type of the partial differential equation. Elliptic partial differential equations lead to boundary value problems, i.e. the boundary conditions have to be defined on all boundaries. Hyperbolic and parabolic partial differential equations have real characteristics and therefore a limited region of influence for which initial conditions must be defined on a non characteristic boundary line (initial value problem). If the region of influence is constrained by boundaries, additional boundary conditions must be defined (Initial-boundary value problem).

- 1. Type (Dirichlet boundary condition)

  $U = g_1(x, y)$ Variable value is defined on the boundary

  e.g. no slip condition $u = 0, \quad v = 0$

- 2. Type (von Neumann boundary condition)

  $\frac{\partial U}{\partial n} = g_2(x, y)$ Normal gradient of the variable is given on the boundary

  e.g. adiabatic wall $q_n = \lambda \frac{\partial T}{\partial n} = 0$

- 3. Type (linear combination of 1. and 2. type)

  $\alpha U + \beta \frac{\partial U}{\partial n} = g_3(x, y)$ Normal gradient and value are combined

  e.g. "slip" stream at the wall for dilute gases $a \frac{\partial u}{\partial n} + u = 0$

- 4. Type Periodic boundary conditions

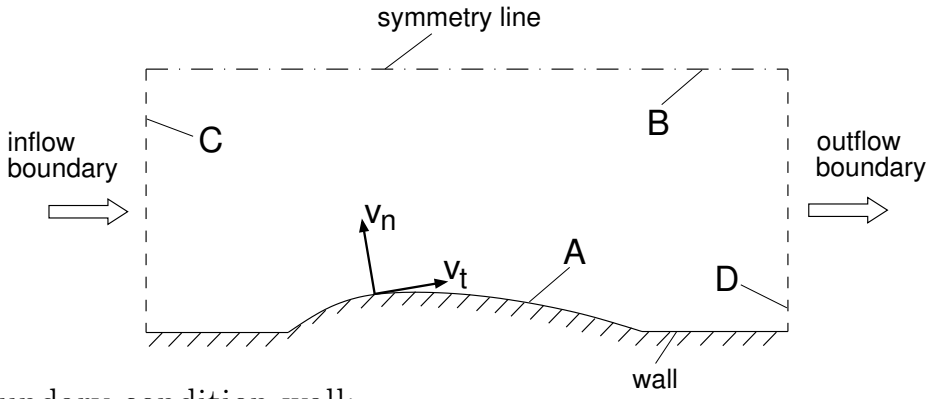
  $U(x_1, y_1) = U(x_2, y_2)$ Boundary values of two integration boundaries C_1 and C_2 equal

  e.g.: Turbine blades

The most important prerequisite for the solution is the formulation of a correct problem, i.e. little changes in the initial or boundary conditions of $O(\varepsilon)$ may only cause little changes in the solution of $O(\varepsilon)$!

Typical boundary conditions in fluid mechanics

Domain of integration:



A. Boundary condition wall:

- inviscid:

$$\left.\begin{array}{l} v_n = 0 \longrightarrow \text{solid wall} \\ \text{respectively} \quad \vec{v} = v_t \longrightarrow \text{Tangential condition} \end{array}\right\} \quad \text{Wall is streamline}$$

Boundary condition for Euler and potential equation.

- viscous:

BC like inviscid $v_n = 0$
+ Additional no slip condition: $v_t = 0$
+ aditional thermal BC: $T = T_w$ isothermal wall
or $q_n = -\lambda \frac{\partial T}{\partial n} = 0$ adiabatic wall

B. Boundary condition symmetry line:

$$\frac{\partial f}{\partial n} = 0 \qquad f = \rho, p, T, v_t$$
$$v_n = 0$$

C.+D. Boundary condition for inflow and outflow boundary:

$\longrightarrow$ Problem dependent, since usually across an unknown flowfield
$\longrightarrow$ Number of necessary boundary condition (i.e. variables that need to be defined) often determined from inviscid method of characteristics

e.g. for a 2-dimensional flow the following definitions are necessary:
$Ma < 1$ Inflow 3 variables (e.g. T_o, p_o, u)
$Ma > 1$ Inflow 4 (all) variables
$Ma < 1$ Outflow 1 variable (e.g. p)
$Ma > 1$ Outflow none

1.2 Classification and characteristic lines of partial differential equations

Characteristic solutions are outstanding solutions of partial differential equations. These solutions are qualified by being independent of neighboring solutions, i.e. the initial value problem cannot be determined uniquely from such a solution curve. Mathematically this means that the derivations cross wise of the solution curve are undetermined.

The slope of the corresponding base curve of the characteristic solution is generally called characteristic. The characteristic lines are independent of the coordinate system and therefore the "characteristic" property of a partial differential equation. The value of the characteristic, real or complex determines the solution behavior of the partial differential equations. It also serves the classification in elliptic, parabolic and hyperbolic partial differential equations.

The characteristic lines define the physical sphere of influence. Real characteristics (hyperbolic and parabolic equations) lead to initial value problems with limited sphere of influence (e.g. Mach cone). Complex characteristics of elliptic equations lead to boundary value problems with no special direction of influence.

Characteristic lines are important for numerical solution methods e.g. for the development of stable and accurate difference schemes and for the modeling of boundary conditions.

1.2.1 First order partial differential equations

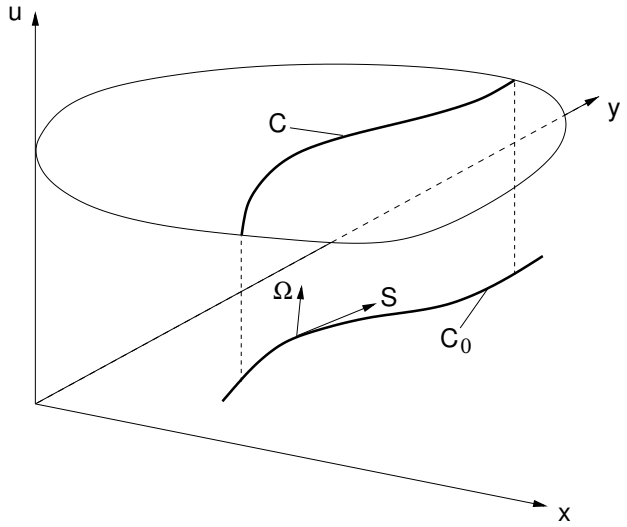
The most straightforward interpretation of the characteristic lines can be given for scalar first order partial differential equations. These equations lead to real characteristics and are therefore of hyperbolic type.

Condition for characteristics

The following equation will be considered for the derivation of the characteristic:

$$a\,u_x + b\,u_y = c$$

It is assumed that the solution $u(x, y)$ on the solution curve C (Initial condition) is known over the basic curve C_0 defined by $\Omega(x, y) = const$. For the presentation of the characteristics it must first be investigated for which basic curves the cross wise derivatives of the solution become undetermined. To achieve this the differential equations are transformed into a coordinate system (S, Ω), where S is tangential to C_0 and Ω perpendicular to C_0.



Solution surface $u(x, y)$

C : Characteristic solution curve
C_0 : char. Basic curve $\Omega = const.$

$\left.\frac{dy}{dx}\right|_{Co}$: char. Directional derivative

For the transformation $(x, y) \rightarrow (S, \Omega)$ and with the derivatives:

$$u_x = S_x u_S + \Omega_x u_\Omega; \qquad u_y = S_y u_S + \Omega_y u_\Omega$$

the differential equation $a u_x + b u_y = c$ is obtained in the new coordinate system :

$$(a\,\Omega_x + b\,\Omega_y)\,u_\Omega + (a S_x + b S_y)\,u_S = c$$

The value of the first bracketed expression, $Q \equiv a\,\Omega_x + b\,\Omega_y$ is determinant for the investigation of the behavior of the cross wise derivative u_Ω:

- a) The value of Q is unequal zero.

  $$Q = a\Omega_x + b\Omega_y \neq 0$$

  In this case the cross wise derivative u_Ω is uniquely defined by the solution. The neighbor solution on a curve $\Omega + \Delta\Omega$ can be continued unambiguously from the solution $u(x, y)$ on $\Omega = const.$.

  $$\rightarrow \quad u(\Omega + \Delta\Omega) = u(\Omega) + u_\Omega(\Omega) \cdot \Delta\Omega + \cdots$$

  The initial value problem is uniquely determined.

- b) The value of Q is zero.

  $$Q = a\,\Omega_x + b\,\Omega_y = 0$$

  In this case the cross wise derivative u_Ω is undetermined ($0 \cdot u_\Omega = ...$). The solution therefore only depends on the derivatives u_S which are tangential to the solution curve. The Initial value problem cannot be continued on a neighboring solution, unambiguously.

  For this case $u(x, y)$ is called *characteristic solution* and the curve $\Omega = const.$ as *characteristic base curve* C_0 of which the derivative $\frac{dy}{dx}$ forms the characteristic. The condition $Q = 0$ is accordingly called the *characteristic condition.*

Characteristic and equation of the characteristic base curve

From the characteristic condition $Q = 0$ and the equation for the base curve $\Omega = const.$

$$\begin{aligned} Q &= a\,\Omega_x + b\,\Omega_y = 0 \\ d\Omega &= \Omega_x\,dx + \Omega_y\,dy = 0 \end{aligned}$$

the slope of the characteristic base curve (=characteristic line) is obtained:

$$\left.\frac{dy}{dx}\right|_{Co} = -\frac{\Omega_x}{\Omega_y} = \frac{b}{a}$$

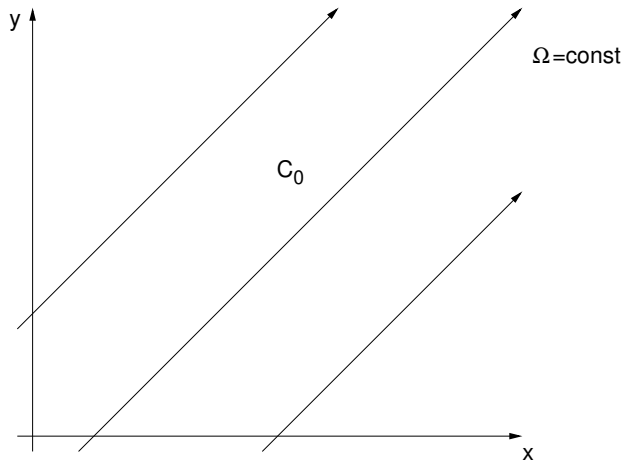
Integration of

$$\frac{d\Omega}{-\Omega_y} = \left.\frac{dy}{dx}\right|_{Co} dx - dy = \frac{b}{a}\,dx - dy = 0$$

with the predefined initial values x_0, y_0 yields the equation of the base curve C_0 :

$$y = y_0 + \frac{b}{a}(x - x_0) \quad .$$

Thus, the characteristic base curves of the equation $au_x + bu_y = c$ form a group of straight lines with the slope b/a.



Characteristic solution

The characteristic solution (conformity condition) is the solution along the characteristic base curve. For the presentation of this special solution the original equation will be transformed in a new coordinate system $\xi(x, y)\,, \tau(x, y)$. One coordinate, in this case $\xi = const.$, represents the characteristic base curve $\Omega = const.$ (in this general case called $\xi = const.$), while the other coordinate can be arbitrarily chosen for first order equations, e.g. $\tau = x$.

$$\begin{aligned} d\xi &= \frac{b}{a}\,dx - dy \\ d\tau &= dx \end{aligned}$$

By application of the chain rule

$$u_x = \xi_x\, u_\xi \,+\, \tau_x\, u_\tau = \frac{b}{a}\, u_\xi \,+\, u_\tau \quad , \qquad u_y = \xi_y\, u_\xi \,+\, \tau_y\, u_\tau = -\, u_\xi$$

the equation $a\, u_x + b\, u_y = c$ yields the transformed equation (normal form of the equation):

$$\left.\frac{\partial u}{\partial \tau}\right|_{\xi = const.} = \frac{c}{a}$$

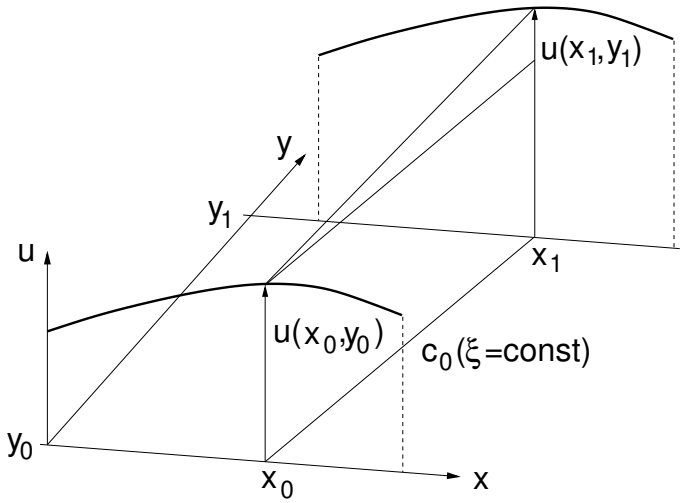
Integration over τ yields the characteristic solution over the base curve $\xi = \frac{b}{a}x - y = const.$:

$$u(\tau, \xi) = \frac{c}{a} \cdot \tau + k\,(\xi)$$

For the initial value $u_0(x_0, y_0)$ on the characteristic base curve $\xi = \xi_0 = \frac{b}{a}x_0 - y_0 = const.$ and with $\tau = x$ one obtains:

$$u(x, y) = \frac{c}{a}\,(x - x_0) + u_0(x_0, y_0)$$

In the special case of $c = 0$ the initial solution on the characteristic line remains constant, i.e. $u(x, y) = u(\xi) = const.$ These types of solutions occur frequently in gas dynamics, e.g. for the Prandtl-Meyer expansion.



1.2.2 Second order partial differential equations

The derivation of the characteristics is performed in analogy to the first order equations. The following equation will be taken as starting point:

$$L(u) = a\,u_{xx} + 2\,b\,u_{xy} + c\,u_{yy} + d(x, y, u, u_x, u_y) = 0$$

The solution $u(x, y)$ and the lower order derivations u_x and u_y are assumed to be known over the base curve C_0, defined by $\Omega(x, y) = const$. For the presentation of the characteristics (highest order cross wise derivatives which are undetermined) the differential equation is transformed in a coordinate system (S, Ω), with S being tangential with C_0 and Ω being perpendicular to C_0.

With the transformation $(x, y) \to (S, \Omega)$ one obtains the derivatives

$$\begin{aligned} u_{xx} &= u_{\Omega\Omega}\,{\Omega_x}^2 + 2\,u_{S\Omega}\,\Omega_x\,S_x + u_{SS}\,{S_x}^2 + u_\Omega\,\Omega_{xx} + u_S\,S_{xx} \\ u_{yy} &= u_{\Omega\Omega}\,{\Omega_y}^2 + \cdots \\ u_{xy} &= u_{\Omega\Omega}\,\Omega_x\Omega_y + \cdots \end{aligned}$$

All derivatives but the second cross wise derivative, $u_{\Omega\Omega}$, are known on the solution curve. This yields the differential equation in the new coordinate system:

$$L(u) = Q \cdot u_{\Omega\Omega} + [\sim]\,u_{\Omega S} + [\sim]\,u_{SS} + [\sim] = 0$$

with

$$Q = a\,{\Omega_x}^2 + 2\,b\,\Omega_x\Omega_y + c\,{\Omega_y}^2$$

For the presentation of the characteristic problem it is decisive that the highest order cross wise derivative, $u_{\Omega\Omega}$ in this case, is undetermined. This is given if the characteristic condition $Q = 0$ is satisfied.
With $Q = 0$ and the base curve $\Omega = const$.

$$\begin{aligned} Q &= a\,{\Omega_x}^2 + 2\,b\,\Omega_x\Omega_y + c\,{\Omega_y}^2 = 0 \\ d\Omega &= \Omega_x dx + \Omega_y dy = 0 \end{aligned}$$

one obtains the characteristic polynomial

$$a\left(\frac{dy}{dx}\right)^2 - 2b\,\frac{dy}{dx} + c = 0$$

The characteristics are found to be square roots of this polynomial:

$$\left.\frac{dy}{dx}\right|_{1,2} = \frac{1}{a}\left(b \pm \sqrt{b^2 - ac}\right)$$

It is decisive for the solution behavior if the characteristics are real or complex. This is determined by the sign of the discriminant Δ:

$$\Delta = b^2 - a \cdot c$$

Depending on the sign of the determinant the type of the second order partial differential equation is classified as hyperbolic, parabolic or elliptic (in analogy to the cone intersection equation $a\,y^2 - 2b\,xy + c\,x^2 = 0$).

- hyperbolic $\quad \Delta > 0 \quad \left.\frac{dy}{dx}\right|_1 \neq \left.\frac{dy}{dx}\right|_2$ real characteristics
- parabolic $\quad \Delta = 0 \quad \left.\frac{dy}{dx}\right|_1 = \left.\frac{dy}{dx}\right|_2$ real double characteristic
- elliptic $\quad \Delta < 0 \quad \left.\frac{dy}{dx}\right|_1 \neq \left.\frac{dy}{dx}\right|_2$ complex characteristics

For systems of first order partial differential equations (e.g. Euler equations) the same classification is used as for the equations of second order.

Canonical or normal form of second order equations

Similar to the derivation of the characteristic solution of first order equations, second order partial differential equations can also be transformed into a typical (normal form) form. In the normal form, given in characteristic coordinates (or combinations) the highest order derivatives are free of coefficients.
For the equation

$$a\,u_{xx} + 2\,b\,u_{xy} + c\,u_{yy} + F\,(u_x, u_y, u, x, y) = 0$$

with the characteristics

$$\frac{dy}{dx} = \frac{b}{a} \pm \frac{1}{a}\sqrt{b^2 - ac}$$

and the abbreviations: $\alpha = \frac{b}{a}$ and $\beta = \frac{1}{a}\sqrt{|b^2 - ac|}$ new coordinates can be introduced

$$\begin{aligned} d\xi_1 &= \alpha\,dx - dy \\ d\eta_1 &= \beta\,dx \end{aligned}$$

The transformation of the equation yields the following normal forms:

$$\begin{aligned} \frac{\partial^2 u}{\partial {\xi_1}^2} - \frac{\partial^2 u}{\partial {\eta_1}^2} + \cdots &= 0 \quad \text{hyperbolic PDE} \\ \frac{\partial^2 u}{\partial {\xi_1}^2} + \frac{\partial^2 u}{\partial {\eta_1}^2} + \cdots &= 0 \quad \text{elliptic PDE} \\ \frac{\partial^2 u}{\partial {\eta_1}^2} + \cdots &= 0 \quad \text{parabolic PDE} \end{aligned}$$

A further normal form exists for hyperbolic equations, it has the characteristic coordinates

$$\begin{aligned} d\xi &= d\xi_1 + d\eta_1 = \left.\frac{dy}{dx}\right|_1 dx - dy \\ d\eta &= d\xi_1 - d\eta_1 = \left.\frac{dy}{dx}\right|_2 dx - dy \end{aligned}$$

This normal form is:

$$\frac{\partial^2 u}{\partial \xi \, \partial \eta} \quad + \quad \cdots\cdots = 0$$

Many equations in fluid mechanics, especially in Cartesian coordinates, occur in their normal form and are to be classified in comparison to the normal forms.

1.2.3 Simplified calculation of the characteristics

From the detailed derivation it can be concluded that the condition $Q = 0$ is sufficient for the determination of the characteristics. Q is obtained from the transformation of the leading Ω derivative. Therefore, in a simplified calculation a variable $u(x, y)$ can be considered as a function of Ω only, i.e. $u(x, y) = u(\Omega(x, y))$. The derivatives, e.g. for x are replaced by $u_x = \Omega_x \cdot u_\Omega \quad , \quad u_{xx} = (\Omega_x)^2 \cdot u_{\Omega\Omega} + \cdots$

The characteristics are then obtained from the characteristic condition $Q = 0$ and from $\Omega = const.$

1. example: First order equation

$$\underline{a\,u_x + b\,u_y = c}$$

$$(a\,\Omega_x + b\,\Omega_y)\,u_\Omega = c$$

$$Q = a\,\Omega_x + b\,\Omega_y = 0$$

$$d\Omega = \Omega_x\,dx + \Omega_y\,dy = 0$$

$$\rightarrow \quad \underline{\left.\frac{dy}{dx}\right|_1 = \frac{b}{a}} \quad \rightarrow \quad \text{hyperbolic}$$

Second example: System of equations (Cauchy-Riemannsche equations)

$$\left.\begin{array}{l} \underline{u_x + v_y = 0} \\ \underline{u_y - v_x = 0} \end{array}\right\} \qquad \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & -\frac{\partial}{\partial x} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0$$

$$Q = \det \begin{vmatrix} \Omega_x & \Omega_y \\ \Omega_y & -\Omega_x \end{vmatrix} = -{\Omega_x}^2 - {\Omega_y}^2 = 0 \quad \rightarrow \quad \left.\frac{\Omega_x}{\Omega_y}\right|_{1,2} = \pm\sqrt{-1} = \pm I$$

$$\rightarrow \quad \underline{\left.\frac{dy}{dx}\right|_{1,2} = \pm I} \quad \rightarrow \quad \text{elliptic}$$

Third example: Wave equation

$$\underline{u_{tt} - {a_o}^2\,u_{xx} = 0}$$

1. Path: $Q = {\Omega_t}^2 - {a_o}^2\,{\Omega_x}^2 = 0 \qquad \rightarrow \quad \left.\frac{\Omega_t}{\Omega_x}\right|_{1,2} = \pm\,a_o$

$$\underline{\left.\frac{dx}{dt}\right|_{1,2} = \pm\,a_o} \qquad \rightarrow \text{hyperbolic}$$

2. Path: Substitution $q = u_t \quad , \quad p = u_x \quad \rightarrow \quad q_x = p_t$

$$\begin{aligned} q_t - {a_o}^2\,p_x &= 0 \\ q_x - p_t &= 0 \end{aligned}$$

$$Q = \det \begin{vmatrix} \Omega_t & -{a_o}^2\,\Omega_x \\ \Omega_x & -\Omega_t \end{vmatrix} = -{\Omega_t}^2 + {a_o}^2{\Omega_x}^2 = 0 \quad \rightarrow \quad \left.\frac{\Omega_t}{\Omega_x}\right|_{1,2} = \pm\,a_o$$

1.3 Basics of numerical solutions

In order to solve partial differential equations, the domain of integration is subdivided in a grid of discrete points in the space of independent variables (physical space, time). At these discrete points the geometrical coordinates and the dependent variables (conserved quantities) are defined. For each grid point the differential equations are approximated by difference equations. These equations and the discretized boundary and initial value conditions yield a system of coupled, algebraic equations that can be solved on a computer.

The numerical solution is inaccurate because of the difference building. It is the aim of the numerical solution to approach the exact solution of the differential problem, i.e. the numerical solution shall converge. A solution is said to be convergent if with decreasing step size the numerical solution turns into the exact solution. Some necessary prerequisites must be granted to obtain convergent solutions, namely consistency and numerical stability of the difference scheme. The foundations for these topics will be laid in this chapter.

1.3.1 Development of consistent difference expressions

In the numerical solution of partial differential equations information is only present at the discrete points. The difference expressions which approximate the differentials at a given point are functions of the surrounding neighbor values.
The development of difference expressions for the dependent variable is performed with a Taylor series expansion around the discrete point. An important prerequisite for this is:

All dependent variables can be locally expanded in a series, i.e. their course is continuous and differentiable to a suitable degree.

In the series expansion the differential is replaced by a difference approximation plus a truncation error τ representing the unconsidered terms of the series.

$$\frac{\partial f}{\partial x} \quad \rightarrow \quad \frac{\Delta f}{\Delta x} + \tau$$

The truncation error is the difference between the differential form and the corresponding difference approximation.

$$\tau \equiv \frac{\partial f}{\partial x} - \frac{\Delta f}{\Delta x}$$

It is an important quantity for the determination of consistency, the accuracy and the solution properties of difference approximations.
Difference approximations are called consistent if they approach the differential for decreasing step sizes, $\Delta x \rightarrow 0$. In this case the truncation error disappears.

$$\lim_{\Delta x \rightarrow 0} \tau = \lim_{\Delta x \rightarrow 0} \left(\frac{\partial f}{\partial x} - \frac{\Delta f}{\Delta x} \right) = 0$$

The truncation error includes the unconsidered higher order derivatives multiplied by powers of the step sizes. For a suitable normalization the derivatives are of order $O(1)$ and the

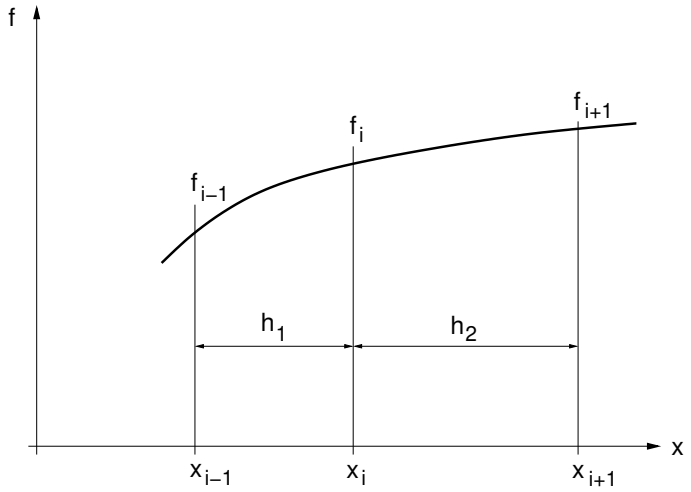
step sizes Δx are much smaller than unity $\Delta x \ll 1$. Therefore, the order of τ and with it the discretization error are governed by the smallest power of the step size, i.e.

$$\tau = O(\Delta x^k) \qquad k > 0$$

Different choices of the considered base points and terms of the series expansion have an impact on the truncation error and therefore on the solution behavior of a difference approximation. As a consequence the transition from differential equation to difference equation is not unique. Several difference approximations exist for a given differential expression. This will be demonstrated for the important first and second order derivatives.

Difference expressions for first and second order derivatives

The difference expressions for $\frac{\partial f}{\partial x}$ or $\frac{\partial^2 f}{\partial x^2}$ of a variable $f(x)$, given at the discrete points x_i shall be constructed ($f(x_i) = f_i$). The maximum number of base points is set to three. The step sizes $h_i = x_i - x_{i-1}$ may be variable.



- Taylor series expansion of $f_{i\pm1}$ around f_i :

(1) $$f_{i+1} = f_i + \left.\frac{\partial f}{\partial x}\right|_i \cdot h_2 + \left.\frac{\partial^2 f}{\partial x^2}\right|_i \cdot \frac{{h_2}^2}{2!} + \left.\frac{\partial^3 f}{\partial x^3}\right|_i \cdot \frac{{h_2}^3}{3!} + \left.\frac{\partial^4 f}{\partial x^4}\right|_i \cdot \frac{{h_2}^4}{4!} + \cdots$$

(2) $$f_{i-1} = f_i - \left.\frac{\partial f}{\partial x}\right|_i \cdot h_1 + \left.\frac{\partial^2 f}{\partial x^2}\right|_i \cdot \frac{{h_1}^2}{2!} - \left.\frac{\partial^3 f}{\partial x^3}\right|_i \cdot \frac{{h_1}^3}{3!} + \left.\frac{\partial^4 f}{\partial x^4}\right|_i \cdot \frac{{h_1}^4}{4!} + \cdots$$

- Approximations for $\frac{\partial f}{\partial x}$ (by combinations of (1) and (2)) :

a) Forward difference

$$\left.\frac{\partial f}{\partial x}\right|_i = \frac{f_{i+1} - f_i}{h_2} + \left(-\frac{\partial^2 f}{\partial x^2}\frac{h_2}{2!} - \frac{\partial^3 f}{\partial x^3}\frac{{h_2}^2}{3!} + \cdots\right) \qquad \tau = O(h_2)$$

b) Backward difference

$$\left.\frac{\partial f}{\partial x}\right|_i = \frac{f_i - f_{i-1}}{h_1} + \left(\frac{\partial^2 f}{\partial x^2}\frac{h_1}{2} - \frac{\partial^3 f}{\partial x^3}\frac{{h_1}^2}{3!} + \cdots\right) \qquad \tau = O(h_1)$$

c) Central difference

$$\left.\frac{\partial f}{\partial x}\right|_i = \frac{f_{i+1}-f_{i-1}}{h_1+h_2} + \left(\frac{\partial^2 f}{\partial x^2}\,\frac{h_1-h_2}{2} - \frac{1}{6}\,\frac{\partial^3 f}{\partial x^3}\,\frac{{h_1}^3+{h_2}^3}{h_1+h_2} + \cdots\right)\tau = O\left(h_1-h_2\right)$$

d) Central difference ($\frac{\partial^2 f}{\partial x^2}$ eliminates, more accurate than c) for $h_2 \neq h_1$)

$$\left.\frac{\partial f}{\partial x}\right|_i = \frac{h_1^2\left(f_{i+1}-f_i\right)+h_2^2\left(f_i-f_{i-1}\right)}{h_1\,h_2\,(h_1+h_2)} + \left(-\frac{1}{6}\,h_1\,h_2\,\frac{\partial^3 f}{\partial x^3} + \cdots\right)\ \tau = O\left(h_1\cdot h_2\right)$$

e) Central difference for constant step sizes $h = h_1 = h_2$

$$\left.\frac{\partial f}{\partial x}\right|_i = \frac{f_{i+1}-f_{i-1}}{2\,h} - \left(-\frac{1}{6}\,h^2\,\frac{\partial^3 f}{\partial x^3} + \cdots\right) \qquad \tau = O\left(h^2\right)$$

- Approximations for $\frac{\partial^2 f}{\partial x^2}$

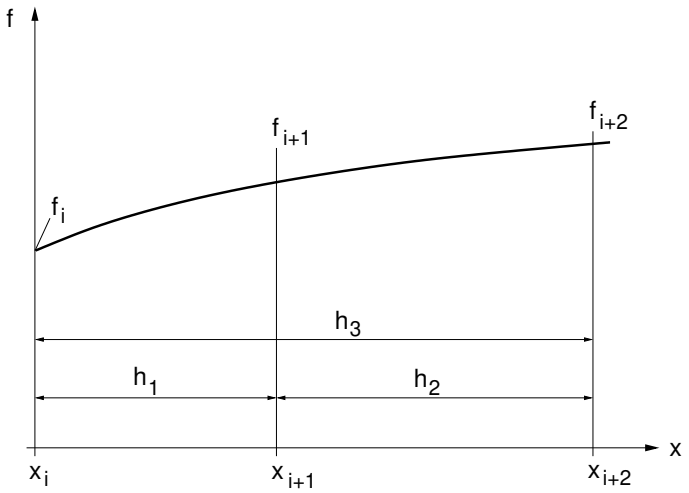
a) Central difference

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_i = \frac{2[(f_{i+1}-f_i)\,h_1-(f_i-f_{i-1})\,h_2]}{h_1\,h_2\,(h_1+h_2)} + \left(\frac{h_1-h_2}{3}\cdot\frac{\partial^3 f}{\partial x^3}+\cdots\right)\tau = O(h_1-h_2)$$

b) Central difference for constant step sizes $h = h_1 = h_2$

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_i = \frac{f_{i+1}-2\,f_i+f_{i-1}}{h^2} - \left(\frac{h^2}{12}\cdot\frac{\partial^4 f}{\partial x^4} + \cdots\right) \qquad \tau = O\left(h^2\right)$$

- Single sided (3 point) approximations for $\frac{\partial f}{\partial x}$ and $\frac{\partial^2 f}{\partial x^2}$ with $h_3 = h_1 + h_2$



Taylor series expansion of f_{i+1}, f_{i+2} around f_i

$$f_{i+1} = f_i + \left.\frac{\partial f}{\partial x}\right|_i\cdot h_1 + \left.\frac{\partial^2 f}{\partial x^2}\right|_i\cdot\frac{{h_1}^2}{2!} + \left.\frac{\partial^3 f}{\partial x^3}\right|_i\cdot\frac{{h_1}^3}{3!} + \left.\frac{\partial^4 f}{\partial x^4}\right|_i\cdot\frac{{h_1}^4}{4!} + \cdots$$

$$f_{i+2} = f_i + \left.\frac{\partial f}{\partial x}\right|_i\cdot h_3 + \left.\frac{\partial^2 f}{\partial x^2}\right|_i\cdot\frac{{h_3}^2}{2!} + \left.\frac{\partial^3 f}{\partial x^3}\right|_i\cdot\frac{{h_3}^3}{3!} + \left.\frac{\partial^4 f}{\partial x^4}\right|_i\cdot\frac{{h_3}^4}{4!} + \cdots$$

a) one sided difference for $\frac{\partial f}{\partial x}$

$$\left.\frac{\partial f}{\partial x}\right|_i = -\frac{\left(f_{i+2}-f_i\right){h_2}^2-\left(f_{i+1}-f_i\right){h_3}^2}{h_2\,h_3\,(h_3-h_2)} + \left(\frac{h_2\,h_3}{6}\cdot\frac{\partial^3 f}{\partial x^3} + \cdots\right)\tau = O\left(h_2 h_3\right)$$

Constant step sizes : $h = h_2 = h_3 / 2$

$$\left.\frac{\partial f}{\partial x}\right|_i = -\frac{f_{i+2} + 3\,f_i - 4\,f_{i+1}}{2\,h} + \left(\frac{h^2}{3} \cdot \frac{\partial^3 f}{\partial x^3} + \cdots\right) \qquad \tau = O\left(h^2\right)$$

b) one sided difference for $\frac{\partial^2 f}{\partial x^2}$

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_i = \frac{2\cdot[(f_{i+2} - f_i)\,h_2 - (f_{i+1} - f_i)\,h_3\,]}{h_2\,h_3\,(h_3 - h_2)} + \left(-\frac{h_3 + h_2}{3} \cdot \frac{\partial^3 f}{\partial x^3} + \cdots\right) \tau = O\left(h_2 + h_3\right)$$

Constant step sizes : $h = h_2 = h_3 / 2$

$$\left.\frac{\partial^2 f}{\partial x^2}\right|_i = \frac{f_{i+2} - 2\,f_{i+1} + f_i}{h^2} + \left(-h \cdot \frac{\partial^3 f}{\partial x^3} + \cdots\right) \qquad \tau = O\left(h\right)$$

Difference schemes

The numerical solution of a differential equation is obtained by replacing the differentials with differences. For each grid point a difference equation is obtained which is generally equal for each point and therefore called difference scheme.
The difference equations of all grid points form a system of algebraic equations that is to be solved. According to the different choices of differences for a derivative, there can be different copulations of the unknowns between the grid points. One generally distinguishes between implicit and explicit difference schemes.

Explicit difference schemes result in solution methods in which the unknown at a grid point is directly determined from known values, since the neighboring points aren't linked (the solution matrix is the identity matrix). The advantage is the straightforward and thus fast solution of the equation system. The disadvantage is a limitation of the step size, caused by numerical instability.

In implicit schemes the unknowns at the neighboring points are coupled (structure of the solution matrix is banded). As a consequence of the coupling a limitation of the step size for stability reasons is generally unnecessary. On the other hand the solution takes a lot more effort since it is performed recursively.

Examples for an implicit and an explicit scheme:

Discretization of the parabolic Fourier equation

$$\frac{\partial u}{\partial t} = \nu\,\frac{\partial^2 u}{\partial x^2} \qquad\qquad \nu = const. > 0$$

Discrete variable:

$$t_n = n \cdot \Delta t \qquad x_i = i \cdot \Delta x \qquad u_i^n = u\left(x_i,\, t_n\right)$$

- Explicit scheme: Time derivative with a forward difference and space derivative with central difference at the point t_n, x_i.

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \nu\,\frac{u_{i+1}^n - 2\,u_i^n + u_{i-1}^n}{\Delta x^2} + O\left(\Delta t, \Delta x^2\right)$$

→ Rearrangement for the unknown u_i^{n+1}

$$\underline{u_i^{n+1}} = u_i^n + \sigma(u_{i+1}^n - 2\,u_i^n + u_{i-1}^n) \qquad \text{with} \quad \sigma = \nu\,\frac{\Delta t}{\Delta x^2}$$

- Implicit scheme: Time derivative with a backward difference and space derivative with central difference at the point t_{n+1}, x_i.

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \nu \frac{u_{i+1}^{n+1} - 2\,u_i^{n+1} + u_{i-1}^{n+1}}{\Delta x^2} + O\left(\Delta t, \Delta x^2\right)$$

$\rightarrow$ Rearrangement for the unknowns $\quad u_{i-1}^{n+1}, \quad u_i^{n+1}, \quad u_{i+1}^{n+1}$

$$\underline{-\sigma\, u_{i-1}^{n+1} + (1 + 2\,\sigma)\, u_i^{n+1} - \sigma\, u_{i+1}^{n+1}} \quad = \quad u_i^n$$

$\Longrightarrow$ Coupled tridiagonal equation system

$\Longrightarrow$ Solution with Gaussian elimination

Solution of a tridiagonal equation system

The solution of tridiagonal equation systems, e.g. resulting from an implicit difference scheme, is performed by Gaussian elimination. The method, especially for tridiagonal systems is also known as an LU decomposition, Thomas or Richtmeyer algorithm.

The tridiagonal equation system for the unknown u_i is:

$$a_i u_{i-1} + b_i u_i + c_i u_{i+1} = R_i \qquad i = 2, \ldots, im-1 \tag{1.1}$$

The boundary conditions for $i = 1$ and $i = im$ are of general form of third type, which includes Dirichlet and gradient boundary conditions.

$$\alpha u + \beta \frac{\partial u}{\partial x} = \gamma$$

The discretization of this boundary conditions, e.g. for $i = 1$ yields $\alpha u_1 + \beta \frac{u_2 - u_1}{\Delta x} = \gamma$ and thus boundary values for $i = 1$ and $i = im$ as

$$u_1 = r_1 u_2 + s_1 \qquad \text{and} \qquad u_{im} = r_{im} u_{im-1} + s_{im}$$

Using a recursion approach equation (1) is reduced to a bidiagonal system (in this case an upper triangular matrix).

$$u_i = E_i u_{i+1} + F_i \tag{1.2}$$

The recursion coefficients E_i and F_i are obtained by substituting equation (2) in the original equation (1) and eliminating u_{i-1}.

$$E_i = \frac{-c_i}{a_i E_{i-1} + b_i} \qquad \text{and} \qquad F_i = \frac{R_i - a_i F_{i-1}}{a_i E_{i-1} + b_i} \tag{1.3}$$

The solution scheme requires at first the calculation of the recursion coefficients for $i = 2 \cdots im$ according to equation (3). The starting values E_1 and F_1 are obtained from the recursion and the boundary condition for $i = 1$.

$$\begin{aligned} u_1 &= E_1\, u_2 + F_1 \\ u_1 &= r_1\, u_2 + s_1 \\ \rightarrow \quad & E_1 = r_1 \qquad F_1 = s_1 \end{aligned}$$

After the calculation of all recursion coefficients E_i , F_i for $i = 2, \cdots im$, the solution of u_i is determined from equation (2). This requires the boundary value u_{im} which is obtained from the recursion approach and the boundary condition for $i = im$.

$$\begin{aligned} u_{im-1} &= E_{im-1}\, u_{im} + F_{im-1} \\ u_{im} &= r_{im}\, u_{im-1} + s_{im} \\ \rightarrow \quad u_{im} &= \tfrac{-r_{im}\, F_{im-1} - s_{im}}{r_{im}\, E_{im-1} - 1} \end{aligned}$$

The final solution is calculated with the recursion approach in equation (2):

$$u_i = E_i\, u_{i+1} + F_i \qquad i = im-1, im-2, \cdots, 1$$

Consistency of difference schemes

The consistency for a difference scheme must be proved as for the single differences. The difference formulation of a partial differential equation must approach the differential equation in the limit $\Delta x,\, \Delta t \to 0$, since only then the numerical solution can approach the analytical solution.

A difference scheme is called consistent with a partial differential equation if for decreasing step sizes $\Delta x,\, \Delta t \to 0$ the difference scheme approaches the partial differential equation, i.e. if the truncation error becomes zero.

For a difference scheme $L_\Delta(u) = 0$ of a partial differential equation $L(u) = 0$ with the exact solution u consistency is granted, if

.
$$\boxed{\lim_{\Delta t, \Delta x \to 0} \left\| L(u) - L_\Delta(u) \right\| = \lim_{\Delta t, \Delta x \to 0} \left\| \tau(u) \right\| = 0}$$

Example : Proof of consistency for the explicit scheme of the Fourier equation

$$\text{PDGl.:} \quad L(u) = \frac{\partial u}{\partial t} - \nu \frac{\partial^2 u}{\partial x^2} = 0$$

$$\text{FDGl.:} \quad L_\Delta(u) = \frac{u_i^{n+1} - u_i^n}{\Delta t} - \nu \frac{u_{i+1}^n - 2\,u_i^n + u_{i-1}^n}{\Delta x^2} = 0$$

Taylor series expansion :

$$u^{n+1} = u^n + \frac{\partial u}{\partial t}\bigg|^n \Delta t + \frac{\partial^2 u}{\partial t^2}\bigg|^n \frac{\Delta t^2}{2} + \frac{\partial^3 u}{\partial t^3}\bigg|^n \frac{\Delta t^3}{6} + \frac{\partial^4 u}{\partial t^4}\bigg|^n \frac{\Delta t^4}{24} + \cdots$$

$$u_{i\pm1} = u_i \pm \frac{\partial u}{\partial x}\bigg|_i \Delta x + \frac{\partial^2 u}{\partial x^2}\bigg|_i \frac{\Delta x^2}{2} \pm \frac{\partial^3 u}{\partial x^3}\bigg|_i \frac{\Delta x^3}{6} + \frac{\partial^4 u}{\partial x^4}\bigg|_i \frac{\Delta x^4}{24} \pm \cdots$$

in FDE:

$$L_\Delta(u) = \frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial t^2}\frac{\Delta t}{2} + O(\Delta t^2) - \nu \left[\frac{\partial^2 u}{\partial x^2} + \frac{1}{12}\Delta x^2 \frac{\partial^4 u}{\partial x^4} + O(\Delta x^4)\right]$$

Truncation error :

$$\tau(u) = L(u) - L_\Delta(u) = -\frac{\partial^2 u}{\partial t^2}\frac{\Delta t}{2} + \nu \frac{\Delta x^2}{12} \cdot \frac{\partial^4 u}{\partial x^4} + O(\Delta t^2, \Delta x^4)$$

Consistency :

$$\lim_{\Delta t, \Delta x \to 0} \tau(u) = 0$$

1.3.2 Numerical stability

Introduction

Numerical stability, resp. instability is a quality of a difference scheme describing the correspondence of the difference solution to small disturbances. Such disturbances are caused by external errors, like e.g. round off errors resulting from the limited numerical precision of a computer. Whether or not these errors stay limited from time step to time step the difference schemes are called stable or unstable.

The difference solution of a differential problem can get unstable, in spite of the stability of the corresponding analytical solution. This originates from the approximation which in principle exactly solves a differential equation changed by the truncation error. For the differential problem $L(\hat{u}) = 0$ and the difference problem $L_\Delta(\hat{u}) = 0$ the modified differential equation is:

$$L_\Delta(\hat{u}) = L(\hat{u}) - \tau = 0$$

The higher order derivatives in the truncation error influence the solution qualities and the included step sizes cause a strong dependency of the stability on the step size (see also stability analysis by Hirt). Therefore, it is important to determine the stability region of a difference scheme.

A simplified stability criterion can be derived as follows:

The exact difference solution at time $t = n\Delta t$ of the difference solution $L_\Delta U = 0$ be U^n and W^n be the current solution (with round off errors) of this equation. The maximum error of the difference $|\,\varepsilon^n\,| = |\,W^n - U^n\,|$ between two time steps $t = n\,\Delta t$ and $t + \Delta t = (n+1)\,\Delta t$ is proportional to a positive constant k

$$\max|\,\varepsilon^{n+1}\,| = k \cdot \max|\,\varepsilon^n\,| \qquad\qquad k > 0$$

The maximum error developing after n steps from the initial error ε^0 is

$$\max|\,\varepsilon^{n+1}\,| = k^n \cdot \max|\,\varepsilon^0\,|$$

A scheme is stable, if the error remains limited for $n \to \infty$, i.e.

$$k \leq 1$$

resp.

$$\max|\,\varepsilon^{n+1}\,| \leq \max|\,\varepsilon^n\,|$$

To investigate the stability of a difference scheme a certain error distribution is inserted in the scheme and checked for the criterion. For a single perturbation the "discrete perturbation theory" is obtained, while for a periodic perturbation the "von Neumann stability analysis" is obtained.

Discrete error perturbation theory

The discrete perturbation theory is an empirical method for the investigation of stability. In this theory a single disturbance ε is defined as initial condition at a grid point. This perturbation overlays an exact solution U. The current solution then becomes:

$$W = U + \varepsilon$$

The time and space dependent behavior of the perturbation is obtained from the solution of the difference equation for subsequent time steps. The scheme becomes unstable, if the maximum modulus of the perturbation increases.
For linear equations the perturbation ε (error) satisfies the same difference equation as the solution of U itself, since

$$L_\Delta \cdot W = \underbrace{L_\Delta \cdot U}_{0} + L_\Delta \cdot \varepsilon = 0 \Rightarrow \quad L_\Delta \cdot \varepsilon = 0$$

Therefore, for linear equations the course of the perturbation can directly be calculated from the difference equation. This is demonstrated with an example:

Example: An explicit scheme for Fourier's equation $u_t = \nu\, u_{xx}$ is considered:

$$\varepsilon_i^{n+1} = \sigma\, \varepsilon_{i-1}^n + (1-2\sigma)\, \varepsilon_i^n + \sigma\, \varepsilon_{i+1}^n \qquad \text{mit} \quad \sigma = \nu\, \frac{\Delta t}{\Delta x^2}$$

The analysis of the error behaviour yields the following results:

- Initial condition $n = 0$
  $\varepsilon_i^0 = \varepsilon$ für $i = is$ $\qquad \varepsilon_i^0 = 0$ für $i \neq is$
- Time step $n = 1$
  $\varepsilon_{is}^1 = (1 - 2\sigma)\, \varepsilon \quad ; \qquad \varepsilon_{is\pm1}^1 = \sigma\, \varepsilon$
  from $\dfrac{\max|\, \varepsilon^1 \,|}{\max|\, \varepsilon^0 \,|} \leq 1$ folgt $|\, \sigma \,| \leq 1$ bzw. $|\, 1 - 2\sigma \,| \leq 1$
  $\rightarrow \quad 0 < \sigma \leq 1$
- Time step $n = 2$
  $\varepsilon_{is}^2 = (1 - 4\sigma + 6\sigma^2)\, \varepsilon \quad ; \quad \varepsilon_{is\pm1}^2 = (2\sigma - 4\sigma^2)\, \varepsilon \quad ; \quad \varepsilon_{is\pm2}^2 = \sigma^2\, \varepsilon$
  $\rightarrow \quad 0 < \sigma \leq 2/3$
- Time step $n \quad (n \rightarrow \infty)$
  $\rightarrow \quad \underline{0 < \sigma \leq 1/2}$ asymptotical stability limit

The behaviour of the perturbation for a stable solution with $\sigma = \frac{1}{2}$ and for an unstable solution with $\sigma = 1$ is presented in the following tables.

• $\underline{\sigma = 1/2}$ Error $\max |\varepsilon_i^n|$ decreases → the solution remains stable

	$i\Delta x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$n\Delta t$	n \ i	0	1	2	3	is=4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0	0
0.5	1	0	0	0	0	$\boxed{\varepsilon}$	0	0	0	0	0	0
1.0	2	0	0	0	$\frac{1}{2}\varepsilon$	0	$\frac{1}{2}\varepsilon$	0	0	0	0	0
1.5	3	0	0	$\frac{1}{4}\varepsilon$	0	$\frac{1}{2}\varepsilon$	0	$\frac{1}{4}\varepsilon$	0	0	0	0
2.0	4	0	$\frac{1}{8}\varepsilon$	0	$\frac{3}{8}\varepsilon$	0	$\frac{3}{8}\varepsilon$	0	$\frac{1}{8}\varepsilon$	0	0	0
2.5	5	0	0	$\frac{1}{4}\varepsilon$	0	$\frac{3}{8}\varepsilon$	0	$\frac{1}{4}\varepsilon$	0	$\frac{1}{16}\varepsilon$	0	0
3.0	6	0	$\frac{1}{8}\varepsilon$	0	$\frac{7}{16}\varepsilon$	0	$\frac{7}{16}\varepsilon$	0	$\frac{1}{32}\varepsilon$	0	$\frac{1}{32}\varepsilon$	0

• $\underline{\sigma = 1}$ Error $\max |\varepsilon_i^n|$ increases → the solution becomes unstable

	$i\Delta x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$n\Delta t$	n \ i	0	1	2	3	is=4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	$\boxed{\varepsilon}$	0	0	0	0	0	0
2	2	0	0	0	ε	$-\varepsilon$	ε	0	0	0	0	0
3	3	0	0	ε	-2ε	3ε	-2ε	ε	0	0	0	0
4	4	0	ε	-3ε	6ε	-7ε	6ε	-3ε	ε	0	0	0
5	5	0	-4ε	10ε	-16ε	19ε	-16ε	10ε	-4ε	ε	0	0
6	6	0	14ε	-30ε	45ε	-51ε	45ε	-30ε	15ε	-5ε	ε	0

Conclusions:

- For a large amount of time steps the scheme remains stable, if

$$0 < \sigma \leq 1/2 \qquad (\text{ conditionally stable scheme })$$

- A time step limitation follows from the stability condition

$$\Delta t_{\max} \leq 1/2 \cdot \frac{\Delta x^2}{\nu}$$

- A decrease in the grid step size causes a square decrease of Δt

$$\Delta t \sim \Delta x^2$$

- For realistic applications the perturbation theory is too costly!

von Neumann stability analysis

J.von Neumann , Los Alamos (1944) in: O'Brien et.al. : Journ. Math.Phys.,29 ,1951

- Instead of a single perturbation, an error function that is periodic in physical space and has a time dependent amplitude (Fourier series)is considered. Such an approach alows the analytical stability investigation, one directly obtains the asymptotical behaviour for $n \to \infty$. The periodical error function is inserted in the difference equation and the temporal behaviour of the amplitude is calculated. A growing amplitude for subsequent time steps indicates an unstable scheme, while a shrinking amplitude indicates a stable scheme.

- The essential limitation is:

  The analysis is only valid for linear initial value problems, i.e. the influence of the boundary conditions is ignored.

- In spite of the limitation the von Neumann analysis is the most commonly used stability analysis for initial value problems. It also obtains useful results for non linear problems, if it is applied to the corresponding linearized difference equation (with fixed coefficients).

Derivation of the method
The derivation is performed for a scalar, linear difference equation in (x, t).

- The error function is formulated as a fourier series, with the function for the amplitude V and a periodic function in physical space e^{Ikx}:

$$\varepsilon\,(x,t) \;=\; \sum_{k_{min}}^{k_{max}} V\,(t,k)\cdot e^{Ikx}$$

  with the wave number $\quad k \;=\; \frac{2\pi}{\lambda} \quad$ and $\quad I \;=\; \sqrt{-1}$

  For the dicrete difference problem with $x = i\,\Delta x$ and $t = n\,\Delta t$ the approach yields

$$\varepsilon_i^n \;=\; \sum_{k_{min}}^{k_{max}} V^n\,(k)\cdot e^{Iki\Delta x} \;=\; \sum_{\Theta_{min}}^{\Theta_{max}} V^n\,(\Theta)\cdot e^{I\Theta i}$$

  Where $\Theta \;=\; k\,\Delta x$ is the wave angle. The wave number k_{min} , k_{max} respectively the wave angles Θ_{min} , Θ_{max} are obtained from the minimum and maximum resolvable wavelength $\lambda_{min} \;=\; 2\,\Delta x$ and $\lambda_{max} \;=\; 2\,L$ of the discrete problem (L= integration length). Therefore, the lower and upper value of the wave angle are $\Theta_{min} = 0$ und $\Theta_{max} = \pi$. For this region the stability needs to be investigated:

- This approach is introduced into the difference equation. A general difference scheme for two time levels is:

$$\sum_{j=-l_1}^{l_2} d_j \cdot u\,(x \,+\, j\Delta x,\, t \,+\, \Delta t) \;=\; \sum_{j=-m_1}^{m_2} c_j \cdot u\,(x \,+\, j\Delta x,\, t)$$

An example is the explicit scheme for $u_t = \nu\, u_{xx}$:

$$u_i^{n+1} = \sigma\, u_{i-1}^n + (1-2\sigma)\, u_i^n + \sigma\, u_{i+1}^n \quad \rightarrow \quad \sum_{j=0}^{0} d_j\, u_{i+j}^{n+1} = \sum_{j=-1}^{1} c_j\, u_{i+j}^n$$

- For linear equation systems the variable u can be replaced by the error ε. With the approach for ε_i^n for the general scheme one obtains:

$$\sum_j d_j \cdot \sum_\Theta V^{n+1}(\Theta) \cdot e^{I\Theta(i+j)} = \sum_j c_j \cdot \sum_\Theta V^n(\Theta) \cdot e^{I\Theta(i+j)}$$

After rearrangement of the sums

$$\sum_\Theta V^{n+1}(\Theta) \cdot \sum_j d_j \cdot e^{I\Theta(i+j)} = \sum_\Theta V^n(\Theta) \cdot \sum_j c_j \cdot e^{I\Theta(i+j)}$$

the equation can be satisfied for each wave angle (i.e. $\sum_\Theta$ is gone).

With the above equations the relation between the amplitudes of a wave angle Θ at the old and the new time level becomes:

$$V^{n+1}(\Theta) = \frac{\sum_j c_j \cdot e^{I\Theta(i+j)}}{\sum_j d_j \cdot e^{I\Theta(i+j)}} \cdot V^n(\Theta)$$

The factor of proportionality between the amplitudes V is the amplification factor G.

$$G(\Theta, \Delta t, \Delta x, c_j, d_j) = \frac{\sum_j c_j \cdot e^{I\Theta(i+j)}}{\sum_j d_j \cdot e^{I\Theta(i+j)}}$$

The stability condition that this factor must satisfy can be found by repetitive (n-fold) application until the initial amplitude $V^0(\Theta)$.

$$V^n(\Theta) = [\, G(\Theta \cdots)]^n \cdot V^0(\Theta)$$

To keep the error limited for $n \rightarrow \infty$, the following condition must hold:

$$\boxed{|\, G(\Theta)\,| \le 1 \qquad \text{for} \qquad 0 \le \Theta \le \pi}$$

This inequality is the stability condition for scalar difference schemes.

- For systems the stability analysis can be derived in an analogous fashion. In this case $\vec{V}$ is the vector of amplitudes of the single variables and $\bar{\bar{G}}$ the amplification matrix. This leads to:

$$\vec{V}^{n+1} = \bar{\bar{G}}\, \vec{V}^n$$

The matrix norm must satisfy the following condition:

$$\|\bar{\bar{G}}^n\| \le const. \qquad \text{für} \quad n \rightarrow \infty$$

For the eigenvalues λ_G of $\bar{\bar{G}}$ the stability criterion requires:

$$\underline{|\, \lambda_G\,| \le 1}$$

Application examples of the von Neumann stability analysis

The von Neumann stability analysis is one of the most important tools in the investigation of numerical solution schemes for initial value problems. Therefore, its application is demonstrated with examples, to show how the analysis can be performed in formal steps.

- 1. Example:

  The stability of an explicit scheme for Fouriers 2–D (heat conduction) equation

  $$u_t \;=\; \nu\,(\,u_{xx} \;+\; u_{yy}\,)$$

  shall be investigated. The explicit difference scheme consists of a forward difference in time and central differences in x and y direction: $u_{i,j}^{n+1} \;=\; u_{i,j}^{n} \;+\; \sigma_x\,(u_{i-1,j}^{n} - 2\,u_{i,j}^{n} + u_{i+1,j}^{n}\,) \;+\; \sigma_y\,(u_{i,j-1}^{n} - 2\,u_{i,j}^{n} + u_{i,j+1}^{n}\,)$

  $$\text{mit } \sigma_x = \tfrac{\nu\,\Delta t}{\Delta x^2} \quad \text{und} \quad \sigma_y = \tfrac{\nu\,\Delta t}{\Delta y^2}$$

  A two dimensional Fourier approach for the variable $u_{i,j}^{n}$ is defined as periodic perturbation (= approach for the error $\varepsilon_{i,j}^{n}$). $u_{i,j}^{n} \;=\; V^n \cdot e^{I(k_x \cdot x + k_y \cdot y)} \;=\; V^n \cdot e^{I(k_x i \Delta x + k_y j \Delta y)} \;=\; V^n \cdot e^{I(\Theta_x i + \Theta_y j)}$

  The amplification factor is obtained from $V^{n+1} = G \cdot V^n$:

  $$G \;=\; 1 \;-\; 2\,\sigma_x\,(\,1 \;-\; cos\,\Theta_x\,) \;-\; 2\,\sigma_y\,(\,1 \;-\; cos\,\Theta_y\,)$$

  The stability condition $|\,G\,| \;\leq\; 1$ has to be proven for $0 \;\leq\; \Theta \;\leq\; \pi$. Since G is usually complex it is more convenient to examine the square modulo instead of the modulo, i.e. $|\,G\,|^2 = (Re\,G)^2 + (Im\,G)^2 \;\leq\; 1$. For the given real expression G one obtains: $|G|^2 = [1 \;-\; 2\,\sigma_x\,(\,1 \;-\; cos\,\Theta_x\,) \;-\; 2\,\sigma_y\,(\,1 \;-\; cos\,\Theta_y\,)]^2 \;\leq\; 1$

  The inequality is satisfied for $0 \;\leq\; \Theta \;\leq\; \pi$, if

  $$\rightarrow\; \underline{\sigma_x \;+\; \sigma_y \quad \leq \quad \tfrac{1}{2}}$$

  The examined scheme is therefore conditionally stable. The time step is limited by:

  $$\Delta t \;\leq\; 1\,/\,(2\,\nu(\tfrac{1}{\Delta x^2} + \tfrac{1}{\Delta y^2}))$$

- 2. Example: In the second example the application of the stability analysis on a system of difference equations will be demonstrated. The initial system of partial differential equations in the variables u and v is:

$$\begin{aligned} u_t + a\,v_x &= 0 \\ v_t + b\,u_x &= 0 \end{aligned}$$

The system can be written in short as:

$$\vec{U}_t + A\,\vec{U}_x = 0 \qquad \text{mit} \quad \vec{U} = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$$

The system can expressed with an explicit scheme, using forward differences in time and central differences in space for both variables equally. This yields:

$$\vec{U}_i^{n+1} = \vec{U}_i^n - A\,\tfrac{\Delta t}{2\Delta x}\,(\vec{U}_{i+1}^n - \vec{U}_{i-1}^n)$$

On the vector the one dimensional Fourier approach is applied:

$$\vec{U}_i^n = \vec{V}^n \cdot e^{I\Theta i} \quad \text{with} \quad \vec{V}^n = \begin{pmatrix} V_u \\ V_v \end{pmatrix}^n$$

The amplification matrix $\overline{\overline{G}}$ is obtained from the rearrangement: $\vec{V}^{n+1} = \overline{\overline{G}}\,\vec{V}^n$:

$$\overline{\overline{G}} = \begin{pmatrix} 1 & -a\frac{\Delta t}{\Delta x}\,I\sin\Theta \\ -b\frac{\Delta t}{\Delta x}\,I\sin\Theta & 1 \end{pmatrix}$$

The stability condition requires that the moduli of the eigenvalues of $\overline{\overline{G}}$ are equal or less than one, i.e. $|\lambda(\overline{\overline{G}})| \leq 1$.

The eigenvalues λ are obtained from:

$$|\,\overline{\overline{G}} - \lambda\,\overline{\overline{E}}\,| = (1-\lambda)^2 + ab\,(\tfrac{\Delta t}{\Delta x}\sin\Theta)^2 = 0$$

The square root of this equation yields:

$$\lambda_{1,2} = 1 \pm I\,\sqrt{ab\,(\tfrac{\Delta t}{\Delta x}\sin\Theta)^2}$$

A closer look at the moduli for $0 \leq \Theta \leq \pi$ shows that at least one eigenvalue exceeds the value one for $a \cdot b > 0$.

$$|\,\lambda_{1,2}\,| = 1 + ab\,(\tfrac{\Delta t}{\Delta x}\sin\Theta)^2$$

Thus this scheme is unstable for $a \cdot b > 0$.

- 3. Example:

  In this example the stability of a scheme with three time levels is investigated. Such a scheme is formulated for a hyperbolic transport equation.

  $$u_t + a\,u_x = 0$$

  Application of central differences in time and space around $t = n\,\Delta t$ yields an explicit scheme $O\,(\Delta x^2, \Delta t^2)$ with the time indices $n-1$ n and $n+1$ (Dufort-Frankel scheme).

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

  resp. $\quad u_i^{n+1} = u_i^{n-1} - C\,(\,u_{i+1}^n - u_{i-1}^n\,) \qquad$ with $C = \frac{a\Delta t}{\Delta x}$

  Using the common Fourier approach

  $$u_i^n = V^n \cdot e^{I\Theta i}$$

  and the condition that the amplifiaction factor is valid for all time levels, i.e.

  $$G = \frac{V^{n+1}}{V^n} = \frac{V^n}{V^{n-1}}$$

  a quadratic, complex equation is obtained for G

  $$G^2 + I\,2C sin\Theta \cdot G = 1$$

  with the solutions:

  $$G_{1,2} = -I\,C\,sin\Theta \pm \sqrt{1 - (C\,sin\Theta)^2}$$

  For both solutions the stability condition $\quad |\,G\,| \leq 1$ for $0 \leq \Theta \leq \pi$ must be examined. The following cases must be distinguished:

  a) $(C\,sin\Theta)^2 > 1 \rightarrow \quad C > 1 \qquad$ imaginary solution

  $$\rightarrow |\,G\,|^2 > 1$$

  unstable scheme for $C > 1$

  b) $(C\,sin\Theta)^2 \leq 1 \rightarrow \quad C \leq 1 \qquad$ real solution

  $$|\,G\,|^2 = (C\,sin\Theta)^2 + (1 - (C\,sin\Theta)^2\,)$$

  $$\rightarrow |\,G\,|^2 = 1$$

  conditionally stable scheme for $\quad C \leq 1$

Stability analysis by Hirt

C.W.Hirt : Heuristic Stability Theorie for Finite Difference Equations, J. of Comp.Phys., vol 2, 1968.

The instability of difference solutions originates from the finite series expansion which in principle leads to the solution of a differential equation which is modified by the truncation error. With the differential problem $L(\hat{u}) = 0$ and the difference problem $L_\Delta(\hat{u}) = 0$ the differential equation of the difference approximation becomes: $L_\Delta(\hat{u}) = L(\hat{u}) - \tau = 0$
The truncation error includes the higher order derivatives, multiplied by the powers of the step sizes. This changes the solution properties of the differential equation of the difference approximation, possibly causing the instability of the solution, although the solution of the original problem remains stable.

The idea of the stability analysis by Hirt is to examine the properties of the differential equation of the difference approximation and to compare it to the behaviour and physical interpretation of known analogous equations (e.g. positive viscosity or characteristics). Therefore, the stability analysis accoring to Hirt delivers a very descriptive characterization of the effects of numerical approximations. However, since a comparative solution is often lacking this method can not always be applied!

The principle of the stability analysis according to Hirt, i.e. the Taylor series expansion of a given difference scheme into a differential equation of fixed $\Delta x, \Delta t \cdots$ and its interpretation is demonstrated for two examples.

- 1. Example

  An explicit scheme with forward differences in time and backward differences in space ("upwind" scheme, if $a > 0$) is considered for the hyperbolic transport equation (model equation for inviscid flows):

  $$u_t + a\,u_x = 0 \qquad a = const.$$

  which yields: $\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\,\frac{u_i^n - u_{i-1}^n}{\Delta x} = 0$

  Using a taylor series expansion for u around the point (x_i, t_n) and substitution of the second order time derivative u_{tt} with the differential equaton, i.e. :

  $$u_{tt} = -a\,u_{xt} = -a\,(u_t)_x = a^2\,u_{xx}$$

  one obtains the differential equation of the difference approximation:

  $$u_t + a\,u_x = \left(a\tfrac{\Delta x}{2} - a^2\tfrac{\Delta t}{2}\right) \cdot u_{xx} + O(\Delta x^2, \Delta t^2)$$

  The originally hyperbolic equation becomes a parabolic equation for fixed step sizes! This equation has the same structure as the convection diffusion equation (model equation for the Navier-Stokes equation): $u_t + a\,u_x = \nu \cdot u_{xx}$

  The solution of this equation is for positive viscosity ν always damped, i.e. the perturbations diminish over time.

  In analogy the factor preceeding the second order derivation of the differential equation of the difference approximation is refered to as the numerical viscosity ν_{num}.

  $$\nu_{num} = a\tfrac{\Delta x}{2} - a^2\tfrac{\Delta t}{2} = a\tfrac{\Delta x}{2} \cdot (1 - C) \qquad \text{with} \quad C = a\tfrac{\Delta t}{\Delta x}$$

From this analogy follows a damped solution for positive numerical viscosity, i.e. stable solution and for the opposite case an excitated flow, i.e. an unstable solution.

From this requirement it follows for stability

$$\begin{aligned} a > 0: \quad & \nu_{num} > 0 \quad \text{for} \quad C = a\tfrac{\Delta t}{\Delta x} \leq 1 \rightarrow \quad \text{conditionally stable} \\ a < 0: \quad & \nu_{num} < 0 \quad \text{for} \quad \text{all } C \qquad \rightarrow \quad \text{unstable} \end{aligned}$$

These conditions also follow from the von Neumann stability analysis.

The numerical viscosity is of major importance for the numerical solution of hyperbolic partial differential equations, like e.g. the Euler equations. On the one hand it is necessary for the numerical stability for the damping of small perturbations (round off errors), on the other hand it imposes quasi viscous effects, like a "smeared out" solution. Therefore, it is one main goal to minimize these effects.

- 2. Example

For the parabolic diffusion equation

$$u_t = \nu\, u_{xx}$$

an explicit scheme is considered:

$$\tfrac{u_i^{n+1} - u_i^n}{\Delta t} = \tfrac{\nu}{\Delta x^2}\,(u_{i+1}^n - 2\,u_i^n + u_{i-1}^n)$$

Taylor series expansion leads to the differential equation of the difference approximation:

$$u_t - \nu\, u_{xx} = -\tfrac{\Delta t}{2}\, u_{tt} + O\,(\Delta x^2, \Delta t^2)$$

The rearranged equation yields:

$$u_{xx} - \tfrac{\Delta t}{2\,\nu}\, u_{tt} = \tfrac{1}{\nu}\, u_t + O\,(\Delta x^2, \Delta t^2)$$

The originally parabolic problem therefore numerically becomes a hyperbolic problem for $\Delta t \neq 0$ (see also wave equation $u_{tt} - a_0^2\, u_{xx} = 0$).

As an analogous equation for comparison the solution of a hyperbolic problem of a wave equation can be considered. The influential domain of this equation is fixed by the characteristics

$\frac{dt}{dx}\,|_{1,2} = \pm\sqrt{\frac{\Delta t}{2\,\nu}}$.

It is necessary for the numerical solution of hyperbolic equations that the numerical domain of influence is equal or greater than the influential region of the characteristics, i.e.: $\frac{\Delta x}{\Delta t} \geq \frac{dx}{dt}\,|_C$

With $\frac{\Delta t}{\Delta x} \leq \sqrt{\frac{\Delta t}{2\,\nu}}$ this delivers a stability definition for the explicit scheme:

$$\underline{\sigma = \nu\tfrac{\Delta t}{\Delta x^2} \leq \tfrac{1}{2}}$$

This relation for the conditionally stable scheme is also obtained by other analysis.

1.3.3 Convergence

The aim of numerical calculations is to approach the exact solution of a partial differential equation as far as possible, where the numerical solution approximates the analytical solution with higher accuracy for refined grids. This is known as convergence of the solution. The proof of convergence for a scheme cannot be performed in the most general form. However, for linear initial value problems it is possible to show that the satisfaction of consistency and stability of a difference scheme is a sufficient condition for convergence of a numerical solution. The linear initial value problem

$$L\,(\widehat{u}) \;=\; 0$$

with the exact solution $\widehat{u}$ and the initial condition

$$B\,(\widehat{u}) \;=\; 0$$

be considered. The difference approximation of this problem with the exact difference solution U be

$$L_\Delta\,(U) \;=\; 0 \quad \text{und} \quad B_\Delta\,(U) \;=\; 0$$

The currently computed solution $W \;=\; U + \varepsilon$ with the round off error ε also satisfies the difference approximation.

$$L_\Delta\,(W) \;=\; 0 \quad \text{und} \quad B_\Delta\,(W) \;=\; 0$$

With the definition of the convergence error $e \;=\; \widehat{u} - U$ and the round off error ε on obtains for the computed solution W:

$$W = U + \varepsilon = \widehat{u} + U - \widehat{u} + \varepsilon = \widehat{u} - e + \varepsilon$$

The convergence of the problem can now be defined as follows: A difference solution converges if for every point $P\,(\vec{x},\, t\,)$ the difference solution approaches the exact differential solution, for step sizes Δx, Δt approaching zero. This yields that e and ε must diminish in this case.

The numerical stability describes the behaviour of the round off errors ε. A difference solution is table, if

$$\max|\,\varepsilon^{n+1}\,| \;=\; k^n \cdot \max|\,\varepsilon^0\,|$$

This means that for a stable scheme ($k \leq 1$) the computed solution W approaches the exact difference solution U (apart from a small initial error ε^0 at $k = 1$). Therefore, it is

$$W \;=\; U \;=\; \widehat{u} - e$$

Consistency requires that the truncation error with the step sizes $\Delta x,\, \Delta t \to 0$ diminishes. For linear equations a relation between consistency and truncation error can be established:

$$L\,(\widehat{u}) \,-\, L_\Delta\,(U) \,=\, L\,(\widehat{u}) \,-\, L_\Delta\,(\widehat{u}) \,+\, L_\Delta\,(\widehat{u}) \,-\, L_\Delta\,(U) \,=\, \tau\,(\widehat{u}) \,+\, L_\Delta\,(e) \,=\, 0$$

This also holds for the discrete initial condition.

$$\tau_B\,(\widehat{u}) \,+\, B_\Delta\, e \,=\, 0$$

For a consitent approximation with $\tau \,\to\, 0\,,\, \tau_B \,\to\, 0$ for $\Delta x,\, \Delta t \,\to\, 0$ and a well posed initial value problem follows:

$$L(e) \,=\, 0 \quad \text{und} \quad B(e) \,=\, 0$$

The solution of this problem leads to

$$e \,=\, \widehat{u} - U \,=\, 0$$

This results in convergence.

The proof has been given by P. Lax in a general form, known as "Lax equivalence theorem".

Lax equivalence theorem

P.D.Lax, R.D.Richtmeyer: Comm. on Pure and Appl. Math., vol 9, 1956

> For a consistent difference approximation of a
> well posed linear initial value problem
> the numerical stability is necessary and sufficient condition
> for the convergence of the solution.

For linear initial value problems the Laxian theorem offers the possibility to replace the proof of convergence by the more straightforward proof of stability and consistency. For non linear initial and boundary value problems the general proof of convergence is missing. Therefore, the proof of conergence according to Lax is often applied to a linearized form of a non linear problem, to enable a prognosis on the usability of a scheme. The Laxian equivalence theorem for the proof of convergence is therefore one of the most important tools for the development of difference schemes.

1.4 Solution methods for elliptic partial differential equations

1.4.1 Introduction

An important class of equations in fluid mechanics is of elliptic type, see also section 1.1 and 1.2 of this script. Equations of this type especially describe steady flows, inviscid (subsonic) as well as viscous flows. Examples are the potential equations, the Poisson equation for the stream function or the pressure, but also equation systems such as the Navier-Stokes equations for steady flow. Typical PDEs of elliptic type, occurring in fluid dynamics, are of the following form

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + f\,(\frac{\partial u}{\partial x}\,,\,\frac{\partial u}{\partial y}\,,\,u\,,\,x\,,\,y) \;=\; 0$$

or they constitute equation systems, like the Cauchy-Riemann differential equations:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \;=\; 0 \quad ; \quad \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \;=\; 0$$

The previously discussed basics of the numerical analysis of partial differential equations, such as the stability analysis and proof of convergence, assumed the PDE to describe an initial value problems. Initial value problems result from a bounded domain of influence because of the real slopes of characteristic lines of parabolic and hyperbolic differential equations. The characteristic lines define the direction of information transport in which the solution develops, e.g. in physical time direction. In contrast to this solution behavior elliptic partial differential equations lead to boundary value problems. For these problems the information in the solution field is transported simultaneously and from all directions. Therefore, boundary values influence values in the flow. This different behavior of information transport for initial value and boundary value problems requires different solution schemes. Because of this initial value problems are also referred to as marching direction schemes, while elliptic boundary value problems are known as field schemes.

Corresponding to the information transport the numerical solution of elliptic boundary value problems requires the direct, simultaneous solution for all grid points of the discretized domain of integration. This can be performed by direct inversion methods, like e.g. the Gaussian elimination algorithm. This algorithm is often most suitable for equation systems of moderate size. For larger systems, like those occurring in the numerical solution of flow problems with high grid point numbers, the memory and computing power requirement of direct methods increases over proportionally. Therefore, such problems are often solved applying iteration methods. These methods directly invert a (easier to invert) part of the solution matrix, while the other part is applied to an approximated solution vector. Starting from an initial solution the solution is performed stepwise until a convergent solution is reached which is numerically defined by an convergence criterion. The memory and computing time requirement per iteration step is essentially smaller than for direct methods. But the overall computational cost until convergence is reached also depends on the number of iterations and thus on the convergence properties of the applied iteration scheme. Therefore, this chapter introduces and discusses some of the most important iteration schemes which are demonstrated for the numerical solution of Poisson's equation.

1.4.2 Discretization of Poisson's equation

For the discussion of iteration schemes for boundary value problems the numerical solution of Poisson's equation is considered. Equations of this type occur in the estimation of the stream function and the pressure for incompressible flows. But their solution also includes Laplace's equation for a diminishing right hand side which describes the potential flow.

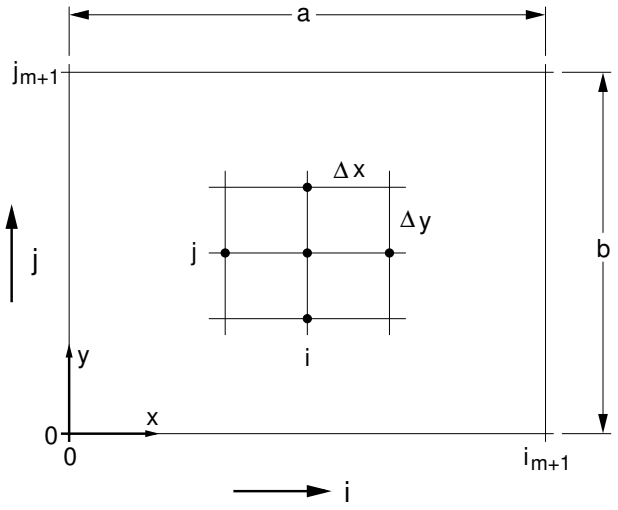
Definition of the boundary value problem

The solution in a rectangular domain of integration D shall be described by Poisson's equation in Cartesian coordinates (x, y).

$$\boxed{\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -f(x, y)}$$

Dirichlet boundary conditions are predefined on the boundary of the domain.

$$u = g(x, y)$$



The discretization of the boundary value problem is performed in a domain with the edge lengths a and b subdivided in $(im+1)$ respectively $(jm+1)$ intervals of constant step size:

$$\Delta x = \frac{a}{im+1} \qquad \Delta y = \frac{b}{jm+1}$$

The second order derivatives of the equation are approximated by central differences of order $O(\Delta x^2)$ and $O(\Delta y^2)$ respectively. The discretized Poisson's equation at a point (i, j) with $1 \le i \le im$ and $1 \le j \le jm$ therefore becomes:

$$\boxed{u_{i,j} - \Theta_x (u_{i-1,j} + u_{i+1,j}) - \Theta_y (u_{i,j-1} + u_{i,j+1}) = \delta^2 f_{i,j}}$$

The following abbreviations have been defined in the above equation:

$$\Theta_x = \frac{\Delta y^2}{2(\Delta x^2 + \Delta y^2)} \qquad \Theta_y = \frac{\Delta x^2}{2(\Delta x^2 + \Delta y^2)} \qquad \delta^2 = \frac{\Delta x^2 \Delta y^2}{2(\Delta x^2 + \Delta y^2)}$$

Together with the boundary conditions a coupled system of $im \cdot jm$ algebraic equations is obtained for which the solution for the discrete points (i, j) must be found.

Matrix - vector presentation

For the discussion of the solution method it is convenient to formulate the system of difference equations in a compact matrix vector notation. To achieve this at first the difference equations for a row $j = const.$ is considered.

$$\begin{array}{lllll}
i=1 & u_{1,j} & -\Theta_x\,(\qquad + \quad u_{2,j}\,) & -\Theta_y\,(\,u_{1,j-1} \;+\; u_{1,j+1}) & = \quad \delta^2\, f_{1,j} \;+\; \Theta_x \cdot u_{0,j} \\
\vdots & & & & \\
i & u_{i,j} & -\Theta_x\,(u_{i-1,j} \;+u_{i+1,j}\,) & -\Theta_y\,(\,u_{i,j-1} \;+\; u_{i,j+1}) & = \quad \delta^2\, f_{i,j} \\
\vdots & & & & \\
i=im & u_{im,j} & -\Theta_x\,(u_{im-1,j}+\qquad\;\;) & -\Theta_y\,(\,u_{im,j-1}+u_{im,j+1}) & = \quad \delta^2\, f_{im,j} \;+\; \Theta_x \cdot u_{im+1,j}
\end{array}$$

The given quantities, $f_{i,j}$ and the boundary values $u_{0,j}\,,\; u_{im+1,j}$ are written to the right hand side. The complete row $j = const.$ with the equations for $1 \le i \le im$ can be written in compact form in a single equation with:

- (im) - dimensional vectors

$$\overrightarrow{U}_j \;=\; \begin{pmatrix} u_{1,j} \\ u_{2,j} \\ \vdots \\ u_{im,j} \end{pmatrix} \qquad \overrightarrow{f}_j \;=\; \begin{pmatrix} f_{1,j} \\ f_{2,j} \\ \vdots \\ f_{im,j} \end{pmatrix} \qquad \overrightarrow{w}_j \;=\; \begin{pmatrix} u_{0,j} \\ 0 \\ \vdots \\ 0 \\ u_{im+1,j} \end{pmatrix}$$

- quadratic matrices of order (im)

$$E_i \;=\; \begin{pmatrix} 1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & 1 \end{pmatrix} \qquad L_i \;=\; \begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ & \cdot & \cdot & & \\ & & & \cdot & 0 \\ & & & 1 & 0 \end{pmatrix} \qquad L_i^T \;=\; \begin{pmatrix} 0 & 1 & & & \\ & 0 & \cdot & & \\ & & \cdot & \cdot & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

Therefore, the system reduces to a column vector with jm elements

$$\begin{array}{lllll}
j=1 & [E_i \;-\; \Theta_x\,(L_i \;+\; L_i^T)]\,\overrightarrow{U}_1 & -\Theta_y\,(\qquad + \quad \overrightarrow{U}_2\,) & = & \delta^2\,(\,\overrightarrow{f}_1 + \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_1 + \tfrac{1}{\Delta y^2}\,\overrightarrow{U}_0) \\
\vdots & & & & \\
j & [E_i \;-\; \Theta_x\,(L_i \;+\; L_i^T)]\,\overrightarrow{U}_j & -\Theta_y\,(\overrightarrow{U}_{j-1} \;+\overrightarrow{U}_{j+1}\,) & = & \delta^2\,(\,\overrightarrow{f}_j + \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_j) \\
\vdots & & & & \\
j=jm & [E_i \;-\; \Theta_x\,(L_i \;+\; L_i^T)]\,\overrightarrow{U}_{jm} & -\Theta_y\,(\overrightarrow{U}_{jm-1}+\qquad\;\;) & = & \delta^2\,(\,\overrightarrow{f}_{jm} + \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_{jm} + \tfrac{1}{\Delta y^2}\,\overrightarrow{U}_{jm+1})
\end{array}$$

The structure of this equations allows a further summation for the complete solution vector $\mathbf{U}$ which includes the (im)-dimensional vectors $\vec{U}_j$ as components.

- compact vectors with (im) - dimensional components

$$\mathbf{U} \;=\; \begin{pmatrix} \overrightarrow{U}_1 \\ \overrightarrow{U}_2 \\ \vdots \\ \overrightarrow{U}_{jm} \end{pmatrix} \qquad \mathbf{F} \;=\; \begin{pmatrix} \overrightarrow{f}_1 \;+\; \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_1 \;+\; \tfrac{1}{\Delta y^2}\,\overrightarrow{U}_0 \\ \overrightarrow{f}_2 \;+\; \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_2 \;+\; 0 \\ \vdots \;+ \\ \overrightarrow{f}_{jm} \;+\; \tfrac{1}{\Delta x^2}\,\overrightarrow{w}_{jm} \;+\; \tfrac{1}{\Delta y^2}\,\overrightarrow{U}_{jm+1} \end{pmatrix}$$

- quadratic matrices of order $(im \cdot jm)$

$$E = \begin{pmatrix} E_i & & & \\ & \cdot & & \\ & & \cdot & \\ & & & E_i \end{pmatrix} \qquad L = \begin{pmatrix} L_i & & & \\ & \cdot & & \\ & & \cdot & \\ & & & L_i \end{pmatrix} \qquad B = \begin{pmatrix} 0 & & & \\ E_i & 0 & & \\ & \cdot & \cdot & \\ & & E_i & 0 \\ & & & E_i \quad 0 \end{pmatrix}$$

The transposed matrix of B is B^T. The matrix L^T corresponds to the Matrix L, but with the elements L_i^T.
With this the complete system of $im \cdot jm$ equations can be written as:

$$\boxed{A \cdot \mathbf{U} \;=\; \delta^2 \, \mathbf{F}}$$

In this equation $\mathbf{U}$ stands for the solution vector, $\mathbf{F}$ represents the vector of known quantities and A is the solution matrix with the components

$$A \;=\; E - \Theta_x \, (L + L^T) - \Theta_y (B + B^T)$$

1.4.3 Principles of iteration schemes

Direct methods invert the complete solution matrix of the difference equation system without iteration. The solution vector is obtained from:

$$\mathbf{U} \;=\; A^{-1} \, \delta^2 \, \mathbf{F}$$

The fundamental algorithm in this method is the Gaussian elimination applied in different variants. For large systems ($im \cdot jm \gg 1$) the requirements for computing time and memory become very high. Additionally, the danger of accumulated round off errors occurs.
Iteration schemes which only perform an inversion of a simplified matrix are more efficient for the solution of large systems. Stable iteration schemes are invulnerable against round off errors as a consequence of the decoupling of the single iteration steps.
To develop an iteration scheme the solution matrix A is split into the components

$$A \;=\; N \;-\; P$$

The sub matrix N has a simplified structure and can be solved directly, while the matrix P is applied to an approximated solution vector. For the system of difference equations $A \cdot \mathbf{U} \;=\; \delta^2 \, \mathbf{F}$ one obtains the recursion, where ν is the iteration counter.

$$N \, \mathbf{U}^{\nu} \;=\; P \, \mathbf{U}^{\nu-1} \;+\; \delta^2 \, \mathbf{F}$$

The correction form of this recursion is build with the difference of the solution vectors $\Delta \mathbf{U}^{\nu} = \mathbf{U}^{\nu} - \mathbf{U}^{\nu-1}$.

$$N \, \Delta \mathbf{U}^{\nu} \;=\; \delta^2 \, \mathbf{F} \;-\; A \, \mathbf{U}^{\nu-1} \;=\; -\, Res(\mathbf{U}^{\nu-1})$$

The equation system $Res(\mathbf{U}^{\nu-1}) \;=\; A \, \mathbf{U}^{\nu-1} - \delta^2 \, \mathbf{F}$ is known as the residual.
The recursion creates a sequence of solution vectors $\mathbf{U}^{\nu}$ starting from an initial vector $\mathbf{U}^{(0)}$.

It is the aim of the iteration to come arbitrarily close to the exact solution, i.e. to reach the convergence of the iteration scheme. In realistic computations a convergence limit is defined for which the iteration is aborted. This limit depends for example on the desired precision. The convergence limit can be fixed in various ways, e.g.:

$$\begin{aligned} max|\,\mathbf{U}^{\nu} - \mathbf{U}^{\nu-1}\,| &\leq \quad \varepsilon_1\, max\,|\mathbf{U}^{\nu-1}| \\ max|\,Res(\mathbf{U}^{\nu})\,| &\leq \quad \varepsilon_2\, max\,|\,Res(\mathbf{U}^{(0)})\,| \end{aligned}$$

1.4.4 Stability and Consistency of iteration schemes

The typical course of an iteration, i.e. the step wise solution beginning with an initial solution, corresponds to the solution of an initial value problem. If the iteration steps are considered as "time steps", iteration scheme can be viewed as initial value problem. Therefore, the formerly discussed consistency and stability investigations can be transferred to an iteration scheme.
The iterative behavior will be presented starting from the iteration rule.

$$N\,\mathbf{U}^{\nu} \;=\; P\,\mathbf{U}^{\nu-1} \;+\; \delta^2\,\mathbf{F}$$

An artificial time is defined $\tau = \nu \cdot \Delta\tau$, where ν is the iteration counter. The solution vector $\mathbf{U}^{\nu}$ is developed by Taylor series expansion for $\nu - 1$:

$$\mathbf{U}^{\nu} \;=\; \mathbf{U}^{\nu-1} + \Delta\tau\,\frac{\partial \mathbf{U}}{\partial \tau}\,\Big|^{\nu-1} \;+ \cdots$$

The quasi time dependent form of the iteration scheme is obtained when the series expansion is inserted in the iteration rule:

$$N\,\Delta\tau\,\frac{\partial \mathbf{U}}{\partial \tau} \;=\; \delta^2\,\mathbf{F} - (N-P)\,\mathbf{U} \;=\; -(A\,\mathbf{U} - \delta^2\,\mathbf{F})$$

The right hand side of the equation represents the discrete Poisson's equation which must be equal zero for the converged ("steady") state ($A \cdot \mathbf{U} \;=\; \delta^2\,\mathbf{F}$).

The investigation of the consistency can be split into the examination of the discrete boundary value problem and the investigation of the iteration scheme. The boundary value problem, i.e. the discretized Poisson's equation at the point (i, j) yields after rearrangement and insertion of the Taylor series in x and y-direction

$$u_{xx} \;+\; u_{yy} \;+\; f \;+\; \frac{\Delta x^2}{12}\,u_{xxxx} \;+\; \frac{\Delta y^2}{12}\,u_{yyyy} \;+\; \cdots \;=\; 0$$

It follows from the equation that the discretization of Poisson's equation is accurate up to order $O\,(\Delta x^2, \Delta y^2)$ and consistent for diminishing step size Δx, Δy. The consistency investigation of the iteration scheme must proof that the converged solution is independent of the iteration scheme. This is the case for the quasi time dependent form of the iteration scheme. The left hand side ($N\,\Delta\tau\,\frac{\partial \mathbf{U}}{\partial \tau}$) diminishes for $\Delta\tau \to 0$ which is equal to $\nu \to \infty$ for a fixed arbitrarily chosen τ. The stability investigation of an iteration scheme can be

performed with analysis for initial value problems. For the examination with the von Neumann stability analysis the variable $u_{i,j}^{\nu}$ is replaced in the iteration scheme by Fourier component

$$u_{i,j}^{\nu} = V^{\nu} \cdot e^{I(k_x x + k_y y)}$$

In the above equation the amplitude V^{ν} stands for the behavior of the perturbation for subsequent time steps. Stability is obtained, when the amplitude doesn't increase, i.e. for

$$|G| \leq 1$$

An example will be postponed to a later section. The convergence of an iteration scheme towards an exact difference solution of the boundary problem can also be estimated by application of the Laxian equivalence theorem. Another quantitative convergence analysis, allowing the comparison of different iteration schemes for Poisson's equation, will be given in a later section.

1.4.5 Presentation of important iteration schemes

Numerous iteration schemes exist for the solution of elliptic equations. They can roughly be divided in so called classic and modern, enhanced schemes.
Classic iteration schemes are e.g.

- Jacobi–iteration
- Gauss-Seidel–point iteration
- Over-relaxed Gauss-Seidel–point iteration
- Gauss-Seidel–line iteration
- Over-relaxed Gauss-Seidel–line iteration
- Alternating line iteration

Modern iteration schemes or concepts are e.g.

- Approximated factorization methods
- Fourier solution methods
- Conjugate gradients
- Multi grid methods

The following section the so called classic methods will be discussed. These are still heavily applied in todays application and often build the foundation for the understanding of enhanced methods.

Jacobi iteration method

The Jacobi method is the most simple iteration scheme. It has a bad rate of convergence. (The rate of convergence, to be defined later, is a measure for the required amount of iteration steps to reduce the initial error to a given limit.) The Jacobi method is often used as a method for comparison, because of its simple structure. Starting from the iteration rule

$$N\,\mathbf{U}^{\nu} \;=\; P\,\mathbf{U}^{\nu-1} \;+\; \delta^2\,\mathbf{F}$$

the most straightforward matrix, the unity matrix E is defined as inversion matrix N for the Jacobi method.

$$N \;=\; E \quad \text{and} \quad P \;=\; N - A \;=\; \Theta_x\,(L \;+\; L^T) \;+\; \Theta_y\,(B \;+\; B^T)$$

Therefore, the matrix formulation of the Jacobi iteration is:

$$E\,\mathbf{U}^{\nu} \;=\; [\,\Theta_x\,(L \;+\; L^T) \;+\; \Theta_y\,(B \;+\; B^T)\,]\,\mathbf{U}^{\nu-1} \;+\; \delta^2\,\mathbf{F}$$

The formulation shows that in the Jacobi iteration the new value $\mathbf{U}^{\nu}$ is calculated from the neighboring old values $\mathbf{U}^{\nu-1}$.

The point wise formulation of the Jacobi method yields an algorithm for the numerical solution steps for $\quad 1 \le i \le im\,;\, 1 \le j \le jm$

$$u^{\nu}_{i,j} \;=\; \Theta_x\,(u^{\nu-1}_{i-1,j} \;+\; u^{\nu-1}_{i+1,j}) \;+\; \Theta_y\,(u^{\nu-1}_{i,j-1} \;+\; u^{\nu-1}_{i,j+1}) \;+\; \delta^2\,f_{i,j}$$

The correction form of the Jacobi method is:

$$\begin{aligned} \Delta\,u^{\nu}_{i,j} \;&=\; -\,Res(u^{\nu-1}_{i,j}) \\ u^{\nu}_{i,j} \;&=\; u^{\nu-1}_{i,j} \;+\; \Delta\,u^{\nu}_{i,j} \end{aligned}$$

with the residual $Res(u^{\nu-1}_{i,j})$ of Poisson's equation

$$Res(u^{\nu-1}_{i,j}) \;=\; u^{\nu-1}_{i,j} \;-\; \Theta_x\,(u^{\nu-1}_{i-1,j} \;+\; u^{\nu-1}_{i+1,j}) \;-\; \Theta_y\,(u^{\nu-1}_{i,j-1} \;+\; u^{\nu-1}_{i,j+1}) \;-\; \delta^2\,f_{i,j}$$

For the stability investigation with the von Neumann analysis the Fourier Ansatz

$$u^{\nu}_{i,j} \;=\; V^{\nu}\,e^{Ik_x x}\cdot e^{Ik_y y} \;=\; V^{\nu}\,e^{I\alpha i}\cdot e^{I\beta j}$$

is introduced to the point wise iteration rule. After rearrangement the amplification factor G is obtained:

$$G \;=\; \Theta_x\,(e^{-I\alpha} + e^{I\alpha}) \;+\; \Theta_y\,(e^{-I\beta} + e^{I\beta}) \;=\; 2\Theta_x\cdot\cos\alpha \;+\; 2\Theta_y\cdot\cos\beta$$

An estimation of the modulo for $0 \le \alpha \le \pi\,;\, 0 \le \beta \le \pi$ yields

$$|G| \;\le\; 1$$

Therefore, the Jacobi method is unconditionally stable.

Gauss Seidel point iteration method

The Gauss Seidel point iteration method also belongs to the more simple iteration schemes. But its convergence rate is two times better compared to the Jacobi method. The method uses the updated values from the neighboring points, as soon as they are available. Because of this the method becomes direction dependent, i.e. it depends on the order of the single steps. If the grid points are updated starting e.g. from $i = 1$ until $i = im$ and from $j = 1$ to $j = jm$ the values for the point (i, j) are to the left, $(i-1, j)$, and below, $(i, j-1)$, they are already on the new iteration level ν. With this the iteration matrices N and P become

$$\begin{aligned} N &= E - \Theta_x L - \Theta_y B \\ P &= N - A = \Theta_x L^T + \Theta_y B^T \end{aligned}$$

The matrix formulation of the Gauss Seidel point iteration scheme is therefore:

$$(E - \Theta_x L - \Theta_y B) \mathbf{U}^{\nu} = (\Theta_x L^T + \Theta_y B^T) \mathbf{U}^{\nu-1} + \delta^2 \mathbf{F}$$

The point wise formulation of the Gauss Seidel method yields the algorithm for the numerical solution steps for $i = 1, \cdots, im\,;\; j = 1, \cdots, jm$

$$u^{\nu}_{i,j} = \Theta_x (u^{\nu}_{i-1,j} + u^{\nu-1}_{i+1,j}) + \Theta_y (u^{\nu}_{i,j-1} + u^{\nu-1}_{i,j+1}) + \delta^2 f_{i,j}$$

The correction form of the method is:

$$\Delta u^{\nu}_{i,j} = - Res(u^{\nu-1}_{i,j}) + \Theta_x \Delta u^{\nu}_{i-1,j} + \Theta_y \Delta u^{\nu}_{i,j-1}$$

$$u^{\nu}_{i,j} = u^{\nu-1}_{i,j} + \Delta u^{\nu}_{i,j}$$

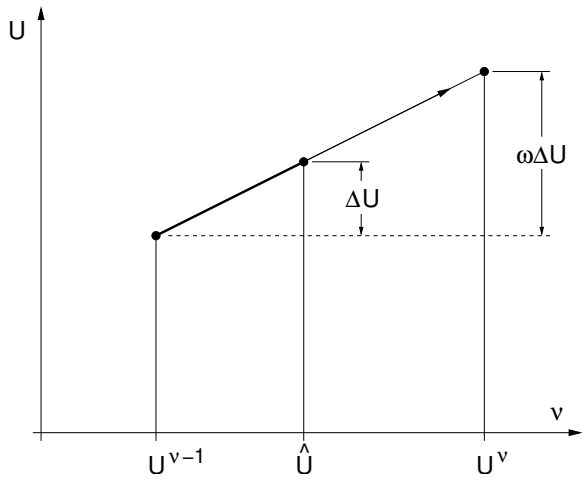
with the residual $Res(u^{\nu-1}_{i,j})$ of Poisson's equation

$$Res(u^{\nu-1}_{i,j}) = u^{\nu-1}_{i,j} - \Theta_x (u^{\nu-1}_{i-1,j} + u^{\nu-1}_{i+1,j}) - \Theta_y (u^{\nu-1}_{i,j-1} + u^{\nu-1}_{i,j+1}) - \delta^2 f_{i,j}$$

The stability investigation for the Gauss Seidel iteration method yields unconditional stability.

Accelerated Gauss Seidel point iteration method

Accelerated iteration methods, also referred to as over relaxed or interpolated methods in the literature, usually display much better convergence rates that the original method. The principle is to use the new value calculated with the iteration rule as intermediate value, called $\widetilde{\mathbf{U}}$ in this case. From this intermediate value $\widetilde{\mathbf{U}}$ and the old value $\mathbf{U}^{\nu-1}$ a new value $\mathbf{U}^{\nu}$ is determined by linear extrapolation. How far this interpolation is performed depends on the acceleration - or relaxation factor ω.



The new value $\mathbf{U}^{\nu}$ determined from the extrapolation is:

$$\mathbf{U}^{\nu} \;=\; \mathbf{U}^{\nu-1} \,+\, \omega\,(\widetilde{\mathbf{U}} - \mathbf{U}^{\nu-1})$$

The application of the accelerated Gauss Seidel point iteration is performed in two steps:
1. Step: Gauss Seidel point iteration for the intermediate value

$$E\,\widetilde{\mathbf{U}} - (\,\Theta_x\,L \,+\, \Theta_y\,B\,)\,\mathbf{U}^{\nu} \;=\; (\,\Theta_x\,L^T \,+\, \Theta_y\,B^T\,)\,\mathbf{U}^{\nu-1} \,+\, \delta^2\,\mathbf{F}$$

2. Step: over-relaxation (extrapolation)

$$E\,\mathbf{U}^{\nu} \;=\; E\,\mathbf{U}^{\nu-1} \,+\, \omega\,E\,(\widetilde{\mathbf{U}} - \mathbf{U}^{\nu-1})$$

Both steps can be summed up by eliminating $\widetilde{\mathbf{U}}$.

$$[\,E \,-\, \omega\,(\Theta_x\,L \,+\, \Theta_y\,B\,)]\;\mathbf{U}^{\nu} \;=\; [(1-\omega)\,E \,+\, \omega\,(\Theta_x\,L^T \,+\, \Theta_y\,B^T\,)]\;\mathbf{U}^{\nu-1} \,+\, \omega\,\delta^2\,\mathbf{F}$$

The point wise formulation of the Gauss Seidel method yields the algorithm for the numerical solution steps for $\quad i = 1, \cdots, im\,;\; j = 1, \cdots, jm$

$$\widetilde{u}_{i,j} \;=\; \Theta_x\,(u^{\nu}_{i-1,j} \,+\, u^{\nu-1}_{i+1,j}) \,+\, \Theta_y\,(u^{\nu}_{i,j-1} \,+\, u^{\nu-1}_{i,j+1}) \,+\, \delta^2\,f_{i,j}$$
$$u^{\nu}_{i,j} \;=\; u^{\nu-1}_{i,j} \,+\, \omega\,(\widetilde{u}_{i,j} \,-\, u^{\nu-1}_{i,j})$$

Both steps can be combined as:

$$u^{\nu}_{i,j} \;=\; (1-\omega)\,u^{\nu-1}_{i,j} \,+\, \omega\,[\Theta_x\,(u^{\nu}_{i-1,j} \,+\, u^{\nu-1}_{i+1,j}) \,+\, \Theta_y\,(u^{\nu}_{i,j-1} \,+\, u^{\nu-1}_{i,j+1}) \,+\, \delta^2\,f_{i,j}]$$

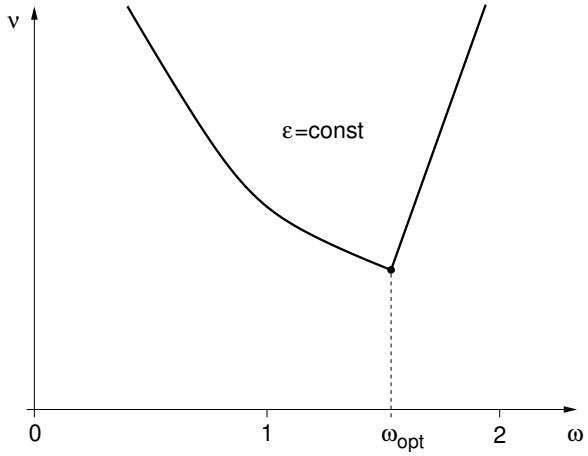
The combined correction form of the method is:

$$\Delta\,u^{\nu}_{i,j} \;=\; -\,\omega\,(Res(u^{\nu-1}_{i,j}) \,-\, \Theta_x\,\Delta u^{\nu}_{i-1,j} \,-\, \Theta_y\,\Delta u^{\nu}_{i,j-1})$$
$$u^{\nu}_{i,j} \;=\; u^{\nu-1}_{i,j} \,+\, \Delta\,u^{\nu}_{i,j}$$

with the residual $Res(u^{\nu-1}_{i,j})$ of Poisson's equation

$$Res(u^{\nu-1}_{i,j}) \;=\; u^{\nu-1}_{i,j} \,-\, \Theta_x\,(u^{\nu-1}_{i-1,j} \,+\, u^{\nu-1}_{i+1,j}) \,-\, \Theta_y\,(u^{\nu-1}_{i,j-1} \,+\, u^{\nu-1}_{i,j+1}) \,-\, \delta^2\,f_{i,j}$$

The stability analysis of the accelerated Gauss Seidel iteration method shows the stability for different values of the relaxation factor $0 \le \omega \le 2$. Values $\omega < 1$ mean under relaxation (sometimes necessary for non linear problems), $\omega = 1$ corresponds to the point wise relaxation while $\omega > 1$ stands for over relaxation.
The number of iterations to reach a given convergence limit ε depending ω, is presented in the figure. The best rate of convergence for Poisson's equation is obtained for a value ω_{opt}, ranging between 1 and 2.



The optimum value ω_{opt} essentially depends on the step sizes. A convergence analysis delivers the optimum value for the discrete Poisson equation with Dirichlet boundary conditions

$$\omega_{opt} \;=\; 2\,\left(1-\pi\cdot\delta\cdot\sqrt{2(1/a^2\;+\;1/b^2\,)}\,\right) \;=\; 2\,\left(1-\pi\cdot\sqrt{2\,(\Theta_x/(im+1)^2\;+\;\Theta_y/(jm+1)^2\,)}\,\right)$$

But for different boundary conditions and coefficients of the differential equation this value might change. In this case the value of ω_{opt} can be obtained by numeric tests. Since the convergence of the method highly depends on ω,the value should be close to the optimum relaxation factor ω_{opt} for effective calculations!

Accelerated Gauss Seidel line iteration method

The rate of convergence rises, the more entries of the solution matrix A are considered in the iteration matrix N. Line iteration methods change the variables on the grid points of a line $x = const.$ or $y = const.$ simultaneously and therefore relates them to the iteration matrix N. This yields a coupled, tri diagonal equation system which is solved by Gaussian elemi nation. The line iteration can be transferred to the principle of the Jacobi, Gauss Seidel and the accelerated Gauss Seidel iteration.

In the following the accelerated Gauss Seidel line iteration shall be explained as an example for line iteration methods. In this example the coupling on a line shall be performed in x-direction, i.e. for $1 \le i \le im$ with $j = const.$. The update of the lines runs from $j = 1$ to $j = jm$. In the first step an intermediate value is determined by the Gauss Seidel line iteration method. In the second step the final value $\mathbf{U}^{\nu}$ is calculated with over relaxation.
1. Step: Gauss Seidel line iteration for the intermediate value

$$[E\;-\;\Theta_x\,(\,L\;+\;L^T\,)]\,\widetilde{\mathbf{U}}\;=\;\Theta_y\,(\,B\,\mathbf{U}^{\nu}\;+\;B^T\,\mathbf{U}^{\nu-1})\;+\;\delta^2\,\mathbf{F}$$

2. Step: Over relaxation (extrapolation)

$$\mathbf{U}^{\nu}\;=\;\mathbf{U}^{\nu-1}\;+\;\omega\,(\widetilde{\mathbf{U}}-\mathbf{U}^{\nu-1})$$

The point wise formulation of the Gauss Seidel line iteration method yields the algorithm for the numeric solution steps for $\quad i = 1,\cdots,im\,;\; j = 1,\cdots,jm.$

$$-\Theta_x\,\widetilde{u}_{i-1,j}\;+\;\widetilde{u}_{i,j}\;-\;\Theta_x\,\widetilde{u}_{i+1,j}\;=\;\Theta_y\,(u^{\nu}_{i,j-1}\;+\;u^{\nu-1}_{i,j+1})\;+\;\delta^2\,f_{i,j}$$

$$u^{\nu}_{i,j}\;=\;u^{\nu-1}_{i,j}\;+\;\omega\,(\widetilde{u}_{i,j}\;-\;u^{\nu-1}_{i,j})$$

The following algorithm is obtained for the correction form :

$$-\Theta_x\,\Delta u^{\nu}_{i-1,j}\;+\;\Delta u^{\nu}_{i,j}\;-\;\Theta_x\,\Delta u^{\nu}_{i+1,j}\;=\;-\,\omega\,(Res(u^{\nu-1}_{i,j})\;-\;\Theta_y\,\Delta u^{\nu}_{i,j-1})$$

$$u^{\nu}_{i,j}\;=\;u^{\nu-1}_{i,j}\;+\;\Delta\,u^{\nu}_{i,j}$$

The Gauss Seidel line iteration method, formulated for the variables of the correction form, leads to the solution of a tri diagonal equation system. This equation system can be solved by Gaussian elimination.

The Gauss Seidel line iteration is stable for the relaxation factor $0 \leq \omega \leq 2$. The optimum value for the relaxation factor of Poisson's equation with Dirichlet boundary conditions is

$$\omega_{opt} = 2\left(1 - \Delta y \cdot \pi \cdot \sqrt{1/a^2 + 1/b^2}\right) \quad \text{for line x-direction}$$
$$\omega_{opt} = 2\left(1 - \Delta x \cdot \pi \cdot \sqrt{1/a^2 + 1/b^2}\right) \quad \text{for line in y-direction}$$

For a better convergence it is beneficial to chose the coupled line in the direction of smaller step sizes.

Line iteration with alternating directions

Line iteration schemes with a line of coupled points transport the information in the direction of the line directly. In the other direction the information transport is only performed step wise from line to line. This slows down the convergence. Therefore, it is more efficient to alternate the direction of the coupling. This leads to the line iteration schemes with alternating directions. This principle can again be transferred and applied to the Jacobi as well as the Gauss Seidel iteration. One of the first of such schemes was published in:

Peaceman , Rachford: SIAM – Journal, 3, 1955.

This method (Alternating Direction Implicit Method (ADI)) is based on the Jacobi line iteration and over relaxation with alternating lines in x and y-direction. In the following the algorithm is presented as an example.

1. Line iteration in x- direction

$$\begin{aligned}[E - \Theta_x (L + L^T)] \cdot \widetilde{\mathbf{U}}^{\nu-1/2} &= \Theta_y (B + B^T) \cdot \mathbf{U}^{\nu-1} + \delta^2 \mathbf{F} \\ \mathbf{U}^{\nu-1/2} &= \mathbf{U}^{\nu-1} + \omega (\widetilde{\mathbf{U}}^{\nu-1/2} - \mathbf{U}^{\nu-1})\end{aligned}$$

2. Line iteration in y- direction

$$\begin{aligned}[E - \Theta_y (B + B^T)] \cdot \widetilde{\mathbf{U}}^{\nu} &= \Theta_x (L + L^T) \cdot \mathbf{U}^{\nu-1/2} + \delta^2 \mathbf{F} \\ \mathbf{U}^{\nu} &= \mathbf{U}^{\nu-1/2} + \omega (\widetilde{\mathbf{U}}^{\nu} - \mathbf{U}^{\nu-1/2})\end{aligned}$$

The point wise formulation yields the following system:

1. Line iteration in x- direction

$$-\Theta_x \widetilde{u}_{i-1,j}^{\nu-1/2} + \widetilde{u}_{i,j}^{\nu-1/2} - \Theta_x \widetilde{u}_{i+1,j}^{\nu-1/2} = \Theta_y (u_{i,j-1}^{\nu-1} + u_{i,j+1}^{\nu-1}) + \delta^2 f_{i,j}$$
$$u_{i,j}^{\nu-1/2} = u_{i,j}^{\nu-1} + \omega (\widetilde{u}_{i,j}^{\nu-1/2} - u_{i,j}^{\nu-1})$$

2. Line iteration in y- direction

$$-\Theta_y \widetilde{u}_{i,j-1}^{\nu} + \widetilde{u}_{i,j}^{\nu} - \Theta_y \widetilde{u}_{i,j+1}^{\nu} = \Theta_x (u_{i-1,j}^{\nu-1/2} + u_{i+1,j}^{\nu-1/2}) + \delta^2 f_{i,j}$$
$$u_{i,j}^{\nu} = u_{i,j}^{\nu-1/2} + \omega (\widetilde{u}_{i,j}^{\nu} - u_{i,j}^{\nu-1/2})$$

This method is stable for all relaxation parameters ω. To improve the rate of convergence an optimized relaxation parameter is used for each direction.

1.4.6 Convergence of iteration schemes

In the convergence investigation of initial value problems (see chapter 3 of this course) it is demanded that the numerical solution approaches the exact solution of the differential equation for diminishing step sizes. In principle this also applies for boundary value problems, like the above considered Poisson equation. If the discretized Poisson equation is solved with a direct method and the system has a unique solution, the consistency of the spatial discretization (truncation error becomes zero) suffices for convergence. This means the exact difference solution approaches the exact Poisson equation for decreasing step sizes Δx and Δy. For the investigated solution of Poisson's equation the convergence of the exact difference solution is given and will be assumed as known in the following.

For the convergence of an iteration scheme with step wise solution of an approximated matrix, it must be ensured that the iterative solution approaches the exact solution of the difference problem for iterating numbers $\nu \to \infty$. This problem is discussed in the following.

A qualitative method for the proof of convergence of an iteration method is given by the Laxian theorem. If an iteration scheme is considered as artificial initial value problem, stability and consistency are sufficient for convergence. This statement doesn't suffice to evaluate iteration schemes. Therefore, a proof of convergence by investigation of the discrete eigenvalue problem of the iteration matrix is presented in this section. This allows the calculation of the rate of convergence with which the methods can be compared quantitatively.

Definitions

The exact difference solution of the discretized Poisson equation is obtained by direct inversion of A.

$$\mathbf{U} \;=\; A^{-1}\,\delta^2\,\mathbf{F}$$

If the solution is performed with an iteration scheme,

$$N\,\mathbf{U}^{\nu} \;=\; P\,\mathbf{U}^{\nu-1} \;+\; \delta^2\,\mathbf{F} \quad \text{with} \quad A = N - P$$

an approximate solution is obtained for each iteration step ν

$$\mathbf{U}^{\nu} \;=\; N^{-1}\,(P\,\mathbf{U}^{\nu-1} \;+\; \delta^2\,\mathbf{F})$$

Convergence of an iteration scheme is obtained, if the solution of the iteration problem approaches the exact solution of the difference problem, i.e.

$$\lim_{\nu\to\infty}\,(\mathbf{U}^{\nu} \;-\; \mathbf{U}) \;=\; 0$$

Defining the convergence error with

$$\mathbf{e}^{\nu} = \mathbf{U}^{\nu} \;-\; \mathbf{U}$$

and insertion in the iteration rule, yields

$$\mathbf{e}^{\nu} \;=\; N^{-1}\,(P\,\mathbf{e}^{\nu-1}) \;=\; M\,\mathbf{e}^{\nu-1} \qquad \text{with} \quad M = N^{-1}\,P$$

for the error. Step wise insertion up to a given initial error $\mathbf{e}^0$ yields

$$\mathbf{e}^\nu \;=\; (M)^\nu\, \mathbf{e}^0$$

It is sufficient for convergence, if the convergence error for $\nu \to \infty$ disappears, ie.e. if

$$\lim_{\nu\to\infty} (M)^\nu \;=\; 0$$

A matrix M satisfying this condition is called convergent.
The modulo of all eigenvalues λ_i of a convergent matrix is smaller than one. This condition is often expressed by the spectral radius $\varrho(M)$, i.e.

$$\varrho(M) \;\equiv\; \max_i\, |\lambda_i| \;<\; 1$$

This yields for an arbitrarily chosen norm$||\cdot||$ für M

$$||\, M\,|| \;<\; 1$$

An estimation of the iteration equation of the convergence error by a norm leads to the inequality

$$||\mathbf{e}^\nu|| \;\leq\; ||(M)||^\nu\; ||\mathbf{e}^0||$$

If the norm of M is replaced by the spectral radius, one obtains the important relation between the convergence error and the spectral radius.

$$||\mathbf{e}^\nu|| \;\leq\; [\varrho(M)]^\nu\; ||\mathbf{e}^0||$$

If for an iterative calculation it is demanded that the amplitude of the is reduced at minimum by a factor 10^{-m}, i.e. $||\mathbf{e}^\nu||/||\mathbf{e}^0|| \;=\; [\varrho(M)]^\nu \;\leq\; 10^{-m}$, the minimum number of required iteration steps is obtained from the inequality.

$$\nu \;\geq\; \frac{m}{-\log\, \varrho(M)} \;=\; \frac{m}{R}$$

The rate of convergence $R = -\log\, \varrho(M)$ is an important measure for the efficiency of a iteration method. The larger the convergence rate R, the better the iteration scheme, i.e. the fewer iterations are necessary to reduce the initial error to 10^{-m}.

Solution of the discrete eigenvalue problem

The eigenvalues of the matrix M are determined from the solution of the
homogeneous eigenvalue problem of the matrix M:

$$M\,\mathbf{W} \;=\; \lambda\, E\,\mathbf{W} \qquad \text{with} \quad \mathbf{W} \;=\; 0 \quad \text{on the boundaries}$$

$\mathbf{W}$ is the solution for the eigenvalue λ and corresponds to the non trivial solution of the boundary value problem.
The solution of the eigenvalue problem will be presented for the example of the solution matrix A.
If one defines $M = A$, the following eigenvalue problem is obtained:

$$A\,\mathbf{W} \;=\; \lambda^A\, E\,\mathbf{W} \qquad \text{with} \quad \mathbf{W} \;=\; 0 \quad \text{on the boundaries}$$

The difference equation at a point (i,j), where $1 \le i \le im\,,\ 1 \le j \le jm$ is

$$w_{i,j} \;-\; \Theta_x\,(\,w_{i-1,j} \;+\; w_{i+1,j}) \;-\; \Theta_y\,(\,w_{i,j-1} \;+\; w_{i,j+1}) \;=\; \lambda^A\, w_{i,j}$$

The linear difference equation can be satisfied by a seperation Ansatz.
(Splitting in a x and y dependent component)

$$w_{i,j} \;=\; \varphi_i \cdot \psi_j \;=\; (a\,e^{I\alpha\,i} + b\,e^{-I\alpha\,i}\,) \cdot (c\,e^{I\beta\,j} + d\,e^{-I\beta\,j}\,)$$

The coefficients a and b are determined from the boundary condition $(w = 0)$.
One obtains for

$$\varphi_i = a\,e^{I\alpha\,i} + b\,e^{-I\alpha\,i} = (a+b)\,\cos\alpha\,i + I\,(a-b)\,\sin\alpha\,i$$

- $i = 0$: $\qquad\qquad \varphi_i = 0 \to 0 = a + b$
- $i = im+1$: $\qquad \varphi_{im+1} = 0 \to 0 = (a-b)\,I\,\sin\,[\alpha(im+1)]$
  $\to$ is satisfied for $\alpha\,(im+1) = p\,\pi$ with $p = (0), 1, 2, \cdots, im, (im+1)$

The boundary conditions for ψ_j are satisfied in an analogous way which yields the solution $w_{i,j}$:

$$w_{i,j} \;=\; \varphi_i \cdot \psi_j \;=\; 2a\,\sin\,(\alpha i) \cdot 2c\,\sin\,(\beta j) \;=\; a\,(e^{I\alpha i} - e^{-I\alpha i}) \cdot c\,(e^{I\beta j} - e^{-I\beta j})$$

$$\text{with} \quad \alpha = \tfrac{p\,\pi}{im+1} \quad p = 1, 2, \cdots, im \quad \text{and} \quad \beta = \tfrac{q\,\pi}{jm+1} \quad q = 1, 2, \cdots, jm$$

This solution $w_{i,j}$ is inserted in the difference equation of the eigenvalue problem.

The eigenvalues of the solution matrix A after rearrangement of the equation yield

$$\lambda^A_{p,q} \;=\; 4\,\Theta_x\,\sin^2\,(\frac{\pi \cdot p}{2(im+1)}) + 4\,\Theta_y\,\sin^2\,(\frac{\pi \cdot q}{2(jm+1)}) \qquad p = 1, 2\cdots, im \quad q = 1, 2, \cdots, jm$$

Where $2\Theta x + 2\Theta y = 1$ and $\sin^2\,(\alpha/2) = \frac{1}{2}\,(1 - \cos\alpha)$ have been applied in the equation.

Convergence of the Jacobi iteration

The calculation of the rate of convergence R is exemplary presented for the Jacobi iteration.

The Jacobi method is defined by

$$N \;=\; E \quad \text{and} \quad P \;=\; N - A \;=\; \Theta_x\,(L \;+\; L^T) \;+\; \Theta_y\,(B \;+\; B^T)$$

The maximum eigenvalue of $M \;=\; N^{-1}\,P$ is required for the calculation of the rate of convergence. If the eigenvalue problem $M\,\mathbf{W} \;=\; \lambda\,E\,\mathbf{W}$ is multiplied by $N = E$ and if P is replaced by $P = E - A$, one obtains

$$A\,\mathbf{W} \;=\; (1 - \lambda^J)\,\mathbf{W}$$

The eigenvalue problem for the complete solution matrix A is

$$A\,\mathbf{W} \;=\; \lambda^A\,\mathbf{W}$$

A comparison directly yields

$$\lambda^J = 1 - \lambda^A$$

Since the eigenvalues for A have already been calculated, one obtains for the Jacobi iteration

$$\lambda^J \;=\; 1 - 4\,\Theta_x\,\sin^2\,(\frac{\pi\cdot p}{2(im+1)}) - 4\,\Theta_y\,\sin^2\,(\frac{\pi\cdot q}{2(jm+1)}) \qquad p = 1,\cdots,im \quad q = 1,\cdots,jm$$

The maximum eigenvalue is obtained for $p = q = 1$. This yields the following spectral radius $\varrho(M^J)$ of the Jacobi iteration:

$$\varrho(M^J) \;=\; 1 \;-\; \Theta_x\,\frac{\pi^2}{(im+1)^2} \;-\; \Theta_y\,\frac{\pi^2}{(jm+1)^2}$$

In this equation the sine has been expanded for small arguments ($im \gg 1$, $jm \gg 1$) ($\sin x \approx x$).

The rate of convergence, $R = -\log\varrho(M^J)$, of the Jacobi method is obtained for $\log x \approx 1 - x$ as

$$R(M^J) \;=\; \pi^2\,\left(\frac{\Theta_x}{(im+1)^2} \;+\; \frac{\Theta_y}{(jm+1)^2}\right) \;=\; \pi^2\,\delta^2\,(\frac{1}{a^2} + \frac{1}{b^2})$$

This yields the minimum number of iterations necessary to reduce the convergence error by 10^{-m}

$$\nu = \frac{m}{R} = f\,(\Delta x, \Delta y, im, jm)$$

If one exemplary assumes that $\Delta x = \Delta y$ and $im = jm$, on obtains

$$\nu = \frac{2m}{\pi^2}\,(im+1)^2 \;\sim\; im^2$$

for the necessary iterations. It shows that with a growing number of grid points the number of necessary iterations grows quadratic!

Comparison of the rates of convergence of iteration schemes

In a similar fashion as has been presented for the Jacobi iteration, the rates of convergence of other iteration schemes can be determined. The rates of convergence of the above discussed methods for the difference solution of Poisson's equation with Dirichlet boundary conditions are presented in the following table.

	R	$R\,/\,R_J$	ω
1) Point iteration			
Jacobi	$R_J = \delta^2 \pi^2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)$	1	1
Gauss–Seidel	$R_{GS} = 2\, R_J$	2	1
accelerated Gauss–Seidel (optimiert)	$R_{BGS} = 2\,\delta\,\pi \sqrt{2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)}$	$\frac{2\sqrt{2}}{\sqrt{R_J}}$	$\omega_{opt} = 2 - R_{BGS}$
2) Line iteration (line in x-direction)			
Jacobi	$R_{LJ} = \frac{\Delta y^2}{2} \pi^2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)$	$1 + \frac{\Delta y^2}{\Delta x^2}$	1
Gauss–Seidel	$R_{LGS} = 2\, R_{LJ}$	$2\,(1 + \frac{\Delta y^2}{\Delta x^2})$	1
beschleunigter Gauss–Seidel (optimiert)	$R_{BLGS} = 2\,\Delta y\,\pi \sqrt{2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)}$	$\frac{2\sqrt{2}}{\sqrt{R_J}} \sqrt{\left(1 + \frac{\Delta y^2}{\Delta x^2} \right)}$	$\omega_{opt} = 2 - R_{BLGS}$
3) Line iteration (line in y-direction)			
Jacobi	$R_{LJ} = \frac{\Delta x^2}{2} \pi^2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)$	$1 + \frac{\Delta x^2}{\Delta y^2}$	1
Gauss–Seidel	$R_{LGS} = 2\, R_{LJ}$	$2\,(1 + \frac{\Delta x^2}{\Delta y^2})$	1
accelerated Gauss–Seidel (optimized)	$R_{BLGS} = 2\,\Delta x\,\pi \sqrt{2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right)}$	$\frac{2\sqrt{2}}{\sqrt{R_J}} \sqrt{\left(1 + \frac{\Delta x^2}{\Delta y^2} \right)}$	$\omega_{opt} = 2 - R_{BLGS}$

The following definitions have been applied:

$$a = (im + 1)\,\Delta x \qquad b = (jm + 1)\,\Delta y \qquad \delta^2 = \frac{\Delta x^2 \cdot \Delta y^2}{2(\Delta x^2 + \Delta y^2)}$$

For the choice of a method for a given problem one should consider:

- The rate of convergence R should be as high as possible, since $\nu = \frac{m}{R}$
- An accelerated scheme with optimized relaxation factor ω_{opt} should be preferred, since the iteration count of non optimized scheme is usually higher and increases quadratic with the number of grid points.
- The computational effort per grid point is in proportion with the iteration count and the number of floating point operations (FLOPs) of a method. Since implicit line iteration schemes require more operations per grid point (around a factor three more), it must be estimated which scheme is more beneficial, line iteration or point iteration.

- For the transformation of boundary conditions in the field (especially other conditions than Dirichlet boundaries) the line iteration schemes are usually more beneficial, because the information transport on a line is performed directly. The best convergence is obtained with line iteration schemes with alternating directions.

For more details for the determination of the rate of convergence see: *E. Isaacson, H. B. Keller: Analyse numerischer Verfahren. Verlag Deutsch, Zürich, 1973*

Chapter 2

Computational Fluid Dynamics II

2.1 Numerical solution of parabolic, partial differential equations

2.1.1 Introduction

Important equations in fluid dynamics are of parabolic type, see also section 1 and 2 of this script. Typical for these equations is that the solution of the characteristic polynomial results in a double solution for the slopes of the characteristic lines (see chapter 3). The solution of parabolic equations describe phenomena, where the information is transported with infinite signal speed in a half space of the independent variables. Considering the general partial differential equation of second order

$$a\,u_{xx} + 2\,b\,u_{xy} + c\,u_{yy} + F\,(u_x, u_y, u, x, y) = 0$$

the slopes of the characteristic lines result in

$$\frac{dy}{dx}|_{1,2} = \frac{b}{a} \pm \frac{1}{a}\sqrt{b^2 - ac}$$

The equation is parabolic, if the discriminant $b^2 - ac$ vanishes. Therefore, the double solution for the slope of the characteristic lines results in $\frac{dy}{dx}|_{1,2} = \frac{b}{a}$. Most equations in fluid dynamics in Cartesian coordinates appear in their normal form, which means that the mixed derivative in the equation above disappears ($b = 0$). The equation is parabolic, if additionally one of the second order derivatives disappears, i.e., if $a = 0$ or $c = 0$. Examples are the heat conduction equation (Fourier equation)

$$u_t = \nu\,u_{xx}$$

or the streamwise momentum equation of the boundary layer equations:

$$uu_x = \nu u_{yy} - vu_y - p_x$$

For the heat conduction equation the domain of influence is defined by its characteristic lines with the slope $\frac{dt}{dx}|_{1,2} = 0$. In the $x - t$ plane for the point P at (x_P, t_P), see Fig. 2.1.1, the domain of influence is thus the half space below the point P defined by $t < t_p$. Therefore, an initial value problem has to be solved in the time direction, for which the solution evolves from an initial condition, e.g. from time $t = 0$ until t_P. Additionally, there is a boundary value problem in x-direction. Due to the second derivative u_{xx}, boundary values have to be specifed at the left and right boundary of the domain, i.e., at x_1 and x_2. The solution of such an initial and boundary value problem is typical of parabolic, partial differential equations and also applies for more complex equation systems such as the Navier-Stokes equations. The solution of parabolic, partial differential equations will be demonstrated in the following by two examples.

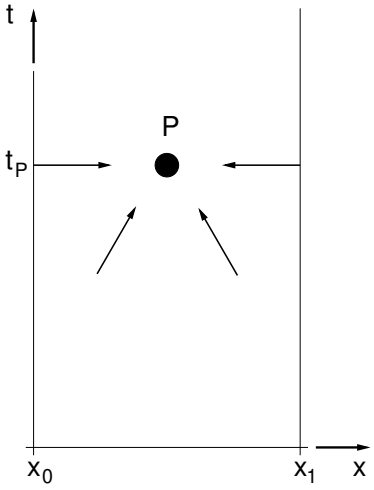


Figure 2.1.1: Domain of influence for the heat conduction equation.

The numerical solution a parabolic equation can be performed with implicit and explicit finte difference methods. The implicit solution allows an information transport with infinite speed in the boundary value problem and the time step is not limited by a stability constraint. An explicit solution method of a parabolic, partial differential equation introduces a finite information propagation speed dependent on the time and spatial step size. Therefore, a hyperbolic partial differential equation is solved by the computational approximation, see section 1.3.2. The explicit scheme usually needs considerably less floating point operations per time step than the implicit scheme, however, the maximum usable time step is limited. Therefore, the ratio of physical time step, which is necessary to resolve the time scales in the physical problem, related to the maximum explicit time step, determine whether an explicit or implicit scheme is compuatationally more efficient.

2.1.2 Numerical solution of the Fourier equation

The temporal development of a simple flow problem, the Couette flow, will be demonstrated by the numerical solution of the Fourier equation. The Couette flow is a steady flow of an incompressible fluid (density ρ, viscosity η) between two parallel plates of infinite extension. At the time $t = 0$, one plate is suddenly accelerated to a constant velocity u_0. The temporal development of the velocity profile between the plates until the asymptotic, steady state solution is reached, can be determined by a solution of the simplified Navier-Stokes Equations.

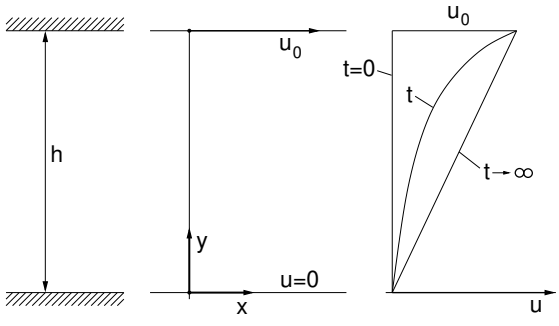


Figure 2.1.2: Domain, boundary conditions and velocity profiles for a Couette flow

Because of the assumption of a fully developed flow between infinitely extended plates the

velocity component normal to the plates vanishes, i.e., $v = 0$. In addition, all gradients in the flow direction vanish, i.e. $\frac{\partial f}{\partial x} = 0$ where $f = u, v, p$. The Navier-Stokes equations for an incompressible fluid then reduce to a single equation, which is identical to the Fourier equation. That is:

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial y^2}$$

The initial conditions for $t = 0$ and boundary conditions for $y = 0$, $y = h$ are:

initial condition at $t = 0$	$u(t = 0, 0 \leq y < h) = 0$	$u(t = 0, y = h) = u_0$
boundary conditions at $y = 0$, $y = h$	$u(t, y = 0) = 0$	$u(t, y = h) = u_0$

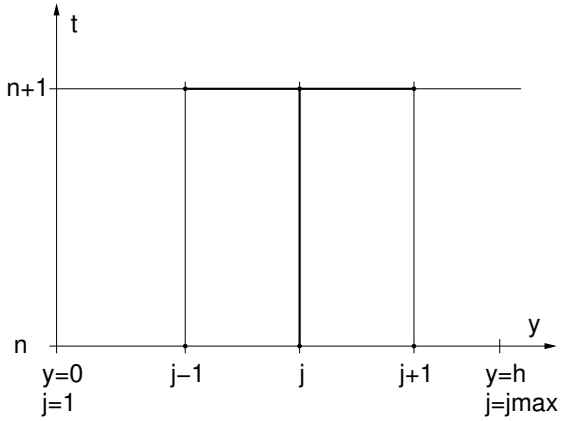
For the validation of the numerical solution an analytical solution for the linear parabolic initial boundary value problem can be derived. That is:

$$u = u_0 \cdot \sum_{n=0}^{\infty} (\text{erfc}(2n\eta_h + \eta) - \text{erfc}(2(n+1)\eta_h - \eta)\ )$$

$$\text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-z^2) dz$$

where $\eta = \frac{y}{2\sqrt{\nu t}}$ and $\eta_h = \frac{h}{2\sqrt{\nu t}}$. The steady solution for $\frac{\partial u}{\partial t} = 0$ results in a linear velocity profile $u = u_0 \cdot \frac{y}{h}$.

The numerical solution occurs using a mesh with $(j_{max} - 1)$ equidistant spatial steps Δy and a time step Δt.



$$\Delta y = \frac{h}{(j_{max} - 1)}$$

$$y = (j - 1)\,\Delta y \quad 1 \leq j \leq jmax$$

$$t = (n - 1) * \Delta t \quad 1 \leq n \leq nmax$$

$$u(y, t) = u_j^n$$

The initial and boundary conditions of the Couette flow in the discrete space become:

initial condition at $t^{n=0} = 0$	$u_1^0 = 0$	$u_{jmax}^0 = u_0$
boundary conditions at $y_{j=1} = 0$, $y_{j=jmax} = h$	$u_1^n = 0$	$u_{jmax}^n = u_0$

The numerical solution of the differential equation can be computed with a general scheme using the parameter Θ, which controls whether an explicit or implicit scheme is obtained. The discretized equation with the numerical diffusion number $\sigma = \nu \frac{\Delta t}{\Delta y^2}$ reads:

$$u_j^{n+1} = u_j^n + \big(1 - \Theta\big)\sigma(u_{j-1}^n - 2u_j^n + u_{j+1}^n) + \Theta\sigma(u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1})$$

For $\Theta = 0$ an explicit scheme with a truncation error on the order of $O(\Delta y^2, \Delta t)$, for $\Theta = 1$ an implicit scheme on the order of $O(\Delta y^2, \Delta t)$ and for $\Theta = 1/2$ the implicit Crank-Nicholson scheme on the order of $O(\Delta y^2, \Delta t^2)$ is obtained.

The implicit solution results in a coupled, tridiagonal equation system for the unknown solution vector $u_j^{n+1}, j = 2, \ldots, jmax-1$. In many cases, especially for non-linear problems, the equation system is not solved for the unknown u_j^{n+1}, but for the correction variable $\Delta u_j^n = u_j^{n+1} - u_j^n$. The advantage using the correction varibale is that the spatial operator, which defines the stationary solution, can be formulated independently of the solution matrix, so that the equation system can be solved using a simplified or iterative solution technique. The discretized equation formulated and sorted for the temporal correction Δu^n, results in:

$$-\Theta\sigma\Delta u_{j-1}^n + (1 + 2\Theta\sigma)\Delta u_j^n - \Theta\sigma\Delta u_{j+1}^n = \sigma(u_{j-1}^n - 2u_j^n + u_{j+1}^n)$$

This differential equation leads to the tridiagonal equation system:

$$a_j\Delta u_{j-1}^n + b_j\Delta u_j^n + c_j\Delta u_{j+1}^n = r_j$$

A direct solution method for the tridiagonal equation system was already discussed in section 1.3.1. After determining the solution for Δu^n, the new variables u_j^{n+1} can be calculated.

$$u_j^{n+1} = u_j^n + \Delta u^n$$

The numerical solution method implemented in a FORTRAN program and results are discussed in the lecture.

The velocity field $\frac{u}{u_0} = f(\frac{y}{H})$ for the Couette flow is illustrated for various time steps n in Fig. 2.1.3, which are computed with the implicit scheme at $\Theta = 1$ and $\sigma = 1$.

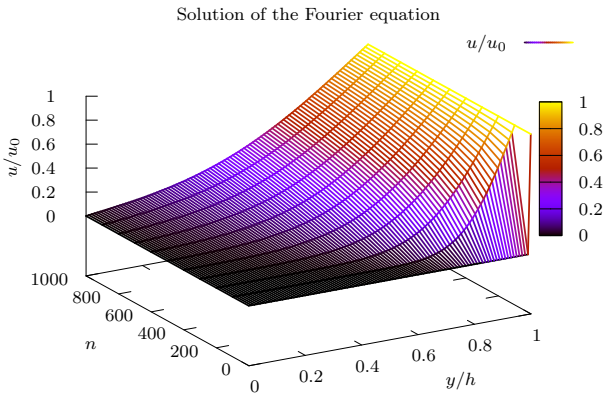


Figure 2.1.3: Numerical solution of the velocity profiles $\frac{u}{u_0} = f(\frac{y}{H})$ for a Couette flow at various time steps n.

2.1.3 Numerical solution of the boundary layer equations

In the following, the numerical solution of the boundary layer equations for two dimensional, incompressible flows will be discussed as an example for systems of parabolic, partial differential equations. The boundary layer approximation is an important approximation of the Navier-Stokes equations since it allows to predict the viscous drag of a streamlined body at a small fraction of the computational costs compared to the Navier-Stokes equations. According to Prandtl, the prediction of a high Reynolds number flow with thin viscous layers attached to a body surface, can be split in the solution of two equation systems, which are easier to solve than the Navier-Stokes equations.

The inviscid exterior flow can be predicted by the potential flow or Euler equations, which results in the pressure distribution around the body. The pressure distribution is then used in the solution of the boundary layer equations to provide the information about the viscous drag on the body. Prerequisites for the boundary layer approximation are high Reynolds numbers $\mathrm{Re} \gg 1$ and an attached boundary layer, where the boundary layer thickness δ is much smaller than the length of the body L. In general it can be assumed that $\delta \sim L/\sqrt{\mathrm{Re}}$. With these assumptions the boundary layer equations can be derived from the Navier-Stokes equations with suitable normalizations of all variables in the limit of $\mathrm{Re} \gg 1$.

The following normalized and dimensionless variables are introduced for two dimensional incompressible boundary layers

$$\begin{aligned} x &= \frac{\bar{x}}{L}, \quad y = \frac{\bar{y}}{L}\sqrt{\mathrm{Re}}, \quad u = \frac{\bar{u}}{u_\infty}, \quad v = \frac{\bar{v}}{u_\infty}\sqrt{\mathrm{Re}}, \\ \eta &= \frac{\bar{\eta}}{\eta_\infty}, \quad \rho = \frac{\bar{\rho}}{\rho_\infty} = 1, \quad p = \frac{\bar{p}}{\rho_\infty u_\infty^2}, \end{aligned}$$

where L, an variable with an overbar $\bar{}$ or the index $_\infty$ denotes a quantity with dimensions. Usually, the flow variables of the undisturbed flow, denoted by the index $_\infty$, are used as reference values. The Reynolds number Re is based on these reference values:

$$\mathrm{Re} = \frac{\rho_\infty u_\infty L}{\eta_\infty}$$

In the limit $\mathrm{Re} \gg 1$ the boundary layer equations in dimensionless formulation are obtained.

$$\begin{aligned} u_x + v_y &= 0 \\ u u_x + v u_y + p_x &= (\eta u_y)_y \\ p_y &= 0 \end{aligned}$$

The viscosity η is a fluid property and often can be considered as a mere function of the temperature, i.e., $\eta = \eta(T)$. For turbulent flows an additional closure assumption for the Reynolds stresses must be introduced. Usually, an eddy viscosity approach is used, in which η is replaced by an effective viscosity η_{eff} consisting of a laminar and a turbulent component, i.e., $\eta_{eff} = \eta_{laminar} + \eta_{turbulent}$. The solution properties of the boundary layer equations are not influenced by such a turbulence closure approach such that the numerical solution presented in this section can also be applied for turbulent boundary layers. For further details on the boundary layer equations and their solutions see e.g. *H. Schlichting: Grenzschichttheorie. Verlag G. Braun, Karlsruhe.*

The pressure p along a direction normal to the wall in the boundary layer is constant (because of $p_y = 0$), its distribution along the streamwise direction $p(x)$ is determined by the inviscid external flow and has to be known. At the edge of the boundary layer $\delta(x)$ the flow velocity $u(x, y = \delta(x))$ in the boundary layer reaches approximately the value of the inviscid flow $u_e(x)$. A smooth transition from the boundary layer solution to the external inviscid flow requires all viscous stresses to vanish at the boundary layer edge. Therefore, all gradients in y-direction must vanish at $\delta(x)$, i.e. $u_y = u_{yy} = \ldots = 0$. The momentum

equation in x-direction then yields a relation between the pressure p and the velocity of the exterior flow u_e at the edge of the boundary layer $y = \delta(x)$:

$$u_e u_{e,x} + p_x = 0$$

Integration of this equations leads to the well known Bernoulli equation, $p + u_e^2/2 = constant$. For the solution of the boundary layer equations, it is convenient to express the pressure by the exterior velocity u_e, which results in the equation system for the unknown velocity components u and v:

$$\begin{aligned} u_x + v_y &= 0 \\ u u_x + v u_y - u_e u_{e,x} &= (\eta u_y)_y \end{aligned}$$

The resulting equation system is of parabolic type and leads to an initial value problem in x-direction (corresponding to the time t in the Fourier equation) and a boundary value problem in y-direction. The solution of the boundary value problem takes two boundary conditions for u and one for v (according to u_{yy} and v_y). The boundary values for the boundary layer can be assumed as following:

- Wall $y = 0$: $u(x, y = 0) = u_W(x)$ and $v(x, y = 0) = v_W(x)$

  With this the following cases can be simulated:

  $u_W = v_W = 0$ solid, no-slip wall

  $v_W < 0$ suction of the boundary layer

  $u_W > 0$, $v_w > 0$ direct blow out of the boundary layer

- Edge $y = \delta$: $u(x, \delta) = u_e(x)$

  In addition the linking condition $u_y(x, \delta) = 0$ can be applied

  to define the edge of the boundary layer δ, e.g.:

  $y = \delta$ when $|\frac{u_e - u}{u_e}| < \epsilon_{edge} \ll 1$

The initial condition for $x = x_0$

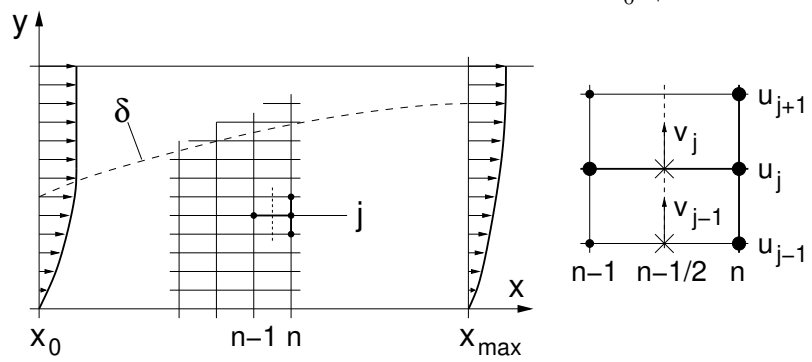
simply requires the velocity profile $u(y)$, i.e.

$$x = x_0 : \qquad u(x_0, y) = u_0(y)$$

The vertical velocity $v(x_0, y)$ is thus uniquely defined.

The solution of the initial boundary value problem of the boundary layer equations is achieved with a so called marching scheme. Starting from the initial value at x_0 a new solution for $x_0 + \Delta x$ is estimated from the boundary layer equations. The newly calculated solution

is then taken as the initial value for the solution at $x_0 + 2\Delta x$ in the next step, etc.



In the following the development of a numerical solution will be demonstrated with a simple but effective implicit solution scheme (Laasoonen scheme). The numerical solution is carried out on a line $x = x_n$ with the initial condition on x_{n-1}. In y-direction with $y_j = (j-1) * \Delta y$ constant step sizes Δy are applied. The amount of points $jmax$ and the value of Δy are fixed through the initial conditions. In the x-direction the boundary layer can change its size, therefore it is evaluated for each step in x-direction. If necessary additional points are added in y-direction. The momentum equations at the point $P(x_n, y_j)$ are expanded with backward differences for u_x and central differences for the y-derivatives. This results in a scheme with an accuracy of order $O(\Delta x, \Delta y^2)$. The discretized momentum equations are:

$$\begin{aligned} R_1 = \quad & u_j^{n-1}\,\frac{u_j^n - u_j^{n-1}}{\Delta x} + v_j^{n-1/2}\,\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta y} - u_e^{n-1}\,\frac{u_e^n - u_e^{n-1}}{\Delta x} \\ - \quad & \big(\eta_{j+1/2}^{n-1}\,\frac{u_{j+1}^n - u_j^n}{\Delta y} - \eta_{j-1/2}^{n-1}\,\frac{u_j^n - u_{j-1}^n}{\Delta y}\big)/\Delta y = 0 \end{aligned}$$

For all emerging coefficients the known values on x_{n-1} are used, except for the vertical velocity v which is newly solved at $x_{n-1/2}$ with the continuity equation. It is therefore beneficial to expand the continuity equation in the point $P(x_{n-1/2}, y_{j-1/2})$.

$$R_2 = \frac{1}{2}\,\big(\frac{u_j^n - u_j^{n-1}}{\Delta x} + \frac{u_{j-1}^n - u_{j-1}^{n-1}}{\Delta x}\big) + \frac{v_j^{n-1/2} - v_{j-1}^{n-1/2}}{\Delta y} = 0$$

The difference equations $R_1 = 0$ and $R_2 = 0$ are combined in the so called residual vector $\vec{Res} = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}$. They form a coupled, algebraic equation system in the unknowns $\vec{V}_j = \begin{pmatrix} u^n \\ v^{n-1/2} \end{pmatrix}$ for the points $(j-1, j, j+1)$. The discretized boundary layer equations can therefore be combined:

$$\vec{Res}(\vec{V}) = 0$$

The residual vector $\vec{Res}(\vec{V})$ connects the variables of the adjacent points $j-1, j, j+1$, i.e.

$$\vec{Res}(\vec{V}) = \vec{Res}(\vec{V}_{j-1}, \vec{V}_j, \vec{V}_{j+1}) = 0$$

The solution is performed iteratively with a Newtonian iteration scheme, since the system is non linearly coupled. Using a Taylor series expansion with the iteration index ν

$$\vec{Res}(\vec{V}^{\nu+1})) = \vec{Res}(\vec{V}^{\nu}) + \frac{\partial \vec{Res}}{\partial \vec{V}}|^{\nu} \cdot (\vec{V}^{\nu+1} - \vec{V}^{\nu}) = 0$$

the following iteration equation is obtained:

$$\sum_{k=j-1}^{j+1} \frac{\partial \vec{Res}}{\partial \vec{V}_k} \cdot (\vec{V}_k^{\nu+1} - \vec{V}_k^{\nu}) \;=\; -\vec{Res}(\vec{V}_j^{\nu})$$

The fully expanded system with $\Delta\vec{V}_k^{\nu} = \vec{V}_k^{\nu+1} - \vec{V}_k^{\nu}$ results in a tridiagonal equation system of the vectors $\Delta\vec{V}^{\nu}$, where the coefficients are 2×2 matrices (block tridiagonal system).

$$\overline{\overline{A}}_j \, \Delta\vec{V}_{j-1}^{\nu} \;+\; \overline{\overline{B}}_j \, \Delta\vec{V}_j^{\nu} \;+\; \overline{\overline{C}}_j \, \Delta\vec{V}_{j+1}^{\nu} \;=\; -\vec{Res}(\vec{V}_j^{\nu})$$

$\overline{\overline{A}}$, $\overline{\overline{B}}$ and $\overline{\overline{C}}$ are the so called Jacobi matrices. Their elements are obtained from the differentiation of the equations $R_1 = 0$ and $R_2 = 0$ in respect to the variables $(u^n, v^{n-1/2})$ for the point $(j-1, j, j+1)$. For example:

$$\overline{\overline{B}}_j = \frac{\partial \vec{Res}}{\partial \vec{V}_j} = \frac{\partial (R_1, R_2)}{\partial (u_j^n, v_j^{n-1/2})} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$b_{11} = \tfrac{\partial R_1}{\partial u_j^n} = u_j^{n-1}/\Delta x + (\eta_{j+1/2}^{n-1/2} + \eta_{j-1/2}^{n-1/2})/\Delta y^2 \quad b_{12} = \tfrac{\partial R_1}{\partial v_j^{n-1/2}} = (u_{j+1}^n - u_{j-1}^n)/(2\Delta y)$$
$$b_{21} = \tfrac{\partial R_2}{\partial u_j^n} = 1/(2\Delta x) \qquad\qquad b_{22} = \tfrac{\partial R_2}{\partial v_j^{n-1/2}} = 1/\Delta y$$

This block tridiagonal system can be solved with Gaussian elimination, analogous to the method presented for scalar equations (Thomas algorithm).

$$\Delta\vec{V}_j^{\nu} \;=\; \overline{\overline{E}}_j \cdot \Delta\vec{V}_{j+1}^{\nu} \;+\; \vec{F}_j$$

By substitution of $\Delta\vec{V}_{j-1}^{\nu}$ the recursion coefficients are obtained as:

$$\begin{aligned} \overline{\overline{E}}_j &= (\overline{\overline{A}}_j \cdot \overline{\overline{E}}_{j-1} + \overline{\overline{B}}_j)^{-1} \cdot (-\overline{\overline{C}}_j) \\ \vec{F}_j &= (\overline{\overline{A}}_j \cdot \overline{\overline{E}}_{j-1} + \overline{\overline{B}}_j)^{-1} \cdot (-\vec{Res} - \overline{\overline{A}}_j \cdot \vec{F}_{j-1}) \end{aligned}$$

The recursion coefficients are calculated stepwise for $j = 2, \cdots, jmax - 1$, starting from the boundary condition $\vec{V}_{j=1} = \begin{pmatrix} u_W^n \\ v_W^{n-1/2} \end{pmatrix}$ at the wall, i.e. $\Delta\vec{V}_{j=1}^{\nu} = 0$ and thus $\overline{\overline{E}}_j = 0$, $\vec{F}_j = 0$.
Therefore, the variables u and v can be determined. For their calculation it must be estimated if the area of integration is big enough, such that $y_{max} = (jmax - 1) * \Delta y > \delta$. The edge of the boundary layer δ is defined by the fact that the velocity u approaches its exterior value u_e apart from a small deviation $\epsilon_{Edge} \sim 10^{-3}$:

$$y = \delta \quad \text{if} \quad |\frac{u_e^n - u_{jmax-1}^{n,\nu+1}}{u_e^n}| < \epsilon_{Edge}$$

with $u_e^n - u_{jmax-1}^{n,\nu+1} = u_e^n - u_{jmax-1}^{n,\nu} - f_{1,jmax-1}$. If this condition doesn't hold a further step

Δy is added, i.e. $jmax$ is set to $jmax = jmax+1$. The recursion coefficients are calculated for this new point and again tested for the edge. If the condition is satisfied the variables for the iteration $\nu + 1$ are determined.
The correction variables $\Delta\vec{V}_j^{\nu}$ are obtained from the recursion of $j = jmax - 1, \cdots, 2$ with the boundary condition $u_{jmax}^n = u_e^n$, with $\Delta\vec{V}_{jmax}^{\nu} = 0$. (A boundary condition for v_{jmax} is not given and not necessary. The variables are determined from:

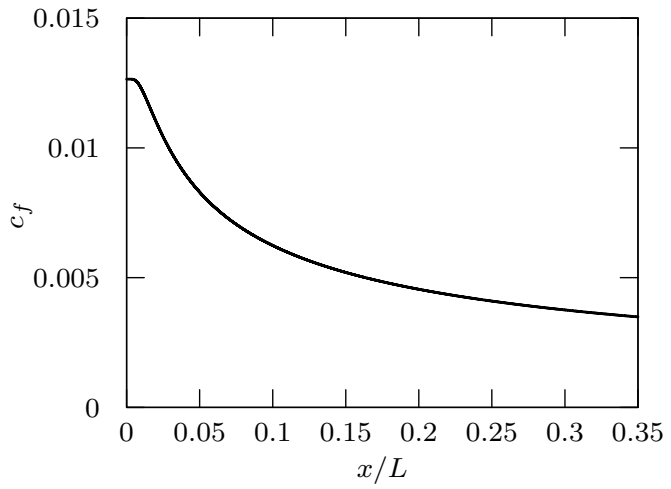
$$\vec{V}_j^{n,\nu+1} = \vec{V}_j^{n,\nu} + \Delta\vec{V}_j^{\nu}$$

The iterative solution at x^n is repeated until the equation $\vec{Res}^{\,n} = 0$ holds up to a given upper limit:

$$max|R_1, R_2| \leq \epsilon_{Res}$$

If the limit is satisfied the calculation for the next x-step can be carried out likewise.
The solution scheme for the implicit solution of a coupled equation system that has been developed in this section, is applied in a similar fashion for the implicit solution of the Euler and Navier-Stokes equations. The detailed solution steps will be dealt with in the corresponding practice with a FORTRAN program and results.

The figure displays the development of the coefficient of friction $c_f = \frac{\tau_w}{\frac{\varrho}{2}U_\infty^2}$ on the length x/L on a flat plate with laminar flow with $Re = 10^6$.



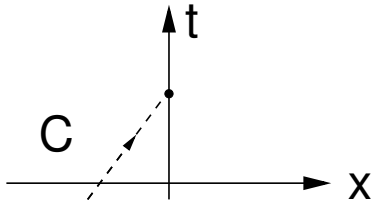
2.2 Numerical solution of scalar hyperbolic, differential equations

2.2.1 Introduction

Hyperbolic, partial differential equations have real characteristics along which the information is transported (characteristic solution, conformity condition). The characteristics determine the area that is influenced by the solution and are therefore decisive for the numerical solution schemes. Their derivation has already been shown in chapter 2.

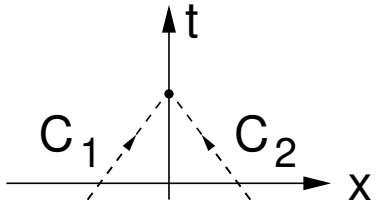
Hyperbolic differential equations can emerge in different forms. An example for first order scalar equations is the convection equation

$$w_t + \lambda\, w_x = 0$$



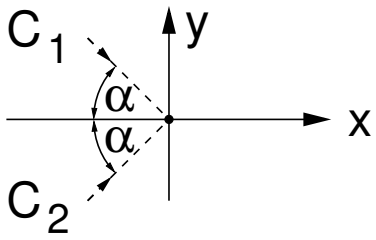
The characteristic $\frac{dx}{dt}|_1 = \lambda$ in this equation corresponds to the slope of the basic characteristic curve.
Examples for scalar equations of second order are the wave equation

$$u_{tt} - a_0^2\, u_{xx} = 0$$



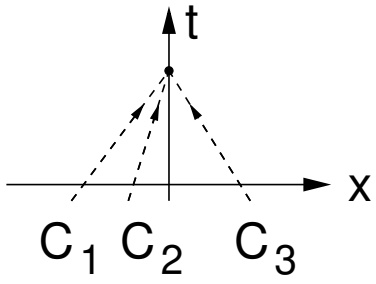
and the perturbation potential equation for supersonic flow :

$$(Ma_\infty^2 - 1)\, \Phi_{xx} - \Phi_{yy} = 0$$



These equations have two real characteristics, in this case $\frac{dx}{dt}|_{1,2} = \pm a_0$ and $\frac{dy}{dx}|_{1,2} = \pm 1/\sqrt{(Ma_\infty^2 - 1)}$ respectively. The region influenced by the solution is limited by the two characteristics.
Systems of differential equations can also be of hyperbolic type, e.g. the time dependent Euler equations for compressible flows (see chapter 1):

$$U_t + F_x = 0$$



This equation system leads to three real characteristics $\frac{dx}{dt}|_1 = u$ and $\frac{dx}{dt}|_{2,3} = u \pm a$. As before the outer characteristics limit the region of influence.

The limited region of influence of hyperbolic differential equations leads to the solution of an initial value problem. Therefore, the solution is developed from a non characteristic curve along the characteristic (method of characteristics). Boundary conditions must be predefined at the boundaries of the region since the domain of integration is usually limited. Similar to equations of parabolic type, this leads to combined initial boundary value problems. The number of necessary boundary conditions depends on the characteristics that point from the boundary to the interior of the solution domain.
For the formulation of a numerical difference scheme for a hyperbolic differential equation it is necessary to capture the region of influence in order to obtain a convergent solution. This is represented by the so called CFL condition which will be introduced in the next chapter.
A further problem results from the numerical solution of hyperbolic differential equations, describing wave transport of constant amplitude along the characteristic lines. Perturbations occur in the course of the numerical calculation caused by discretization and round off errors. Those perturbations evolve in the whole body and overlap the exact solution.

This can be avoided by introducing so called damping terms which numerically suppress the perturbations. Those numerical damping terms can either be added to the difference scheme like in the central difference scheme, or they are already included in the discretization, as it is the case in the upwind scheme. This issue will be discussed in a following section.
The most important discretization schemes for differential equations of hyperbolic type will be presented for a scalar model equation. The solution of the system of Euler equations will be discussed in an own chapter. The different forms and solution properties and numerical solution schemes will also be investigated.

2.2.2 Courant–Friedrichs–Lewy (CFL) condition

The CFL condition must be satisfied by the formulation of difference schemes for hyperbolic partial differential equations. The condition goes:

> *For the convergence of the numerical solution of initial value problems for hyperbolic partial differential equations it is necessary that the numerical domain of dependence of a difference scheme encloses the domain of dependence of the differential equation*

The dependency region of the differential equation is defined by the characteristics, whereas the computational dependency domain of the difference scheme is determined by the step size (stencil).The CFL condition demands to choose the step size such that the characteristics are inside of the stencil. With this the characteristic solution is completely captured. In a time dependent equation the characteristic is given as $\frac{dx}{dt}\big|_C = \lambda$ and the computational region of dependence is given by the step size ratio $\frac{\Delta x}{\Delta t}$. The CFL condition is then expressed with:

$$\frac{\Delta x}{\Delta t} \geq \frac{dx}{dt}\big|_C = \lambda$$

This condition is often represented as the non dimensional Courant number C:

$$C = \frac{dx}{dt}\big|_C \cdot \frac{\Delta t}{\Delta x} = \lambda \frac{\Delta t}{\Delta x} \leq 1$$

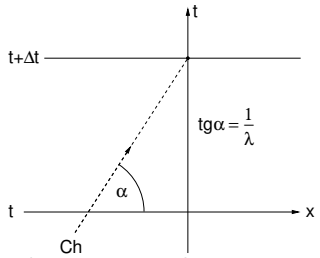
The Courant number C can be understood as the ratio of exact information rate $\frac{dx}{dt}\big|_C = \lambda$ to computational information rate $\frac{\Delta x}{\Delta t}$.

To estimate the Courant number for equations with multiple characteristics, e.g. the Euler equations with $\frac{dx}{dt}\big|_C = (u\,,\, u+a\,,\, u-a)$, the biggest absolute value must be chosen, i.e.:

$$C = max\left(\left|\frac{dx}{dt}\big|_C\right|\right) \cdot \frac{\Delta t}{\Delta x} = (|u|+a)\frac{\Delta t}{\Delta x} \leq 1$$

In the following the meaning of the CFL condition will be shown with a scalar model equation.

Example:



The scalar convection equation will be considered:

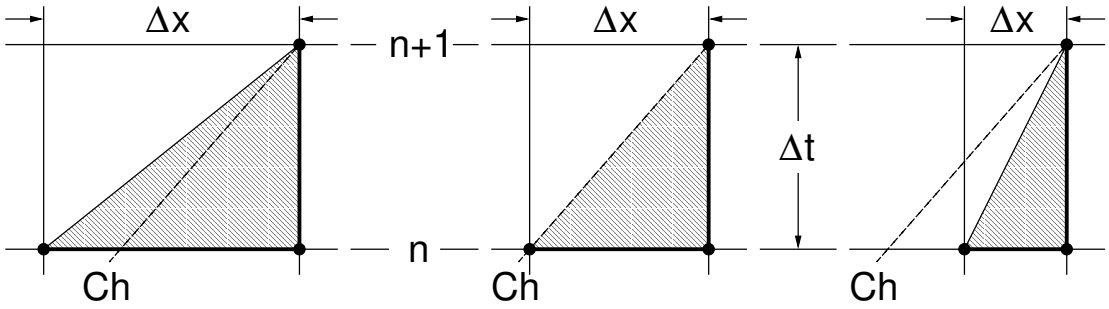
$$w_t + \lambda\, w_x = 0 \quad \text{with} \quad \lambda = const. \quad > 0$$

The characteristic of this equation is given as $\frac{dx}{dt}|_C = \lambda$ and the exact solution is given as $w(x,t) = w(x-\lambda t)$. The region of influence is the straight line $x - \lambda t = const.$ The Courant number is defined as $C = \lambda \frac{\Delta t}{\Delta x}$.

Difference scheme:

a) explicit scheme, backward difference for w_x

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_i^n - w_{i-1}^n}{\Delta x} = 0$$



CFL condition satisfied	CFL condition satisfied	CFL condition not satisfied
$\frac{\Delta x}{\Delta t} > \frac{dx}{dt}\|_C = \lambda$	$\frac{\Delta x}{\Delta t} = \frac{dx}{dt}\|_C = \lambda$	$\frac{\Delta x}{\Delta t} < \frac{dx}{dt}\|_C = \lambda$
$C < 1$	$C = 1$	$C > 1$

von Neumann stability analysis stable scheme for $C \leq 1$

b) explicit scheme, forward difference for w_x

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^n - w_i^n}{\Delta x} = 0$$

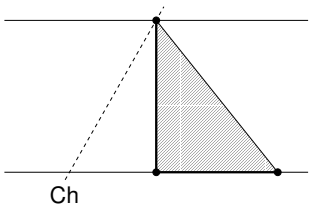
CFL condition: not satisfied
Neumann stability analysis unstable scheme for all C

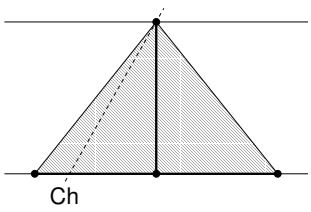
c) explicit scheme, central difference for w_x

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^n - w_{i-1}^n}{2\Delta x} = 0$$

CFL condition: satisfied for $C \leq 1$
von Neumann stability analysis unstable scheme for all C

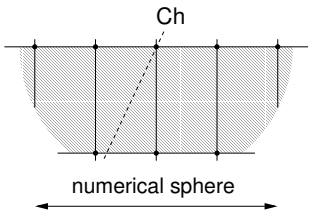
$\Longrightarrow$i.e. CFL condition only necessary, but not sufficient!





d) implicit scheme, central difference for w_x

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^{n+1} - w_{i-1}^{n+1}}{2\Delta x} = 0$$



CFL condition: always satisfied

$\Longrightarrow$values w^{n+1} are linked by x

von Neumann stability analysis absolutely stable for all C

2.2.3 Numerical damping

The numerical damping describes the dissipative effects of the numerical discretization error. The dissipation leads to a smoothing (smearing out) of the solution, similar to the effects of viscosity. There is a big impact on numerical solutions of hyperbolic partial differential equations since the exact, characteristic solution only allows wave transport, but no dissipation. Therefore, the numerical damping in the discretization can cause a distorted solution.

Unwanted effects of the numerical damping
on the numerical solution of hyperbolic problems are e.g.:

- smearing out of the solution
- artifical vortex production and decay
- artifical entropy changes

Therefore, the numerical damping should be reduced to a minimum.

On the other hand, each numerical calculation leads to perturbations, e.g. caused by round off errors which can be amplified and overlay the exact solution. In order to avoid the spreading of these perturbations in the whole solution domain, they must be damped

out in the course of the computation. In this respect the numerical damping can be used. Desired effects of numerical damping

on the numerical solution of hyperbolic problems are:

- Damping of numerical perturbations in the solution domain

An accurate numerical solution can therefore not be obtained without a certain damping of the discretization.

Numerical damping should be as small as possible and as big as necessary

It is important to follow this demand to achieve stability and accuracy for the numerical solution. But it also takes very good numerical knowledge and experience.

Again a scalar model equation will be considered to discuss the effects of numerical damping. Example of the numerical discretization error: The scalar convection equation

will be considered:

$$w_t + \lambda\, w_x \;=\; 0 \qquad \text{where} \quad \lambda = const. > 0$$

The exact solution for a periodic test function $w(x,t) \;=\; V(t) \cdot e^{Ikx}$ where the wave number is $k = 2\pi/\lambda$ is:

$$w(x,t) = V_0 \cdot e^{Ik(x-\lambda\Delta t)}$$

As a common property of hyperbolic equations the solution describes the transport along the characteristic base curve $(x - \lambda\Delta t) = const$, but it doesn't describe a variation of the amplitude, i.e. $V(t) = V_0$.

To enable a comparison a numerical solution, calculated with an explicit upwind scheme, shall be examined.

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} \;+\; \lambda\, \frac{w_i^n - w_{i-1}^n}{\Delta x} \;=\; 0$$

Taylor expansion of the difference scheme around $x_i = i\cdot\Delta x$ and $t_n = n\cdot\Delta t$ and substitution with the differential equation $w_{tt} = \lambda^2\, w_{xx} + \cdots$ leads to the partial differential equation of the numerical approximation (see also chapter 3):

$$w_t + \lambda\, w_x \;=\; c_2\Delta x \cdot w_{xx} \;+\; c_3\Delta x^2 \cdot w_{xxx} \;-\; c_4\Delta x^3 \cdot w_{xxxx} \;+\; \cdots$$

The solution of the periodic test function for this equation is:

$$w(x,t) = V_0 \cdot e^{Ik(x-\lambda\Delta t)} \cdot e^{-Ic_3\Delta x^2 k^3\cdot\Delta t} \cdot e^{-(c_2\Delta x k^2 + c_4\Delta x^3 k^4)\cdot\Delta t}$$

This solution demonstrates the effects of typical discretization errors, which occur in similar appearance in nearly all numerical methods for hyperbolic equations.

The possible error types are:

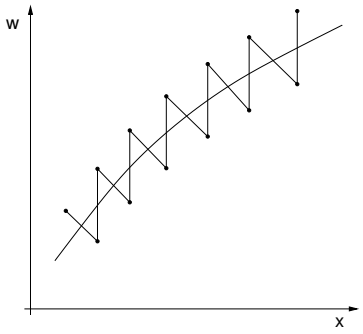
- dispersive errors $\sim e^{-I(c_3 \Delta x^2 k^3)\cdot \Delta t}$ of the term $\sim c_3 \Delta x^2 \, w_{xxx}$
  These errors cause a phase shift , i.e. a deviation of the characteristic ground curve without an influence on the amplitude.

- Dissipative errors $\sim e^{-(c_2 \Delta x k^2)\cdot \Delta t}$ of the term $\sim c_2 \Delta x \, w_{xx}$
  These errors have an impact on the amplitude and have similar effects as the friction terms. Since the term is of $O(\Delta x)$, it should be avoided by using higher order discretizations.

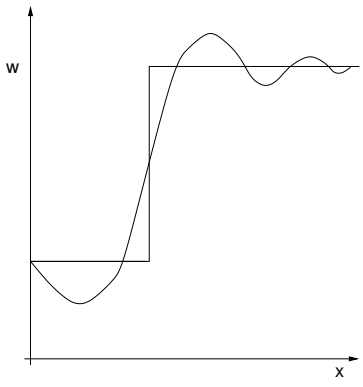
- Dissipative errors $\sim e^{-(c_4 \Delta x^3 k^4)\cdot \Delta t}$ of the term $\sim c_4 \Delta x^3 \, w_{xxxx}$
  This error has also an effect on the amplitude, but it is limited to the high frequency parts of the solution, since the amplitudes are strongly dependent on the wave number ($\sim k^4$!). Therefore, this term is often referred to as high frequency damping term.

The different effects of the dissipative discretization error are used in the formulation of numerical damping to suppress perturbations.

<u>High frequency damping terms:</u>
Round off errors cause fluctuations between adjacent grid points, i.e. they are short waved perturbations. To smooth out those perturbations, terms resulting from a discretized fourth order derivation are used. *High frequency damping term:* $D^{(4)} = \varepsilon^{(4)} \Delta x^3 \, w_{xxxx}$
The often constant factor $\varepsilon^{(4)}$ serves the purpose of adapting and minimizing the damping.



<u>Shock damping terms:</u>
In solutions of hyperbolic equations discontinuities, like shock waves, can occur. If these are embedded in the solution area very strong changes in the variables over few step sizes occur. This causes large discretization errors, especially in the non linear terms. The solution displays strong deviations near the discontinuity. To smooth out these deviations a strong damping term is needed. As a common means the term $\sim \Delta x w_{xx}$ which is similar to the friction term, is used. It can be fine tuned by the factor $\varepsilon^{(2)}$, depending on the solution. This term diminishes in regions of small deviations.



Shock damping term: $D^{(2)} = \varepsilon^{(2)} \Delta x \, w_{xx}$

The damping terms that control the numerical perturbations are either added to the scheme (central schemes) or they are (for suitable discretizations) already included (upwind schemes). In the context of this course some examples will be presented in chapter

8. Further details of the formulation won't be discussed here. These formulations are the object of recent numerical developments and can be found in specialist literature.

2.2.4 Important difference schemes for the scalar convection equation

The scalar convection equation:

$$w_t + \lambda w_x = 0$$

will be used to display numerical solution schemes and their analysis. This equation is associated to the important Euler equations, since its structure reassembles the characteristic form of the Euler equations. The formerly discussed basics lead to the following demands concerning solution schemes:

1. Convergence
   According to the Laxian theorem convergence can be proved for linear initial value problems, i.e. *consistency + stability = convergence*. Such a proof is often not given for non linear equations and boundary value problems. The validity of the solution is often proved by comparison with other solutions or experiments.

2. CFL condition
   The necessary demand that the numerical region of dependency is bigger or equal to the characteristic region, can be uniquely satisfied for scalar equations with only a single characteristic. Problems arise if multiple characteristics with alternating signs occur, as it is the case with the Euler equations. Special formulations of the original equations and the discretization must be used to take into account the different characteristic directions of expansion (E.g. "Flux vector splitting " $\Longrightarrow$see chapter 7 and 8 on Euler equations).

3. Accuracy
   In applications of numerical computations a minimal accuracy of second order ($O(\Delta^2)$ should be applied for the spatial discretization to avoid the numerical viscosity effects of the discretization error $\sim w_{xx}$ in smooth areas.

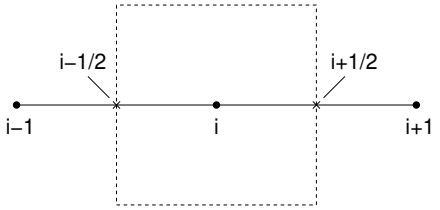
4. Non oscillatory solutions
   To obtain non oscillatory solutions in smooth areas, short waved error components must be suppressed by suitable high frequency damping terms.

5. Discontinuities
   Discontinuities (e.g. shock waves) are solutions of non linear hyperbolic equations. They occur especially in hypersonic flows. Discontinuities that are embedded in the domain of integration can not be exactly (as a jump) dissolved by difference schemes, since those schemes demand continuously differentiable equations. Numerically however, such discontinuities can be expressed without oscillations within few step sizes by applying suitable so called "shock capturing" methods. This includes the consideration of the characteristic direction of expansion and the formulation of shock damping terms.

Several difference schemes can be applied to satisfy this demands. A so called conservative formulation, like it is necessary for the Euler equations, will be used for the discrete formulation of the convection equation $w_t + \lambda w_x = 0$.

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1/2}^n - w_{i-1/2}^n}{\Delta x} = 0$$

The location $i \pm 1/2$ corresponds to the location of the cell surface between the points i and $i \pm 1$ where the Euler fluxes must be formulated.



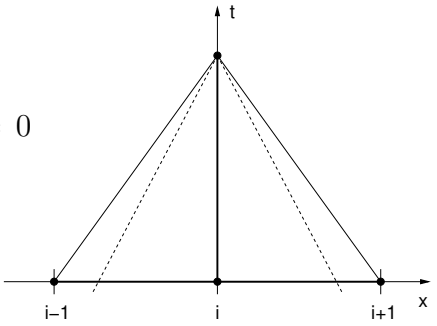
Central schemes

Central schemes are often used schemes for the solution of the Euler and Navier Stokes equations. A spatial central difference can be obtained for the formerly discussed conservative scheme by formulating mean values:

$$w_{i\pm1/2} = \frac{1}{2}\left(w_i + w_{i\pm1}\right)$$

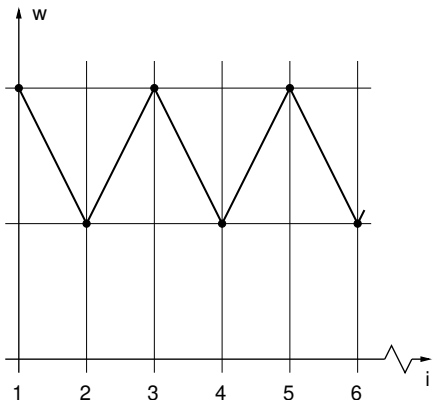
This leads to the following central scheme:

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^n - w_{i-1}^n}{2\Delta x} = 0$$



Typical properties of central schemes are:

- Second order approximation in space $O(\Delta x^2)$ with only 3 points.
- Capturing of upstream and downstream influences, of positive and negative characteristics
- Central differences cause decoupling of even and odd numbered points. E.g. the stationary solution $w_{i+1} = w_{i-1}$ has two decoupled solutions, namely $w_1 = w_3 = w_5 = \cdots$ and $w_2 = w_4 = w_6 = \cdots$. A difference between the two solutions which can be expressed as ε leads to an oscillating overall solution.
- Central space differences do not have a dissipative truncation error. $\frac{w_{i+1} - w_{i-1}}{2\Delta x} = w_x + w_{xxx} \cdot \frac{\Delta x^2}{6} + \cdots$
- Therefore, high frequency damping terms $\sim w_{xxxx}$ must be added to the central difference to damp short waved perturbations.



$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1} - w_{i-1}}{2\Delta x} + D^{(4)}(w) = 0$$

The fourth order damping is often formulated as:

$$D^{(4)}(w) = \varepsilon^{(4)} \frac{1}{\Delta t} \Delta x^4 \cdot w_{xxxx} = \varepsilon^{(4)} \frac{1}{\Delta t} \cdot (w_{i+2} - 4w_{i+1} + 6w_i - 4w_{i-1} + w_{i-2})$$

Where $\varepsilon^{(4)}$ is a constant, with a usual magnitude around $O(10^{-2})$
In the following central schemes this damping term won't be displayed separately.

Explicit central schemes

a) Basic central scheme

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^n - w_{i-1}^n}{2\Delta x} = 0$$

Consistency $w_t + \lambda w_x = -w_{tt} \frac{\Delta t}{2} - \lambda w_{xxx} \frac{\Delta x^2}{6} + \cdots = O(\Delta t, \Delta x^2)$

Stability : (von Neumann analysis) $\rightarrow$ instable !

(In spite of the satisfied CFL condition at $C \leq 1$)

Explanation of the instability by the Hirtsche stability analysis

with $w_{tt} = -\lambda w_{xt} = \lambda^2 w_{xx}$ one obtains $w_t + \lambda w_x = -\lambda^2 \frac{\Delta t}{2} w_{xx} - \lambda w_{xxx} \frac{\Delta x^2}{6} + \cdots$

I.e. negative numerical viscosity $\nu_{num} = -\lambda^2 \frac{\Delta t}{2}$ which has an activating effect on the flow.

b) Lax - Keller scheme

To obtain a simple but stable scheme the value w_i^n in the scheme a) will be replaced by the mean value $\bar{w}_i^n = \frac{w_{i+1}^n + w_{i-1}^n}{2}$.

$$\frac{w_i^{n+1} - \frac{1}{2}(w_{i+1}^n + w_{i-1}^n)}{\Delta t} + \lambda \frac{w_{i+1}^n - w_{i-1}^n}{2\Delta x} = 0$$

Consistency : $w_t + \lambda w_x = -w_{tt} \frac{\Delta t}{2} + \frac{\Delta x}{2} \cdot \frac{\Delta x}{\Delta t} w_{xx} - \lambda w_{xxx} \frac{\Delta x^2}{6} + \cdots = O(\Delta t, \Delta x)$

Stability : stable for $C = |\lambda| \frac{\Delta t}{\Delta x} \leq 1$

Explanation of stability by Hirtsche stability analysis

$w_t + \lambda w_x = \lambda^2 \frac{\Delta t}{2} (1/C^2 - 1) w_{xx} + \cdots$

$\Longrightarrow \nu_{num} = \lambda^2 \frac{\Delta t}{2} (1/C^2 - 1) \geq 0$ for $C \leq 1$

$\Rightarrow$ Rarely applied scheme, since $O(\Delta x)$!

c) Lax - Wendroff scheme

The Lax Wendroff scheme is an exact and stable central scheme, it is therefore used as a starting point for further schemes. The destabilizing term $\lambda^2 \frac{\Delta t}{2} w_{xx}$ of the scheme a) is compensated by an additional, equally sized term.

The derivation is done with the Taylor series expansion for w_i^{n+1}: $w_i^{n+1} = w_i^n + w_t|_i^n \Delta t + w_{tt}|_i^n \frac{\Delta t^2}{2} + O(\Delta t^3) = w_i^n - \lambda w_x|_i^n \Delta t + \lambda^2 w_{xx}|_i^n \frac{\Delta t^2}{2} + O(\Delta t^3)$

The additional term $\sim w_{xx}$ uses a central discretization, such that the spatial accuracy $O(\Delta x^2)$ can be preserved and the temporal accuracy $O(\Delta t^2)$ is increased. This leads to the Lax - Wendroff scheme as :

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^n - w_{i-1}^n}{2\Delta x} - \lambda^2 \frac{\Delta t}{2} \frac{w_{i-1}^n - 2w_i^n - w_{i+1}^n}{\Delta x^2} = 0$$

Consistency :

$w_t + \lambda w_x = -\lambda w_{xxx} \frac{\Delta x^2}{6} - \frac{\Delta t^2}{6} w_{ttt} + \cdots = \underline{O(\Delta t^2, \Delta x^2)}$

Stability : stable for $C = |\lambda| \frac{\Delta t}{\Delta x} \leq 1$

Disadvantage The additional term requires costly matrix operations for equations systems like e.g. the Euler equations $U_t + F_x = 0$. With the Jacobian $\bar{\bar{A}} = \frac{\partial F}{\partial U}$ the additional term becomes
$\Longrightarrow \frac{\Delta t}{2} U_{tt} = -\frac{\Delta t}{2} F_{xt} = -\frac{\Delta t}{2} (F_t)_x = -\frac{\Delta t}{2} (\bar{\bar{A}} U_t)_x = +\frac{\Delta t}{2} (\bar{\bar{A}} F_x)_x$

It is therefore more beneficial for equation systems to execute the scheme in two steps. This results in the so called predictor - corrector scheme by Mac Cormack.

d) Predictor - corrector scheme (Mac Cormack, 1969)

The predictor - corrector scheme by Mac Cormack is an often applied scheme for the solution of the Euler and Navier Stokes equations. For linear equation systems the two step scheme has the same features like the Lax-Wendroff scheme c), but it doesn't require additional matrix operations for equation systems. The stability and consistency behavior is equal to the Lax-Wendroff scheme. The Predictor - Corrector scheme can be obtained by substitution of the variable $\tilde{w}$ from the first step in the second step.

1st Step (Predictor step)

$$\tilde{w}_i = w_i^n - \lambda \frac{\Delta t}{\Delta x} (w_i^n - w_{i-1}^n)$$

2nd Step (Corrector step)

$$w_i^{n+1} = \frac{1}{2}(w_i^n + \tilde{w}_i) - \frac{1}{2} \lambda \frac{\Delta t}{\Delta x} (\tilde{w}_{i+1} - \tilde{w}_i)$$

The forward and backward discretizations for the steps can be exchanged. The extension on two and three dimension can be performed in the same way.

e) Runge-Kutta scheme

The Runge-Kutta scheme for the solution of initial value problems of ordinary differential equations can be transferred to partial differential equations. For the integration according to Runge-Kutta the semi discrete differential equation is formulated as

$$\frac{\partial w}{\partial t} = -Res(w)$$

where the residual $Res(w)$ represents the discretized operator of the spatial derivatives. This could be the spatial operator of the complete Euler equations, or as in the present case, the operator of the scalar convection equation.

$$Res(w) = \lambda \frac{w_{i+1} - w_{i-1}}{2\Delta x} = \lambda w_x + O(\Delta x^2)$$

The solution for a time step Δt is performed in several explicit steps which are marked with the index k. There are many different variants for the multi stage formulation. The following scheme has been proved useful for the solution of partial differential equations of fluid dynamics (minimal memory requirements). For the integration domain between $t = n\,\Delta t$ and $t + \Delta t = (n+1)\,\Delta t$ follows:

$$\begin{array}{lcll} w_i^{(0)} & = & w_i^n & \\ w_i^{(1)} & = & w_i^{(0)} - \alpha_1 & \cdot \Delta t \cdot Res\,(w^{(0)}) \\ \vdots & & & \\ w_i^{(k-1)} & = & w_i^{(0)} - \alpha_{k-1} & \cdot \Delta t \cdot Res\,(w^{(k-2)}) \\ w_i^{(k)} & = & w_i^{(0)} - \alpha_k & \cdot \Delta t \cdot Res\,(w^{(k-1)}) \\ \vdots & & & \\ w_i^{(n+1)} & = & w_i^{(N)} & \end{array}$$

The number of steps N is chosen between 3 and 5.

Consistency : In the following the consistency shall be shown with a 3 step scheme $(N = 3)$. The variables $w^{(k)}$ of the intermediate steps are eliminated, which leads to the differential equation, assuming the linear space operators $Res(w)$, i.e. $Res(a + b) = Res\,a + Res\,b$. The differential equation results in

$$w^{n+1} = w^n - \alpha_3\Delta t Res(w^n) + \alpha_3\alpha_2\Delta t^2 Res(Res(w^n)) - \alpha_3\alpha_2\alpha_1\Delta t^3 Res(Res(Res(w^n)))$$

Using a Taylor series expansion

$$w^{n+1} = w^n + w_t^n\,\Delta t + w_{tt}^n\,\frac{\Delta t^2}{2} + w_{ttt}^n\,\frac{\Delta t^3}{6} + \cdots$$

and the original equations

$$\begin{array}{rcl} w_t^n & = & -Res(w^n) = -(\lambda\, w_x + O(\Delta x^2)) \\ w_{tt}^n & = & (-Res(w^n))_t = -Res(w_t^n) = Res(Res(w^n)) \\ w_{ttt}^n & = & -Res(Res(Res(w^n))) \end{array}$$

one obtains the differential equation of the difference approximation:

$$w_t + \lambda\, w_x = (\alpha_3 - 1)\, w_t^n + (\alpha_3\alpha_2 - \frac{1}{2})\, w_{tt}^n\,\Delta t + (\alpha_3\alpha_2\alpha_1 - \frac{1}{6})\, w_{ttt}^n\,\Delta t^2 + O(\Delta t^3) + O(\Delta x^2)$$

This equation shows the consistency of the Runge Kutta scheme for $\alpha_3 = 1$ in time. The temporal accuracy depends on the choice of the other α factors. The coefficients

($\alpha_k \leq 1$) can be formulated such that the truncation error in time is minimized. Therefore, the temporal accuracy is of order $O(\Delta t^N)$. In the above example this leads to the values $\alpha_k = \frac{1}{3}, \frac{1}{2}, 1$ or in general:

$$\alpha_k = \frac{1}{N-k+1} \quad \text{where} \quad k = 1, 2, \cdots N$$

Stability :

The stability factor $G = \frac{V^{n+1}}{V^n}$ according to the von Neumann stability analysis can also be obtained by insertion of the intermediate steps. Using the Fourier expansion $w_i^{(k)} = V^{(k)}\, e^{I\theta i}$ and the abbreviation $\Delta t Res(w^{(k)}) \rightarrow V^{(k)}\, e^{I\theta i}\, \frac{\lambda \Delta t}{\Delta x} I \sin\theta = V^{(k)}\, e^{I\theta i}\, z$ one obtains for the 3 step scheme:

$$G = 1 - \alpha_3\, z + \alpha_3\alpha_2\, z^2 - \alpha_3\alpha_2\alpha_1\, z^3$$

The estimation of the stability limit is rather costly and generally performed numerically. An interesting possibility in this respect is the optimization of the coefficients α_k for maximum stability. The theoretical stability limit for the N step scheme results in:

$$C_{max} = (|\lambda| \frac{\Delta t}{\Delta x})_{max} = N - 1$$

An often applied set of coefficients for maximum stability of a central 5 step scheme of order $O(\Delta t^2, \Delta x^2)$ and $C_{max} = 4$ is: $\alpha_k = 0.25\,,\ 0.166\,,\ 0.375\,,\ 0.5\,,\ 1$

The Runge Kutta method in the present formulation is today one of the most applied explicit solution schemes for the Euler and Navier Stokes equations of compressible flows.

Implicit central schemes

In contrast to explicit schemes implicit schemes are free of time step limitations due to numerical instabilities. On the other hand their computational cost per time step is substantially higher. Therefore, implicit schemes are normally chosen for calculations with noticeable larger time steps than explicit schemes. This is often the case when convergence to a stationary solution is demanded.

a) Implicit scheme for one dimensional equations

An implicit scheme is obtained when the space operator for the new time step $(n+1)\Delta t$ is formulated. For the scalar convection equations such a scheme is:

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1}^{n+1} - w_{i-1}^{n+1}}{2\Delta x} = 0$$

This scheme leads to a tridiagonal equation system for w^{n+1} with $C = \lambda \frac{\Delta t}{\Delta x}$.

$$[-C/2] \cdot w_{i-1}^{n+1} + [1] \cdot w_i^{n+1} + [C/2] \cdot w_{i+1}^{n+1} = w_i^n$$

The solution of such a coupled equation system can be performed with the well known Gaussian elimination (Thomas-Algorithmus).

Consistency: $w_t + \lambda w_x = \frac{\Delta t}{2} w_{tt} - \lambda w_{xxx} \frac{\Delta x^2}{6} + \cdots = O(\Delta t, \Delta x^2)$

Stability : unlimited stability for all C

Implicit schemes can often be found in the literature in the so called correction formulation, with the definition of the correction variables as :

$$\Delta w_i^n = w_i^{n+1} - w_i^n$$

with the additional definition of the difference operator δ_x

$$\delta_x w_i = \frac{w_{i+1} - w_{i-1}}{2\Delta x}$$

the implicit scheme can be written as

$$\begin{aligned} (1 + \Delta t \lambda \delta_x) \Delta w_i^n &= -\Delta t \lambda \delta_x w_i^n \\ w_i^{n+1} &= w_i^n + \Delta w_i^n \end{aligned}$$

The solution of the tridiagonal equation system is performed for the correction variable.

b) Implicit schemes for multi dimensional equations

The formulation for two or three dimensions is carried out in a similar fashion like the one dimensional case. The implicit scheme for two dimensions can be obtained by defining the two dimensional convection equation:

$$w_t + \lambda_x w_x + \lambda_y w_y = 0$$

and discretization in y direction with central differences:

$$\delta_y w_{i,j} = \frac{w_{i,j+1} - w_{i,j-1}}{2\Delta y}$$

. When using the correction variables

$$\Delta w_{i,j}^n = w_{i,j}^{n+1} - w_{i,j}^n$$

the two dimensional implicit equation is given as:

$$\begin{aligned} (1 + \Delta t (\lambda_x \delta_x + \lambda_y \delta_y)) \Delta w_{i,j}^n &= -\Delta t (\lambda_x \delta_x + \lambda_y \delta_y) w_{i,j}^n \\ w_{i,j}^{n+1} &= w_{i,j}^n + \Delta w_{i,j}^n \end{aligned}$$

The solution of two dimensional, coupled equation systems has a high memory and computation requirement for the generally large computational grids in fluid dynamical calculations. Thus, one often tries to solve these multi dimensional systems approximatively.
One possible approach for such a solution is the application of iteration schemes like they are used for elliptical equations (e.g. Gauss Seidel method). Such iterative schemes become more and more common in the solution of the Euler and Navier Stokes equations. One prerequisite for such iterative schemes is a diagonal dominant solution matrix. Such matrices normally result from the discretization of upwind schemes which lead to very effective solution methods for steady problems. Iteration schemes are usually not suitable for time accurate problems, since they are not consistent in time.
Another possibility for the approximative solution is the method of approximate factorization which also allows a time consistent solution and central discretization.

c) Implicit approximate factorization (Beam, Warming 1970)

The approximative factorization enables the decomposition of multi dimensional, implicit schemes in a sequence of one dimensional steps. Since each one dimensional step can be solved as a tridiagonal computational cost can be saved. Starting point for this method is the above formulated two dimensional, implicit scheme. The (implicit) left hand side can approximatively be decomposed in two factors:

$$(1 + \Delta t\,\lambda_x\,\delta_x)\,(1 + \Delta t\,\lambda_y\,\delta_y)\,\Delta w^n_{i,j} = (1 + \Delta t\,\lambda_x\,\delta_x + \Delta t\,\lambda_y\,\delta_y + O(\Delta t^2))\,\Delta w^n_{i,j}$$

If the last term is neglected (factorization error $O(\Delta t^2)$), the implicit scheme can be written as:

$$(1 + \Delta t\,\lambda_x\,\delta_x)\,(1 + \Delta t\,\lambda_y\,\delta_y)\,\Delta w^n_{i,j} = -\Delta t\,(\lambda_x\,\delta_x + \lambda_y\delta_y)\,w^n_{i,j}$$

If one defines the temporary occurring variable $\Delta\tilde{w}^n_{i,j}$:

$$\Delta\tilde{w}^n_{i,j} = (1 + \Delta t\,\lambda_y\,\delta_y)\,\Delta w^n_{i,j}$$

the scheme for a time step can be subdivided in several smaller steps:

1. Step
$$(1 + \Delta t\,\lambda_x\,\delta_x)\,\Delta\tilde{w}^n_{i,j} = -\Delta t\,(\lambda_x\,\delta_x + \lambda_y\delta_y)\,w^n_{i,j}$$
A tridiagonal equation system is solved in x-direction for $\Delta\tilde{w}^n_{i,j}$.
2. Step
$$(1 + \Delta t\,\lambda_y\,\delta_y)\,\Delta w^n_{i,j} = \Delta\tilde{w}^n_{i,j}$$
A tridiagonal equation system is solved in y-direction for $\Delta\tilde{w}^n_{i,j}$.
3. Step
$$w^{n+1}_{i,j} = w^n_{i,j} + \Delta w^n_{i,j}$$

The Method of approximative factorization is an important implicit solution method for systems like e.g. the Euler equations. A drawback of this method is the factorization error which causes a decreased convergence speed for big time steps. Therefore, in realistic computations only maximum CFL numbers of $C = O(10)$ can be applied.

Upwind schemes

Upwind schemes, also known as advective schemes, are useful in spatial discretizations where the difference approximation is built single-sided from the direction of the characteristic (Information from the "wind direction" $\Longrightarrow$upwind). This results in a better representation of the characteristic behavior of hyperbolic equations than using central differences. Therefore, upwind schemes are increasingly used in the simulation of gas dynamic flows with numerical solutions of the Euler equations.

Typical features of upwind schemes are:

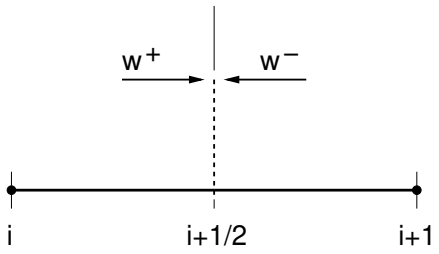
- The spatial discretization is different for positive and negative characteristics.
- Equation systems of multiple characteristics with different signs require the separation of the flux formulation according to the different domains of influence and the corresponding upwind discretization (see Euler equations $\Longrightarrow$Flux splitting).
- Upwind schemes of first order result in non oscillatory solutions also for discontinuities. But the solution is too inexact (smeared out) because of the truncation error $O(\Delta x)$.
- Higher order upwind schemes can be constructed with extrapolation methods.
- Because of their one-sided difference formulation, upwind schemes include dissipative truncation errors $\sim w_{xx}$ (only for first order discretization) and parts of $\sim w_{xxxx}$. Artificial damping terms are unnecessary, the degree of damping is fixed by the discretization.
- Higher order upwind schemes require additional discretization elements to enable the oscillation free, exact representation of discontinuities (TVD methods, Limiter,...). For details please refer to the specialized literature.

a) First order upwind schemes

Starting point is the conservative discretization of the scalar convection equation $w_t + \lambda\, w_x = 0$:

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \lambda \frac{w_{i+1/2}^n - w_{i-1/2}^n}{\Delta x} = 0$$

Since the discretization depends on the sign of the characteristic, in the following the values with $i \pm 1/2$ are declared as $w_{i\pm1/2}^+$ for positive characteristics and $w_{i\pm1/2}^-$ for negative characteristics. For characteristics $\frac{dx}{dt} = \lambda > 0$ the values $w_{i\pm1/2}$ are replaced by the node values from the direction of the (left) characteristic.



$$w_{i+1/2}^+ = w_i \qquad \text{and} \qquad w_{i-1/2}^+ = w_{i-1}$$

This results in an explicit scheme with a spatial backward difference.

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \frac{\lambda}{\Delta x}\,(\, w_i^n - w_{i-1}^n\,) = 0$$

In analogy to negative characteristics $\frac{dx}{dt} = \lambda < 0$, the values $w_{i\pm1/2}$ are replaced by the node values of the direction of the (right) characteristic, i.e.

$$w_{i+1/2}^- = w_{i+1} \qquad \text{and} \qquad w_{i-1/2}^- = w_i$$

This leads to an explicit scheme with a spatial forward difference.

$$\frac{w_i^{n+1} - w_i^n}{\Delta t} + \frac{\lambda}{\Delta x}\,(\, w_{i+1}^n - w_i^n\,) = 0$$

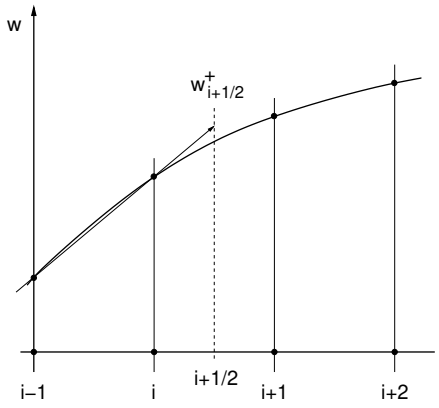
Consistency $\quad w_t + \lambda\, w_x = -\, w_{tt}\, \frac{\Delta t}{2} + |\lambda| \cdot w_{xx}\, \frac{\Delta x}{2} + \cdots = O\,(\Delta t, \Delta x)$

Stability $\quad$ stable for $\quad C = \lambda\, \frac{\Delta t}{\Delta x} \le 1$

b) Higher order upwind schemes

Higher order upwind schemes are obtained by extrapolation of the variables over several values on the "cell walls" $i \pm 1/2$. The sought after values $w_{i\pm1/2}^{\pm}$ can be represented by a polynomial, built from the neighboring values. For a second or third order scheme four nodes are required for the polynomial. $\quad w_{i+1/2}^+ = P\,(\, w_{i-2},\, w_{i-1},\, w_i,\, w_{i+1})$

$$w_{i+1/2}^- = P\,(\, w_{i-1},\, w_i,\, w_{i+1},\, w_{i+2})$$



An often applied polynomial formulation by *van Leer* is based on a Legendre polynomial. This extrapolation approach goes as follows:

$$\begin{aligned}(w_{i+1/2})^+ &= w_i + \frac{1}{4}\varphi \cdot [(1+\kappa)(w_{i+1} - w_i) + (1-\kappa)(w_i - w_{i-1})] \\ (w_{i+1/2})^- &= w_{i+1} - \frac{1}{4}\varphi \cdot [(1+\kappa)(w_{i+1} - w_i) + (1-\kappa)(w_{i+2} - w_{i+1})]\end{aligned}$$

These formulations are inserted in the conservative discretization scheme. Two parameters are used to setup the polynomial.

- The parameter φ allows to switch the scheme from first order ($\varphi = 0$) to at least second order ($\varphi = 1$).
  Note: The parameter φ is used in so called TVD schemes to regulate the numerical damping term in order to enable a very high resolution of shock waves. In this case φ is the so called limiter function which varies, depending on the neighboring gradients, between the values 1 and 0.

- The scheme is defined by the discretization parameter κ. The following formulations for the scheme are possible:

$$\begin{array}{llll} \varphi = 0 \quad . & \Rightarrow & O(\Delta x) & \text{1st order upwind} \\ \\ \varphi = 1 \quad \kappa = -1 & \Rightarrow & O(\Delta x^2) & \text{full upwind} \\ \varphi = 1 \quad \kappa = 0 & \Rightarrow & O(\Delta x^2) & \text{"half" upwind} \\ \varphi = 1 \quad \kappa = 1/3 & \Rightarrow & O(\Delta x^3) & \text{"half" upwind} \\ \varphi = 1 \quad \kappa = 1 & \Rightarrow & O(\Delta x^2) & \text{central} \end{array}$$

To check the accuracy of the higher order upwind scheme a truncation error analysis can be applied. This shall be demonstrated with the calculation of the truncation error of the difference $\frac{w^+_{i+1/2} - w^+_{i-1/2}}{\Delta x}$ for $\lambda > 0$. The above extrapolation relation is applied to the values $w^+_{i\pm 1/2}$ and the single differences are represented by Taylor series expansion. This leads to the following relation for the spatial truncation error $\tau = (w_x - \frac{w^+_{i+1/2} - w^+_{i-1/2}}{\Delta x})$:

$$\tau = (1-\varphi)\frac{\Delta x}{2} w_{xx} - [1 - \frac{3}{2}\varphi(1-\kappa)]\frac{\Delta x^2}{6} w_{xxx} - [3\varphi(1-\kappa) - (1-\varphi)]\frac{\Delta x^3}{24} w_{xxxx} + \ldots$$

This relation shows that the difference is of first order accuracy for $\varphi \neq 1$ and that the numerical viscosity $\nu_{num} = (1-\varphi)\frac{\Delta x}{2}$ is controlled by the parameter φ. The difference is for all parameter κ of at least second order accuracy if $\varphi = 1$. Third order accuracy is obtained if the term $\sim w_{xxx}$ vanishes, i.e$\kappa = \frac{1}{3}$.

This formulation of upwind schemes is the foundation for many modern solution methods for the Euler equations.

The upwind discretization can be implemented in explicit, as well as in implicit schemes. These solution schemes correspond to the methods that were presented for central differences. The Runge Kutta scheme is often applied for the explicit method while the method of approximate factorization and iterative schemes are used for implicit schemes.

2.2.5 Scalar, hyperbolic equations of second order

Besides the first order equations of the type of the convection equation, the hyperbolic, partial differential equations of second order have an important role in fluid dynamics. Two examples for scalar equations of second order are the wave equation

$$u_{tt} - a_0^2 \, u_{xx} = 0$$

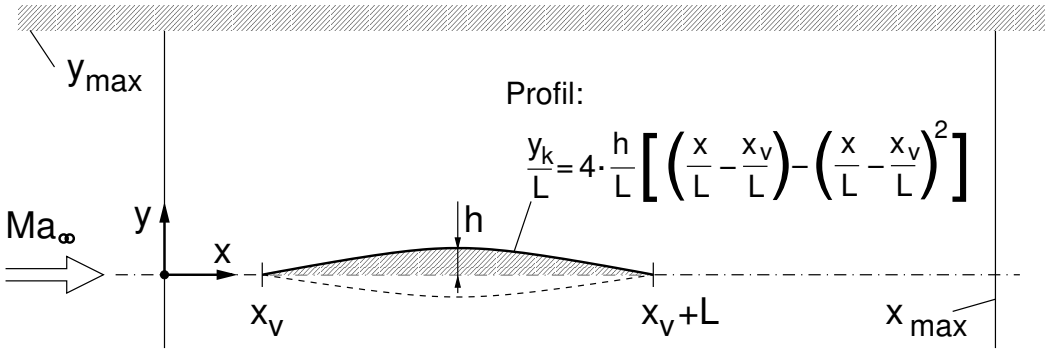
and the small disturbance equation in 2D for sub and supersonic flow:

$$(Ma_\infty^2 - 1) \, \Phi_{xx} \; - \; \Phi_{yy} = 0$$

In the following the solution of the hyperbolic small perturbation equation for hypersonic flows is considered to demonstrate the numerical solution of such equations.

Formulation of a flow problem of the small disturbance equation

A difference solution of the small disturbance equation shall be applied to solve the hypersonic flow ($Ma_\infty > 1$) around a symmetric profile with the contour $y_k(x)$ and the depth L. The profile is placed in a channel of height $2y_{max}$ with straight walls. Only one half of the flow problem needs to be considered, since the problem is axis symmetric in the channel. The domain of integration is displayed in the figure.



Assuming small width of the profile, i.e. $h \ll L$, and therefore small perturbations caused by the profile in the initial flow, the flow can be described by the small disturbance equation.

Which is:

$$-\beta^2\,\varphi_{xx} + \varphi_{yy} = 0 \quad \text{with} \quad \beta^2 = Ma_\infty^2 - 1 \quad \text{und} \quad Ma_\infty > 1$$

The velocities and the pressure parameter are functions of the disturbance potential φ. I.e.

$$\frac{u - u_\infty}{u_\infty} = \varphi_x \quad , \quad \frac{v}{u_\infty} = \varphi_y \quad , \quad c_p = -2 \cdot \varphi_x$$

The hyperbolic small disturbance equation describes an initial boundary value problem with the x-coordinate as direction of the initial value problem.
Because of the second order derivative φ_{xx} the initial values at $x = 0$ require the definition of two flow properties at the inflow, in this case:

$$x = 0: \qquad u = u_\infty \quad \rightarrow \varphi_x(0,y) = 0 \quad \text{und} \quad v = 0 \quad \rightarrow \varphi_y(0,y) = 0$$

The boundary conditions for $y = 0$ (symmetry axis) and $y = y_{max}$ (channel wall) for the flow problem are:

$$\begin{array}{lcccc} y & = & 0 & x_v \le x \le x_v + L & \varphi_y(x,0) = \frac{dy_k}{dx} = y_k'(x) \\ y & = & 0 & x < x_v \; x > x_v + L & \varphi_y(x,0) = 0 \\ y & = & y_{max} & 0 \le x \le x_{max} & \varphi_y(x,y_{max}) = 0 \end{array}$$

An exact solution of the flow can be obtained by using characteristics for $y_{max} \rightarrow \infty$ (free initial flow). The characteristics for the equation are:

$$\frac{dy}{dx} = \pm\frac{1}{\beta} = \pm\sqrt{\frac{1}{Ma_\infty^2 - 1}} = \pm\, tan\,\alpha \qquad \alpha = \text{Mach angle}$$

By application of a transformation $d\xi = dy - \frac{1}{\beta}\,dx$ and $d\eta = dy + \frac{1}{\beta}\,dx$ one obtains the normal form of the potential equation as:

$$\frac{\partial^2\varphi}{\partial\xi\partial\eta} = 0 \quad \text{with the solution} \quad \varphi(x,y) = \varphi_1(\xi) + \varphi_2(\eta).$$

For the solution of the upper side of the profile one obtains with the boundary condition $\varphi_y = \varphi_\xi \cdot 1 = y_k'(x)$ the potential φ and the pressure parameter

$$\varphi(x,y) = \varphi(\xi) = \varphi\left(y - \frac{1}{\beta}x\right) \quad \text{and} \quad c_p = -2\,\varphi_x = -2\,\varphi_\xi \cdot \left(-\frac{1}{\beta}\right) = \frac{2}{\beta}\,y_k'(x)$$

Numerical solution of the small disturbance equation

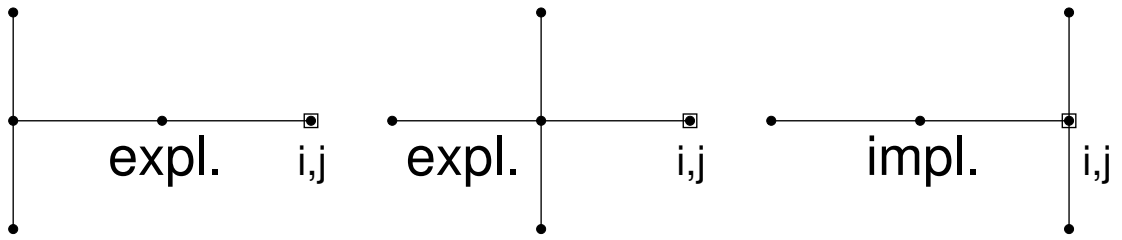
The problem is solved in the domain of integration $0 \le x \le x_{max}$ and $0 \le y \le y_{max}$ for a Cartesian grid with constant step sizes Δx and Δy. This leads to $x_i = (i-1)\,\Delta x$ and $y_j = (j-1)\,\Delta y$.
The discretization of the small disturbance equation in point (i,j) is done by using central differences for φ_{yy}.

$$\varphi_{yy}|_{i,j} \rightarrow \frac{1}{\Delta\, y^2}\,(\,\varphi_{i,j-1} - 2\,\varphi_{i,j} + \varphi_{i,j+1}\,) + O(\Delta y^2) \equiv (\,\delta_{yy}\,\varphi\,)_{i,j} + O(\Delta y^2)$$

and backwards differences for φ_{xx}

$$\varphi_{xx}|_{i,j} \to \frac{1}{\Delta x^2} \left(\varphi_{i-2,j} - 2\,\varphi_{i-1,j} + \varphi_{i,j} \right) + O(\Delta x) \equiv \left(\delta_{xx}\,\varphi \right)_{i,j} + O(\Delta x)$$

Depending on how the difference $(\delta_{yy}\,\varphi)$ is situated in relation to the point (i, j), different explicit or implicit schemes of order $O(\Delta x, \Delta y^2)$ are obtained:



$$-\beta^2 \left(\delta_{xx}\varphi \right)_{i,j} + \left(\delta_{yy}\varphi \right)_{i-2,j} = 0 \qquad -\beta^2 \left(\delta_{xx}\varphi \right)_{i,j} + \left(\delta_{yy}\varphi \right)_{i-1,j} = 0 \qquad -\beta^2 \left(\delta_{xx}\varphi \right)_{i,j} + \left(\delta_{yy}\varphi \right)_{i,j} = 0$$

The complete description of the discrete problem also requires the discretization of the initial and boundary conditions.
The boundary conditions are gradient conditions of type $\varphi_y = y'_k(x)$. An approximation of the boundary value for φ is obtained by substituting the gradient with a difference using the next inner point.
I.e. $\varphi_y|_1 = y'_k(x_i)$ for the profile at $y = 0\,(j = 1)$. An approximation is found by applying Taylor series expansion around $j = 1$ for φ of the nearest point to the wall, $j = 2$:

$$\varphi_2 = \varphi_1 + \varphi_y\,|_1\,\Delta y + \varphi_{yy}\,|_1\,\Delta y^2/2 + \cdots$$

In this case, for simplicity the Taylor expansion shall be aborted after the second term:

$$\varphi_{i,1} = \varphi_{i,2} - y'_k\,(x_i) \cdot \Delta y$$

A higher order can be achieved by substitution of the third term with the potential equation, i.e. $\varphi_{i,1} = \varphi_{i,2} - y'_k\,(x_i) \cdot \Delta y + \beta^2 \cdot \frac{\Delta y^2}{2}\,(\delta_{xx}\,\varphi)_{i,1}$
In analogy the discretization can be performed for the boundary values on the symmetry axis and the channel wall with $y'_k\,(x_i) = 0$.
The initial condition at $x = 0$ requires $\varphi_x = 0$ and $\varphi_y = 0$. Integration of $\varphi_y = 0$ results in $\varphi = const. = 0$. Therefore, by using a first order boundary approximation for $\varphi_x = 0$, the following initial value condition on the first columns $i = 1$ and $i = 2$ is obtained:

$$x = 0: \quad \varphi_x = 0 \qquad \Rightarrow \varphi_{1,j} = 0, \quad \varphi_{2,j} = 0$$

In the following the solutions of an explicit and an implicit scheme will be discussed in more detail. Because of the initial value problem in direction of the x axis the marching schemes can be formulated in this direction. Starting from the initial value condition at the points $i = 1$ and $i = 2$ the solution for $i = 3$ is calculated. The solution at $i = 2$ and $i = 3$ respectively represents the initial value condition for the next point.

Explicit scheme

After insertion of the differences of the defining equation for the unknown $\varphi_{i,j}$, the second explicit scheme

$$-\beta^2\,(\,\delta_{xx}\,\varphi\,)_{i,j}\;+\;(\,\delta_{yy}\,\varphi\,)_{i-1,j}\;=\;0$$

results in

$$\varphi_{i,j}\;=\;2\,\varphi_{i-1,j}\;-\;\varphi_{i-2,j}\;+\;C^2\;(\,\varphi_{i-1,j-1}\;-\;2\,\varphi_{i-1,j}\;+\;\varphi_{i-1,j+1})$$

The abbreviation $C\;=\;|\,\frac{1}{\beta}\,|\cdot\frac{\Delta x}{\Delta y}$ is the Courant number. The CFL condition is a necessary condition that must hold for the explicit scheme. The demand that the characteristics of the equation lie inside the numerical region of influence results in the following constraints for the Courant number (see figure):

$$C\;=\;|\,\frac{1}{\beta}\,|\cdot\frac{\Delta x}{\Delta y}<1$$

The CFL condition is proved by the stability analysis according to von Neumann which is sufficient for a linear and consistent initial value problem. Therefore, the scheme is stable for:

$$\Delta x\;\leq\;\beta\cdot\Delta y$$

Implicit scheme

After transformation, the implicit scheme

$$-\beta^2\,(\,\delta_{xx}\,\varphi\,)_{i,j}\;+\;(\,\delta_{yy}\,\varphi\,)_{i,j}\;=\;0$$

results in a tridiagonal equation system

$$\left[\,C^2\,\right]\,\varphi_{i,j-1}\;+\;\left[\,-\,2\,C^2\;-\;1\right]\,\varphi_{i,j}\;+\;\left[\,C^2\,\right]\,\varphi_{i,j+1}\;=\;\varphi_{i-2,j}\;-\;2\,\varphi_{i-1,j}$$

by using the common abbreviations:

$$a_j\,\varphi_{i,j-1}\;+\;b_j\,\varphi_{i,j}\;+\;c_j\,\varphi_{i,j+1}\;=\;RS_j$$

The solution with the Thomas algorithm, applying the recursion approach

$$\varphi_j\;=\;E_j\cdot\varphi_{j+1}\;+\;F_j$$

results, after substitution in the difference equation, in the following coefficients:

$$E_j\;=\;\frac{-\,c_j}{a_j\,E_{j-1}\;+\;b_j}\quad,\qquad F_j\;=\;\frac{RS_j\;-\;a_j\,F_{j-1}}{a_j\,E_{j-1}\;+\;b_j}$$

The boundary values for E and F at $j=1$ are obtained with the boundary value $\varphi_1\;=\;\varphi_2\;-\;y'_k\cdot\Delta y$ ant the recursion approach $\varphi_1\;=\;E_1\,\varphi_2\;+\;F_1$ as:

$$E_1\;=\;1\quad\text{und}\quad F_1\;=\;-\,y'_k\cdot\Delta y$$

Therefore, the recursion coefficients for $j = 2, \cdots, jmax - 1$ can be determined. The boundary value φ_{jmax} is needed to calculate the new values φ_j. It can be obtained from the boundary value $\varphi_{jmax} = \varphi_{jmax-1}$ and the recursion approach: $\varphi_{jmax-1} = E_{jmax-1}\,\varphi_{jmax} + F_{jmax-1}$ as:

$$\varphi_{jmax} = \frac{F_{jmax-1}}{1 - E_{jmax-1}}$$

With this the new variables for $j = jmax - 1, \cdots, 2$ can be calculated and the solution can be continued for a new value x_i. The CFL condition and the stability analysis for the implicit scheme do not impose constraints for the Courant number $C = |\frac{1}{\beta}| \cdot \frac{\Delta x}{\Delta y}$ therefore, the scheme is totally stable. The detailed solution and a corresponding FORTRAN program will be presented in the course.

2.3 Formulation of the Euler equations

2.3.1 Introduction

The Euler equations describe the conservation of mass, momentum and energy in an inviscid flow free of heat transfer. The Euler equations are simplifications of the Navier Stokes equations, obtained by neglecting friction and heat terms. Orthogonal forces on the body, i.e. pressure forces can be determined with the Euler equations, because of those simplifications. The pressure in the inviscid flow has physical meaning as long as the influence of the friction on the pressure distribution can be neglected. This is the case for e.g. high Reynolds number boundary layer flows, since the pressure is determined by the inviscid free stream pressure. This allows the calculation of lift and wave drag by solution of the Euler equations, in important areas of the flow field. The Euler equations are valid without constrains for subsonic, transonic and supersonic flows. They also admit the calculation of gas dynamic processes. This general usability is the reason for the importance of the numerical solution of the Euler equations in the project aerodynamic for aeronautics.
From the mathematical point of view the time dependent Euler equations form a system of non linear, hyperbolic, partial differential equations. The different formulations of the Euler equations have an important role in the solutions and will therefore be discussed in the following section.
Nonlinear, hyperbolic equations fall into two different solution types, i.e. continuous and discontinuous solutions. Continuous solutions are smooth, differentiable solutions. They can be determined with the method of characteristics. Discontinuous solutions are volatile solutions, like the shock waves. The general solution for this, by the integral, conservative formulation of the Euler equation will be shown in a further section.
The numerical solution schemes for the Euler equations are based on the methods which have already been presented for the scalar equations. What makes matters a bit more difficult is the fact that several characteristics with sometimes different signs must be considered, and the methods must be capable of capturing continuous and discontinuous solutions. This requires special numerical treatment of the flux formulations, including numerical damping. Examples for several important discretizations will be discussed.

2.3.2 Different forms of the Euler equations

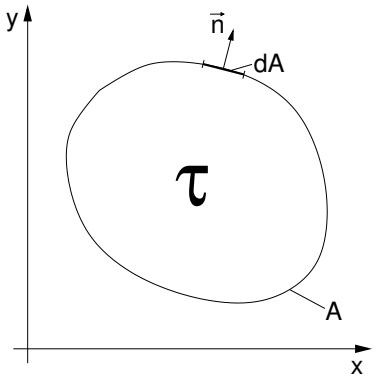
The calculation of the inviscid, compressible flow of a thermal and caloric ideal gas by the Euler equations requires the following equations:

- Conservation equations for mass, momentum and energy
- thermal state equation : $p = \rho R T$
- caloric state equation : $c_v = c_v(T), \quad c_p = c_p(T), \quad R = c_p - c_v$
- initial and boundary values

In the following the different formulations of the conservation equations will be considered.

Integral form

Since this form directly describes the balance in a control volume, it can be viewed as the original physical form of the Euler equations. For a closed element of volume $\tau(t)$ and the surface $A(t)$ the conservation of mass (ρ), momentum ($\rho\vec{v}$) and energy (ρE) per volume unit yields in an initial system:



$$\text{Mass :} \quad \int_\tau \frac{\partial \rho}{\partial t}\, d\tau + \oint_A \rho\, \vec{v} \cdot \vec{n}\, dA = 0$$

$$\text{Momentum :} \quad \int_\tau \frac{\partial \rho \vec{v}}{\partial t}\, d\tau + \oint_A [\, \rho\, \vec{v}\, \vec{v} + p\, I\,] \cdot \vec{n}\, dA = 0$$

$$\text{Energy :} \quad \int_\tau \frac{\partial \rho E}{\partial t}\, d\tau + \oint_A [\, \rho E \vec{v} + p\, \vec{v}] \cdot \vec{n}\, dA = 0$$

E describes the total energy $E = e + \vec{v}^2\,/\,2$, I is the unity tensor and $\vec{n}$ describes the normal vector on the surface A.
The three conservation equations can be combined in a system:

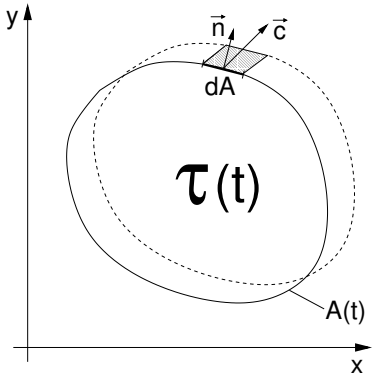
$$\int_\tau \frac{\partial U}{\partial t}\, d\tau + \oint_A \vec{H} \cdot \vec{n}\, dA = 0$$

where U describes the vector of conserved properties and $\vec{H}$ represents the generalized flux vector across the control surface.

$$U = \begin{pmatrix} \rho \\ \rho\, \vec{v} \\ \rho E \end{pmatrix} \qquad \vec{H} = \begin{pmatrix} \rho\, \vec{v} \\ \rho\, \vec{v}\, \vec{v} + p\, I \\ \rho \vec{v}\, E + p \vec{v} \end{pmatrix}$$

Conservation equations in the relative system

It is often convenient to formulate the conservation equations in a moving system to calculate unsteady problems. For the derivation it is assumed that the control volume τ moves with the velocity $\vec{c}$ relative to the reference system, fixed in space. The temporal change of the conserved quantities in the control volume τ, moving with the velocity $\vec{c}$ is in this case equal to the temporal change in the reference system (local acceleration), plus the transport of conservation properties caused by the shift of the control volume.



$$\frac{d}{dt}\int\limits_{\tau(t)} U\,d\tau \;=\; \lim_{\Delta t\to 0}\frac{1}{\Delta t}\left[\int\limits_{\tau(t+\Delta t)} U(t+\Delta t)\,d\tau - \int\limits_{\tau(t)} U(t)d\tau\right] \;=\; \int\limits_{\tau(t)}\frac{\partial U}{\partial t}\,d\tau + \oint\limits_{A} U\,\vec{c}\cdot\vec{n}\,dA$$

The temporal derivative in the moving system is defined as:

$$\frac{d}{dt} \;=\; \frac{\partial}{\partial t} \;+\; \vec{c}\cdot\nabla$$

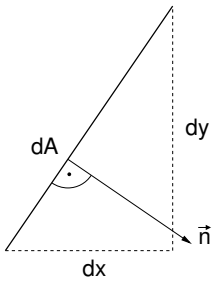
Therefore, one obtains the following conservation equation in the relative system:

$$\frac{d}{dt}\int\limits_{\tau} U\,d\tau \;+\; \oint\limits_{A} (\,\vec{H} \;-\; U\,\vec{c}\,)\cdot\vec{n}\,dA \;=\; 0$$

For the special case of a coordinate system moving with the stream velocity $\vec{c}$, i.e. $\vec{c} = \vec{v}$, the temporal derivative becomes the already known substantial derivative $\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{v}\cdot\nabla$.

Integral form in Cartesian coordinates

The integral form of the equations in a two dimensional, Cartesian coordinate system $(x,\, y,\, t)$ for a control volume $\tau \;=\; \tau\,(x,y)$ can be found when the vectors are regarded component wise.



$$\vec{v} = \begin{pmatrix} u \\ v \end{pmatrix} \quad , \quad \vec{H} = \begin{pmatrix} F \\ G \end{pmatrix} \quad , \quad \vec{c} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad , \quad \vec{n}\,dA = \begin{pmatrix} dy \\ -\,dx \end{pmatrix}$$

One obtains the integral form of the conservation equations

$$\frac{d}{dt}\int\limits_{\tau} U\,d\tau \;+\; \oint\limits_{A} (\,F \;-\; U\,c_x\,)\,dy \;-\; \oint\limits_{A} (\,G \;-\; U\,c_y\,)\,dx \;=\; 0$$

where U describes the vector of the conserved properties and F and G describe the Cartesian components of the flux vector.

$$U = \begin{pmatrix} \rho \\ \rho\, u \\ \rho\, v \\ \rho\, E \end{pmatrix} \quad F = \begin{pmatrix} \rho\, u \\ \rho\, u^2 + p \\ \rho\, u\, v \\ u\,(\rho\, E + p) \end{pmatrix} \quad G = \begin{pmatrix} \rho\, v \\ \rho\, v\, u \\ \rho\, v^2 + p \\ v\,(\rho\, E + p) \end{pmatrix}$$

Divergence form

A conservative, differential form (divergence form) of the Euler equations can be found by application of the Gaussian theorem.

$$\oint_A \vec{H} \cdot \vec{n}\, dA \quad = \quad \int_\tau \nabla \cdot \vec{H}\, d\tau$$

If in the integral form of the Euler equations the surface integral is replaced by a volume integral and a vanishing integrand is demanded for arbitrary volumes, the divergence form of the Euler equations is obtained.

$$\boxed{\frac{\partial U}{\partial t} + \nabla \cdot \vec{H} = 0}$$

The differential forms of the conservation of mass, momentum and energy are:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &+ \nabla \cdot (\rho\, \vec{v}) = 0 \\ \frac{\partial \rho\, \vec{v}}{\partial t} &+ \nabla \cdot (\rho\, \vec{v}\, \vec{v} + p\, I) = 0 \\ \frac{\partial \rho\, E}{\partial t} &+ \nabla \cdot \vec{v}\,(\rho\, E + p) = 0 \end{aligned}$$

Divergence form for Cartesian coordinates

The equations in a two dimensional, Cartesian coordinate system (x, y, t) can be found by component wise splitting of the vectors:

$$\vec{v} = \begin{pmatrix} u \\ v \end{pmatrix} \quad , \quad \vec{H} = \begin{pmatrix} F \\ G \end{pmatrix} \quad , \quad \vec{c} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad , \quad \nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}$$

This results in the divergence form of the conservation equations in Cartesian coordinates as follows:

$$\boxed{\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} = 0}$$

where U describes the vector of the conserved properties and F and G describe the Cartesian components of the flux vector.

$$U = \begin{pmatrix} \rho \\ \rho\, u \\ \rho\, v \\ \rho\, E \end{pmatrix} \quad F = \begin{pmatrix} \rho\, u \\ \rho\, u^2 + p \\ \rho\, u\, v \\ u\,(\rho\, E + p) \end{pmatrix} \quad G = \begin{pmatrix} \rho\, v \\ \rho\, v\ u \\ \rho\, v^2 + p \\ v\,(\rho\, E + p) \end{pmatrix}$$

Quasi conservative form and Jacobian matrices

For the development of numerical solution methods and for the analysis of the conservative equations, Jacobi matrices for the fluxes are required. Jacobi matrices represent the functional relation between the single flux components and the components of the vector of the conservation values. For the divergence form

$$U_t + F_x + G_y = 0$$

with the fluxes F and G as a function of the conservation values U

$$F(U) = \begin{pmatrix} F_1(U) \\ F_2(U) \\ F_3(U) \\ F_4(U) \end{pmatrix} \quad G(U) = \begin{pmatrix} G_1(U) \\ G_2(U) \\ G_3(U) \\ G_4(U) \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}$$

one obtains the Jacobi matrices $\overline{\overline{A}}$ and $\overline{\overline{B}}$ of the fluxes.

$$\overline{\overline{A}} = \frac{\partial F}{\partial U} = \frac{\partial\,(F_1, F_2, F_3, F_4)}{\partial\,(U_1, U_2.U_3, U_4)} \quad \text{where} \quad a_{k\,l} = \frac{\partial F_k}{\partial U_l} \quad (k,\, l = 1, \cdots 4)$$

$$\overline{\overline{B}} = \frac{\partial G}{\partial U} = \frac{\partial\,(G_1, G_2, G_3, G_4)}{\partial\,(U_1, U_2.U_3, U_4)} \quad \text{where} \quad b_{k\,l} = \frac{\partial G_k}{\partial U_l} \quad (k,\, l = 1, \cdots 4)$$

The quasi conservative form is formulated using the Jacobi matrices. Those equations no longer have divergence form, nevertheless, the conservation values remain the dependent variables. Therefore, the change in the fluxes is expressed by a change in the conservation values, i.e.:

$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial U} \cdot \frac{\partial U}{\partial x} = \overline{\overline{A}}\,\frac{\partial U}{\partial x} \quad \text{and} \quad \frac{\partial G}{\partial y} = \frac{\partial G}{\partial U} \cdot \frac{\partial U}{\partial y} = \overline{\overline{B}}\,\frac{\partial U}{\partial y}$$

Thus, the quasi conservative Form is obtained from the divergence form and yields:

$$\boxed{U_t + \overline{\overline{A}}\,U_x + \overline{\overline{B}}\,U_y = 0}$$

A special case is given by the Euler equations of a perfect gas. The non linear fluxes of the Euler equations are in this case homogeneous functions of first order in respect to the

conservation values. This means that the fluxes become linear functions of the conservation values, coupled by the Jacobi matrix.

$$F(U) = \frac{\partial F}{\partial U} \cdot U = \overline{\overline{A}}\, U \quad \text{and} \quad G(U) = \frac{\partial G}{\partial U} \cdot U = \overline{\overline{B}}\, U$$

Example:
In the following the derivation of the Jacobi matrices for the one dimensional Euler equation shall be demonstrated. Considered are the Euler equations $U_t + F_x = 0$ with the conservative variables:

$$U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix}$$

The flux $F(U)$ results in

$$F = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(\rho E + p) \end{pmatrix} = \begin{pmatrix} U_2 \\ U_2^2/U_1 + p(U) \\ U_2/U_1 \cdot (U_3 + p(U)) \end{pmatrix}$$

with the pressure $p(U) = (\kappa - 1)\,[U_3 - 1/2 \cdot U_2^2/U_1]$.
For the Jacobi Matrix $\overline{\overline{A}}$ with the elements $a_{kl} = \frac{\partial F_k}{\partial U_l}$ one obtains

$$\overline{\overline{A}} = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{(3-\kappa)}{2}\, u^2 & (3-\kappa)\, u & \kappa - 1 \\ u\left((\kappa-1)\, u^2 - \kappa\, E\right) & \kappa\, E - \frac{3}{2}(\kappa - 1)\, u^2 & \kappa\, u \end{pmatrix}$$

Non-conservative forms

All formerly discussed forms of the Euler equations like the integral and the divergence form are based on the conservation values U as dependent variable, i.e. mass, momentum and energy. They therefore directly represent the conservation laws of fluid mechanics. If one chooses to select dependent variables which are non-conservative, the resulting equations are called non-conservative forms. The non conservative forms can not be formulated in integral or divergence form, because there will always occur variable dependent coefficients in the differentials. For the numerical solution of the Euler equations these forms are of minor importance, but they often lead to informations on the solution properties in a more straightforward fashion. Many of the non-conservative forms can be formulated in Lagrangian notion which is used for the description of the flow in a moving system. The time derivative becomes in this case identical with the substantial derivative.

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \vec{v} \cdot \nabla \quad \text{or in } (x, y, t) \qquad \frac{d}{dt} = \frac{\partial}{\partial t} + u\,\frac{\partial}{\partial x} + v\,\frac{\partial}{\partial y}$$

In the following two examples of non-conservative forms will be given. The equations will be formulated in vector notation, i.e. in Cartesian coordinates (x, y, t).

- Dependent variable $V = (\rho, \vec{v}, E)^T$:

$$\begin{aligned} \frac{d\rho}{dt} &+ \rho \nabla \cdot \vec{v} &&= 0 \\ \frac{d\vec{v}}{dt} &+ 1/\rho \nabla p &&= 0 \\ \frac{dE}{dt} &+ 1/\rho \nabla \cdot (p\vec{v}) &&= 0 \end{aligned} \qquad \begin{aligned} \frac{d\rho}{dt} &+ \rho (u_x + v_y) &&= 0 \\ \frac{du}{dt} &+ 1/\rho \, p_x &&= 0 \\ \frac{dv}{dt} &+ 1/\rho \, p_y &&= 0 \\ \frac{dE}{dt} &+ 1/\rho \, ((pu)_x + (pv)_y) &&= 0 \end{aligned}$$

- Dependent variable $V = (\rho, \vec{v}, p)^T$:

$$\begin{aligned} \frac{d\rho}{dt} &+ \rho \nabla \cdot \vec{v} &&= 0 \\ \frac{d\vec{v}}{dt} &+ 1/\rho \nabla p &&= 0 \\ \frac{dp}{dt} &+ \rho a^2 \nabla \cdot \vec{v} &&= 0 \end{aligned} \qquad \begin{aligned} \frac{d\rho}{dt} &+ \rho (u_x + v_y) &&= 0 \\ \frac{du}{dt} &+ 1/\rho \, p_x &&= 0 \\ \frac{dv}{dt} &+ 1/\rho \, p_y &&= 0 \\ \frac{dp}{dt} &+ \rho a^2 (u_x + v_y) &&= 0 \end{aligned}$$

Characteristic form

The Euler equations are a hyperbolic system of partial differential equations with real characteristics. The characteristic form of the Euler equations is a special variant of the non-conservative equations, built with the characteristic variable. According to the definition of the characteristic solution, i.e. a solution independent from the neighbor solution, one obtains a decoupled system for the Euler equations. The solution for each equation is the characteristic solution for the corresponding characteristic line. This characteristic form is the starting point for the development of the method of characteristics and the numerical difference method of the Euler equations.
The hyperbolic system can be transformed in the characteristic form by diagonalization of the system matrix. The system matrix is the matrix holding the coefficients of the derivatives of the space derivatives. The time derivatives are multiplied with the unity matrix, i.e. they don't have coefficients. The eigenvalues λ_i of the system matrix become identical with the characteristic, directional derivatives, i.e. $\lambda_i = \frac{dx}{dt}|_i$. A derivation of the characteristic form can only be achieved for the one dimensional, time dependent Euler equations. A complete diagonalization of multi dimensional equations is not possible, since the diagonal transformation of the matrices preceding the spatial derivatives can vary for each direction. The non-conservative form of the Euler equations is the starting point for the diagonalization. If the conservative divergence form shall be diagonalized, the quasi

conservative form has to be built first. With the vector of the dependent variables V of the system matrix $\overline{\overline{A}}$ and the unity matrix I one obtains the following equation:

$$I\,V_t \,+\, \overline{\overline{A}}\,V_x \,=\, 0$$

The diagonal transformation of the matrix $\overline{\overline{A}}$ into a diagonal matrix Λ with the real eigenvalues λ_i as diagonal entries, can be performed with the eigenvector matrix T und its inverse T^{-1}.

$$\Lambda \,=\, T^{-1}\,\overline{\overline{A}}\quad T \qquad \text{where} \qquad \Lambda \,=\, \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

The eigenvalues λ_i of the matrix $\overline{\overline{A}}$ are $\lambda_i \,=\, \frac{dx}{dt}|_i \,=\, (u, u+a, u-a), \qquad i \,=\, 1,2,3.$
The eigenvalues can be calculated from the determinant

$$|\overline{\overline{A}} \,-\, \lambda_i\, I\,| \,=\, 0$$

The calculation of the eigenvector matrix $T \,=\, (\vec{x}_1,\, \vec{x}_2,\, \vec{x}_3)$, whose columns consist of the eigenvectors $\vec{x}_i$ of the eigenvalues λ_i, results from the solution of the equation system

$$(\,\overline{\overline{A}} \,-\, \lambda_i\, I\,)\,\vec{x}_i \,=\, 0 \qquad i = 1,2,3$$

The transformation of the Euler equations into the characteristic form is achieved by left hand multiplication with T^{-1}.

$$\begin{aligned} T^{-1} \quad & V_t \,+\, T^{-1}\,\overline{\overline{A}}\,T\,T^{-1}\,V_x \,=\, 0 \\ T^{-1} \quad & V_t \,+\, \Lambda\,T^{-1}\,V_x \,=\, 0 \end{aligned}$$

Using the definition of the characteristic variables W

$$dW \,=\, T^{-1}\,dV$$

the characteristic form is obtained:

$$W_t \,+\, \Lambda\,W_x \,=\, 0$$

or fully written:

$$\frac{\partial w_i}{\partial t} \,+\, \lambda_i\,\frac{\partial w_i}{\partial x} \,=\, 0 \qquad i \,=\, 1,2,3$$

The derivation of the characteristic form will be demonstrated for two different original forms of the Euler equation. The matrix operations for the non-conservative form are easier to perform, but the derivation of the conservative form is nevertheless important for the development of upwind schemes.

1. Example:

As a starting point an easy to solve, non-conservative form with $V = (\rho, u, p)$ is taken.

$$\begin{array}{lcllcl} \rho_t & + & u\,\rho_x & + & \rho\,u_x & = 0 \\ u_t & + & u\,u_x & + & (1/\rho)\,p_x & = 0 \\ p_t & + & u\,p_x & + & \rho\,a^2\,u_x & = 0 \end{array}$$

The combined system yields:

$$V_t + \overline{\overline{a}}\, V_x = 0 \qquad \text{,where} \quad V = \begin{pmatrix} \rho \\ u \\ p \end{pmatrix} \quad \overline{\overline{a}} = \begin{pmatrix} u & \rho & 0 \\ 0 & u & 1/\rho \\ 0 & \rho a^2 & u \end{pmatrix}$$

From $|\, \overline{\overline{a}} - \Lambda \,| = 0$ one obtains the eigenvalues: $\lambda_1 = u\,,\ \lambda_2 = u + a\,, \lambda_3 = u - a.$ The eigenvectors $\vec{x}_i = (x_i^1,\, x_i^2,\, x_i^3)^T$ with $i = 1, 2, 3$ are determined by the equation system $(\overline{\overline{a}} - \lambda_i\, I\,)\, \vec{x}_i = 0.$

$$\begin{array}{ccccccc} (u - \lambda_i)\, x_i^1 & + & \rho x_i^2 & + & 0 & = & 0 \\ 0 & + & (u - \lambda_i)\, x_i^2 & + & 1/\rho\, x_i^3 & = & 0 \\ 0 & + & \rho\, a^2\, x_i^2 & + & (u - \lambda_i)\, x_i^3 & = & 0 \end{array}$$

The equation system for $x_i^1,\, x_i^2,\, x_i^3$ is undetermined, therefore each component can be arbitrarily chosen.

$$\begin{array}{llll} \lambda_1 = u & x_1^1 = -\frac{1}{a^2} & \text{(chosen)} & \to x_1^2 = 0 \quad , \quad x_1^3 = 0 \\ \lambda_2 = u + a & x_2^3 = \frac{1}{2} & \text{(chosen)} & \to x_2^2 = \frac{1}{2\rho a} \quad , \quad x_2^1 = \frac{1}{2a^2} \\ \lambda_3 = u - a & x_3^3 = \frac{1}{2} & \text{(chosen)} & \to x_3^2 = -\frac{1}{2\rho a} \quad , \quad x_3^1 = \frac{1}{2a^2} \end{array}$$

The eigenvector matrices $T = T(\vec{x}_1, \vec{x}_2, \vec{x}_3)$ and T^{-1} result in:

$$T = \begin{pmatrix} -\frac{1}{a^2} & \frac{1}{2a^2} & \frac{1}{2a^2} \\ 0 & \frac{1}{2\rho a} & -\frac{1}{2\rho a} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad T^{-1} = \begin{pmatrix} -a^2 & 0 & 1 \\ 0 & \rho a & 1 \\ 0 & -\rho a & 1 \end{pmatrix}$$

The characteristic variables are obtained from $dW = T^{-1}\, dV$

$$dW = \begin{pmatrix} d\, w_1 \\ d\, w_2 \\ d\, w_3 \end{pmatrix} = T^{-1} \cdot \begin{pmatrix} d\, \rho \\ d\, u \\ d\, p \end{pmatrix} = \begin{pmatrix} -a^2\, d\rho + dp \\ \rho a\, du + dp \\ -\rho a\, du + dp \end{pmatrix}$$

This yields the characteristic form of the Euler equations

$$\frac{\partial w_i}{\partial t} + \lambda_i \frac{\partial w_i}{\partial x} = 0$$

Written in separate equations one obtains

$$\begin{array}{lll} (p_t - a^2\, \rho_t) & + u & (p_x - a^2\, \rho_x) = 0 \\ (p_t + \rho a\, u_t) & + (u + a) & (p_x + \rho a\, u_x) = 0 \\ (p_t - \rho a\, u_t) & + (u - a) & (p_x - \rho a\, u_x) = 0 \end{array}$$

2. Example

As a starting point the divergence form of the one dimensional Euler equations is taken.

$$U_t + F_x = 0 \quad \text{with} \quad U = \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix} \qquad F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho u E + up \end{pmatrix}$$

The divergence form must be transformed into the quasi conservative form prior to the derivation of the characteristic form (see above).

$$U_t + \overline{\overline{A}}\, U_x = 0 \quad \text{with} \quad \overline{\overline{A}} = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{(3-\kappa)}{2} u^2 & (3-\kappa)\, u & \kappa - 1 \\ u\left((\kappa-1)\, u^2 - \kappa\, E\right) & \kappa\, E - \frac{3}{2}(\kappa-1)\, u^2 & \kappa\, u \end{pmatrix}$$

The diagonalization of the Jacobian Matrix $\overline{\overline{A}}$ with an eigenvector matrix R is carried out in the same fashion with:

$$\Lambda = R^{-1}\, \overline{\overline{A}}\, R \qquad dW = R^{-1}\, dU \qquad \Lambda = \begin{pmatrix} u & 0 & 0 \\ 0 & u+a & 0 \\ 0 & 0 & u-a \end{pmatrix}$$

With the abbreviations:

$a_1 = -\frac{1}{a^2}$, $a_2 = \frac{1}{2a^2}$, $b_1 = \frac{\kappa-1}{2} M^2$, $b_2 = \kappa - 1$, $M = \frac{u}{a}$ the eigenvector matrix R of the conservative form yields:

$$R = \begin{pmatrix} a_1 & a_2 & a_2 \\ a_1 u & a_2(u+a) & a_2(u-a) \\ a_1 u^2/2 & a_2(H+au) & a_2(H-au) \end{pmatrix} \qquad R^{-1} = \begin{pmatrix} a^2(b_1-1) & -a(b_2\, M) & b_2 \\ a^2(b_1 - M) & -a(b_2 M - 1) & b_2 \\ a^2(b_1+M) & -a(b_2 M + 1) & b_2 \end{pmatrix}$$

The characteristic form is equal to the one calculated in the first example.

$$W_t + \Lambda\, W_x = 0 \quad \text{with} \quad dW = R^{-1}\, dU = \begin{pmatrix} -a^2\, d\rho + dp \\ \rho a\, du + dp \\ -\rho a\, du + dp \end{pmatrix}$$

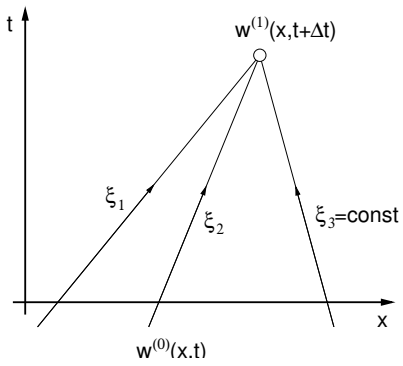
The characteristic form of the Euler equations is the starting point for the method of characteristics and the development of difference schemes, especially upwind schemes. The latter will be discussed in a special section. The method of characteristics uses the solution along a characteristic for the calculation of a flow. The characteristic form is transformed into characteristic coordinates ($\Longrightarrow$canonical form) and integrated. The application of a coordinate transformation $d\xi_i = dx - \lambda_i\, dt$ and $d\tau_i = dt$ yields:

$$\frac{\partial w_i}{\partial t} + \lambda_i \frac{\partial w_i}{\partial x} = 0 \quad \Longrightarrow \quad \frac{dw_i}{d\tau_i} = 0$$

This results in

$$dw_i = 0 \quad \text{along} \quad d\xi_i = dx - \lambda_i\, dt = 0$$

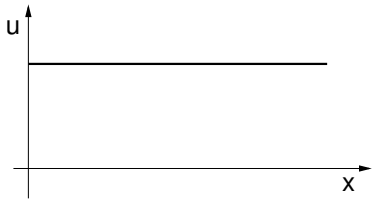
New values on a neighboring point $P(x,t)$ can be determined for a given initial condition $w_i^{(o)}$ on a non characteristic curve.



2.3.3 Discontinuous solutions of the Euler equations

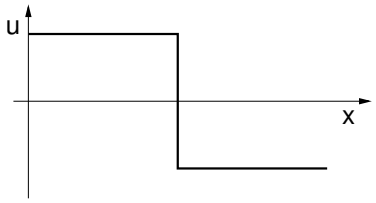
There are two types of solutions for nonlinear, hyperbolic, partial differential equations, continuous solutions and discontinuous solutions. According to the demands concerning the differentiability of the solution one speaks of strong and weak solutions describing continuous and discontinuous solutions of the equations, respectively (see e.g. *Courant, Hilbert: Mathematische Methoden der Physik*). The continuous solution can be calculated from the conservative as well as from the non conservative form, e.g. with the characteristics. The discontinuous solution which describes a jump in the variables, can only be calculated from the conservative form. The two varying solution types can be demonstrated with a straightforward example.

Example: The nonlinear, hyperbolic model equation $u_t + uu_x = 0$ will be considered. The steady, continuous solution of this non-conservative form results from:



$$uu_x = 0 \quad \rightarrow \quad u_x = 0 \quad \rightarrow \quad u = \text{const.} \quad \rightarrow \text{continuous}$$

On the other hand, considering the steady solution of the conservative form $u_t + (\frac{u^2}{2})_x = 0$ of the same equation, one obtains by integration:



$$(\frac{u^2}{2})_x = 0 \quad \rightarrow \quad u^2 = \text{const.} \quad \rightarrow \quad u = \pm\sqrt{\text{const.}} \quad \rightarrow \text{unsteady}$$

The solution of the conservative form therefore describes a jump in the variables!

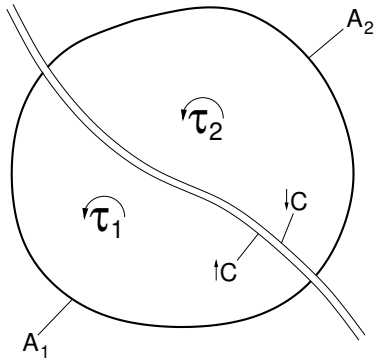
The occurrence of the discontinuity can also be explained by taking a closer look on the characteristics. The characteristics of nonlinear, hyperbolic equations, have gradients which depend on the solution itself, i.e. the gradient shifts through the solution domain (in this example $\frac{dx}{dt} = u$). As a consequence the intersection between characteristics becomes possible. In this case, the solution is no longer definite. One obtains a discontinuous solution.
An equal solution behavior, on a more complex level, is displayed by the Euler equations which form a non linear system of hyperbolic equations. The inviscid flow, described by the Euler equations, can contain various types of discontinuities.
The best known discontinuity is the shock wave which describes the rapid compression of a gas. The discontinuous solution yields for this case the jump conditions across the shock wave, often called Rankine-Hugoniot relation. Such shock waves can also occur in steady, super sonic flows, e.g. in a Laval nozzle or along a profile. They can also occur as an unsteady phenomenon, in a subsonic flow, e.g. in a shock tube.
A different type of discontinuity is the contact discontinuity which separates gases of different elements species and state (see shock tube). A further discontinuity occurs from the separation surface between two flows of different tangential velocity. This so called

tangential discontinuity occurs for instance in the "inviscid" wake of a profile. The pressure and normal velocity are constant across the tangential and the contact discontinuity. These examples show the importance of unsteady solutions for flow problems. From the numerical point of view these unsteady solutions can be quite challenging, since most numerical solution schemes demand the differentiability of the solution (Taylor series expansion) which is not granted in this case. This issue has already been addressed in the section on numerical damping. Examples of solutions will follow in a preceeding chapter. Essential for the numerical solution are the following conditions:

> The calculation of discontinuous flows requires the solution of the conservative equations!
>
> It is important for the calculation of the flow to know the exact solution of the discontinuity.

Unsteady solutions of the Euler equations

A general solution for a moving discontinuity for the Euler equations can be determined in a straightforward fashion. One considers the control volume τ which is divided by a discontinuity C, across which the variables are discontinuous. The velocity of the discontinuity is $\vec{c}$. The laws of conservation require the conservation of mass, momentum and energy in each sub volume τ_1 and τ_2 and in the entire volume $\tau = \tau_1 + \tau_2$. The resulting condition is the discontinuous solution of the jump condition.
The conservation is described by the integral form of the Euler equations for an arbitrary domain, in a system, moving with the velocity $\vec{c}$.



$$\frac{d}{dt}\int_{\tau} U\,d\tau + \oint_{A} (\vec{H} - U\,\vec{c})\cdot\vec{n}\,dA = 0$$

The conservation equations are formulated for each control volume:

Overall volume $\tau = \tau_1 + \tau_2$ with the surface $A = A_1 + A_2$:

$$\frac{d}{dt}\int_{\tau_1+\tau_2} U\,d\tau + \oint_{A_1+A_2} (\vec{H} - U\,\vec{c})\cdot\vec{n}\,dA = 0$$

Sub volume τ_1 with the surface $A_1 + \uparrow C$:

$$\frac{d}{dt}\int_{\tau_1} U\,d\tau + \int_{A_1} (\vec{H} - U\,\vec{c})\cdot\vec{n}\,dA + \int_{\uparrow C} (\vec{H} - U\,\vec{c})\cdot\vec{n}\,dA = 0$$

Sub volume τ_2 with the surface $A_2 + \downarrow C$:

$$\frac{d}{dt}\int\limits_{\tau_2} U\,d\tau \;+\; \int\limits_{A_2} (\vec{H} \;-\; U\,\vec{c})\cdot\vec{n}\,dA \;+\; \int\limits_{\downarrow C} (\vec{H} \;-\; U\,\vec{c})\cdot\vec{n}\,dA \;=\; 0$$

The balance across the two sub volumes must be equal the overall volume. This yields the requested jump condition across the discontinuity C;

$$\int\limits_{\downarrow C} (\vec{H} - U\,\vec{c})_2\cdot\vec{n}\,dA + \int\limits_{\uparrow C} (\vec{H} - U\,\vec{c})_1\cdot\vec{n}\,dA \;=\; \int\limits_{\downarrow C} \Big\{(\vec{H} \;-\; U\,\vec{c})_2 \;-\; (\vec{H} \;-\; U\,\vec{c})_1\Big\}\cdot\vec{n}\,dA \;=\; 0$$

Such discontinuous solutions are often written with the definition "discontinuity in function f", i.e. $[f] \;\equiv\; f_2 \;-\; f_1$. Therefore, one obtains the following general solution across the discontinuity:

$$\boxed{\Big[\vec{H} \;-\; U\,\vec{c}\Big]\cdot\vec{n}\,dA \;=\; 0}$$

To clarify this, the relation for a Cartesian coordinate system (x, y, t) will be given. The components of the vectors are:

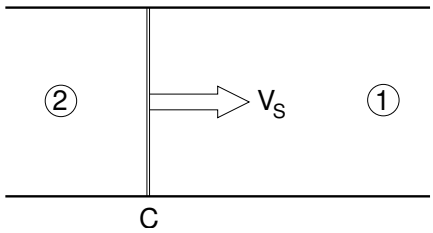
$$\vec{H} \;=\; \begin{pmatrix} F \\ G \end{pmatrix} \quad , \quad \vec{c} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad , \quad \vec{n}\,d\,A \;=\; \begin{pmatrix} d\,y \\ -\,d\,x \end{pmatrix}$$

For the jumping condition $\Big[\vec{H} \;-\; U\,\vec{c}\Big]\cdot\vec{n}\,dA \;=\; 0$ one obtains:

$$\left\{ (\,F \;-\; Uc_x\,)\frac{dy}{dx}|_C \;-\; (\,G \;-\; Uc_y\,) \right\}_2 \;=\; \left\{ (\,F \;-\; Uc_x\,)\frac{dy}{dx}|_C \;-\; (\,G \;-\; Uc_y\,) \right\}_1$$

With this the jump condition for instance across a curved shock can be formulated.

Example:
One dimensional, running shock with shock velocity v_s:



$\Longrightarrow c_x = v_s \quad , \quad c_y = 0 \quad , \quad \frac{dy}{dx}|_C \;=\; \infty$

The jump condition becomes $[\,F \;-\; U\cdot v_s] \;=\; 0$ and yields the conservation across the shock:

$$\begin{array}{lcl} \{\rho(u - v_s)\}_2 & = & \{\rho(u - v_s)\}_1 \\ \{\rho u(u - v_s) + p\}_2 & = & \{\rho u(u - v_s) + p\}_1 \\ \{\rho(u - v_s)E + up\}_2 & = & \{\rho(u - v_s)E + up\}_1 \end{array}$$

For a predefined state 1 with velocity v_s the state 2 behind the shock wave can be determined.

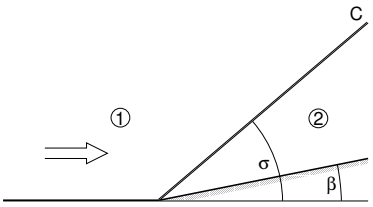
Example:
Two dimensional steady shock with shock angle σ

$\Longrightarrow \vec{c} = 0 \quad , \quad \frac{dy}{dx}|_C = tan\,\sigma$

The jump condition is $\{F \cdot tan\,\sigma - G\}_2 = \{F \cdot tan\,\sigma - G\}_1$

and yields the conservation across the shock wave:

$$\begin{aligned}
\{\rho u \tan\sigma - \rho v\}_1 &= \{\rho u \tan\sigma - \rho v\}_2 \\
\{(\rho u^2 + p)\tan\sigma - \rho uv\}_1 &= \{(\rho u^2 + p)\tan\sigma - \rho uv\}_2 \\
\{\rho uv \tan\sigma - (\rho v^2 + p)\}_1 &= \{\rho uv \tan\sigma - (\rho v^2 + p)\}_2 \\
\{\rho uH \tan\sigma - \rho vH\}_1 &= \{\rho uH \tan\sigma - \rho vH\}_2
\end{aligned}$$



2.4 Numerical solution of the Euler equations

The presentation of important solution methods will be carried out for the one dimensional, time dependent Euler equations. All discussed schemes are based on the conservative discretization of the conservation equations for a small, discrete control volume. One distinguishes between space and time discretization. The space discretization describes the change in the fluxes across the control volume. One important space discretization of the Euler equations, the central flux formulation, including the damping terms and the upwind formulation with characteristic flux splitting, is introduced. The time discretization, essentially defines the solution method. Some starting points for the formulation of explicit and implicit schemes for the time discretization are presented. The numerical solution of the Euler equations shall be demonstrated with an unsteady flow problem, i.e. the flow in a shock tube.

The transfer of those solution methods on multi dimensional problems is in general possible without great effort, since for this case principally a quasi one dimensional problem can be formulated in each coordinate direction. The procedure for the space discretization is demonstrated with two dimensional Cartesian and curved grids.

2.4.1 Formulation of the one dimensional Euler equations

The conservation of mass, momentum and energy in an inviscid flow is described by the time dependent Euler equations which are presented for the one dimensional case. These equations can be applied in integral as well as in divergence form for the numerical solution. The integral form is:

$$\int_\tau U_t \, d\tau + \oint_A F \cdot dy = 0$$

while the conservative divergence form is:

$$U_t + F_x = 0$$

where U describes the vector of the conservative variables and $F(U)$ represents the flux vector.

$$U = \begin{pmatrix} \rho \\ \rho\, u \\ \rho\, E \end{pmatrix} \qquad F = \begin{pmatrix} \rho\, u \\ \rho\, u^2 + p \\ u\,(\rho\, E + p) \end{pmatrix}$$

In the following a caloric and thermal ideal gas is assumed. The pressure p and speed of sound a in such a gas are:

$$p = (\kappa - 1)\,\rho\,(E - \frac{1}{2}u^2) \qquad a = \sqrt{\kappa \frac{p}{\rho}} \qquad \kappa = \frac{c_p}{c_v} = const.$$

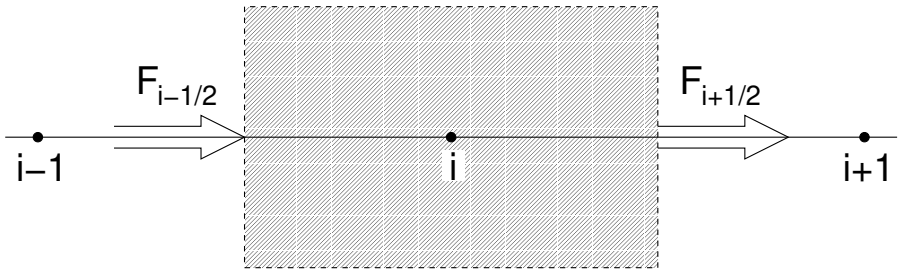
2.4.2 Spatial discretization of the fluxes

Conservative discretization

The domain of integration in the $x-t$ plane is divided by a grid with $t_n = (n-1)\cdot\Delta t$ and $x_i = (i-1)\cdot\Delta x$. The space step size Δx is assumed constant and can be obtained from the integration length x_{max} and the maximum number of points $imax$. With this the space step size becomes $\Delta x = \frac{x_{max}}{imax-1}$.
The time step size Δt can either be defined by the numerical stability or by the demands on the accuracy of the solution. In general it is governed by the predefined Courant number C, i.e.

$$\Delta t = C \cdot \frac{\Delta x}{|\lambda|_{max}} \quad \text{where} \quad |\lambda|_{max} = \max_i(|u|, |u-a|, |u+a|) = \max_i(|u|+a)$$

The maximum value of the Courant number C depends on the method (stability). For unsteady problems, the Courant number is limited even for implicit schemes because of the time accuracy. Often it is only of order $C = O(1)$.



Conservative discretization requires the discretized equations to also comply with the law of conservation. Starting from the integral conservation equations one defines a control volume $\tau = \Delta x \cdot \Delta y$ with the center point i, in which the conservation equations are formulated. The variables are averaged over the control volume τ. Therefore, one obtains constant values inside a cell. This yields

$$\int_\tau U_t d\tau \longrightarrow \frac{\Delta U_i}{\Delta t} \cdot \Delta x \Delta y$$

for the temporal change. Where $\frac{\Delta U_i}{\Delta t}$ is the discretized time derivative at the point i which will be defined later by the solution method.
The temporal change of the conservative variables is in balance with the change of the fluxes over the control volume. The location of the surface normal $dA = \Delta y$ of the fluxes in x–direction is labeled with $i \pm 1/2$. Geometrically they are assumed between the point i and the points $i \pm 1$. (The value of Δy doesn't matter for the one dimensional problem.) In the one dimensional case the flux balance results in:

$$\oint_A F \cdot \vec{n}\, dA \longrightarrow (F_{i+1/2} - F_{i-1/2})\Delta y$$

Therefore, one obtains the following discretized integral form:

$$\frac{\Delta U_i}{\Delta t} \cdot \Delta x \Delta y \; + \; (F_{i+1/2} - F_{i-1/2}) \Delta y \; = \; 0$$

Division by the volume $\tau = \Delta x \cdot \Delta y$ yields the difference form of the Euler equations.

$$\boxed{\frac{\Delta U_i}{\Delta t} \; + \; \frac{F_{i+1/2} - F_{i-1/2}}{\Delta x} \; = \; 0}$$

This difference form can also be obtained from the divergence form of the Euler equations, by difference building:

$$U_t \longrightarrow \frac{\Delta U_i}{\Delta t} \qquad F_x \longrightarrow \frac{F_{i+1/2} - F_{i-1/2}}{\Delta x}$$

This straightforward example shows that the integral as well as the divergence form of the Euler equations can result in the same difference form, provided that the law of conservation also holds for the discretized form. This also applies for multi dimensional problems. It is especially important for multi dimensional problems that the metric is properly fixed by the conservative discretization, as has been shown.

The conservative divergence form is the starting point for the different solution schemes. The next subsection deals with the different space discretizations for the fluxes $F_{i\pm1/2}$ like central and upwind schemes.

Numerical flux function

The fluxes $F_{i\pm1/2}$ at the wall of the control volume are functions of the conservative variables of the neighboring checkpoints. Therefore, for the points $i \pm 1/2$ they must be approximated by interpolation polynomials. Because of this the discrete flux is also known as Numerical flux function or numerical flux $\tilde{F}_{i\pm1/2}$. The numerical flux for the point $i + 1/2$ is for instance:

$$\tilde{F}_{i+1/2} = \tilde{F}_{i+1/2}(U_{i-1}, U_i, U_{i+1}, U_{i+2})$$

It is the aim of the space discretization to find a numerical approximation for the numerical fluxes $\tilde{F}_{i\pm1/2}$, at the walls of the control volume τ, such that the scheme is consistent in space.
Consistency means that the change of the numerical flux in the limit $\Delta x \to 0$ leads to the divergence of the exact flux F, i.e.:

$$\lim_{\Delta x \to 0} \frac{\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}}{\Delta x} = F_x$$

Different space discretizations like the central upwind scheme have already been discussed in the preceeding chapter (chap.6) on hyperbolic, scalar equations. In principle it is possible to apply the therein presented schemes to the Euler equations. However, this requires some considerations. First the Euler equations form a system, second their fluxes $F(U)$ are nonlinear functions and finally multiple characteristics exist (eigenvalues) with possibly

different signs. Therefore, the definiteness of the space discretization is lost. There are various possibilities of consistent discretizations. This is especially true for upwind schemes.

In the following the characteristic form of the Euler equations will serve as a starting point to present a scheme for the derivation of numerical flux formulations. This form consists of a system of decoupled, hyperbolic equations, similar to the formerly discussed scalar model equations. In addition the characteristic form allows an unambiguous splitting in reference to the sign of the characteristic and therefore an unambiguous formulation of upwind schemes. Suitable formulations of the numerical fluxes can therefore be obtained by reverse transformation of the discretized characteristic form back to the conservative form. The initial form for this consideration is given by the divergence form of the Euler equations

$$U_t \, + \, F_x \, = \, 0$$

For this a consistent approximation of the following form shall be determined:

$$\boxed{\frac{\Delta U_i}{\Delta t} \, + \, \frac{\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}}{\Delta x} \, = \, 0}$$

The Euler equations can, as already described in chapter 7, be transformed into the characteristic form by using a diagonal transformation:

$$\overline{\overline{A}} = \frac{\partial F}{\partial U} \qquad \Lambda \, = \, R^{-1} \, \overline{\overline{A}} \, R \qquad dW \, = \, R^{-1} \, dU \qquad \Lambda \, = \, \begin{pmatrix} u & 0 & 0 \\ 0 & u+a & 0 \\ 0 & 0 & u-a \end{pmatrix}$$

$$W_t \, + \, \Lambda \, W_x \, = \, 0$$

This characteristic equation system can be discretized analogous to the conservative form.

$$\boxed{\frac{\Delta W_i}{\Delta t} \, + \, \Lambda \, \frac{W_{i+1/2} - W_{i-1/2}}{\Delta x} \, = 0}$$

The different schemes which have been derived for scalar, hyperbolic equations, can be applied to this discretized, characteristic form.
The tracing back to the conservative form can be achieved by conservative reverse transformation. For the transformation of the difference equation it is assumed that the field of characteristics is locally fixed. This means that all coefficients, e.g. Λ, R, A, can be assumed to be constant at a point (x_i, t_n). Thus the transformation relation can also be defined for the differences Δf:

$$\Delta W = R^{-1} \Delta U \quad , \quad A = R \Lambda R^{-1} \quad , \quad \Delta F = A \, \Delta U$$

Central schemes

Central schemes of order $O(\Delta x^2)$ are obtained by forming the average for the cell wall, e.g. for $W_{i+1/2}$ the average is:

$$W_{i+1/2} = \frac{1}{2} \, (W_i + W_{i+1})$$

This yields a central scheme of the form

$$\frac{\Delta W_i}{\Delta t} + \Lambda \frac{W_{i+1} - W_{i-1}}{2\Delta x} = 0$$

The approximative reverse transformation with locally constant matrices back to the conservative form leads to the numerical flux of a central scheme, depending on the splitting:

$$\tilde{F}_{i+1/2} = F(\frac{U_i + U_{i+1}}{2}) \quad \text{res.} \quad \tilde{F}_{i+1/2} = \frac{F(U_i) + F(U_{i+1})}{2}$$

Both forms are consistent in the limit $\Delta x \to 0$ and approximate the central difference in the linear case. Differences occur for the non linear flux function. In practice both forms are used.

Additional numerical damping terms are required for central schemes to damp the numerical oscillations, as has already been discussed in chapter 6. Solutions of the Euler equations which also include discontinuities, e.g. shock waves, require, apart from the *high frequency damping terms* $D^{(4)} \sim U_{xxxx}$, damping terms to suppress non linear fluctuations close to the shock solution (i.e.*shock damping terms* $D^{(2)} \sim U_{xx}$). For the uniform, conservative presentation of a numerical flux formulation one defines the damping terms similar to the fluxes:

$$D_i(U) = \frac{d_{i+1/2}(U) - d_{i-1/2}(U)}{\Delta x}$$

Thus, the numerical flux of a central schemes including the damping terms is:

$$\boxed{\tilde{F}_{i+1/2} = \frac{1}{2}\left(F(U_i) + F(U_{i+1})\right) + d^{(4)}_{i+1/2}(U) - d^{(2)}_{i+1/2}(U)}$$

For completeness a detailed formulation of the damping terms like they often occur in computations is given.
(A. Jameson, W. Schmidt, E. Turkel: Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes. AIAA paper AIAA-81-1259, 1981)

- The shock damping term $D^{(2)} \sim U_{xx}$ suppresses non linear oscillations near the shock solution

$$d^{(2)}_{i+1/2}(U) = \frac{\Delta x}{\Delta t}\,\varepsilon^{(2)}_{i+1/2}\,(U_{i+1} - U_i)$$

This term causes a strong dissipation which is necessary for an oscillation free solution across a shock wave. This term is governed by the pressure change to restrict the dissipation to the area of the shock wave. Thus, one gets a small dissipation for weak changes in the pressure gradient.

$$\varepsilon^{(2)}_{i+1/2} = \kappa^{(2)}\,\max(\nu_i, \nu_{i+1}) \quad \text{with} \quad \nu_i = \frac{|p_{i-1} - 2p_i + p_{i+1}|}{p_{i-1} + 2p_i + p_{i+1}}$$

The damping term is fine tuned by the constant variable $\kappa^{(2)}$ which usually has a value $\kappa^{(2)} = O(1)$.

- The high frequency damping term $D^{(4)} \sim U_{xxxx}$ suppresses short waved oscillations in smooth solutions, e.g. those occurring because of round off errors.

$$d^{(4)}_{i+1/2}(U) = \frac{\Delta x}{\Delta t}\,\varepsilon^{(4)}_{i+1/2}\,(U_{i+2} - 3U_{i+1} + 3U_i - U_{i-1})$$

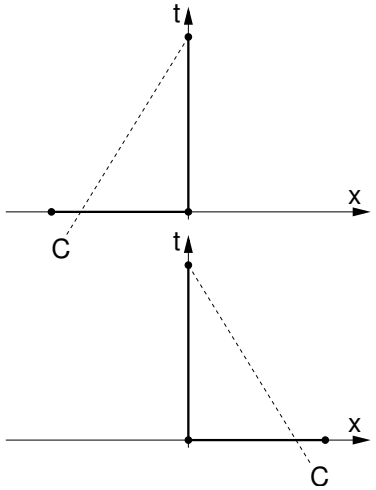
Close to the shock solution, the damping term causes a broadening of the shock, therefore it is suppressed in where the shock damping term becomes large. This procedure is called blended damping.

$$\varepsilon^{(4)}_{i+1/2} = \max\,(0, (\kappa^{(4)} - \varepsilon^{(2)}_{i+1/2})$$

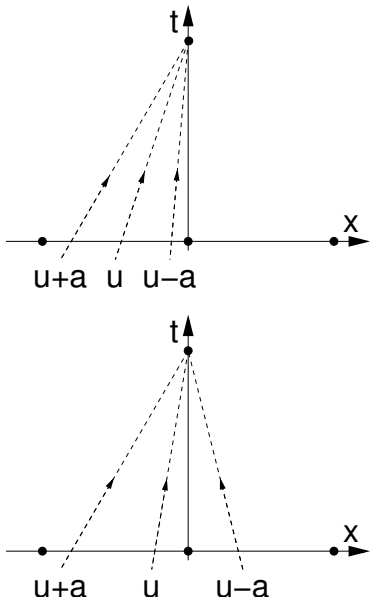
The constant variable $\kappa^{(4)} = O(10^{-2})$ serves the tuning of the high frequency damping term.

Upwind schemes

Upwind schemes take into account the characteristic direction of influence, when forming the differences. The space differences are formed from the direction of which the information is transported along the characteristics. This ensures a better numerical representation of the exact, characteristic solution.
This also applies for the numerical solution of the Euler equations. Upwind schemes are especially advantageous, if strong gas dynamic wave phenomena like shocks and expansions govern the flow field. This yields from the fact that they often show a better numerical resolution of the flow field. The necessity to take into account non linear fluxes for different characteristics makes the formulation of upwind schemes for the Euler equations much more complex than for instance for scalar, hyperbolic equations.



In time and space the Euler equations have three different real characteristics (eigenvalues) $\lambda_i = \frac{dx}{dt}|_i = u\,,\ u-a\,,\ u+a$. In supersonic flows ($u > a$) all characteristics have the same sign, but in subsonic flows one characteristic always has a sign different from the two remaining characteristics. As a consequence, different discretizations (forward and backward differences) are necessary for an upwind scheme, in the subsonic case. For the conservative flux F this relation is not clearly given since the flux is composed of components of varying characteristics. Therefore, a suitable splitting for the numerical flux must be found, such that the single components of the directions of influenced can be considered ($\Longrightarrow$flux splitting).



For the characteristic form of the Euler equations a unique splitting in components with positive and negative eigenvalues (directions of influence) and a formulation of upwind differences is possible. Starting from this fact, a reverse transformation back to the conservative form can yield the approximative splitting of the fluxes F in components with positive and negative eigenvalues and the formulation of upwind differences.
Varying conservative upwind schemes are possible because of the nonlinearity. Two major concepts, when deriving the upwind scheme, shall be discussed in the following without going into too much detail for the context of this course. These concepts are the flux vector splitting and the flux difference splitting.

Starting point for the derivation is the characteristic form of the Euler equations. Assuming the eigenvalue matrix Λ to contain positive as well as negative eigenvalues a splitting in corresponding components $\Lambda^{\pm}$ can be performed.

$$\Lambda = \Lambda^+ + \Lambda^- \quad \text{with} \quad \Lambda^+ > 0\ ,\ \Lambda^- < 0$$

With this the characteristic form is:

$$W_t + \Lambda^+ W_x + \Lambda^- W_x = 0$$

The discretization according to the principle of the upwind schemes is carried out with backward differences for Λ^+ and forward differences for Λ^-. Thus, i.e.:

$$\boxed{\frac{\Delta W_i}{\Delta t} + \Lambda^+ \frac{W_i - W_{i-1}}{\Delta x} + \Lambda^- \frac{W_{i+1} - W_i}{\Delta x} = 0}$$

This discretized form is the starting point for the various conservative flux splitting schemes.

Flux-Vector Splitting

The aim of the flux-vector splitting is to divide the numerical flux $\tilde{F}$ in components $F^{\pm}$ which solely include the positive and negative eigenvalues respectively. Their sum must represent the original flux, i.e. $F = F^{+} + F^{-}$. Lefthand multiplication of the discrete characteristic form with R and

$$R\Lambda^{\pm}\Delta W = R\Lambda^{\pm}R^{-1}\,R\Delta W = A^{\pm}\,\Delta U = \Delta F^{\pm}(U)$$

results in the conservative form:

$$\frac{\Delta U_i}{\Delta t} + \frac{F^{+}(U_i) - F^{+}(U_{i-1})}{\Delta x} + \frac{F^{-}(U_{i+1}) - F^{-}(U_i)}{\Delta x} = 0$$

This difference equation is formulated with the flux function $\tilde{F}_{i\pm1/2}$ for the numerical flux, for instance at the cell wall $i + 1/2$ of the control volume, which is then given as:

$$\tilde{F}_{i+1/2} = F^{+}(U_i) + F^{-}(U_{i+1})$$

For the introduction of a more general notion for the cell walls $i \pm 1/2$ lefthand extrapolated values $U^{+}_{i+1/2}$ are defined for positive eigenvalues Λ^{+}, where

$$U^{+}_{i+1/2} = U_i \quad , \quad U^{+}_{i-1/2} = U_{i-1}$$

The same is done for the righthand extrapolated values $U^{-}_{i+1/2}$ for negative eigenvalues Λ^{-}:

$$U^{-}_{i+1/2} = U_{i+1} \quad , \quad U^{-}_{i-1/2} = U_i$$

Using these definitions $U^{\pm}$ the numerical flux for the cell walls $i + 1/2$ becomes:

$$\boxed{\tilde{F}_{i+1/2} = F^{+}(U^{+}_{i+1/2}) + F^{-}(U^{-}_{i+1/2})}$$

Again, because of the nonlinearity of the flux function, various consistent splittings of the flux vector are possible.

The concept introduced by Steger and Warming for the Flux- Vector Splitting exploits the above given conservative reverse transformation. The split fluxes are obtained from:

$$F^{\pm}(U) = A^{\pm}\,U = R\Lambda^{\pm}R^{-1}\,RW$$

The positive and negative eigenvalues are defined as follows:

$$\Lambda^{\pm} = \frac{1}{2}\left(\Lambda \pm |\Lambda|\right)$$

The relations for the fluxes $F^{\pm}$ can therefore be determined from corresponding matrix and vector multiplications. Since the concept from Steger and Warming is not applied very often in the literature, further details won't be discussed at this point. *Steger, J.L., Warming, R.F.: Flux-vector splitting of the inviscid gas dynamic equations with applications to finite-difference methods. J. Comp. Phys., vol 40, pp 263-293, (1981)*

Much more often the concept of van Leer for Flux-Vector Splitting is used for the solution of the Euler equations.
van Leer, B.: Flux-vector splitting for the Euler equations. Lecture Notes in Physics vol. 170, pp. 507-512, (1982).
In this concept the split fluxes $F^{\pm}$ are approximated by a polynomial approach using the Mach number $Ma = \frac{u}{a}$. The following conditions are considered in this concept. First the eigenvalues or the corresponding Jacobian matrices $\frac{\partial F^+}{\partial U}$ and $\frac{\partial F^-}{\partial U}$ must be positive and negative, respectively. Second the overall flux must be conserved ($F = F^+ + F^-$) and finally there must be a strict transition for F^+ and F^- at Mach number $Ma = \frac{u}{a} = \pm 1$. It could be verified in many applications that this concept allows effective implicit schemes and satisfying resolutions of shock phenomena. The formulation of the fluxes $F^{\pm}$ will be given in the following without derivation. Some additional modifications have also been added.
Schwane, R., Hänel, D.: An implicit flux-vector splitting for the computation of viscous hypersonic flow. AIAA-paper No. 89-0274, (1989).

The fluxes for subsonic flow $-a \leq u \leq a$ are given as:

$$\begin{aligned} F_1^{\pm} &= \pm 1/4 \rho a (u/a \pm 1)^2 \\ F_2^{\pm} &= F_1^{\pm} \cdot (u + \frac{p}{\rho a}(-u/a \pm 2)) \qquad \text{für} \quad -a \leq u \leq a \\ F_3^{\pm} &= F_1^{\pm} \cdot H_t \end{aligned}$$

For supersonic flow, i.e. $u > a$ respectively $u < -a$ the flux is not split, since all eigenvalues have the same sign.

$$\begin{aligned} F^+ &= F \quad F^- = 0 \qquad \text{für} \quad u > a \\ F^+ &= 0 \quad F^- = F \qquad \text{für} \quad u < -a \end{aligned}$$

In this case the speed of sound is $a = \sqrt{\kappa \frac{p}{\rho}}$ and the total enthalpy is $H_t = E + \frac{p}{\rho}$. In general

an extension to higher accuracy of the upwind scheme is necessary in order to get a more accurate description of the flow in applied numerical calculations. The upwind scheme with Flux-Vector Splitting, like presented above, uses three points in space $i-1, i, i+1$ which only leads to a scheme of first order accuracy $O(\Delta x)$. Thus the scheme is highly dissipative.
An improvement can be obtained by using a polynomial for the approximation of the left and righthand side extrapolated values $U^{\pm}_{i+1/2}$ instead of using the direct neighboring values U_i and U_{i+1}. This polynomial is based on several points and extrapolated onto the cell walls. For a second order scheme it is sufficient to linearly extrapolate the variables onto the cell wall. For third order accuracy a second order extrapolation must be used. In general a polynomial with four basis points is used for the left respectively righthand side extrapolated values $U^{\pm}_{i+1/2}$, i.e.

$$\begin{aligned} U^+_{i+1/2} &= P^+ (\, U_{i-2},\, U_{i-1},\, U_i,\, U_{i+1},\,) \\ U^-_{i+1/2} &= P^- (\, U_{i-1},\, U_i,\, U_{i+1},\, U_{i+2}) \end{aligned}$$

An often applied polynomial formulation for upwind schemes was introduced by *van Leer* (MUSCL extrapolation $\Longrightarrow$Monotonic Upstream Schemes for Conservation Laws).
van Leer, B.: Towards the ultimate conservative difference scheme V. A second-order sequel to Godunov's method. J. Comp. Phys. vol.32, pp.101-136, (1979).
This formulation has already been mentioned in chapter 6. For the conservative variables U the formulation yields:

$$\begin{aligned}(U^+_{i+1/2})^+ &= U_i + \frac{1}{4}\,\varphi_i \cdot [(1+\kappa)(U_{i+1}-U_i) + (1-\kappa)(U_i - U_{i-1})] \\ (U^-_{i+1/2})^- &= U_{i+1} - \frac{1}{4}\,\varphi_{i+1} \cdot [(1+\kappa)(U_{i+1}-U_i) + (1-\kappa)(U_{i+2} - U_{i+1})]\end{aligned}$$

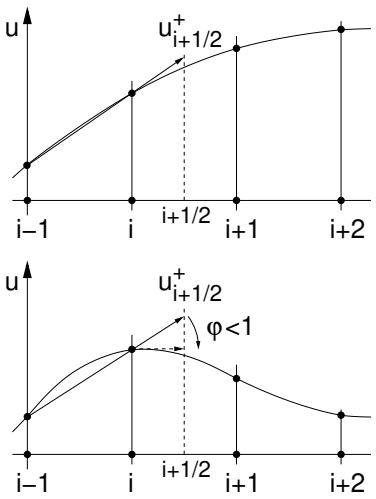
The numerical flux is formulated with these extrapolated values.

$$\tilde{F}_{i+1/2} = F^+(U^+_{i+1/2}) + F^-(U^-_{i+1/2})$$

With the aid of the parameter φ and the discretization parameter κ the scheme can be varied.

$$\begin{array}{llll} \varphi = 0 \quad . & \Rightarrow & O(\Delta x) & \text{1. order upwind} \\ \\ \varphi = 1 \quad \kappa = -1 & \Rightarrow & O(\Delta x^2) & \text{full upwind} \\ \varphi = 1 \quad \kappa = 0 & \Rightarrow & O(\Delta x^2) & \text{"half" upwind} \\ \varphi = 1 \quad \kappa = 1/3 & \Rightarrow & O(\Delta x^3) & \text{"half" upwind} \\ \varphi = 1 \quad \kappa = 1 & \Rightarrow & O(\Delta x^2) & \text{central} \end{array}$$

The parameter φ plays an important role for higher order upwind schemes, especially for shock capturing. When $\varphi = 0$, one obtains a first order scheme and thus a strongly dissipative behavior, while for $\varphi = 1$ the scheme is of higher order accurate. This feature can be exploited by tuning φ to suppress numerical oscillations.
For strong changes of U, e.g. close to shocks and especially close to extrema, the extrapolation of $U^{\pm}$ can result in an over or undersized value for the cell wall. This would cause unwanted oscillations of the numerical solution. The extrapolation is limited by decreasing the parameter φ to avoid this effect. Therefore, the parameter φ is also called flux limiter or slope limiter. The limitation of the extrapolation is locally achieved from the solution itself, by expressing the flux limiter with the change of U on the left and right side of the point i.



$$\varphi_i = \varphi(U_{i+1} - U_i\,,\, U_i - U_{i-1})$$

The flux limiter φ in the above definition varies between the values $\varphi = 0$ and $\varphi = 1$, i.e. between a first order and an at least second order scheme. A first order scheme generates a numerical dissipation $\sim U_{xx}$ and thus suppresses oscillations. This dissipation, therefore, supports the geometrical limitation. As a consequence the solution is smooth in areas of

strong fluctuations, e.g. shock waves. Outside of these areas the solution is governed by higher order schemes.
The flux limiter φ is a nonlinear function of the change in the variables. The formulation of the dependence on the change can mathematically be performed according to the theory of monotonic difference functions. A very successful method in this respect is the assumption of total variation diminishing (TVD). Essential methods for the development of high resolution difference schemes are based on these theories (see *Harten, Osher, Sweby etc.*)
In spite of the fact that the rigorous derivation, according to these theories is only valid for scalar, one dimensional equations, their application on systems and multi dimensional equations has lead to improved schemes. The flux limiter according to *van Albada,van Leer* and according to *Roe* will be given as examples for often applied flux limiters φ_i developed from the TVD theory. Using the abbreviations $\Delta^+ = U_{i+1} - U_i$ and $\Delta^- = U_i - U_{i-1}$ the flux limiters are:

$$\begin{aligned} \varphi_i|_{Alb} &= \frac{2\,\Delta^+ \cdot \Delta^-}{(\Delta^+)^2 + (\Delta^-)^2} \\ \varphi_i|_{Roe} &= \begin{cases} \min\left(|\frac{2\,\Delta^+}{\Delta^+ + \Delta^-}|\,,\, |\frac{2\,\Delta^-}{\Delta^+ + \Delta^-}|\,,\, 1\right) \\ 0 \qquad \text{falls} \quad sign(\Delta^+) \neq sign(\Delta^-) \end{cases} \end{aligned}$$

The following steps need to be performed for the formulation of the described upwind schemes with Flux Vector Splitting, starting from a known distribution U_i:

- Calculate the flux limiter φ_i
- Calculate the left and righthand extrapolated values $U^{\pm}_{i+1/2}$
- Calculate the split fluxes $F^{\pm}(U^{\pm}_{i+1/2})$
- Calculate the numerical flux $\tilde{F}_{i+1/2} = F^+(U^+_{i+1/2}) + F^-(U^-_{i+1/2})$
  and implement it in the given solution method.

Instead of using conservative variables, the extrapolation can also be performed with a set of different variables like for instance ρ, u, p. The presented formulation shows one possible approach, other different variants for Flux Vector Splitting can be found in the literature.

Flux-Difference Splitting

The concept of Flux-Difference Splitting was first originally introduced by *Roe* in:
Roe, P.L.: Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes. J. Comp. Phys., Vol.43 (1981).
In contrast to the flux vector concept, this method splits the numerical flux in a central formulated flux and an additional upwind term. The latter is added to the central flux, such that an increasing value of the characteristic leads to a unilateral upwind formulation for the numerical flux. A further effect of the upwind term is the damping of the numerical oscillations which occur in central difference schemes.

Starting point for the derivation of this concept is the discretized, characteristic form of the Euler equations:

$$\frac{\Delta W_i}{\Delta t} + \Lambda^+ \frac{W_i - W_{i-1}}{\Delta x} + \Lambda^- \frac{W_{i+1} - W_i}{\Delta x} = 0$$

The positive and negative eigenvalues $\Lambda^\pm$ are replaced by

$$\Lambda^\pm = \frac{1}{2}\left(\Lambda \pm |\Lambda|\right)$$

Restructuring of the equation yields the following difference equation:

$$\frac{\Delta W_i}{\Delta t} + \frac{1}{2}\Lambda \frac{(W_{i+1} + W_i) - (W_i + W_{i-1})}{\Delta x} - \frac{1}{2}|\Lambda| \frac{(W_{i+1} - W_i) - (W_i - W_{i-1})}{\Delta x} = 0$$

The first space difference corresponds to a central difference, while the second represents the upwind term. For $\Lambda > 0$ a backward difference is obtained for the discretization and for $\Lambda < 0$ a forward difference is obtained.
The conservative form

$$\frac{\Delta U_i}{\Delta t} + \frac{\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}}{\Delta x} = 0$$

can again be obtained by lefthand multiplication of the eigenvector matrix R. Then, the numerical flux results in

$$\boxed{\tilde{F}_{i+1/2} = \frac{1}{2}\left(F(U_i) + F(U_{i+1})\right) - \frac{1}{2}\left|A(\bar{U}_{i+1/2})\right|\left(U_{i+1} - U_i\right)}$$

Structurally the matrix $|A|$ corresponds to the Jacobian $A = \frac{\partial F}{\partial U} = R\Lambda R^{-1}$, except that it was constructed using the absolute eigenvalues, i.e.:

$$|A| = R\,|\Lambda|\,R^{-1}$$

This matrix has to be formulated at the cell wall $i + 1/2$. Therefore it is formed by mean values $\bar{U}_{i+1/2} = \bar{U}(U_i, U_{i+1})$. The arithmetic mean value $\bar{U}_{i+1/2} = \frac{U_i + U_{i+1}}{2}$ is suitable for this. A better shock resolution can be obtained with an averaging scheme according to *Roe*. The mean values are determined such that they satisfy the shock conditions for a steady shock between the points i and $i + 1$. This yields $\bar{U}_{i+1/2}$ such that

$$A(\bar{U}_{i+1/2}) \cdot (U_{i+1} - U_i) = F_{i+1} - F_i$$

From this the mean values of the velocity $\bar{u}$ and the total enthalpy $\bar{H}_t$ result in:

$$\bar{u}_{i+1/2} = \frac{D\,u_{i+1} + u_i}{D+1} \quad , \quad \bar{H}_{ti+1/2} = \frac{D\,H_{ti+1} + H_{ti}}{D+1} \quad , \quad D = \sqrt{\rho_{i+1}\,/\,\rho_i}$$

For $H_t = \frac{\kappa\,p}{(\kappa-1)\,\rho} + \frac{1}{2}u^2$ and $a^2 = \kappa\frac{p}{\rho}$, all necessary values for the calculation of $|A|$ are obtained.

In the case that an eigenvalue λ of the upwind matrix $|A|$ becomes very small (close to zero) the formulation of the numerical flux becomes problematic. The component of the upwind term $|A|\Delta U$ representing this eigenvalue gets lost and the difference equation becomes non dissipative for this component since only the central difference remains. For this case the direction of the change of entropy across a discontinuity is not fixed which can cause unphysical solutions, e.g. expansion shocks. In addition the elements $|\lambda_k|$ of the diagonal matrix $|\Lambda|$ can not be differentiated at $\lambda_k = 0$ which makes the convergence worse, especially for implicit schemes. Because of these reasons, when determining $|A|$, the matrix $|\Lambda|$ is replaced by an approximate diagonal matrix $Q(\lambda)$ which can be differentiated for $\lambda_k = 0$. An example for such a matrix with a method dependent constant, the so called entropy correction factor $\delta = O(10^{-1})$ is :

$$|\Lambda| \;\rightarrow\; Q(\lambda) = diag \begin{cases} \frac{1}{2}(\frac{\lambda^2}{\delta} + \delta) & |\lambda| < \delta \\ |\lambda| & |\lambda| \geq \delta \end{cases}$$

The upwind matrix $|A|$ is formed with this function, i.e.:

$$|A| = R\,Q(\lambda)\,R^{-1}$$

In the presented form the numerical flux $\tilde{F}_{i+1/2}$ in the Flux-Difference Splitting results in a first order upwind scheme as can be easily verified by considering the amount of base points, namely three.

For realistic applications a generalization for higher order accuracy is necessary.

The MUSCL-extrapolation like it has been presented for the Flux Vector Splitting is a possible option. Therefore, the values U_i and U_{i+1} in the numerical flux are replaced by left and right handedly extrapolated values $U^+_{i+1/2}$ and $U^-_{i+1/2}$, using the above given extrapolation scheme. This yields the numerical flux of the Flux-Difference Splitting

$$\tilde{F}_{i+1/2} = \frac{1}{2}(F(U^+_{i+1/2}) + F(U^-_{i+1/2})) - \frac{1}{2}\big|A(\bar{U}_{i+1/2})\big|\,(U^-_{i+1/2} - U^+_{i+1/2})$$

The modified flux approach according to *Harten* is another possibility to increase the accuracy of the scheme. This approach is most often applied for the Flux-Difference Splitting. *Harten, A.: On a Class of High Resolution Total-Variation-Stable Finite-Difference Schemes. SIAM J. Numer. Anal., Vol.21 (1983).*
In the modified approach an additional flux term $g_{i+1/2}$ is added to the already existing first order numerical flux.

$$\tilde{F}_{i+1/2} = \frac{1}{2}(F(U_i) + F(U_{i+1})) - \frac{1}{2}\left|A(\bar{U}_{i+1/2})\right|(U_{i+1} - U_i) \; + \; g_{i+1/2}$$

This flux term is chosen such that the first order upwind term is compensated which results in a scheme of second order accuracy. Close to shocks, where upwind differences and increased numerical dissipation is necessary for a better shock resolution, the flux term $g_{i+1/2}$ must diminish, such that the original first order upwind scheme is retained. The formal condition for the additional flux term is therefore:

$$\begin{array}{lcccll} g_{i+1/2} & \rightarrow & \frac{1}{2}\,|A|_{i+1/2}\,(U_{i+1} - U_i) & \rightarrow & \text{in smooth areas} & O(\Delta x^2) \\ g_{i+1/2} & \rightarrow & 0 & \rightarrow & \text{for extreme values, shocks} & O(\Delta x) \end{array}$$

The tuning of the flux term $g_{i+1/2}$ is carried out with a limiter function φ, as has been already described for the Flux-Vector Splitting.
In a strongly simplified form, the flux term $g_{i+1/2}$ can be formulated as follows:

$$g_{i+1/2} = \varphi_{i+1/2} \cdot \frac{1}{2}\,|A|_{i+1/2}\,\Delta^+ Q_i \qquad 0 \le \varphi \le 1$$

Examples for the limiter function φ are already given for the Flux-Vector Splitting. For strong changes in the variables (extrema, shocks) φ approaches zero, while for minor changes φ is close to one.

The presented simplified form shows the essential principle for the modified flux approach in higher order Flux-Difference schemes. For real world applications, further refinements are necessary. A presentation of different formulations of this concepts can for instance be found in:

Yee, H. C., Warming, R.F., Harten A.: Implicit Total Variation Diminishing (TVD) Schemes for Steady-State Calculations. J. of Comput. Physics, vol 57, pp 327-360, (1985).

Yee, H. C.: A Class of High-Resolution Explicit and Implicit Shock-Capturing Methods. Lecture Series 1989-04, von Karman Institute for Fluid Dynamics, Rhode-St-Genese, Belgium, (1989).

2.4.3 Time discretization (Solution methods)

In the preceeding section the space discretization of the fluxes were considered, and presented in the general conservative difference equation for the numerical flux function. This equation is the starting point for the different solution schemes which are defined by their discretization in time. These schemes essentially vary in the way they perform the discretization of the time derivative U_t and by the way the space discretization is done (e.g. implicit, explicit).
One can decide between schemes whose time discretization is dependent on or independent from the time discretization. If a time dependency exists, the steady solution also depends on the chosen time step. An example for this is the Mac Cormack scheme (Lax-Wendroff Method). This dependency is often disadvantageous, therefore most often schemes with independent time and space discretizations are used. Advantages are: time step independent, steady solutions, possible improvements of the convergence because of the application of iterative schemes for steady solutions and separate treatment of space and time discretization. The examples that will be discussed for this case in the following are the explicit Runge-Kutta method and implicit schemes in correction notation.

Mac Cormack's method, 1969

Mac Cormack's method is an example for schemes whose time discretization and space discretization are not independent from each other.
Mac Cormack's method is one of the first successful schemes for the solution of the Navier-Stoles and Euler equations. It is an often used scheme also today. This scheme has the same features for scalar equations as the Lax-Wendroff scheme (see chapter 6). It therefore corresponds to a central, explicit scheme of order $O(\Delta t^2, \Delta x^2)$.

The two step method for the Euler equations $U_t + F_x = 0$ is given as:

1. Step (Predictor):
$$\widetilde{U}_i = U_i^n - \frac{\Delta t}{\Delta x}\left(F(U_i) - F(U_{i-1})\right)$$

2. Step (Corrector step):
$$U_i^{n+1} = \frac{1}{2}\left(\widetilde{U}_i + U_i^n\right) - \frac{1}{2}\frac{\Delta t}{\Delta x}\left(F(\widetilde{U}_{i+1}) - F(\widetilde{U}_i)\right)$$
$\widetilde{U}_i$ is a temporary variable that is located between t_n and t_{n+1}. Since the scheme results in a central formulation, damping terms for the high frequency damping and for shock waves must be added. This can be performed for instance in a third step. Taking the variables U_i^{n+1} from the corrector step as temporary variable $\widetilde{\widetilde{U}}_i$, the third step can be formulated as:

3. Step (Damping):
$$U_i^{n+1} = \widetilde{\widetilde{U}}_i - D^{(4)}\left(\widetilde{\widetilde{U}}\right) + D^{(2)}\left(\widetilde{\widetilde{U}}\right)$$
The damping fourth and second order damping terms are equal to the above defined terms.

According to the linear stability analysis the scheme is stable for

$$C \;=\; (|u| + a) \cdot \frac{\Delta t}{\Delta x} \;\leq\; 1$$

in practice values for C a little smaller than 1 are used.
Without changes the scheme can also be applied to multi dimensional problems, by adding the fluxes and damping terms in the other coordinate directions correspondingly.

The Mac Cormack scheme is an effective and simple scheme for unsteady problems. A disadvantage is given in regard of steady solutions, since in this scheme the time discretization and space discretization is coupled, such that the latter also depends on the time step. Inserting the predictor step in the corrector step and assuming constant Jacobian matrices A, one obtains approximatively the one step Lax-Wendroff scheme:

$$\frac{U^{n+1} - U^n}{\Delta t} \;+\; \frac{F_{i+1} - F_{i-1}}{2\Delta x} \;-\; \frac{\Delta t}{2}\, A\, \frac{F_{i-1} - 2\,F_i + F_{i+1}}{\Delta x^2} = 0$$

As one can easily verify, the steady solution, i.e. if $U_i^{n+1} = U_i^n$, still depends on the time step Δt. This can result in accuracy problems for steady solutions.

Runge-Kutta time stepping schemes

The Runge-Kutta time stepping scheme is at the moment the perhaps most often applied explicit scheme for the solution of the conservation equations of compressible fluids. One of the first applications of this method for the Euler equations was published in:
A. Jameson, W. Schmidt, E. Turkel: Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge-Kutta Time-Stepping Schemes. AIAA paper AIAA-81-1259, 1981

With some modifications the Runge-Kutta scheme is widely applicable. Some applications from the literature are known for which the one and multi dimensional Euler and Navier-Stokes equations with central upwind discretization and for steady and also time accurate solutions were solved. The method has already been presented in chapter 6 for scalar equations. For equation systems it can be applied in equal fashion. Starting point is the conservative difference scheme, for which a solution U^{n+1} at time t_{n+1} as a function of the initial condition U^n is wanted.

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} \;+\; \frac{\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}}{\Delta x} \;=\; 0$$

The $\tilde{F}_{i\pm 1/2}$ are the central or upwind discretized numerical fluxes. The residual vector $Res_i(U)$ is introduced for abbreviation. It represents the steady space operators of the single equations and can also consist of multi dimensional components. For the one dimensional case the residual is defined as:

$$Res_i(U) = \frac{\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}}{\Delta x}$$

Thus the difference equations that needs to be solved are:

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} \;=\; -Res_i(U)$$

In analogy to the original Runge-Kutta scheme for ordinary differential equations, the integration for a time step Δt is performed in several explicit sub steps, marked with the step index k. Because of the minimal memory requirement, the following N–step scheme for the solution of partial differential equations in fluid dynamics has been proven successful:

$$\begin{aligned}
U_i^{(0)} &= U_i^n \\
U_i^{(1)} &= U_i^{(0)} - \alpha_1 \cdot \Delta t \cdot Res_i\left(U^{(0)}\right) \\
&\vdots \\
U_i^{(k-1)} &= U_i^{(0)} - \alpha_{k-1} \cdot \Delta t \cdot Res_i\left(U^{(k-2)}\right) \\
U_i^{(k)} &= U_i^{(0)} - \alpha_k \cdot \Delta t \cdot Res_i\left(U^{(k-1)}\right) \\
&\vdots \\
U_i^{n+1} &= U_i^{(N)}
\end{aligned}$$

The number of steps N is normally chosen between 3 and 5. The coefficients ($\alpha_k \leq 1$) can be determined such that, the truncation error in time becomes minimal, i.e. the largest time accuracy or order $O(\Delta t^N)$ is reached. One obtains from a Taylor series expansion:

$$\alpha_k = \frac{1}{N-k+1} \quad \text{with} \quad k = 1, 2, \cdots N$$

Another possibility is to optimize the coefficients for maximum stable time steps. The theoretically upper bound for stability is:

$$C_{max} = \min_i \left(\left(|u_i| + a \right) \frac{\Delta t}{\Delta x_i} \right) = N - 1$$

A typical set of coefficients for a central five step scheme of $O(\Delta t^2)$, proven to be suitable is:

$$\alpha_k = 0.25\,,\, 0.166\,,\, 0.375\,,\, 0.5\,,\, 1 \quad \text{for} \quad C_{max} = 4$$

and for upwind differences

$$\alpha_k = .059\,,\, .14\,,\, .273\,,\, .5\,,\, 1. \quad \text{for} \quad C_{max} = 3.5$$

Using the presented algorithm and suitable coefficients, the explicit solution for time accurate as well as asymptotical steady problems can be obtained.

The limitation of the time step Δt or the Courant number C, respectively is ineffective for steady problems, because of numerical instabilities. Since no time accuracy of the scheme is demanded in this case, the artificial acceleration of the scheme can be used to increase the convergence speed toward the steady solution. Two important possibilities will be presented in the following:

The application of local time steps allows for a given Courant number the maximum possible time step for each grid point. In the time accurate calculation a single time step is chosen, i.e. the smallest time step in the whole domain of integration because of stability

considerations. In contrast to this with local time stepping the time step varies between the grid points. The value of the time steps is determined from the local value of the spatial step and the maximum eigenvalue:

$$\Delta t_i = C \, \frac{\Delta x_i}{(|u| + a)_i}$$

The solution in the unsteady state is no longer consistent in time. However, the steady solution $Res(U) = 0$ is not influenced by the time step, since $U^{(k)}$ approaches U^{n+1} in the limit. The local time stepping allows the largest possible numerical propagation velocity on each grid point which leads to an acceleration of the calculation for steady solutions by using the local stability bounds of an explicit scheme. The advantage is especially large, if the space step size vary significantly.

For steady solutions, the implicit residual smoothing permits a higher Courant number C than the one dictated by stability considerations, i.e. C_{expl}. To achieve this, the residual $Res(U^{(k)})$, from the k th Runge-Kutta step, is implicitly averaged, such that the numerical propagation speed is increased and the local distribution of the residual is smoothed. A straightforward smoothing rule is given by the implicit formulated diffusion equation (Fourier equation) with the smoothed residual $\bar{Res}^k$ as variable.

$$\bar{Res}_i^k - \varepsilon \left(\bar{Res}_{i-1}^k - 2\bar{Res}_i^k + \bar{Res}_{i+1}^k \right) = Res_i^{(k-1)}$$

The new value U^k in the k th step of the Runge-Kutta scheme is calculated from the smoothed residual $\bar{Res}_i^k$, i.e.

$$U_i^{(k)} = U_i^{(0)} \; - \; \alpha_k \cdot \; \Delta t \cdot \bar{Res}_i^k$$

Using the following abbreviation: $\delta_{xx} \bar{Res}_i^k = (\bar{Res}_{i-1}^k - 2\bar{Res}_i^k + \bar{Res}_{i+1}^k)$ the k th smoothing and Runge-Kutta step can be combined as:

$$\begin{aligned} (1 - \varepsilon \, \delta_{xx}) \, \bar{\Delta U}^k &= - \alpha_k \cdot \; \Delta t \cdot Res_i \, (U^{(k-1)}) \\ U_i^{(k)} &= U_i^{(0)} \; + \; \bar{\Delta U}_i^k \end{aligned}$$

The solution of the scalar, tri diagonal system is performed with the Thomas algorithm and requires for equation systems a relatively small amount of computation time.
The smoothing parameter ε is chosen according to numerical reasons, i.e. such that the faster convergence towards a steady state is obtained. A stability analysis of the Runge-Kutta method with residual smoothing results in unlimited stability for:

$$\varepsilon \geq \frac{1}{4} \left[(\frac{C}{C_{expl}})^2 - 1 \right]$$

Good convergence rates in practical calculations were achieved for Courant numbers around two to three times bigger than their explicit value C_{expl} and by using values for ε from the above relation, when applying the equality sign.

Annotation: For multi dimensional problems, smoothing is applied in each direction, e.g. for 2-D:

$$\begin{aligned}(1-\varepsilon\,\delta_{xx})\,\bar{\bar{\Delta U}}^k &= -\alpha_k\cdot\ \Delta t\cdot Res_i\,(U^{(k-1)})\\ (1-\varepsilon\,\delta_{yy})\,\bar{\Delta U}^k &= \bar{\bar{\Delta U}}^k\\ U_i^{(k)} &= U_i^{(0)}\ +\ \bar{\Delta U}_i^k\end{aligned}$$

Implicit schemes for the Euler equations

For implicit schemes the variables of the space differences become unknowns at the time level t_{n+1}. This results in a equation system, coupled in the position space. The solution matrix can be solved directly with an elimination algorithm or approximated by an iteration. The advantage of the implicit formulation is the generally unlimited numerical stability which allows the choice of bigger time steps. For most cases the implicit scheme of the non linear conservation equations is essentially more robust for heavy fluctuations of the flow field, e.g. strong shock waves in hypersonic flows. On the other hand the essentially more complex algorithm and the bigger computation effort per time step for the solution of the equation system is a big drawback.
The method for the development of an implicit scheme has already been presented in the preceeding chapter for scalar model equations. The implicit solution of the system of the conservative Euler equations can be carried out analogous. In contrast to the scalar equations, a coupling of the variables U occurs, between the single equations of the system and in addition to the coupling in position space. This coupling occurs because of the dependency of the fluxes $F = F\,(U)$ on the conservative variable U. For an implicit scheme of the Euler equations the dependency of the flux F on the variables U is considered in the Jacobian matrices of the fluxes. The development of implicit schemes for such systems will be shown in the following.
For the conservative discretized Euler equations $U_t + F_x$ the approach of an implicit scheme with backward differences $O(\Delta t)$ in time is:

$$\frac{U_i^{n+1}\ -\ U_i^n}{\Delta t}\ +\ \frac{\widetilde{F}_{i+1/2}^{n+1}\ -\ \widetilde{F}_{i-1/2}^{n+1}}{\Delta x}\ =\ 0$$

The numerical flux $\widetilde{F}_{i\pm1/2}^{n+1}\ =\ \widetilde{F}\,(\,U_{i\pm1/2}^{n+1})$ is a function of the unknown U^{n+1} of the adjacent grid points. The numerical flux can be defined by central upwind formulation, including damping terms. To present a scheme of general formulation, the residual vector $Res\,(U)$ is defined which represents the sum of all space derivatives. For the above scheme the residual is:

$$Res\,(U)\ =\ \frac{\widetilde{F}_{i+1/2}\ -\ \widetilde{F}_{i-1/2}}{\Delta x}$$

With this the implicit scheme can be written as follows:

$$\frac{U_i^{n+1}\ -\ U_i^n}{\Delta t}\ +\ Res\,(U^{n+1})\ =\ 0$$

To present the dependence of the residual on U^{n+1}, it will be expanded in a Taylor series for the time t_n.

$$Res\,(U^{n+1}) \;=\; Res\,(U^n) \;+\; \frac{\partial Res(U^n)}{\partial t}\,\Delta t \;+\; \cdots$$

The time derivative of the residual can be presented as temporal development of the variable U:

$$\frac{\partial Res(U)}{\partial t}\,\Delta t \;=\; \frac{\partial Res(U)}{\partial U}\;\cdot\frac{\partial U}{\partial t}\,\Delta t \;=\; \frac{\partial Res(U)}{\partial U}\,(U^{n+1} \;- U^n) \;+\; \cdots$$

Introducing the definition of the correction variables $\Delta U^n \;\equiv\; U^{n+1} \;-\; U^n$ one obtains the implicit scheme:

$$\frac{\Delta U_i^n}{\Delta t} \;+\; \frac{\partial Res(U)}{\partial U}\;\cdot \Delta U \;=\; -\,Res(U^n)$$

The Jacobian matrix $\frac{\partial Res(U)}{\partial U}$ describes the dependence of the residual on the conservative variable U on the single grid points. According to the definition of the residual the matrix is composed of components of the numerical fluxes.

$$\frac{\partial Res(U)}{\partial U}\Delta U^n \;=\; \frac{1}{\Delta x}\left(\frac{\partial \widetilde{F}_{i+1/2}}{\partial U}\,\Delta U^n_{i+1/2} \;-\; \frac{\partial \widetilde{F}_{i-1/2}}{\partial U}\,\Delta U^n_{i-1/2}\right)$$

The components $\frac{\partial \widetilde{F}}{\partial U}$ correspond to the known Jacobian matrices of the fluxes for the conservative variables. In the following the derivation of an implicit method for the conservative Euler equations will be shown with a central scheme:

Example:
The numerical flux for a central scheme is:

$$\widetilde{F}_{i+1/2} \;=\; \frac{1}{2}\,(F\,(U_i) \;+\; F\,(U_{i+1}))$$

The dependence of the numerical flux $\widetilde{F}_{i+1/2} = \widetilde{F}(U_i, U_{i+1})$ on U thus results in

$$\frac{\partial \widetilde{F}_{i+1/2}}{\partial U}\,\Delta U^n \;=\; \frac{\partial \widetilde{F}\,(U_i, U_{i+1})}{\partial (U_i, U_{i+1})}\,\Delta U^n \;=$$

$$\frac{1}{2}\left(\frac{\partial F(U_i)}{\partial U_i}\,\Delta U_i^n \;+\; \frac{\partial F(U_{i+1})}{\partial U_{i+1}}\,\Delta U_{i+1}^n\right) \;=\; \frac{1}{2}\left(\overline{\overline{A}}_i \cdot \Delta U_i^n \;+\; \overline{\overline{A}}_{i+1} \cdot \Delta U_{i+1}^n\right)$$

Using an analogous expansion for $\widetilde{F}_{i-1/2}$ and by application of the residual definition $Res(U) = \frac{1}{2\Delta x}\,(F(U_{i+1}) \;-\; F(U_{i-1}))$ the following implicit scheme is obtained:

$$\frac{\Delta U_i^n}{\Delta t} \;+\; \frac{1}{2\Delta x}\left(\overline{\overline{A}}_{i+1} \cdot \Delta U_{i+1}^n \;-\; \overline{\overline{A}}_{i-1} \cdot \Delta U_{i-1}^n\right) \;=\; -\,Res(U^n)$$

Sorting the unknowns ΔU , yields a block tri-diagonal equation system of the form:

$$\overline{\overline{a}}_i \cdot \Delta U_{i-1}^n \;+\; \overline{\overline{b}}_i \cdot \Delta U_i^n \;+\; \overline{\overline{c}}_i \cdot \Delta U_{i+1}^n \;=\; -\,Res(U^n)$$

$$\text{where} \qquad \overline{\overline{a}}_i \;=\; -\,\frac{1}{2\Delta x}\overline{\overline{A}}_{i-1} \quad , \quad \overline{\overline{b}}_i \;=\; \frac{1}{\Delta t}\cdot I \quad , \quad \overline{\overline{c}}_i \;=\; \frac{1}{2\Delta x}\overline{\overline{A}}_{i+1}$$

This equation system is similar to the one obtained for scalar equations with the difference that the coefficients $\overline{a}, \overline{b}, \overline{c}$ are block matrices of the rank of the Jacobian matrices $\overline{\overline{A}}$ (in this case 3x3 matrices). The solution of the equation system is performed by Gaussian elimination (Thomas algorithm). Inserting the recursion into the scheme

$$\Delta U_i^n \;=\; \overline{\overline{E}}_i \cdot \Delta U_{i+1}^n \;+\; \vec{F}_i$$

yields the recursion matrix $\overline{\overline{E}}_i$ and the vector $\vec{F}_i$

$$\overline{\overline{E}}_i \;=\; (\overline{\overline{a}}_i\,\overline{\overline{E}}_{i-1} \;+\; \overline{\overline{b}}_i)^{-1}\cdot\overline{\overline{C}}_i \quad \text{and} \quad \overline{\overline{F}}_i \;=\; (\overline{\overline{a}}_i\,\overline{\overline{E}}_{i-1} \;+\; \overline{\overline{b}}_i)^{-1}\left(-Res(U^n) \;-\; \overline{\overline{a}}_i\cdot\overline{\overline{F}}_{i-1}\right)$$

The coefficients for $i = 2, 3, \cdots, i_{max}$ can be determined for the boundary conditions at $i = 1$. The new variables are calculated with the boundary conditions at $i = i_{max}$:

$$\begin{aligned} \Delta U_i^n &= \overline{\overline{E}}_i \cdot \Delta U_{i+1} \;+\; \vec{F}_i \\ U_i^{n+1} &= U_i^n \;+\; \Delta U_i^n \end{aligned}$$

This example demonstrates the principle of the solution of an implicit scheme for systems of differential equations.
The application of higher order upwind schemes or central schemes with damping generally leads to a position operator which is constructed from five grid points, i.e. with U_{i-2}, U_{i-1}, U_i, U_{i+1}, U_{i+2}. With this the residual becomes

$$Res\,(U) \;=\; Res\,(U_{i-2},\, U_{i-1},\, U_i,\, U_{i+1},\, U_{i+2})$$

In analogy to the preceeding derivation an equation system emerges. This equation system is coupled by five variables and is known as a penta-diagonal system:

$$\overline{\overline{d}}_i\,\Delta U_{i-2} \;+\; \overline{\overline{a}}_i\,\Delta U_{i-1} \;+\; \overline{\overline{b}}_i\,\Delta U_i \;+\; \overline{\overline{c}}_i\,\Delta U_{i+1} \;+\; \overline{\overline{e}}_i\,\Delta U_{i+2} \;=\; -\,Res(U^n)$$

The solution of such a system can also be performed by Gaussian elimination and should be carried out for time accurate calculations. But the effort is higher than for a tri-diagonal system.
If time accuracy is not required, like e.g. for steady solutions, one therefore often simplifies the implicit operator (= the lefthand side of the difference equation). This is a valid measure, since the steady solution $Res\,(U^n) = 0$ is obtained if the correction variables ΔU^n diminish. Thus, the steady solution becomes independent from the way in which the implicit scheme converges towards $\Delta U^n = 0$. An often applied approximation for the implicit operator is to formulate its space differences with a first order scheme, while approximating the space operators in the residual $Res(U^n)$ which define the accuracy of the solution, with higher order differences. The first order space operators for the lefthand side are only a function of the variables located at $i-1$, i , $i+1$ which results in a simplified tri-diagonal equation system of the following form

$$\overline{\overline{a}}_i \cdot \Delta U_{i-1}^n \;+\; \overline{\overline{b}}_i \cdot \Delta U_i^n \;+\; \overline{\overline{c}}_i \cdot \Delta U_{i+1}^n \;=\; -\,Res\,(U_{i-2},\, U_{i-1},\, U_i,\, U_{i+1},\, U_{i+2},\,)$$

An example for this is the above presented implicit scheme for a central difference with application of the damping terms $D^{(2)}$ and $D^{(4)}$, as described above. The numerical flux for the residual $Res(U)$ is:

$$\widetilde{F}_{i+1/2} = \frac{1}{2}\left(F\left(U_i\right) + F\left(U_{i+1}\right)\right) + d^{(4)}(U) + d^{(2)}(U)$$

Since the damping term $D^{(4)} = d^{(4)}_{i+1/2} - d^{(4)}_{i-1/2}$ includes the values on five grid points, the damping terms $D^{(4)}$ and $D^{(2)}$ in the implicit component are replaced by an approximated term $D^{(2)}_I$. This leads to a simplified flux for the creation of the implicit operator:

$$(\widetilde{F}_{i+1/2})_{impl} = \frac{1}{2}\left(F\left(U_i\right) + F\left(U_{i+1}\right) - \varepsilon_I \frac{1}{\Delta t}(U_{i+1} - U_i)\right)$$

The Jacobian matrices become:

$$\frac{\partial \widetilde{F}_{i+1/2}}{\partial U}\Delta U = \frac{1}{2}\left(\overline{\overline{A}}_i \cdot \Delta U_i + A_{i+1} \cdot \Delta U_{i+1}\right) - \varepsilon_I/\Delta t\ (\Delta U_{i+1} - \Delta U_i)$$

The upwind schemes are treated correspondingly. For them the residual is formulated with a high order upwind discretization, while the construction of the implicit operator is performed with the corresponding first order scheme.

An extension of the implicit scheme for the Euler equations to multiple dimensions is most often performed approximatively, as described for the scalar equations. An important method in this respect is the method of approximated factorization and the iteration schemes applying the Gauss-Seidel point or line iteration. For details please refer to the specialized literature.

2.4.4 Simulation of a one dimensional flow problem – shock tube flow

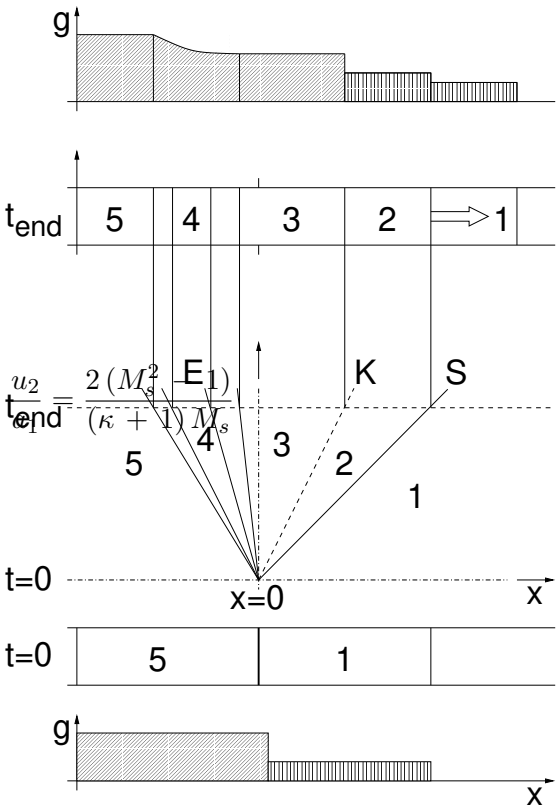
An often used test case for the presentation and validation of different solution schemes for the Euler equations is the numerical calculation of the flow in a shock tube. This test case involves essential flow phenomena of compressible, inviscid flows like shock waves, discontinuities and expansion waves.

The physical problem includes the flow and the wave phenomena in a shock tube. In a straight tube, separated by a membrane, the lefthand part is filled with gas of state (5) while the righthand part is filled with gas of state (1). After the burst of the membrane a shock wave S propagates in the low pressure part, followed by the contact discontinuity K which forms the separation surface between the to gases in the high and low pressure parts. The pressure in the high pressure region is reduced by an unsteady expansion wave. For the following example it is assumed that equal gases are chosen for the high and low pressure parts.

The analytical calculation is performed for the given states (1) and (5) with the coupling of the states across the propagating shock wave, across the discontinuity and across the expansion wave until the high pressure region.
For the shock propagating with velocity V_s the jumping conditions $[H - V_s\,U]_1^2 = 0$ are valid. After rearrangement of the jumping conditions one obtains state(2) as a function of the yet unknown shock Mach number $M_s = \frac{v_s}{a_1}$ (Rankine-Hugoniot relation). E.g.:

$$\frac{p_2}{p_1} = \frac{2\,\kappa\,M_s^2 - (\kappa - 1)}{\kappa + 1}, \quad \frac{\rho_2}{\rho_1} = \frac{(\kappa + 1)\,M_s^2}{2 + (\kappa - 1)\,M_s^2},$$

The contact surface moves as a material boundary with the flow velocity u_2. The jumping condition $[H - u_2\,U]_2^3 = 0$ leads to constant pressure and velocity across the contact surface, i.e. $p_3 = p_2$ and $u_3 = u_2$. All other properties change across the discontinuity depending on the initial state.
From state (3) behind the contact surface until state (5) in the high pressure part the flow is isentrop and the total enthalpies H_5 and H_3 are equal, i.e.:



$$H_5 = c_p\,T_5 = H_3 = c_p\,T_3 + u_3^2/2 \quad , \qquad \frac{p_5}{p_3} = \left(\frac{T_5}{T_3}\right)^{\frac{\kappa}{\kappa-1}} = \left(\frac{\rho_3}{\rho_5}\right)^{\frac{\kappa-1}{\kappa}}$$

By coupling of the relations the shock Mach number from the states (1) to (5) can be calculated. For equal gases ($\kappa_5 = \kappa_1 = \kappa$) one obtains the shock Mach number M_s by iteration:

$$\frac{p_5}{p_1} = \left(1 + \frac{2\kappa}{\kappa+1}\left(M_s^2 - 1\right)\right) \Big/ \left(1 - \frac{a_1}{a_5} \cdot \frac{M_s^2 - 1}{\frac{\kappa+1}{\kappa-1} M_s}\right)^{\frac{2\kappa}{\kappa-1}}$$

With this the states (2) and (3) can completely be calculated.
Inside the unsteady expansion wave, i.e. state (4), isentropy is conserved, while the equality of the total enthalpies is not conserved. The state in the expansion wave (4) is calculated according to the method of characteristics. The characteristic solution (see chapter 7) is:

$$dp \pm \rho\, a\, du = 0 \qquad \text{für} \qquad \frac{dx}{dt} = u \pm a$$

By application of the isentropy relation $dp = a^2\, d\rho$ and the equation for a perfect gas $p = \rho RT$ these solutions can be integrated. One obtains:

$$\frac{2}{\kappa - 1}\, a \pm u = const \qquad \text{für} \qquad \frac{dx}{dt} = u \pm a$$

The characteristics $\frac{dx}{dt} = u - a$ describe straight lines with of the equation $x - (u-a)t = 0$, coming from the membrane origin ($t = 0$, $x = 0$). On these lines $\frac{2}{\kappa-1}(a-u) = const$ holds. The lines intersect with characteristics of $\frac{dx}{dt} = u + a$ which emerge from the lefthand state (5). For them the following applies:

$$\frac{2}{\kappa - 1}\, a + u = \frac{2}{\kappa - 1} \cdot a_5$$

From the equation of a straight line $x - (u - a) \cdot t = 0$ one therefore obtains the states in the expansion wave, e.g.

$$\frac{u}{a_5} = \frac{2}{\kappa+1}\left(1 + \frac{x}{a_5\, t}\right) \quad , \qquad \frac{a}{a_5} = \frac{\kappa - 1}{\kappa + 1}\left(1 + \frac{x}{a_5\, t}\right)$$

The other gas properties can be determined from the isentropy relation. The limiting characteristics of the expansion wave are $x - a_5 t = 0$ on the lefthand side and $x - (u_3 - a_3)t = 0$ on the righthand side.

The numerical solution is performed for initial states, taken from a test case formulated in the literature:

Yee, H. C.: A Class of High-Resolution Explicit and Implicit Shock-Capturing Methods. Lecture Series 1989-04, von Karman Institute for Fluid Dynamics, Rhode-St-Genese, Belgium, (1989).

The following initial conditions for $t = 0$ are given:

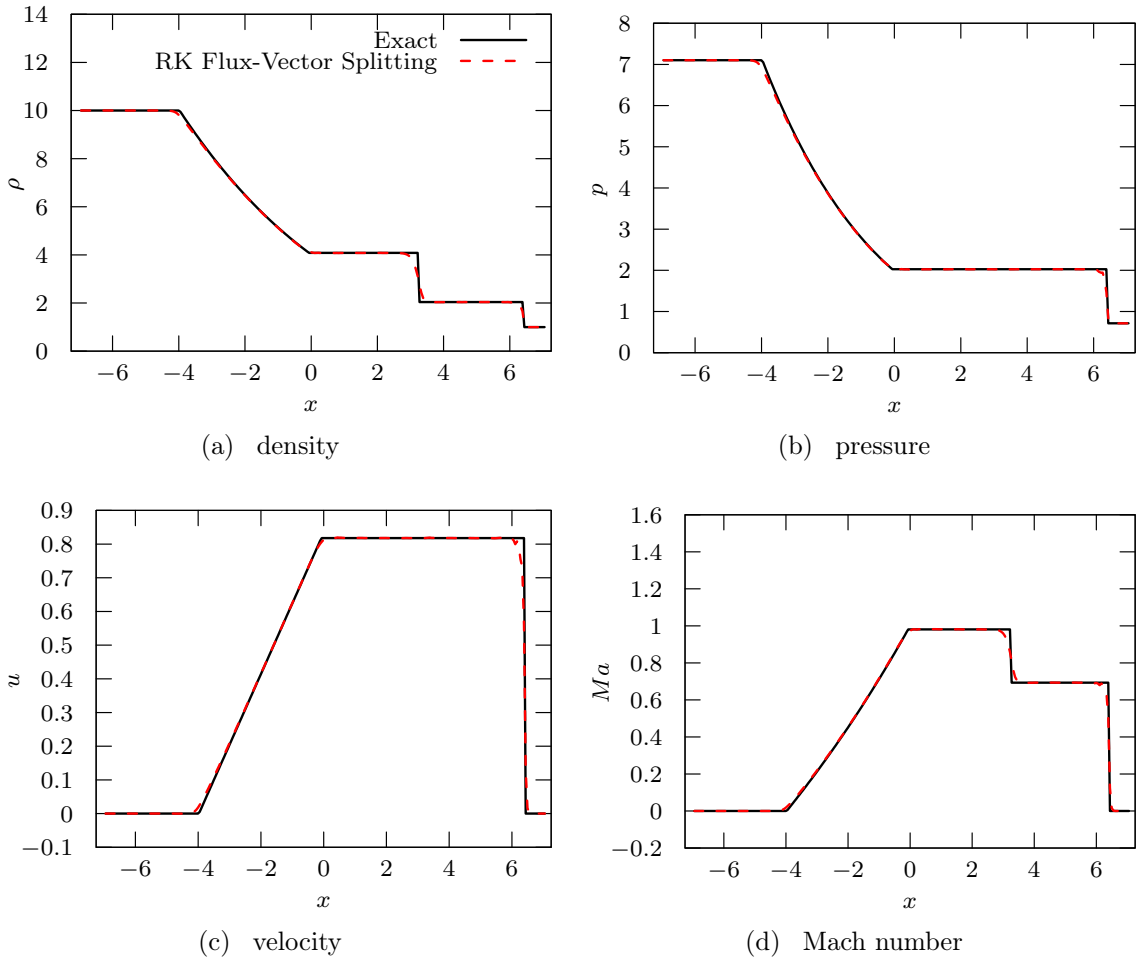
$$\begin{array}{llllll} x \le 0: & u_5 = 0 & \rho_5 = 1,4\ \frac{kg}{m^3} & T_5 = 2438\ \ K & p_5 = 9,88 \cdot 10^5\ \frac{N}{m^2} \\ x > 0 & u_1 = 0 & \rho_1 = 0,14\ \frac{kg}{m^3} & T_1 = 2452\ \ K & p_1 = 9,93 \cdot 10^4\ \frac{N}{m^2} \end{array}$$

The domain of integration is $-7m \leq x \leq +7m$. The numerical discretization is performed with $imax = 141$ grid points. This yields a constant step size $\Delta x = 0,1m$. The presentation of the results is done after a time $t_{end} = 4 \cdot 10^{-3} sec$. At this time the shock and expansion wave still reside inside the domain of integration.
Fixed walls are given as boundary conditions on the left and righthand side. The condition for a fixed wall is $u = 0$. By application of the momentum equations this yields $p_x = 0$ which is numerically approximated by substitution of the wall value with the next inner point.
The time step Δt is calculated from the Courant number C, which must be defined for the different schemes.

$$\Delta t = C \cdot \Delta x \,/\, \max_i |\lambda| \qquad \text{where} \qquad \max_i |\lambda| = \max_i (\,|u| + a)$$

The solution of the Euler equations is carried out with various, explicit schemes to demonstrate some typical solution behaviors.

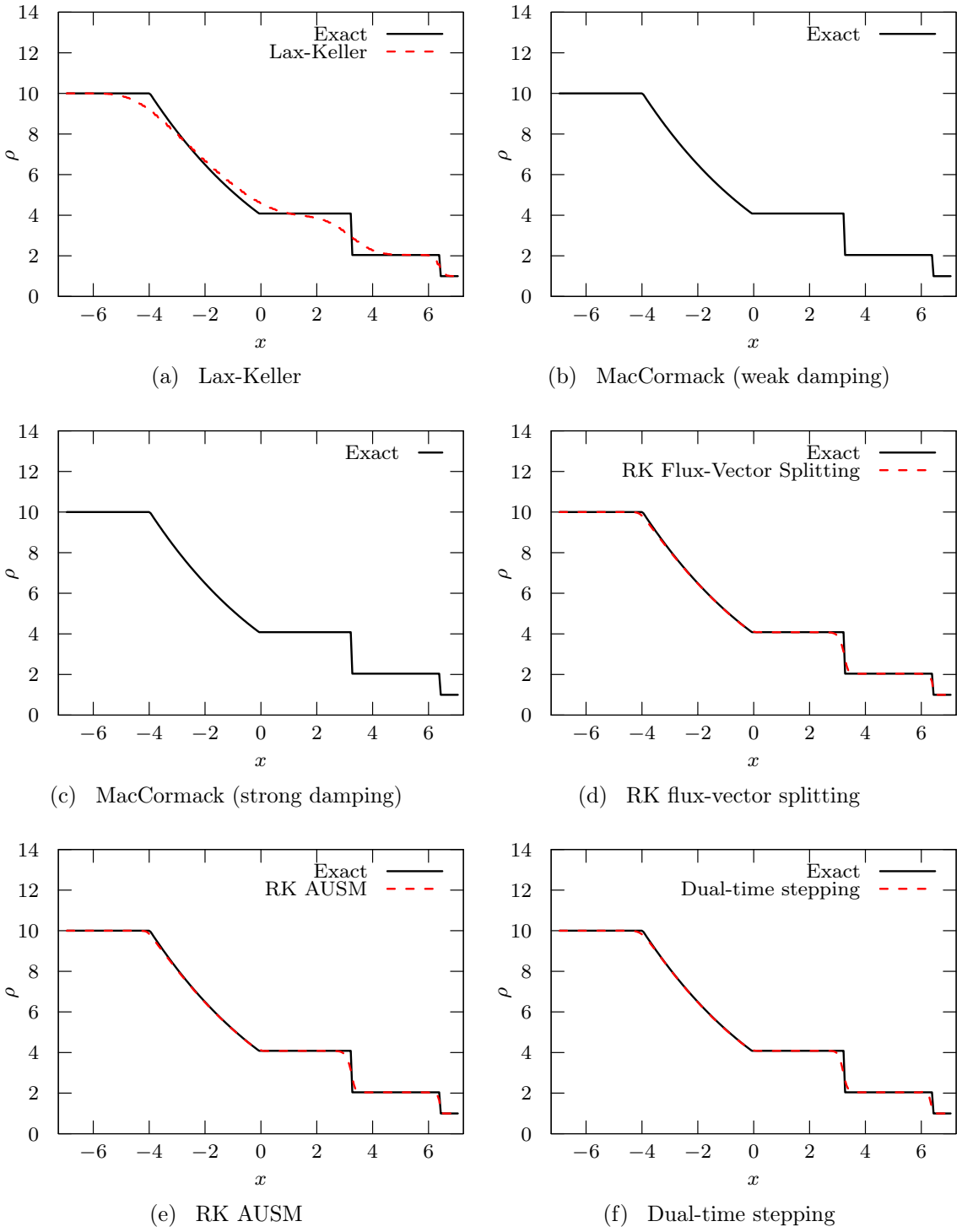
The following figures present results from the numerical solution of the shock tube problem with the Euler equations, as described above. The circles correspond to the numerical solution at the single grid points, while the solid line presents the exact solution along the variable length x at the given time t_{end}, after the membrane burst.
The figures 2.4.1 -a) to d) show respectively, density, pressure, Mach number and velocity for the shock tube problem. With these properties the typical wave phenomena like expansion waves, contact surfaces and shock waves can be recognized. It can also be observed that the pressure and velocity are constant across the contact surface. The numerical solution was in this case obtained with a Runge-Kutta scheme applying a Courant number $C = 1.5$. The space discretization has been performed with a Flux-Vector splitting scheme with limiter function according to van Albada and van Leer, in the above described fashion.
The figures 2.4.2 a) to d) show the course of the density in order to express the solution qualities of different methods.
Figure 2.4.2 a) shows the solution according to the Lax-Keller scheme as an example for a fist order accurate scheme, $O(\Delta x, \Delta t)$ and $C = 1$. The effects of an excessive numerical viscosity, leading to a smearing out of the discontinuities can be clearly seen from its course.
In the figures 2.4.2 b) and c) the (central) Mac Cormack scheme was applied with $C = 0.8$. The scheme displays very strong oscillations without damping, it even gets instable in the given case. Applying a weak shock damping term $\varepsilon^{(2)} = 0.1$ in figure b), the solution remains stable, but still displays noticeable oscillations. Stronger dampers $\varepsilon^{(2)} = 0.25$ and additional high frequency damping $\varepsilon^{(4)} = 0.05$ in figure c) result in a smoother solution, but still not free of oscillations. These courses present the typical behavior of central schemes which require thorough treatment of the damping terms if discontinuities are involved.
In figure d) the solution of the method applying Flux-Vector splitting with limiter function is presented once more for a comparison of the two methods. The course of the solution is smooth and captures the discontinuities very accurately. The good resolution of discontinuities is in general a typical quality for upwind schemes with TVD behavior, like the presented examples for Flux-Vector and Flux-Difference splitting. However, it should be noted that the computational effort of such schemes is essentially higher than for central schemes.



(a) density (b) pressure

(c) velocity (d) Mach number

Figure 2.4.1: Solution of the Euler equations for the shock tube problem obtained with the Runge-Kutta flux-vector splitting scheme. Course of density, pressure, velocity, and Mach number along the x-coordinate at a time t_{end}.



(a) Lax-Keller

(b) MacCormack (weak damping)

(c) MacCormack (strong damping)

(d) RK flux-vector splitting

(e) RK AUSM

(f) Dual-time stepping

Figure 2.4.2: Solution of the Euler equations for the shock tube problem. Course of the density along the x-coordinate at a time t_{end}. a) Lax-Keller scheme with $C = 1$. b) MacCormack scheme with $C = 0.8$, $\varepsilon^{(2)} = 0.1$ c) MacCormack scheme with $C = 0.8$, $\varepsilon^{(2)} = 0.25$, $\varepsilon^{(4)} = 0.05$ d) Runge-Kutta scheme with flux-vector splitting and van Albada limiter e) Runge-Kutta scheme with Advection Upstream Splitting Method (AUSM) f) Dual-time stepping scheme with artificial time derivative.

2.4.5 Space discretization in multiple dimensions

The numerical solution of the one dimensional Euler equations as treated above is the starting point for the solution of the equations in two and three dimensions. In most schemes a quasi one dimensional flux discretization is formulated for each coordinate direction. Therefore, one obtains a scheme which in principle consists of the superposition of one dimensional discretizations.
The following considerations are limited to the two dimensional physical space, to keep matters simple. The extension to three dimensions, can in general be achieved in an analogous way.
The Euler equations in general integral or divergence formulation are:

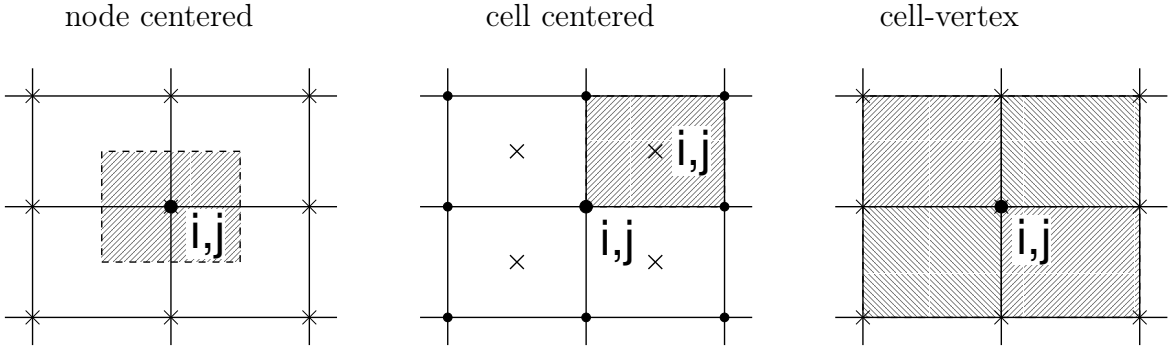
$$\int_\tau U_t \, d\tau + \oint_A \vec{H} \cdot \vec{n} \, dA = 0 \qquad \text{resp..} \qquad U_t + \nabla \cdot \vec{H} = 0$$

A two dimensional Cartesian coordinate system x, y, t is used as reference system. With the cartesian components of the flux vector $\vec{H} = \binom{F}{G}$, the nabla operator $\nabla = \binom{\frac{\partial}{\partial x}}{\frac{\partial}{\partial y}}$ and the vector of the surface normal $\vec{n} \, dA = \binom{dy}{-dx}$ one obtains the conservation equations in integral, respectively in divergence formulation:

$$\int_\tau U_t \, d\tau + \oint_A F \cdot dy - \oint_A G \cdot dx = 0 \quad \text{respectively} \quad U_t + F_x + G_y = 0$$

Both forms are a starting point of the conservative discretization of the Euler equations. The formulation of the conservation of mass, momentum and energy in the discrete space is performed with a small control volume $\tau = 0(\Delta x \cdot \Delta y)$ of the computational grid.

The definition of a control volume for a given grid is an essential step in the conservative space discretization. In multiple dimensions this can be done in various ways. The most important configurations for the control volume are:

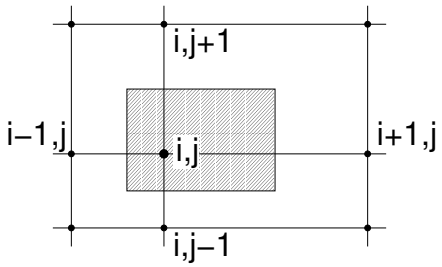


In the node-centered scheme the variables U and the space coordinates (x, y) are defined at the same grid point. The boundary of the control volume is chosen in the middle between two neighboring points.
In the cell centered scheme only the space coordinates are defined at the grid points. The variables U are assumed in the center of the cell defined by four grid points. Therefore, the

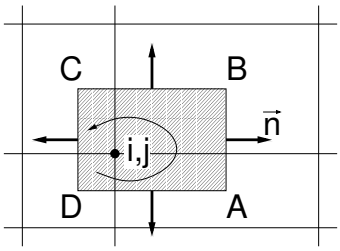
indexing of the variables $U_{i,j}$ and the geometry $x_{i,j}, y_{i,j}$ doesn't apply for the same place. The cell-vertex scheme is in principle composed from four control volumes for the calculation of the values $U_{i,j}$ in the central point $x_{i,j}, y_{i,j}$. Variables and coordinates are defined on the same grid.
All three configurations are commonly applied in the literature. They differ in the formulation of the numerical flux and the boundary conditions, without any essential advantages or disadvantages of one of the configurations. The different control volumes can be formulated in Cartesian as well as in general curved coordinates. (Similar configurations can be applied for triangulated grids with triangle shaped cells. But the formulation for this case leaves the scope of this course.) For the following considerations the node-centered scheme of the control volume serves as an example.

Space discretization in two dimensions on Cartesian grids

For the discretization in a Cartesian space, one chooses a grid who's points are oriented along the axis directions. The following index notation for the grid points is used: $i = 1, 2, \cdots, i_{max}$ for the x-direction and $j = 1, 2, \cdots, j_{max}$ for the y-direction. The step size between the grid points are no longer assumed to be constant.



The discretization with the integral form of the Euler equations offers a straightforward physical interpretation of the discretization. The conservation laws are directly applied on a small, finite value τ, therefore this method is referred to as the "Finite-Volume Method".
For the formulation of the temporal change of the conservative variables U in the volume τ one considers the variables $U_{i,j}$ as spatially averaged across τ. For the control volume τ_{ABCD} one obtains:



$$\int_{\tau} U_t \, d\tau \quad \rightarrow \quad \frac{\Delta U_{i,j}}{\Delta t} \cdot \tau_{ABCD}$$

The temporal change is in equilibrium with the fluxes normal to the surface of τ. The fluxes across a cell boundary of the control volume are assumed to be piecewise constant. Positive signs are determined according to the outward facing surface normal. For a given mathematically positive turn direction for the four sides of the control volume the flux integral yields:

$$\oint \vec{H} \cdot n \, dA \approx \hat{H}_{AB} + \hat{H}_{BC} + \hat{H}_{CD} + \hat{H}_{DA}$$

This relation can be simplified for a Cartesian element, since then the surface normal always

coincides with the coordinate direction.

$$\oint_A F\,dy \;-\; \oint_A G\,dx \;\approx\; F_{AB}\,\Delta y_{AB} \;+\; F_{CD}\,\Delta y_{CD} \;-\; G_{BC}\,\Delta x_{BC} \;-\; G_{DA}\,\Delta x_{DA}$$

According to the turn direction it is:

$$\Delta y_{AB} \;=\; y_B \;-\; y_A \quad , \qquad \Delta x_{BC} \;=\; x_C \;-\; x_B \quad , \cdots$$

The discrete integral form of the Euler equations therefore becomes:

$$\frac{\Delta U_{i,j}}{\Delta t} \cdot \tau_{ABCD} \;+\; F_{AB}\,\Delta y_{AB} \;+\; F_{CD}\,\Delta y_{CD} \;-\; G_{BC}\,\Delta x_{BC} \;-\; G_{DA}\,\Delta x_{DA} \;=\; 0$$

For the formulation of an algorithm it is in general beneficial to introduce an indexing with i, j. This yields:

$$\Delta x_i = \frac{x_{i+1} \;-\; x_{i-1}}{2} = \Delta x_{DA} = -\Delta x_{BC} \quad , \quad \Delta y_j = \frac{y_{j+1} \;-\; y_{j-1}}{2} = \Delta y_{AB} = -\Delta y_{CD}$$

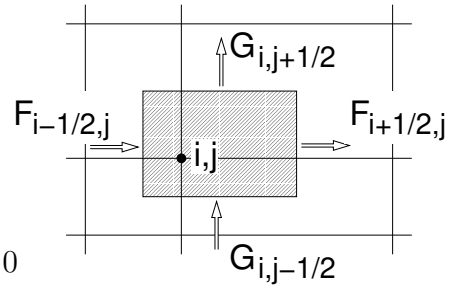
$$\tau_{i,\,j} \;=\; \tau_{ABCD} \;=\; \Delta x_i \cdot \Delta y_j$$

$$\begin{aligned} \tilde{F}_{i+1/2,\,j} \;&=\; F_{AB} \quad , \quad \tilde{F}_{i-1/2,\,j} \;=\; F_{CD} \\ \tilde{G}_{i+1/2,\,j} \;&=\; G_{BC} \quad , \quad \tilde{G}_{i,\,j-1/2} \;=\; G_{DA} \end{aligned}$$

With these relations one obtains the following formulation for the integral form:

$$\frac{\Delta U_{i,j}}{\Delta t}\,\Delta x_i \cdot \Delta y_j \;+\; \left(\tilde{F}_{i+1/2,\,j} \;-\; \tilde{F}_{i-1/2,\,j}\right)\Delta y_j \;+\; \left(\tilde{G}_{i,\,j+1/2} \;-\; \tilde{G}_{i,\,j-1/2}\right)\Delta x_i \;=\; 0$$

For the discretization of the divergence form the space derivatives are replaced by differences of the numerical fluxes $\tilde{F}_{i\pm1/2,\,j}$ and $\tilde{G}_{i,\,j\pm1/2}$ at the cell walls. These fluxes are assumed to be known, for the time being. This yields the discretized form:



$$\frac{\Delta U_{i,j}}{\Delta t} + \frac{\tilde{F}_{i+1/2,\,j} \;-\; \tilde{F}_{i-1/2,\,j}}{\Delta x_i} + \frac{\tilde{G}_{i,\,j+1/2} \;-\; \tilde{G}_{i,\,j-1/2}}{\Delta y_j} \;=\; 0$$

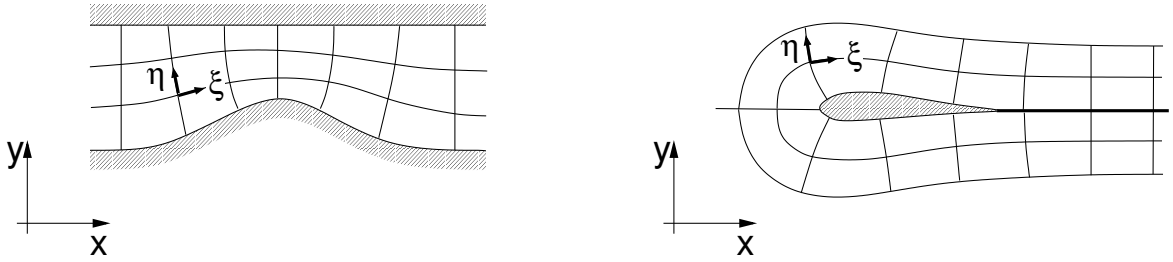
This method is often referred to as "Finite Difference Method" because of the direct difference formulation.
The same difference formulation can be obtained from the above discretized integral form by division with $\tau_{i,\,j} = \Delta x_i \cdot \Delta y_j$ Therefore, the discretized, two dimensional Euler equations have the same form, like it was derived for the one dimensional equation. In contrast to this the numerical flux functions in x-direction and y-direction, i.e. $\tilde{F}_{i\pm1/2,\,j}$ and $\tilde{G}_{i,\,j\pm1/2}$,

respectively must be defined. This is performed in analogy to the one dimensional case by using interpolation polynomials in the concerning direction.
For instance the central discretized flux with artificial damping results in:

$$\begin{aligned}\tilde{F}_{i+1/2,\,j} &= F\,(\frac{U_{i,j}+U_{i+1,j}}{2}) + d_x^{(4)}(U) - d_x^{(2)}(U)\\ \tilde{G}_{i,\,j+1/2} &= G\,(\frac{U_{i,j}+U_{i,j+1}}{2}) + d_y^{(4)}(U) - d_y^{(2)}(U)\end{aligned}$$

The damping terms $d_x(U)$ and $d_y(U)$ are replaced with second and fourth order differences in x- respectively in y-direction.

Space discretization on curved grids



Computational fluid dynamics applications most often require the calculation of a flow around curved surfaces like e.g. wing profiles, turbine blades or fuselages. When Cartesian grids are used the surface contours lie somewhere between the grid points and the discretization of the boundary conditions requires interpolation. This often leads to errors. Therefore, one usually applies body fitted grids. With this a grid line, or a grid surface in the three dimensional case, becomes identical to the body contour. Along such a grid line the boundary conditions can be formulated unambiguously . This contour fitted configuration of the grid leads to a curved grid configuration inside the domain of integration which is in general not orthogonal.

The principle method for the discretization of the conservation equations in such a body fitted grid has already been described for the conservative discretization on one and two dimensional, Cartesian grids. The grid generation, i.e. the distribution of grid points in the domain of integration is not trivial compared to Cartesian grids. This is because the increased number of degrees of freedom for the grid points which results from the general curved configuration. An essential principle in this respect is that the discretization error for the conservation equations and the boundary conditions is kept small. Some aspects for the configuration of the grid are:

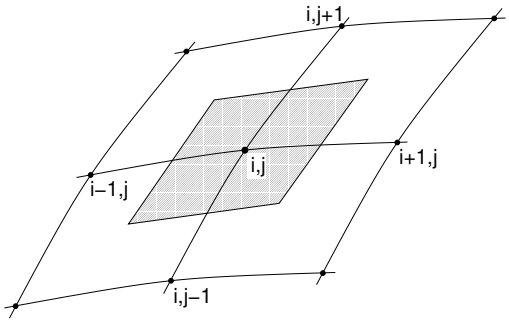
- Given boundary contours must be captured geometrically correct.
- The grid should be more dense in areas of strong changes of the unknowns, e.g. at the edge or close to a shock wave.

- The distortion of the grid cells should not be too strong, since the angle between to cell boundaries also has an impact on the discretization error. Nearly orthogonal grids are best in this respect.

- Some geometrical configurations, especially in three dimensions, can include grid singularities. A typical example for this phenomenon is the origin ($r = 0$) of a grid formulated in cylindrical coordinates. Close to these singularities irregular cells occur which lead to large discretization errors. Such cells must be discretized in a special way, or they can often be circumvented or reduced by applying a different grid configuration.

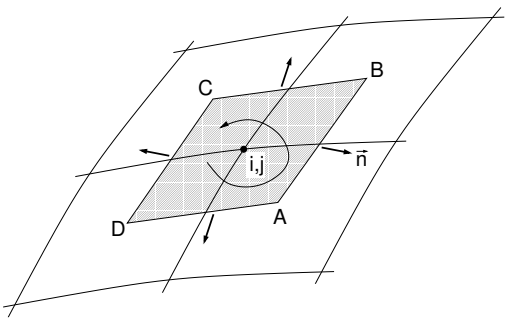
The generation of body fitted grids can be performed with various methods. In simple cases the point distribution can be described with algebraic relations, e.g. the definition of a function for a grid line between two boundaries. For more complex contours special grid generation algorithms are applied which for instance construct the point coordinates from the solution of elliptic equations. The presentation of such grid generation schemes can be found in the literature.
Thompson, J.F., Warsi, Z., Mastin, C.W.: Numerical Grid Generation, Foundations and Applications. Pub. North-Holland, (1985).
Weatherill, N.P.: Mesh Generation in Computational Fluid Dynamics. VKI Lecture Series on Comp. Fluid Dynamics, von Karman Institute for Fluid Dynamics, Rhode-Saint-Genese, March 1989, (1989).

For the discretization on curved grids a grid is assumed who's grid line directions are defined by new coordinates, e.g. in this case ξ and η and the indexing $\xi_i = i \cdot \Delta\xi$ and $\eta_j = j \cdot \Delta\eta$. The conservative discretization on such a grid requires the definition of a control volume τ around a grid point (i, j). Various configurations have already been introduced. For the following discussion the node centered configuration of the control volume is assumed. In this configuration the coordinates $x_{i,j}$, $y_{i,j}$ and the variables $U_{i,j}$ are defined on the same grid points. The control volume τ is formed in the middle between the point i, j and the neighboring points.



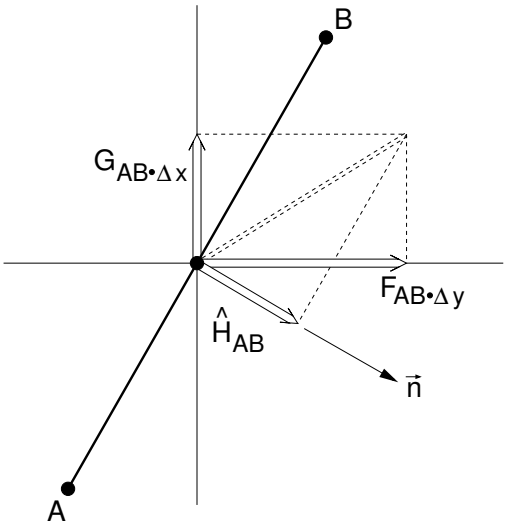
The discretization on curved grids can be performed starting from the integral form as well as starting from the divergence form. Both methods are commonly found in the literature. The correct conservative formulation of these forms lead to the same difference approximation.

The discretization with the integral form of the Euler equations offers a straightforward physical interpretation of the discretization. The conservation laws are directly applied on a small, finite volume τ, therefore this method is referred to as the "Finite-Volume Method".
Starting point are the integral conservation equations



$$\int_\tau U_t \, d\tau + \oint_A F \cdot dy - \oint_A G \cdot dx = 0$$

The control volume $\tau = \tau_{ABCD}$ is defined by the cornering points $ABCD$. For the node centered configuration the coordinates of the cell corner points are obtained as arithmetic mean values of the coordinates of the surrounding grid points, e.g.



$$x_B = (x_{i,j} + x_{i+1,j} + x_{i+1,j+1} + x_{i,j+1})/4$$

$$y_B = (y_{i,j} + y_{i+1,j} + y_{i+1,j+1} + y_{i,j+1})/4$$

The control volume (= surface in the two dimensional case) τ_{ABCD}, can be determined with the scalar product of the space vectors $\vec{r} = \binom{x}{y}$, e.g. in the form:

$$\tau_{ABCD} = \frac{1}{2}(\vec{r}_B - \vec{r}_D) \times (\vec{r}_C - \vec{r}_A) = \frac{1}{2}((x_B - x_D)(y_C - y_A) - (x_C - x_A)(y_B - y_D))$$

The positive defined normal vectors of a cell wall are facing outwards. For the estimation of the flux integral a mathematically positive turn direction is assumed.

The temporal change of the conservative properties in the volume τ_{ABCD} becomes:

$$\int_\tau U_t \, d\tau \quad \rightarrow \quad \frac{\Delta U_{i,j}}{\Delta t} \cdot \tau_{ABCD}$$

The temporal change is in equilibrium with the fluxes normal to the surface of τ. In general form the flux integral for a given mathematically positive turn direction yields:

$$\oint \vec{H} \cdot n \, dA \approx \sum_{k=1}^{4} (H \cdot \vec{n} \Delta A)_k = \hat{H}_{AB} + \hat{H}_{BC} + \hat{H}_{CD} + \hat{H}_{DA}$$

The fluxes $\hat{H}_k$ are the normal projection of the physical fluxes, multiplied with the surface element ΔA of each cell wall. They are obtained by the Cartesian components of the flux

$\vec{H} = \binom{F}{G}$ and the surface normal vector $\vec{n}\Delta A = \binom{\Delta y}{-\Delta x}$ as:

$$\hat{H}_{AB} = F_{AB}\,\Delta y_{AB} - G_{AB}\,\Delta x_{AB} \quad , \quad \hat{H}_{BC} = F_{BC}\,\Delta y_{BC} - G_{BC}\,\Delta x_{BC} \quad \text{etc.}$$

According to the turn direction it is:

$$\Delta y_{AB} \;=\; y_B \;-\; y_A \quad , \qquad \Delta x_{BC} \;=\; x_C \;-\; x_B \quad , \cdots$$

Care must be taken that the sign convention is fixed by the outward facing surface normal and the mathematically positive turn direction. The flux components, e.g. F_{AB}, G_{AB} are determined with interpolation polynomials between the neighboring points, like it has already been described.

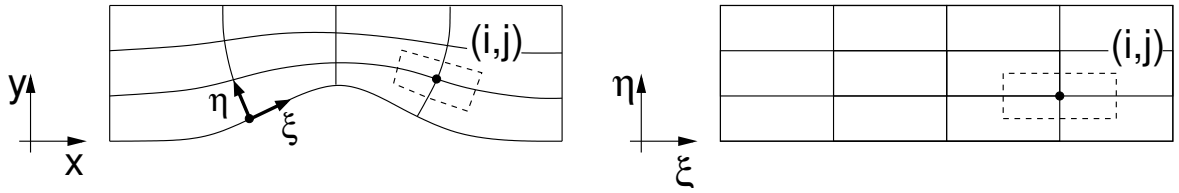
The discretized Euler equations for the point (i, j) with the control volume τ_{ABCD} thus become:

$$\boxed{\frac{\Delta U_{i,j}}{\Delta t} \cdot \tau_{ABCD} + \hat{H}_{AB} \;+\; \hat{H}_{BC} \;+\; \hat{H}_{CD} \;+\; \hat{H}_{DA} = 0}$$

The discretization of the divergence form (Finite Difference Method) first requires the transformation of the equations from Cartesian coordinates (x, y, t)

$$U_t \;+\; F_x \;+\; G_y \;=\; 0$$

into the new curved coordinate system (ξ, η, τ). When the physical plane (x, y) is displayed in the new coordinates (ξ, η), the originally curved domain of integration is given as rectangular domain with "equidistant" steps $\Delta\xi$ and $\Delta\eta$. In the transformed form the discretization can be performed for a given control volume.



For the transformation of the equations it is assumed that each point (x, y, t) in physical space is uniquely represented by a point in the transformed space (ξ, η, τ). For a grid, fixed in time this yields:

$$x = x\,(\xi, \eta) \quad , \quad y = y\,(\xi, \eta) \quad , \quad t = \tau$$

The derivatives of a function $f(x, y, t)$ in the new coordinates can be determined with the chain rule:

$$\begin{aligned} f_\xi &= f_x\,x_\xi \;+\; f_y\,y_\xi \\ f_\eta &= f_x\,x_\eta \;+\; f_y\,y_\eta \\ f_\tau &= f_t \end{aligned}$$

This yields the derivatives in Cartesian coordinates:

$$\begin{aligned} f_x &= \frac{1}{J}\,(+y_\eta\, f_\xi \,-\, y_\xi\, f_\eta) \\ f_y &= \frac{1}{J}\,(-x_\eta\, f_\xi \,+\, x_\xi\, f_\eta) \\ J &= x_\xi\, y_\eta \,-\, x_\eta\, y_\xi \end{aligned}$$

Formally, this transformation can also be performed with functional (Jacobian) determinants. In this case one obtains:

$$J = \frac{\partial(x,y)}{\partial(\xi,\eta)} = \begin{vmatrix} x_\xi & y_\xi \\ x_\eta & y_\eta \end{vmatrix} = x_\xi y_\eta \,-\, x_\eta y_\xi$$

$$f_x = \frac{1}{J}\,\frac{\partial(f,y)}{\partial(\xi,\eta)} = \frac{1}{J}\begin{vmatrix} f_\xi & y_\xi \\ f_\eta & y_\eta \end{vmatrix} = \frac{1}{J}\,(y_\eta f_\xi \,-\, y_\xi f_\eta) \qquad f_y = \frac{1}{J}\,\frac{\partial(x,f)}{\partial(\xi,\eta)} = \frac{1}{J}\,(-x_\eta f_\xi \,+\, x_\xi f_\eta)$$

If the Cartesian derivatives in the divergence form are replaced with these relations, one obtains:

$$U_t \cdot J \,+\, (y_\eta\, F_\xi \,-\, y_\xi\, F_\eta) \,+\, (-x_\eta\, G_\xi \,+\, x_\xi\, G_\eta) \,=\, 0$$

The equations in this form are not conservative. By rearrangement of the terms, like e.g.

$$(y_\eta\, F)_\xi \,-\, (y_\xi\, F)_\eta \,=\, y_\eta\, F_\xi \,-\, y_\xi\, F_\eta \,+\, F\,\underbrace{(y_{\eta\xi} \,-\, y_{\xi\eta})}_{0}$$

the divergence form can be regained in the new coordinates. Therefore, the transformed divergence form of the Euler equations becomes:

$$U_t \cdot J \,+\, (y_\eta\, F \,-\, x_\eta\, G)_\xi \,+\, (\,-\, y_\xi\, F \,+\, x_\xi\, G)_\eta \,= 0$$

The expressions in brackets are the contra variant flux components, multiplied with the surface which correspond to the fluxes normal to the cell walls. For these fluxes the abbreviations $\widehat{F}$ and $\widehat{G}$ are introduced.

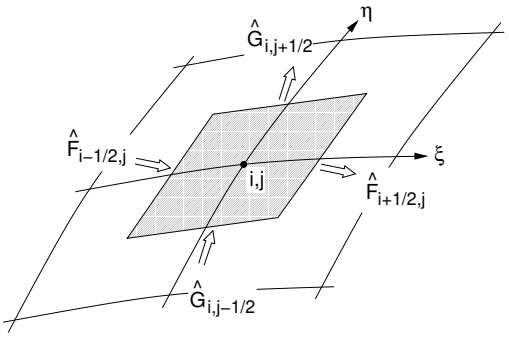
$$\boxed{U_t \cdot J \,+\, \widehat{F}_\xi \,+\, \widehat{G}_\eta \,= 0}$$

The discretization of the equations is carried out in the transformed space (ξ, η) on an equidistant grid with $\xi = i\Delta\xi$ and $\eta = j\Delta\eta$. Using the definition of the numerical fluxes $\tilde{F}_{i\pm1/2,j}$ and $\tilde{G}_{i,j\pm1/2}$ which represent an approximation of the contra variant fluxes $\widehat{F}$ and $\widehat{G}$ a formulation is obtained in analogy to the one and two dimensional, Cartesian examples.

$$\boxed{\frac{\Delta U_{i,j}}{\Delta t} \cdot J \,+\, \frac{\tilde{F}_{i+1/2,j} \,-\, \tilde{F}_{i-1/2,j}}{\Delta\xi} \,+\, \frac{\tilde{G}_{i,j+1/2} \,-\, \tilde{G}_{i,j-1/2}}{\Delta\eta} \,=\, 0}$$

The numerical fluxes of a general, two dimensional grid include the physical flux components f and G and the metric terms, like x_ξ and x_η.

$$\begin{aligned} \tilde{F}_{i\pm1/2,j} &= (+y_\eta\, F \,-\, x_\eta\, G)_{i\pm1/2,j} \\ \tilde{G}_{i,j\pm1/2} &= (-x_\xi\, F \,+\, y_\xi\, G)_{i,j\pm1/2} \end{aligned}$$



Their geometrical meaning and discretization becomes obvious when compared with the discretized integral form (Finite Volume Method). Since both cases are based on the same control volume, both discrete forms must yield the same results. The coordinates (ξ, η) serve only the assignment of the directions. For a comparison $\Delta\xi = 1$ and $\Delta\eta = 1$ are therefore arbitrarily chosen.

First, one recognizes that the Jacobian determinant J represents the control volume, i.e.

$$J_{i,j} = \tau_{i,j} = \tau_{ABCD}$$

The cell walls AB and CD or BC and DA respectively, correspond to the positions $i\pm1/2, j$, and $i, j \pm 1/2$. This yields for instance:

$$\tilde{F}_{i+1/2,j} \;=\; (+y_\eta\, F \,-\, x_\eta\, G)_{i+1/2,j} = \hat{H}_{AB} = \Delta y_{AB}\, F_{AB} - \Delta x_{AB}\, G_{AB}$$

Therefore, the metric coefficients $(x_\eta, y_\eta)_{i+1/2,j}$ describe the change of the coordinates along the cell wall AB, i.e. along η for $\xi = const$. The numerical flux $\tilde{F}_{i+1/2,j}$ is the flux multiplied with the surface normal to the cell wall AB.
In a similar way the other components can be interpreted. The sign conventions must be taken into account. For the integral form the sign is determined by the surface normal, for the divergence form it is determined by the positive ξ and η direction.
This completely defines the conservative discretization in a general, curved grid. The choice of the two starting forms, the integral or the divergence form, is most often a subjective decision. Both choices can be found in the literature. The integral formulation is often more graphic, while the divergence formulation is mathematically more formal.

For the complete space discretization the numerical fluxes $\tilde{F}_{i\pm1/2,j}$ and $\tilde{G}_{i,j\pm1/2}$ must be formulated as functions of the conservative variable U at the grid points (i, j). The formulation is performed in analogy to the one dimensional problems. The flux function is fixed for each cell wall by a one dimensional interpolation along the coordinate crosswise to the cell wall. Therefore, the various one dimensional approaches like central fluxes, Flux-Vector and Flux-Difference splitting can be transfered.

Example:
As an example a central formulation of the discrete space operator $Res(U)$ for a point (i, j)

shall be presented. The residual $Res(U)$ is defined as:

$$Res(U)_{i,j} \;=\; \frac{\tilde{F}_{i+1/2,j} \;-\; \tilde{F}_{i-1/2,j}}{\Delta\xi} \;+\; \frac{\tilde{G}_{i,j+1/2} \;-\; \tilde{G}_{i,j-1/2}}{\Delta\eta}$$

where

$$\begin{aligned} \tilde{F}_{i\pm1/2,j} &= (+y_\eta\, F \;-\; x_\eta\, G)_{i\pm1/2,j} \\ \tilde{G}_{i,j\pm1/2} &= (-x_\xi\, F \;+\; y_\xi\, G)_{i,j\pm1/2} \end{aligned}$$

The vector of the conservation variables U and the Cartesian components F and G of the flux vector are:

$$U \;=\; \begin{pmatrix} \rho \\ \rho\, u \\ \rho\, v \\ \rho\, E \end{pmatrix} \quad F \;=\; \begin{pmatrix} \rho\, u \\ \rho\, u^2 \;+\; p \\ \rho\, u\, v \\ u\,(\rho\, E \;+\; p) \end{pmatrix} \quad G \;=\; \begin{pmatrix} \rho\, v \\ \rho\, v\; u \\ \rho\, v^2 \;+\; p \\ v\,(\rho\, E \;+\; p) \end{pmatrix}$$

A central difference is obtained by algebraic averaging of the cell wall values from the neighboring grid values. It is e.g.:

$$U_{i+1/2,j} = \frac{1}{2}\,(U_{i,j} + U_{i+1,j}) \quad , \quad U_{i,j+1/2} = \frac{1}{2}\,(U_{i,j} + U_{i,j+1})$$

Replacing the variables in the fluxes by the averaged conservative properties yields the numerical fluxes:

$$\begin{aligned} \tilde{F}_{i+1/2,j} &= +y_\eta|_{i+1/2,j}\, F(U_{i+1/2,j}) \;-\; x_\eta|_{i+1/2,j}\, G(U_{i+1/2,j}) \\ \tilde{G}_{i,j+1/2} &= -x_\xi|_{i,j+1/2}\, F(U_{i,j+1/2}) \;+\; y_\xi|_{i,j+1/2}\, G(U_{i,j+1/2}) \end{aligned}$$

The metric coefficients ($\Delta\xi = 1$ and $\Delta\eta = 1$) for a node centered control volume are:

$$\begin{aligned} x_\eta|_{i+1/2,j} &= \Delta x_{AB} = \frac{1}{4}\,(x_{i,j+1} + x_{i+1,j+1} - x_{i,j-1} - x_{i+1,j-1}) \\ x_\xi|_{i,j+1/2} &= -\Delta x_{BC} = \frac{1}{4}\,(x_{i+1,j} + x_{i+1,j+1} - x_{i-1,j} - x_{i-1,j}) \end{aligned}$$

The calculation of the other flux components can be performed in an analogous way. Therefore, the residual $Res(U)$ is completely formulated. (For simplicity, the damping terms were omitted. They are defined for the single directions in analogy to the one dimensional case.)

The discrete Euler equations can therefore be combined as;

$$\frac{\Delta U_{i,j}}{\Delta t} \cdot J \;+\; Res(U)_{i,j} = 0$$

The solution of this equation system can be performed with explicit or implicit solution schemes which have been described for the scalar model equation i.e. for the one dimensional Euler equations.

LITERATURE ABOUT COMPUTATIONAL FLUID DYNAMICS

Books:

[1] C.Y.Chow: An Introduction to Computational Fluid Mechanics,
J. Wiley & Sons, New York, 1979

[2] G.Jordan-Engeln, F.Reutter: Numerische Mathematik für Ingenieure.
BI-Taschenbuch Bd. 104

[3] E.Isaacson, H.B.Keller: Analyse numerischer Verfahren,
Verlag H.Deutsch, Zürich, Frankfurt, 1973

[4] D.Marsal: Die numerische Lösung partieller Differentialgleichungen.
B I Wissenschaftsverlag, 1976

[5] R.d.Richtmeyer, K.W.Morton: Difference Methods for Initial-Value Problems,
Wiley & Sons, NY, 1967

[6] P.J.Roache: Computational Fluid Dynamics.
Hermosa publishers, Albuquerque, 1976

[7] R.Peyret, T.D.Taylor: Computational Methods for Fluid Flow.
Springer Verlag, NY, Heid., Berlin, 1983

[8] G.D.Smith: Numerische Lösung von partiellen Differentialgleichungen,
Studienausgabe, Vieweg-Verlag, Braunschweig, 1970

[9] R.Courant, D.Hilbert: Methoden der mathematischen Physik,
Teil II , Springer Verlag, 1968

[10] C. A. Fletcher: Computational Techniques for Fluid Dynamics. Vol. I: Fundamental and General Techniques, Vol. II: Specific Techniques for Different Flow Categories. Springer Verlag, 1988

[11] C. Hirsch: Numerical Computation of Internal and External Flows. Vol. I: Fundamentals of Numerical Discretization, Vol. II: wird publiziert. J. Wiley & Sons, 1988