A KINETIC SCHEME FOR THE NAVIER-STOKES EQUATIONS AND HIGH-ORDER METHODS FOR HYPERBOLIC CONSERVATION LAWS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF AERONAUTICS AND
ASTRONAUTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Georg May

September 2006

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Antony Jameson Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Brian J. Cantwell

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Juan J. Alonso

Approved for the University Committee on Graduate Studies.

Abstract

This dissertation revolves around algorithm development in the context of numerical methods for hyperbolic conservation laws and the compressible Navier-Stokes equations, with particular emphasis on unstructured-mesh algorithms. Three distinct topics may be identified.

A New Kinetic Scheme for the Compressible Navier-Stokes Equations: Kinetic numerical schemes are based on the discretization of a probability density function, whose moments define the dependent variables of the problem. In the context of fluid flow such schemes have a natural basis rooted in the kinetic theory of gases. A significant advantage of kinetic schemes is that they allow a compact, completely mesh-independent discretization of the Navier-Stokes equations, which is attractive in the context of general unstructured meshes. In this dissertation a new kinetic scheme for the compressible Navier-Stokes equations is presented, based on the popular Xu-Prendergast BGK scheme. In comparison, a dramatic reduction in computational cost is achieved, while also improving the physical modeling with respect to the underlying gas-kinetic theory, and clarifying the analysis of the multidimensional formulation. Another significant contribution is convergence acceleration for steady-state problems. For the first time a functional multigrid formulation for a BGK-based scheme on unstructured meshes is presented.

Contributions to the Development of the Spectral Difference Method: In many industrial applications numerical simulation on unstructured meshes has been dominated by schemes of relatively low order of accuracy. Often, however, the mesh requirements for low-order methods become prohibitive, while high-order methods potentially produce higher accuracy with fewer degrees of freedom. Traditionally this has been

achieved only for modeling complex physics in relatively simple domains. This motivates the formulation of high-order methods for unstructured meshes in order to make complex domains amenable. The Spectral Difference Method offers a remarkably simple alternative to such high-order schemes for unstructured meshes as the Discontinuous Galerkin and Spectral Volume Method. Significant contributions to the development of the scheme have been made as part of this dissertation, including stability analysis, formulation for the compressible Navier-Stokes equations, data limiting methodologies, and convergence acceleration using h/p-multigrid for two-dimensional unstructured meshes.

Gibbs-Complementary Reconstruction for Nonlinear Hyperbolic Equations: Nonlinear hyperbolic conservation laws admit solutions that develop discontinuities in finite time, even if the initial data is smooth. In particular for high-order numerical schemes, the treatment of discontinuities is highly problematic. For (pseudo-) spectral methods high-order filtering is often used to stabilize the solution, which ensures convergence, but not in the pointwise sense. According to the theory of Gibbs-complementary reconstruction, it is possible to extract the same order of convergence from the spectral coefficients of a non-smooth function as for a smooth one, despite the presence of the Gibbs phenomenon. In this work the concept is utilized in the context of high-order numerical methods. A stable high-order scheme may contain enough information to recover the full order of accuracy, even if it is not monotone. The information is extracted from the spectral coefficients of the solution by re-projection onto a functional space endowed with certain properties (called a Gibbs-complementary space), recovering *pointwise* high-order accuracy, even at discontinuities. This dissertation includes a proof of concept validating the technique on nonlinear hyperbolic PDE, such as the Burgers equation and the Euler equations.

Acknowledgment

First and foremost I wish to thank my academic adviser Professor Antony Jameson for inspiring and encouraging me to pursue my research interests, and for supporting me throughout my time as a graduate student. I am extremely fortunate to have access to his numerous numerical software packages, technical papers, and personal advice. This thesis would not have been possible without Antony Jameson's legacy in scientific computing.

I also wish to thank Professors Antony Jameson and Juan Alonso for providing a wealth of computational resources, and everyone at ACL/ADL for their help, Friday beers, and countless hours of company in a basement room without windows.

I have been fortunate to have received a fellowship from the Department of Aeronautics and Astronautics, and a Stanford Graduate Fellowship, which have provided me with financial independence. I was also supported by the DoE ASCI project at Stanford University and Professor Antony Jameson through a research assistantship, for which I am very grateful.

I would also like to thank Professors Juan Alonso, Sanjiva Lele, Brian Cantwell, Doron Levy, and Antony Jameson for serving on my oral defense committee and/or my dissertation reading committee. I greatly appreciate their time and effort.

Lastly, but most importantly I would like to thank my parents Georg sr. and Rosina, my sister Petra and her beautiful daughter Maren, and my girlfriend Mélli for believing and caring, and for everything else.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Conservation laws arise naturally in mathematical models of physics in which conserved quantities, such as energy, momentum, and mass are of interest. In particular hyperbolic or nearly hyperbolic conservation laws receive undiminished attention among researchers in applied mathematics. The numerical solution of nonlinear hyperbolic conservation laws is particularly challenging, because such equations admit solutions that develop discontinuities in finite time, even if the initial data is smooth. A well-known example is the formation of shock waves in gas dynamics.

Numerical simulation of physics governed by conservation laws is ubiquitous in many scientific environments and industrial applications, and is furthermore often desired for rather complex geometries, often involving different scales. This has led to increased popularity of unstructured meshes, which simplify the generation of numerical meshes considerably.

This dissertation is the result of research in numerical algorithm development, geared in particular toward improving the status quo in the realm of (hyperbolic) conservation laws with emphasis on unstructured meshes.

1.1 Motivation and Contributions

The general motivation behind this research is improving the status quo in the numerical treatment of conservation laws. Particular points of emphasis are numerical algorithms for the solution of conservation laws on arbitrary meshes (in particular the Navier-Stokes equations), hyperbolic equations and their inherent tendency to admit discontinuous solutions, and solutions of conservation laws to high order of accuracy. In detail one may identify the following three topics.

1.1.1 A New Kinetic Scheme for the Navier-Stokes Equations

Mesh-Transparent Algorithms

Meshing is one of the major bottlenecks in scientific computing. For example, in the numerical simulation of high-Reynolds-number flow in complex geometry the cost of mesh generation often outweighs the cost of computing the solution. The use of unstructured and Cartesian meshes has alleviated the problem somewhat, but has also created the situation that today many different mesh topologies are used: Body-fitted structured meshes, (hybrid) unstructured meshes, and Cartesian meshes with cut cells.

Research efforts have been undertaken to devise computational tools that treat all these mesh topologies as arbitrary polyhedral meshes [49]. This motivates the development of algorithms which do not depend on the mesh topology. For the special case of the Navier-Stokes equations, this has been achieved to some degree of success for the convective terms, but not for the viscous terms. In this work the development of mesh-independent numerical algorithms is approached via kinetic formulation of numerical schemes.

The Kinetic Approach

Kinetic schemes for the Navier-Stokes equations have advantages in the compactness of the discretization stencil, offering a stable, mesh-independent discretization, and a sound physical reasoning rooted in gas-kinetic theory. The main result reported in this dissertation is a new kinetic scheme, based on the popular Xu-Prendergast BGK scheme [56]. The new formulation dramatically reduces the incurred computational expense and also improves the physical modeling with respect to the underlying gas-kinetic theory. Furthermore, the analysis of the multidimensional formulation has been clarified. Another significant aspect is convergence acceleration for steady-state problems. For the first time a functional multigrid formulation for a kinetic scheme on unstructured meshes is presented, which makes the kinetic route more competitive for steady-state problems.

1.1.2 High-Order Methods for (Hyperbolic) Conservation Laws on Unstructured Meshes

The Current Status Quo

In most industrial applications of computational physics, simulations on unstructured meshes have been dominated by low order schemes. The simulation of compressible fluid flow for complex geometry is one notable example. In many disciplines, however, extremely high accuracy is required, for example in LES to avoid the masking of subgrid-scale dissipation by numerical dissipation, and in problems of aeroacoustics and electromagnetics, where one needs to track wave propagation over long distances. The resulting mesh requirements for low-order methods are often prohibitive, while high-order methods promise to provide higher accuracy with fewer degrees of freedom. Traditionally this has been achieved for modeling complex physics in relatively simple domains. The popularity of low order methods stems in part from the fact that in industrial applications the accuracy requirements are often not excessive ("engineering accuracy"). Furthermore, low-order schemes are very robust, while pertaining convergence acceleration techniques for steady-state problems are mature and well understood.

However, the promise of high-order methods to provide superior performance at lower computational cost is compelling and motivates the further research to extend the range of applicability to complex computational domains (which motivates the formulation for unstructured meshes).

The Development of the Spectral Difference Scheme

A significant contribution to the development of the Spectral Difference Method, which was originally proposed by Liu, Vinokur, and Wang [46], is reported in this dissertation. The Spectral Difference Method is a collocation-based high-order method designed for hyperbolic conservation laws on unstructured meshes. It is remarkably simple in formulation, compared to such high-order schemes as the Discontinuous Galerkin (DG) [9], and the Spectral Volume method [76], while sharing the property of arbitrary accuracy by means of a local, high-order reconstruction on unstructured mesh elements. Contributions reported in this thesis include linear and nonlinear stability analysis, and the extension to viscous problems, such as advection-diffusion systems and the Navier-Stokes equations. Furthermore, a flexible data limiting methodology for two-dimensional unstructured meshes is developed, as well as convergence acceleration techniques via h/p-Multigrid formulation.

1.1.3 An Alternative to Shock Capturing for High-Order Schemes

Capturing discontinuities is one of the most important design criteria for numerical schemes for nonlinear hyperbolic partial differential equations (PDE). In particular for high-order schemes, the treatment of discontinuities is highly problematic.

A viable path is the local reduction of the order of the reconstruction, sacrificing local resolution, to avoid oscillations. While this approach is very widespread and accepted, and also offers relatively simple nonlinear stability proofs, based for instance on TVD theory [26, 27], it is certainly interesting to investigate alternatives that avoid the reduction of accuracy.

Gibbs-Complementary Reconstruction

A technique to ensure a global, pointwise, high-order accuracy in the solution of hyperbolic PDE even in the presence of discontinuities, based on the concept of Gibbs-complementary reconstruction, pioneered by Gottlieb and coworkers [24], has been produced as part of this thesis. The theory of Gibbs-Complementary reconstruction has established that it is possible to extract the same order of convergence from the spectral coefficients of a non-smooth function as for a smooth one. Furthermore, following Lax [44], one can argue that as long as no explicit reduction of the order of approximation is carried out, a convergent high-order scheme still contains the necessary accuracy, even if it is not monotone. The problem is thus reformulated as a problem of information retrieval: The general idea is to not enforce monotonicity of the solution during the computation, but rather extract the necessary information from the spectral coefficients by re-projection of the solution onto a Gibbs-complementary functional space endowed with certain properties. This in fact allows the recovery of pointwise accuracy, even in the presence of discontinuities. This dissertation includes a proof of concept validating the technique on nonlinear hyperbolic PDE, such as the Burgers equation and the Euler equations.

1.2 Organization of the Thesis

In chapter 2 a short overview of the mathematical models that are used in this work is presented. This chapter is admittedly short, but it is hoped that it serves to orient the reader in the relevant framework of physical gas-dynamics, conservation laws, hyperbolic equations, and their numerical treatment. The following three chapters address the research topics identified in section 1.1 above, and can be read independently of each other. Chapter 3 introduces the new BGK-based kinetic scheme for the Navier-Stokes Equations. In chapter 4 the contributions toward the development of the Spectral Difference Scheme are reported, while the Gibbs-Complementary treatment of discontinuities in the numerical solution of hyperbolic equations is presented in chapter 5. A short summary and outlook is provided in chapter 6.

1.3 A Word on Notation

Generally the convention of summation over repeated indices applies, i.e. $a_i b_i$ means $\sum a_i b_i$. However, often the summation sign is explicitly added to clarify the notation, in particular for double summations.

Vector notation is generally omitted. It is felt that it can almost always be inferred from the context whether a variables is scalar or a vector. On rare occasions, where this is not the case, boldface, e.g. $\mathbf{u}_i$, is used to identify vectors.

Other than that, there are no special conventions used in this work, and the meaning of all variables and symbols is explained as they appear.

Chapter 2

Mathematical Models

An overview of the mathematical models that form the background of the thesis shall be provided in this section. Relevant elements include the kinetic theory of gases, which has considerable influence of the derivation of the numerical schemes discussed in chapter 3, the mathematical theory of hyperbolic conservation laws and the numerical treatment of such equations, which is highly relevant for all subsequent chapters.

Surely it is not advisable to attempt a self-contained account of the history and theory of those fields. A rather short overview is given instead, frequently preferring conciseness over completeness, while referring the interested reader to the relevant literature. It is attempted to convey a coherent line of thought that first follows the path from kinetic theory to special classes of macroscopic equations governing fluid flow in sections 2.1 and 2.2. Subsequently these equations are put in a broader context by introducing general conservation laws and hyperbolic equations in section 2.3. The numerical treatment of such equations is discussed in sections 2.4 and 2.5. Finally a full circle is completed when in section 2.6 it is demonstrated how the concepts from kinetic theory may be used to solve the macroscopic equations numerically.

2.1 Kinetic Gas Theory

The kinetic schemes discussed in chapter 3 are rather heavily influenced by the kinetic theory of gases and the physical interpretation of relevant statistical concepts. The following discussion is restricted to mono-atomic gases mostly because of notational convenience. For polyatomic gases internal degrees of freedom cause minor differences in the formalism, see for example [72].

2.1.1 The Distribution Function

In kinetic gas theory a statistical description of fluid flow is employed, based on a probability density function $f(\mathbf{x}, \mathbf{u}, t)$. Here $\mathbf{x}$ is the Cartesian coordinate-vector, and $\mathbf{u} = (u, v, w)^T$ are the molecular velocities. The distribution function gives, depending on the normalization, the specific mass of particles having velocity between $\mathbf{u}$ and $\mathbf{u} + \mathbf{du}$. Macroscopic flow variables may be obtained by taking moments of the distribution function, which thus defines the entire fluid state. The state vector of specific mass, momentum and energy of the fluid can be written

$$q(x, y, z, t) = \begin{pmatrix} \rho \\ \rho U \\ \rho V \\ \rho W \\ E \end{pmatrix} = \int \phi \, f(x, y, z, u, v, w, t) \, d\Xi \ , \tag{2.1}$$

where $d\Xi = du\,dv\,dw$, and ρ is the density of the fluid. The Cartesian components of the velocity vector $\mathbf{U}$ are denoted by U, V and W, respectively, and E is the total energy. Frequently we shall also use index notation to identify velocity components, i.e. $\mathbf{U} = U_i e_i$, where e_i with $i = 1, \ldots, 3$ is a unit vector. For mono-atomic gases the vector ϕ can be written

$$\phi = \left(\ 1, u, v, w, \tfrac{1}{2} u_i u_i \ \right)^T \ . \tag{2.2}$$

The elements of the vector are often called the *collisional invariants* because they represent molecular mass, momentum, and energy, which are conserved during molecular collisions. Under the assumption that only binary collisions in the gas occur, the mean distance between collisions is much greater that the molecular spacing, and intermolecular forces are negligible (which essentially corresponds to the usual assumptions for an ideal gas), the evolution of the distribution function is given by the Boltzmann equation

$$\frac{\partial f}{\partial t} + u_i \frac{\partial f}{\partial x_i} = J(f(u,v,w), f(u',v',w')) \ . \tag{2.3}$$

Here J is the collision integral over $du'dv'dw'$, whose functional form shall be omitted here, as it is not needed in the further discussion. Details regarding the derivation of the Boltzmann equation, as well as the expression for the collision integral can be found in [68]. A consequence of the conservation of the collisional invariants is that we have

$$\int \phi \, J \, d\Xi = 0 \ . \tag{2.4}$$

The Boltzmann equation is an integro-differential equation. Because of its high dimensionality and the complicated collision integral it is hardly tractable for most practical problems, both analytically and numerically. It is shown below that approximate solutions can be generated, and in particular well know macroscopic equations of fluid flow, such as the Euler and Navier-Stokes equations can be derived as low-order approximations. A special solution can be obtained for the case $J = 0$. It can be shown that this condition is satisfied by the Maxwellian distribution function, which may be written:

$$g = \rho \left(\frac{\lambda}{\pi} \right)^{3/2} e^{-\lambda (U_i - u_i)\cdot(U_i - u_i)} \ , \qquad \lambda = \frac{\rho}{2p} \ . \tag{2.5}$$

If the distribution function is given by the Maxwellian, $f = g$, and the macroscopic variables in Eq. (2.5) do not depend on the coordinates and time, the Boltzmann equation is satisfied trivially. In this sense Eq. (2.5) may be regarded as the distribution corresponding to thermodynamic equilibrium. However, the collision integral vanishes even if the macroscopic variables are not constant. Such *locally* Maxwellian solutions are also admitted. In this case the left-hand-side of the Boltzmann equations imposes additional constraints, which can be condensed into macroscopic transport equations (the Euler equations), which is the subject of section 2.1.2. The important notion of local thermodynamic equilibrium that is found so often in fluid mechanics and thermodynamics, comes from these locally Maxwellian solutions.

Much research has been devoted to the simplification of the Boltzmann equation over the course of the last century. Bhatnagar, Gross, and Krook replaced the collision integral with a simple relaxation term, thus arriving at the BGK equation [5]:

$$\frac{\partial f}{\partial t} + u_i \frac{\partial f}{\partial x_i} = -\frac{f - f_0}{\tau} \ . \tag{2.6}$$

In a rough rationalization the BGK equation can be thought of as modeling the distribution function under the assumption that the flow will be driven to local equilibrium on a time scale τ, the collision time. The BGK equation represents a drastic simplification both mathematically (it is a linear hyperbolic PDE and possesses a closed-form analytic solution) and physically. Nevertheless, it has some remarkable properties that are revealed in the derivation of the macroscopic equations.

2.1.2 Establishing the Macroscopic Equations

Noting Eq. (2.4), it is a matter of formal integration of the Boltzmann equation against the collisional invariants, Eq. (2.2), to obtain a set of equations for the macroscopic variables:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho U_j}{\partial x_j} = 0 \tag{2.7}$$

$$\frac{\partial \rho U_i}{\partial t} + \frac{\partial (\rho U_i U_j + P_{ij})}{\partial x_j} = 0 \tag{2.8}$$

$$\frac{\partial \rho E}{\partial t} + \frac{\partial (\rho U_j E + U_k P_{kj} + q_j)}{\partial x_j} = 0 \ . \tag{2.9}$$

However, the tensor P_{ij} and the vector q_j remain specified only by the kinetic expressions

$$\int u_i u_j f \, d\Xi = P_{ij} - p\delta_{ij} \tag{2.10}$$

$$\int \frac{1}{2} u_i u^2 f \, d\Xi = q_i + u_j P_{ij} + \rho u_i E \ , \tag{2.11}$$

which can only be resolved once the distribution function is known. A popular approach to the closure of the macroscopic equations is given by a series expansion of the distribution function along the lines of perturbation theory. The so-called Chapman-Enskog expansion is perhaps the best known such technique, and can be applied to both the Boltzmann and the BGK equation leading to the Euler, Navier-Stokes, and higher-order macroscopic equations. This section focuses on the BGK equation. The algebra is far less complex for the Chapman-Enskog expansion of the BGK equation compared to the Boltzmann equation, while the idea is precisely the same, and yields remarkably similar results. An outline of the Chapman-Enskog expansion applied to the Boltzmann equation is found in Appendix A.

Consider a characteristic molecular velocity u_{ref}, molecular time scale τ_{ref}, which may be chosen as the average time between molecular collisions, and a characteristic length of the system under consideration, L. The nondimensional BGK equation may be written as in Eq. (2.6), with τ replaced by ϵ, where

$$\epsilon = \frac{u_{\mathrm{ref}}\tau_{\mathrm{ref}}}{L} = \frac{\lambda_{\mathrm{ref}}}{L} = \mathrm{Kn}\ . \tag{2.12}$$

Here λ_{ref} is the mean free path, and Kn is the Knudsen number. The Chapman-Enskog expansion can be carried out in the Knudsen number as expansion parameter, which leads to an interpretation as a perturbation around the equilibrium for small Knudsen numbers. It is obvious then that the gas-kinetic description is superior to macroscopic low-order approximations for large Knudsen numbers. This is the case, whenever the molecular free path becomes very large, i.e. in rarefied flow, or whenever shocks are present, because in this case the local length scale of the system becomes small. Here the dimensional form of the equations is preferred, and we treat ϵ as a small parameter. One may choose $\epsilon = \tau$, where τ is the characteristic time used in Eq. (2.6).

Assuming that the distribution function depends on the coordinates and time only through the macroscopic variables, i.e. [1]

$$f = f(u, q(x,t), \nabla q(x,t), \nabla^2 q(x,t), \nabla^3 q(x,t), \dots)\ , \tag{2.13}$$

an expansion of the following form is postulated:

$$f = f_0(u,q) + \epsilon f_1(u,q,\nabla q) + \epsilon^2 f_2(u,q,\nabla q,\nabla^2 q) + \dots \quad . \tag{2.14}$$

Similarly one expands the (as yet unknown) set of macroscopic equations as

$$\frac{\partial q}{\partial t} = \Phi_0(q) + \epsilon\Phi_1(q,\nabla q) + \epsilon^2\Phi_2(q,\nabla q,\nabla^2 q) + \dots \quad . \tag{2.15}$$

Formally applying the chain rule,

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial q}\frac{\partial q}{\partial t} + \frac{\partial f}{\partial \nabla q}\frac{\partial \nabla q}{\partial t} + \dots\ , \tag{2.16}$$

---

[1] The superscripts of the nabla operator mean degrees of total differentiation, in the multi-index sense.

both Eq. (2.14) and Eq. (2.15) can be substituted into the Boltzmann or BGK equation, which leads, after matching like coefficients in ε, to a hierarchy of equations. For the BGK equation, f_0 is given by the Maxwellian distribution by definition, while the terms for $k \geq 1$ can be written

$$\sum_{l+m=k-1} \left(\frac{\partial f_l}{\partial q} \Phi_m + \frac{\partial f_l}{\partial \nabla q} \nabla \Phi_m + \frac{\partial f_l}{\partial \nabla^2 q} \nabla^2 \Phi_m + \dots \right) + u_i \frac{\partial f_{k-1}}{\partial x_i} = -f_k \ . \tag{2.17}$$

In particular, we have for $k = 1$:

$$\frac{\partial f_0}{\partial q} \Phi_0 + u_i \frac{\partial f_0}{\partial x_i} = -f_1 \ . \tag{2.18}$$

Note that by assumption

$$\frac{\partial f_l}{\partial \nabla^j q} = 0 \ , \qquad \text{for } j > l \ . \tag{2.19}$$

A straightforward computation reveals that the macroscopic variables as defined in Eq. (2.1) can be obtained from the Maxwellian

$$q(x,y,z,t) = \int \phi \, f_0(x,y,z,u,v,w,t) \, d\Xi \ . \tag{2.20}$$

Since the parameter ϵ is free this necessitates

$$\int \phi \, f_k(x,y,z,u,v,w,t) \, d\Xi = 0 \ , \qquad k > 0 \ . \tag{2.21}$$

This in turn leads to

$$\int \phi_i \frac{\partial f_0}{\partial q_j} \, d\Xi = \delta_{ij} \tag{2.22}$$

$$\int \phi_i \frac{\partial f_k}{\partial q_j} \, d\Xi = 0 \ , \qquad \int \phi_i \frac{\partial f_k}{\partial \nabla q_j} \, d\Xi = 0 \ , \qquad \dots \ , \quad k > 0 \ . \tag{2.23}$$

The constraint (2.21) means that the left-hand side in Eq. (2.17) vanishes upon integration against the collisional invariants for all k. For the first order term, Eq. (2.18), one finds immediately Φ_0:

$$\Phi_0 = -\frac{\partial}{\partial x_i} \int \phi u_i f_0 d\Xi \ . \tag{2.24}$$

Because of the conditions (2.22) and (2.23) the same condition holds for $k > 0$. Hence the macroscopic equations may be written.

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x_i} \sum_k \epsilon^k \int \phi u_i f_k \, d\Xi = 0 \ , \tag{2.25}$$

i.e.

$$\Phi_k = -\frac{\partial}{\partial x_i}\int \phi u_i f_k \, d\Xi \ . \tag{2.26}$$

The infinite sum in Eq. (2.25) may be truncated at any value of k. In particular for $k = 0$ only f_0 appears, and one may verify by direct integration that Eq. (2.25) gives the Euler equations. The term Φ_0, obtained by Eq. (2.24), may be substituted into Eq. (2.18), to obtain f_1 by direct calculation:

$$f_1 = -f_0 \left\{ c_j \left(\frac{c^2}{2RT} - \frac{5}{2} \right) \frac{\partial \ln T}{\partial x_j} + \frac{1}{RT} \left(c_i c_j - \frac{1}{3} c^2 \delta_{ij} \right) \frac{\partial U_i}{\partial x_j} \right\} \ , \tag{2.27}$$

where $c_i = u_i - U_i$, and R is the specific gas constant, appearing in the equation of state[2]

$$p = \rho RT. \tag{2.28}$$

It may be verified that Eq. (2.27) is equivalent to

$$f_1 = -\left(\frac{\partial f_0}{\partial t} + u_j \frac{\partial f_0}{\partial x_j} \right) \ , \tag{2.29}$$

if one uses the already established Euler equations to eliminate the time derivatives of the Maxwellian. The term Φ_1 can then be obtained from (2.26). Using $k = 0, 1$ in Eq. (2.25), the Navier-Stokes equations are obtained, where it is understood that the transport coefficients are not completely specified at this point, since the parameter ϵ is still free. It can be appreciated how higher order terms can be computed successively by going back and forth between Eq. (2.17) and Eq. (2.26).

For the Boltzmann equation the method is analogous, but calculating the functional form of the f_k is much more tedious, since they cannot be determined by direct calculation as in Eq. (2.17). This is so, because the collision integral necessitates the solution of an inhomogeneous integral equation in each step, see Appendix A. Remarkably, the term f_1 has the same functional form as in Eq. (2.27), although different coefficients. This fact leads to the often quoted assertion that both the BGK and the Boltzmann equation allow the derivation of the Navier-Stokes equations. This is indeed true, although one has to understand that the differences in the coefficients of the first nonequilibrium state do affect

---

[2] The assumption of an ideal gas is implicit in the kinetic model used here. The equation of state can then be easily derived under a simple kinetic model

the viscous transport coefficients. In particular, note that the ratio between the two terms in the f_1 state coming from the BGK equation, Eq. (2.27), is fixed. Upon integrating, this means that the ratio of viscous momentum transport (proportional to the velocity gradients) and heat flux (proportional to the temperature gradient), is fixed as well. Thus both can only be simultaneously adjusted, using ϵ, to recover empirical values. This issue is detailed in the next section, and needs to be addressed if a numerical scheme is based on the Chapman-Enskog expansion of the BGK equation.

2.2 The Navier-Stokes Equations

A remarkable success of kinetic theory has been that in the macroscopic limit the empirically derived Euler and Navier-Stokes equations could be recovered. The Navier-Stokes equations are based on $f = f_0 + \varepsilon f_1$, where f_1 can be either Eq. (2.27) or the corresponding distribution based on the Chapman-Enskog expansion of the Boltzmann equation, Eq. (A.5) in Appendix A. The Navier-Stokes equations have the same functional form in both cases, obtained from Eq. (2.25).

We have $P_{ij} - p\delta_{ij} = \tau_{ij}$ in Eq. (2.10). The tensor τ_{ij} can be written

$$\tau_{ij} = \mu \left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) + \lambda \delta_{ij} \frac{\partial U_k}{\partial x_k} \ . \tag{2.30}$$

If the Chapman-Enskog expansion is based on the BGK equation the viscosity coefficient μ is directly related to the expansion parameter ϵ, which was chosen equal to the collision time in section 2.1.2

$$\mu = \tau p \ . \tag{2.31}$$

The second coefficient of viscosity becomes $\lambda = -2/3\mu$ for mono-atomic gases, and $\lambda = -2/5\mu$ for diatomic gases. For the expansion based on the Boltzmann equation, free parameters in the viscosity remain, which can be resolved by applying more specific kinetic models of the molecular collision process [41]. For air one often uses the Sutherland law which corresponds to a kinetic model that assumes a weak attraction force between molecules [62], and gives good agreement with experimental data for not too high pressures and densities:

$$\frac{\mu}{\mu_0} = \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \frac{T_0 + A}{T + A} \ , \tag{2.32}$$

where the constant A is usually taken as $A = 100.3\text{K}$, and T_0 is a reference temperature corresponding to the viscosity μ_0.

For the heat flux vector q the Chapman-Enskog expansion recovers Fourier's law of heat conduction,

$$q_i = -\kappa \frac{\partial T}{\partial x_i} \ , \tag{2.33}$$

where T is the temperature of the fluid, for which $dT = c_v dE_{\text{int}}$, and κ is the thermal conductivity. E_{int} is the internal energy, $E_{\text{int}} = E - \frac{1}{2}\rho U_i U_i$. Again, the expansion based on the Boltzmann equation allows one to adjust the coefficients such that the heat conduction coefficient recovers empirical Prandtl numbers Pr, where

$$\text{Pr} = \frac{\mu c_p}{\kappa} \ . \tag{2.34}$$

The Prandtl number is a material constant for many fluids over a wide range of operating conditions. For air it takes the value of $\text{Pr} \approx 0.72$. In the Navier-Stokes equations based on the BGK equation, however, the ratio between the transport coefficients is fixed, see Eq. (2.27), and the Prandtl number is in fact always $\text{Pr} = 1$.

Finally the Navier-Stokes equations may be written in conservative form:

$$\frac{\partial q}{\partial t} + \nabla \cdot \{F_c - F_v\} = 0 \ , \tag{2.35}$$

where q is the state vector of conserved variables defined in Eq. (2.1), and $F_c = (F_{c_1}, F_{c_2}, F_{c_3})^T$ is the convective flux vector:

$$F_{c_i} = \begin{pmatrix} \rho U_i \\ \rho U_i U_1 + p\delta_{i1} \\ \rho U_i U_2 + p\delta_{i2} \\ \rho U_i U_3 + p\delta_{i3} \\ U_i(E + p) \end{pmatrix} \ . \tag{2.36}$$

The viscous flux $F_v = (F_{v_1}, F_{v_2}, F_{v_3})^T$ may be written as

$$F_{v_i} = \begin{pmatrix} 0 \\ \tau_{1i} \\ \tau_{2i} \\ \tau_{3i} \\ U_j \tau_{ij} - q_i \end{pmatrix} . \tag{2.37}$$

The assumption of a Newtonian fluid leading to the empirical derivation of the Navier-Stokes equations, with stress tensor and heat flux given by Eqns. (2.30) and (2.33), respectively, thus has a sound basis in kinetic theory and it is also clear that the Navier-Stokes equations are a low-order approximation to the more accurate kinetic models. Nevertheless, they give a good description of fluid flow for many applications. However, if one wishes to simulate flows with high degree of non-equilibrium, such as shock structure computations, higher order approximations must be used.

2.3 Conservation Laws and Hyperbolic Equations

In the macroscopic realm of continuum mechanics, both the Euler and Navier-Stokes equations can be derived as conservation laws, where different physical models are used to express the conservation of mass, momentum, and energy. One may generally define a conservation law as a balance equation for some quantity q, which is considered conserved in the sense that the integral of q over a control volume Ω changes only due to transport (or *flux*) of q through the boundaries of that domain, $\partial\Omega$. Mathematically this notion is most naturally expressed in integral form:

$$\frac{d}{dt}\int_{\Omega} q \, dV + \int_{\partial\Omega} f \cdot d\Gamma = 0, \tag{2.38}$$

where Γ is some suitable parametrization of $\partial\Omega$, and f is the *flux function.* In general, q is an m-vector, which, in a three-dimensional problem means that $f \in \mathbb{R}^{3\times m}$. Formally applying the divergence theorem for sufficiently smooth functions the generic conservation law (2.38) may be re-written

$$\frac{\partial q}{\partial t} + \nabla \cdot f = 0 \ , \tag{2.39}$$

where the integration disappears entirely, due to the fact, that the domain Ω is arbitrary, and the integrand (2.39) must vanish everywhere.

Consider the quasi-linear form of (2.39):

$$\frac{\partial q}{\partial t} + A_i \frac{\partial q}{\partial x_i} = 0 \ , \tag{2.40}$$

where $A_i = D_q f_i$ is the *flux Jacobian.* Eq. (2.39) is called hyperbolic if any real combination $\xi_i A_i$ has only real eigenvalues, and m linearly independent eigenvectors. Such hyperbolic conservation laws are associated with physical transport processes, which can be illustrated very easily in the scalar, one-dimensional case:

$$\frac{\partial q}{\partial t} + \frac{\partial f}{\partial x} = 0 \ . \tag{2.41}$$

In this case Eq. (2.40) becomes

$$\frac{\partial q}{\partial t} + c(q) \frac{\partial q}{\partial x} = 0 \ . \tag{2.42}$$

The gradient of q projected onto the direction $(t, x) = (1, c)$ vanishes, which means that q is transported in this direction. Consider the linear equation, where c is a constant. One may readily verify that Eq. (2.42) admits solutions of the form

$$q(x, t) = q_0(x - ct) \ , \tag{2.43}$$

where $q_0 = q(x, t = 0)$. The solution is thus transported along parallel straight lines $x - ct$ in the x-t plane, which are called *characteristics.*

One-dimensional systems can always be written as decoupled equations of the form (2.42), by diagonalizing the flux Jacobian via similarity transformation (the Jacobian matrix has real eigenvalues and a complete set of eigenvectors). The transformed solution variables are called Riemann Invariants. While for multidimensional systems such a decoupling is not in general possible, the inherent nature of hyperbolic systems is always wave propagation along characteristic lines.

Consider again Eq. (2.42). In the nonlinear case the characteristics are no longer parallel and may intersect. At that point the simple notion of the solution being transported along the characteristics obviously suffers a crisis, since it leads to multivalued solutions. Anticipating the formation of discontinuities in this situation, making the differential form (2.39) useless (at least locally), one resorts to the concept of weak solutions for which Eq. (2.39) is integrated against a smooth test function with compact support, and one obtains after integration by parts:

$$\int_{\mathbb{R}\times\mathbb{R}^+} \left(q\frac{\partial\phi}{\partial t} + f\frac{\partial\phi}{\partial x} \right) dxdt - \int_{\mathbb{R}} q(x,0)\phi(x,0)\, dx = 0 \ . \tag{2.44}$$

The solution is thus no longer required to be differentiable. From here on it should be understood that generally the notion of hyperbolic conservation laws admitting discontinuous solutions is meant in the sense of weak solutions. In the continuous case these are obviously identical to the conventional "strong" solutions.

While mathematically there is little guidance as to what circumstances admit discontinuous solutions, one may use physical concepts of entropy to motivate the admissibility conditions for discontinuous solutions. In particular a convex entropy function $\mathcal{U}(q)$ with $(\mathcal{U})'' > 0$ may be defined, such that

$$\mathcal{F}(q)' = \mathcal{U}(q)'f'(q) \tag{2.45}$$

holds, which actually means that Eq. (2.39) holds for smooth functions with q and f replaced by $\mathcal{U}$ and $\mathcal{F}$. Only such solutions are called admissible for which in the non-smooth case

$$\int_{\mathbb{R}\times\mathbb{R}^+} \left(\mathcal{U}\frac{\partial\phi}{\partial t} + \mathcal{F}\frac{\partial\phi}{\partial x} \right) dxdt + \int_{\mathbb{R}} \mathcal{U}(q(x,0))\phi(x,0)\, dx \geq 0 \tag{2.46}$$

holds. While the physical rationale condensed in Eq. (2.46) may not be completely obvious at this point, further discussion is deferred to the literature (see, e.g. [18] for a discussion of the mathematical notion of entropy, and the relation to the usual physical definitions). Here we are merely interested in recalling that weak solutions are not unique, and the entropy condition may be used to identify the physically meaningful solutions. The two most important aspects particular to hyperbolic conservation laws that enormously influence the design of numerical schemes are thus

- Transport processes are the nature of hyperbolic conservation laws in any dimension. Many numerical schemes aim to track the flow of information along the characteristics by adapting the discretization stencil accordingly (upwinding).

- The formation of shockwaves has to be expected for nonlinear hyperbolic conservation laws. Any numerical scheme for such equations must be equipped to deal with this situation.

2.4 Conservative Difference Schemes

The physical concept of conservation leads to the mathematical abstraction of conservation laws. In this section this concept is carried over from section 2.3 and used to introduce numerical approximations to equations of the form (2.39). Consider the one-dimensional case:

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = 0 \tag{2.47}$$

for $x \in I$. In this section boundary conditions are of no concern, so we simply assume $I = \mathbb{R}$ and initial data with compact support, or $I \subset \mathbb{R}$ with periodic boundary conditions. Consider a partition of the domain into intervals $I_j = (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$, and a sequence t^n, for $n = 0, \dots, N$. Define the numerical solution as piecewise constant on $I_j \times [t^n, t^{n+1}]$, and denote the solution for any (x, t) in $I_j \times [t^n, t^{n+1}]$ as u_j^n.

Definition 2.4.1. *Let $\Delta t^n = t^{n+1} - t^n$ and $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$. A one-step, explicit scheme for equation (2.47) is said to be conservative if it can be written*

$$u_j^{n+1} = u_j^n - \lambda \left(h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}} \right) \ , \tag{2.48}$$

where $\lambda = \Delta t^n / \Delta x_j$, and the numerical flux $h_{j+\frac{1}{2}}$ is a consistent approximation to the flux $f(x_{j+\frac{1}{2}})$ in Eq. (2.47).

Intuitively, such a formulation for the numerical scheme seems rational, due to the consistency with the conservation law in integral form (2.38). In particular for weak solutions of hyperbolic equations, the integral form is more fundamental than the differential form, as discussed in section 2.3, and it is the basis for the pertaining mathematical framework, including the formulation of shock jump conditions. Indeed, non-conservative schemes often have incorrect shock speeds (for a simple example, see [45]).

The conservation form is also at the root of some now classic results in nonlinear stability theory for conservation laws, also stemming from the mathematical theory of shock waves as formulated by Lax [43]. It is known that the total variation of the solution to Eq. (2.47), defined by

$$TV = \int_{\mathbb{R}} \left| \frac{\partial u}{\partial x} \right| \, dx \tag{2.49}$$

is nonincreasing. This has led to the theory of total-variation diminishing (TVD) theory [26, 27], which attempts to establish a discrete analog to (2.49).

Definition 2.4.2. *A numerical scheme is said to be TVD if the numerical solution satisfies* $TV^{n+1} \leq TV^n$ *for all* n*, where the total variation* TV *is defined by* $TV^n = \sum_j |u^n_{j+1} - u^n_j|$

Harten and others have given many useful results that use the concept of total variation to prove stability of difference schemes [26, 27, 53, 54]. A now classic result is given by

Theorem 2.4.1 (Harten)**.** *Let* h *be a Lipschitz-continuous numerical flux. Then scheme (2.48) can be written as*

$$u^{n+1}_j = u^n_j + C^+_{j+\frac{1}{2}}(u^n_{j+1} - u^n_j) - C^-_{j-\frac{1}{2}}(u^n_j - u^n_{j-1}) \ , \tag{2.50}$$

where the coefficients $C^{\pm}_{j+\frac{1}{2}}$ *are functions of* u*. The Scheme is TVD if for all* j

$$C^{\pm}_{j+\frac{1}{2}} \geq 0 \ , \tag{2.51}$$

$$C^+_{j+\frac{1}{2}} + C^-_{j+\frac{1}{2}} \leq 1 \ . \tag{2.52}$$

Proof. see [27] □

Remark 2.4.1. *in the semi-discrete analog of (2.48) the TVD condition is a property of the spatial operator alone. In this case* $C^{\pm}_{j+\frac{1}{2}} \geq 0$ *is the only condition to be satisfied by the spatial discretization on the right-hand side of the resulting ODE.*

Remark 2.4.2. *TVD implies convergence. For instance convergence in* L_1 *sense to a weak solution can be shown, which, if the scheme is consistent with the entropy inequality (2.46), can be strengthened to convergence to the unique entropy solution [28, 27].*

The form of the numerical flux h can be quite general. Here we consider only three-point flux functions, e.g. $h_{i+\frac{1}{2}} = h(u_{i+1}, u_i)$. Consistency requires $h(u, u) = f$, and for global conservation one must have $h(u, v) = -h(v, u)$. A particular class of flux functions are represented in so-called E-schemes, due to Osher, which allow proof of convergence to the correct entropy solution [53]. E-schemes can be established by a practical condition on the numerical flux $h_{j+\frac{1}{2}}$:

$$\operatorname{sgn}(u_{j+1}-u_j)(h_{j+\frac{1}{2}}-f(u)) \leq 0 \ , \tag{2.53}$$

for u between u_j and u_{j+1}. Of particular importance is the class of monotone (nonincreasing in its first argument and nondecreasing in its second) and Lipschitz-continuous numerical flux functions. For such flux functions, which are a subset of E-fluxes, stability estimates along the lines of theorem 2.4.1 are quite easily obtained. Example for monotone fluxes include

- Godunov Flux:
$$h(u,v) = \begin{cases} \min_{v\leq\xi\leq u} f(\xi) & v \leq u \\ \max_{v\geq\xi\geq u} f(\xi) & v > u \end{cases} \tag{2.54}$$
- Lax-Friedrich Flux:
$$h(u,v) = \frac{1}{2}\left\{f(u)+f(v)-\alpha\,(u-v)\right\} \tag{2.55}$$
where $\alpha \sim \max_\xi f'$ and the $\max_\xi$ can be taken locally, i.e. ξ between u and v, or globally over the whole solution.

Generally, the TVD condition restricts schemes to first order accuracy. However, the theory can be applied to high-resolution and even high-order schemes as well, see section 4.4.4 in this thesis and [9], with the provision that enforcing the TVD condition to guarantee stability and convergence, locally degrades accuracy to first order (at steep gradients and critical points), while high-order accuracy is retained in noncritical smooth regions.

The TVD concept does not trivially extend to multidimensional schemes. An alternative viewpoint, the concept of Local-Extremum-Diminishing (LED) has been given by Jameson [35]:

Definition 2.4.3. *A semi-discrete scheme of the form*

$$\frac{du_j}{dt} = \sum_{k\neq j} c_{jk}(u_k-u_j) \tag{2.56}$$

is called Local-Extremum-Diminishing (LED) if $c_{jk} \geq 0$ *for all* k *in the discretization stencil for node* j.

Such a scheme is stable in maximum norm, since a maximum cannot increase, and a minimum cannot decrease, which holds for multiple dimensions. In one dimension, a discretization that is LED is also TVD in the semi-discrete sense.

2.5 Finite-Volume Schemes

Finite-Volume Schemes shall be discussed here with particular emphasis on unstructured numerical meshes. Consider a conservation law, such as the Navier-Stokes equations, in a domain $\Omega \subset \mathbb{R}^d$ in integral form (2.38), along with a decomposition of the domain into N nonoverlapping control volumes, denoted V_i, for $i = 1 \ldots N$, such that $\bigcup_i V_i = \Omega$, which we call an unstructured numerical mesh. Finite-volume schemes are based on the application of the conservation law for each control volume V_i. Upon introducing the volume average to second order accuracy,

$$\overline{q}_i = \frac{1}{V_i} \int_{V_i} q \, dV \ , \tag{2.57}$$

the conservation law may thus be written

$$\overline{q}_i(t_1) - \overline{q}_i(t_0) + \frac{1}{V_i} \int_{t_0}^{t_1} \int_{\partial V_i} H \cdot \, d\Gamma_i \, dt = 0 \ , \qquad i = 1, \ldots, N \ , \tag{2.58}$$

where the key numerical step is to introduce the numerical flux function H, which approximates the true flux f in a consistent manner, along the lines discussed in section 2.4 for general conservative difference schemes. In fact, if boundary conditions are neglected, the finite-volume scheme in one dimension becomes formally identical to Eq. (2.48) in Definition 2.4.1, which means that whether such a scheme is considered a finite difference or a finite-volume scheme is entirely a matter of interpretation. In higher-dimensions, and for general meshes, however, the formulation (2.58) is not readily applied to general difference schemes (which are often not conservative).

In practice, the multidimensional numerical flux H is most often based on the projected flux $f \cdot n$, associated with the solution of the one-dimensional Riemann problem in direction n, where n is the local normal to the boundary ∂V. For this reason one frequently denotes the multdimensional numerical flux as $H = H(q^r, q^l; n)$. For example, the Lax-Friedrich flux becomes

$$H(q^r, q^l, n) = \frac{1}{2}\left\{\left(F(q^l) + F(q^r)\right)\cdot n - \alpha_n\left(q^r - q^l\right)\right\} \ , \tag{2.59}$$

where α_n is proportional to the spectral radius of the local projected flux Jacobian, for example $\alpha_n \propto |q_n| + c$ in the case of the Euler equations, where q_n is the local projected velocity $q \cdot n$, and c is the local speed of sound.

It is obvious that this form is particularly well suited for approximating conservation laws. If $H(u, u, n) = f(u) \cdot n$ and $H(u, v, n) = -H(v, u, -n)$, one has local consistency with the conservation law, and global conservation in the sense that Eq. (2.58) is satisfied for the whole domain Ω (with F replaced by some suitable H on the boundary).

2.6 Kinetic Schemes

Most applications of computational fluid dynamics are based on the solution of the Euler or Navier-Stokes equations. Both can be derived as low-order approximations to the Boltzmann or BGK equation, which govern the evolution of a single gas distribution function describing the entire fluid state, as outlined in section 2.1.

The gas-kinetic description of flow physics offers an interesting alternative in the construction of numerical schemes. If a distribution function is constructed up to the Navier-Stokes level of accuracy, which is usually accomplished on a simple two-point discretization stencil, then in a finite-volume context, the complete Navier-Stokes fluxes can be obtained by taking suitable moments (see below in this section). This means that the terms which give rise to the stress tensor (2.30) and heat flux vector (2.33) are computed as natural parts of a gas distribution function on the same stencil as convective fluxes, avoiding additional viscous discretization. This facilitates mesh-transparent formulation of numerical schemes considerably, which makes the gas-kinetic route particularly attractive for the implementation on general unstructured meshes.

Here we shall outline the fundamental concepts behind kinetic schemes. If it is assumed that the local direction x and corresponding phase-space velocity u are normal to the face (one can work with rotated coordinates at each face), one obtains for the numerical flux on the boundary of cell i, upon formally integrating Eq. (2.6) or (2.3) against ϕ and comparison with Eq. (2.58):

$$H \cdot d\Gamma = \int u\phi f(\vec{x}_l, \vec{u}, t)\, d\Xi\, d\Gamma \ . \tag{2.60}$$

Here $\vec{x}_l = (x_l, y_l, z_l)^T$ is the face coordinate. In deriving Eq. (2.60) we have made use of the fact, that the moments with ϕ of the right-hand side of both the Boltzmann and the BGK equation vanish.

The task remains to provide a valid approximation to the time-dependent flux function at the face. To illustrate a very basic approach to this problem, consider the BGK equation (2.6). A two-step approach may be used in the derivation of kinetic schemes, considering the collisionless equation as a starting point of the discretization, and subsequently modeling the collision process by enforcing the conservation constraint, i.e. [3]

$$\int \phi(f - f_0) d\Xi = 0 \ . \tag{2.61}$$

Consider the one-dimensional problem for simplicity. The collisionless kinetic equation may be used as a starting point for discretization by considering for some fixed $u \in \mathbb{R}$ and $(x,t) \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [t^n, t^{n+1}]$ the initial value problem

$$\begin{aligned} &\partial_t f + u\partial_x f = 0 \\ &f(x, t^n; u) = \psi(x; u) \ , \end{aligned} \tag{2.62}$$

where for example $\psi(x;u) = f_0(x, t^n; u) - \tau f_1(x, t^n; u)$ and $f_1 = \partial_t f_0 + u\partial_x f_0$ is the first order BGK-based Chapman-Enskog expansion of the distribution function at time $t = t^n$. If Eq. (2.62) is discretized as

$$f^{n+1} = f^n - \frac{u\Delta t}{\Delta x}\left\{ f^n_{i+\frac{1}{2}} - f^n_{i-\frac{1}{2}} \right\} \ , \tag{2.63}$$

and the numerical flux for the distribution function is computed by upwinding based on u, i.e.

$$f_{i+\frac{1}{2}} = \begin{cases} f_i & u > 0 \\ f_{i+1} & u < 0 \end{cases} \ , \tag{2.64}$$

---

[3] This does not necessarily depend on kinetic gas theory. The BGK equation used with unspecified collisional invariants may be viewed as general kinetic model equation

a kinetic scheme can readily be written by noting Eq. (2.61):

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x} \left\{ F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}} \right\} \ , \tag{2.65}$$

where $Q = 1/\Delta x_i \int q \, dx$, and $F = F^+ + F^-$, where

$$F^+_{i+\frac{1}{2}} = \int_{u>0} u\phi f_i \, d\Xi \qquad , \qquad F^-_{i+\frac{1}{2}} = \int_{u<0} u\phi f_{i+1} \, d\Xi \ . \tag{2.66}$$

More generally, one may approximate the flux using time-dependent reconstructed distribution functions to the left and right of the interface $i + \frac{1}{2}$, which may be constructed by a more sophisticated approach than the one shown here. In particular it is shown in chapter 3 how the analytical solution of the BGK equation may be used to aid in the construction of an approximation to a local time-dependent distribution function for the Navier-Stokes equations.

Fluxes computed in this manner will be valid provided that the reconstructed distribution function is consistent with the exact distribution function at least to the Navier-Stokes order of accuracy. The fundamental task of a gas-kinetic scheme is thus to construct such a distribution function, which will be outlined in chapter 3. Numerical fluxes can then be computed according to Eq. (2.60), and finally Eq. (2.58) may be integrated in time to advance the numerical solution.

Chapter 3

A New Kinetic Multigrid Method for CFD

3.1 Introduction

There is a rich body of literature on kinetic schemes for conservation laws. A brief theoretical introduction can be found in [18]. The benefit of kinetic formulation and representation of conservation laws is not limited to the construction of kinetic schemes, although this is our primary interest here. The kinetic approach may also be used to simplify the theoretical analysis of conservation laws, which, in particular for scalar conservation laws, has led to useful results [55]. An early example for a kinetic scheme for inviscid fluid flow is the equilibrium flux method by Pullin [57] that made use of the concept of flux splitting outlined in section 2.6. Deshpande proposed a similar scheme for the Euler equations, based on the collisionless Boltzmann equation [13], i.e. $J = 0$ in Eq. (2.3). This scheme has been dubbed the kinetic flux-vector splitting (KFVS) scheme, a term which has subsequently also been used by Chou and Baganoff [8] who proposed a KFVS scheme for the Navier-Stokes equations. A scheme based on the BGK equation was proposed by Prendergast and Xu [56], and was later analyzed and modified by Xu et al. [75, 73]. The BGK equation contains a simple model of the molecular collision process, and possesses an analytic solution, which can be directly used to guide the construction of numerical fluxes.

Recently Ohwada has given an analysis of the scheme [52], which established and quantified its consistency and accuracy. We refer to this scheme simply as the BGK scheme. It is the starting point for the modified scheme proposed in section 3.3.

Gas-kinetic schemes have considerable potential on general unstructured meshes, for which one aims to employ discretization techniques which do not depend on the mesh topology, i.e. are mesh transparent. For conventional schemes this is straight forward only for the convective terms of the Navier-Stokes equations, while the viscous discretization typically depends on the mesh topology and also the choice of control volume in a finite-volume scheme (cell-centered vs. cell-vertex schemes). The BGK scheme allows a stable viscous discretization on a universal next-neighbor stencil regardless of the mesh topology. The overall formulation of the scheme presented in subsequent chapters is designed for general meshes of any type, and results for different types of meshes in two and three dimensions are shown.

The only obvious disadvantage of the original BGK scheme is the fact that it is comparatively expensive in terms of floating point operations per flux computation compared to other finite-volume schemes, in particular in a multidimensional formulation [74]. The alternative scheme proposed here is specifically designed to address this issue. The proposed modifications of the scheme significantly reduce the computational cost, in particular for three-dimensional flow. Furthermore, the physical modeling is improved by the new scheme. In the original BGK scheme the constraint in Eq. (2.21) for the first non-equilibrium state f_1 in the Chapman-Enskog expansion is not satisfied, while the new scheme constructs all nonequilibrium states by explicitly enforcing this constraint.

The issue of convergence acceleration has not been given much attention in the context of kinetic schemes, which have not been widely used for steady-state problems. A multigrid formulation for the modified BGK scheme on general meshes is presented, which dramatically accelerates convergence to a steady state. Together with the proposed modifications this makes the BGK route viable for nontrivial three-dimensional computations.

This chapter is organized as follows. In section 3.2 we briefly recapitulate the derivation of the original BGK scheme, before we present and motivate the new scheme in section 3.3. Next, the multidimensional formulation for kinetic schemes is analyzed in section 3.4, in particular with regard to the viscous terms. After discussing algorithmic details, such as the multigrid methodology, in section 3.5, we present a variety of numerical results for both structured and unstructured meshes in section 3.6.

3.2 The BGK Finite Volume Method

As outlined in section 2.6, the main task of a kinetic finite volume scheme is to generate an approximation to a gas distribution function at cell interfaces to evaluate the numerical flux. The purpose of this section is to show how the BGK scheme constructs such a distribution function using the analytical solution of the BGK equation.

We initially restrict the analysis to one dimension to demonstrate the essential steps. Subsequently the multidimensional extension is considered.

In order to approximate the Euler or Navier-Stokes equations via a gas-kinetic formulation, one effectively approximates a modified gas-kinetic equation along with initial conditions, obtained by means of Chapman-Enskog expansion, based on the Boltzmann or BGK equation, see section 2.1.2 and [51, 52]. If the starting point is the BGK equation, the modified kinetic equation can be approximated by

$$\partial_t f + u\partial_x f = -f_1/\tau \ , \tag{3.1}$$

with the initial condition given by the first order Chapman-Enskog distribution

$$f(x,u,0) = f_0(x,u,0) + \tau f_1(x,u,0) \ . \tag{3.2}$$

The state f_0 is given by the Maxwellian distribution, see Eq. (2.5), while the first nonequilibrium state is given in case of the BGK equation by Eq. (2.27) or Eq. (2.29). Expanding the analytical solution of this equation for small Δt, one can write to the order of $\tau\Delta t^2$ in truncation error:

$$f(x,u,t) = f_0(x,u,0) - ut\partial_x f_0(x,u,0) + \tau f_1(x,u,0) - tf_1(x,u,0) \ . \tag{3.3}$$

The truncation error in this simplification is the same as the one incurred in the approximated modified kinetic equation (3.1) with respect to the Chapman-Enskog expansion corresponding to the Navier-Stokes level. See [51, 52] for details. As shown in subsequent sections, the BGK scheme approximates this distribution function at cell interfaces, given the macroscopic variables to the left and right of the face. In the following we assume without loss of generality that any such interface is located at $x_0 = 0$. At the face, a logical left cell is associated with $x < 0$ and a logical right cell with $x > 0$. We assume that cell-averaged states Q^1 and Q^2 and piecewise constant gradients ∂_x^l, ∂_x^r are available in both cells. One can then construct the left and right states Q^l, Q^r at the face as

$$Q^l = Q^1 + (x_0 - x_1)\partial_x^l\, q \qquad , \qquad Q^r = Q^2 + (x_0 - x_2)\partial_x^r\, q\ . \tag{3.4}$$

In practice the deviations from the cell-averaged state will have to be limited by some appropriate method in order to preserve monotonicity. In this work Venkatakrishnan's limiter [67] has been used for unstructured meshes.

The Xu-Prendergast BGK scheme can be obtained by first solving the BGK equation (2.6) analytically:

$$f(x,u,t) = \frac{1}{\tau}\int_0^t f_0(x',u,s)e^{-\frac{-(t-s)}{\tau}}ds + \ e^{-\frac{t}{\tau}} f(x-ut,u,0)\ , \tag{3.5}$$

where $x' = x - u(t-s)$. Eq. (3.5) represents a nonlinear average between the initial state $f(x,u,0)$ and an equilibrium state $f_0(x,u,t)$. The weight on the equilibrium state increases with t, which reinforces the interpretation of the BGK equation as modeling the relaxation to a local equilibrium state on a time scale τ. This analytical solution can be used to construct a distribution function at a cell interface using the reconstructed left and right states.

3.2.1 Approximation of the Initial State

The initial state, i.e. the second term in Eq. (3.5), must be approximated in a way consistent with the Navier-Stokes equations, based on the theory of Chapman-Enskog expansion. In the Xu-Prendergast BGK scheme the KFVS discretization is chosen, which uses the information from either side of the face by introducing a kinetic splitting. As $t > 0$, the characteristic line $x_0 - ut$ falls to the left of the interface for $u > 0$, and to its right for $u < 0$. Recalling that $x_0 = 0$, the distribution function can thus be approximated along the characteristic ut as

$$f(-ut,u,0) = \begin{cases} f_0^l - ut\partial_x f_0^l + \tau f_1^l \ , & u > 0 \\ f_0^r - ut\partial_x f_0^r + \tau f_1^r \ , & u < 0 \ . \end{cases} \tag{3.6}$$

The splitting of the phase space is introduced so as to respect the gas-kinetic characteristic lines during reconstruction. It can thus be viewed as gas-kinetic upwinding. The terms $f_1^{l/r}$ are first order terms in the Chapman-Enskog expansion of the BGK equation, see Eq. (2.27). The alternative notation, Eq. (2.29), can be used to efficiently compute f_1 states. As noted in section 2.1.2, higher order terms f_k, where $k > 0$, must vanish upon integration over the velocity space against the collisional invariants. Hence, after direct differentiation to compute the space derivative in Eq. (2.29), i.e. differentiating Eq. (2.5) and using the chain rule, according to

$$\partial_x f_0 = \partial_q f_0 \, \partial_x q \ , \tag{3.7}$$

one may enforce this constraint directly by computing

$$\int \phi \partial_t f_0^{l/r} d\Xi = - \int \phi u \partial_x f_0^{l/r} \, d\Xi \ , \tag{3.8}$$

and adding the spatial and temporal gradients to give the f_1 state. A more detailed description of the computation of the derivatives is given in Appendix B. This procedure is equivalent to eliminating the time derivatives using the Euler equations. Note that for the initial state, *both* the state and the derivatives are discontinuous.

We note here that in this implementation the spatial gradients are modified slightly to effectively make the f_1 state identical to the one corresponding to the Chapman-Enskog expansion of the Boltzmann equation, and ensure that both the coefficient of viscosity and the Prandtl number have the correct value [1]. Explicit formulas for this procedure can be found in Appendix B, see also the discussion in section 2.1.2, Appendix A and [62]. In contrast to our approach, the original BGK scheme uses an “a posteriori” correction of the Prandtl number [73], which shall not be detailed here. This minor point is the only difference in our implementation of the original BGK scheme and the standard version presented in [73].

---

[1] recall that the Chapman-Enskog expansion of the BGK equation fixes the Prandtl number at unity

For convenience we introduce the notation

$$\begin{aligned}
\widehat{f_0}(u) &= f_0^l H(u) + f_0^r\left(1 - H(u)\right) \ , && (3.9)\\
\widehat{\partial_x f_0}(u) &= \partial_x f_0^l H(u) + \partial_x f_0^r\left(1 - H(u)\right) \ , && (3.10)\\
\widehat{f_1}(u) &= f_1^l H(u) + f_1^r\left(1 - H(u)\right) \ . && (3.11)
\end{aligned}$$

3.2.2 Approximation of the Equilibrium State

To approximate the equilibrium state under the integral sign in Eq. (3.5), a relaxation state $\bar{Q}$ is postulated with a corresponding distribution $\bar{f}_0$ at the face:

$$\bar{f}_0(-ut,u,t) = \bar{f}_0(0,u,0) - ut\partial_x^l \bar{f}_0 H(u) - ut\partial_x^r \bar{f}_0(1-H(u)) + t\widetilde{\partial_t f_0} \ , \tag{3.12}$$

where H(x) is the Heaviside function with $H(x) = 1$ for $x \geq 0$ and $H(x) = 0$ otherwise.

The equilibrium distribution function at the face and $t = 0$ is obtained by imposing the constraint

$$\int \phi \bar{f}_0 \, d\Xi = \int_{u>0} \phi f_0^l \, d\Xi + \int_{u<0} \phi f_0^r \, d\Xi \ , \tag{3.13}$$

where f_0^l and f_0^r represent Maxwellians evaluated with the left and right states, Q^l and Q^r. For the macroscopic state $\bar{Q}$ corresponding to the relaxation function $\bar{f}_0$ no gradients are available from the initial reconstruction. One may use the approximation

$$\begin{array}{c} \partial_x^l \bar{q} = \frac{\bar{Q}-Q^1}{||x_0-x_1||} \\ \partial_x^r \bar{q} = \frac{Q^2-\bar{Q}}{||x_0-x_2||} \end{array} \ . \tag{3.14}$$

Notice that the gradients have thus changed from the initial reconstruction, reflecting the modeling of the collision process leading to the relaxation state $\bar{Q}$. Note also that while the new gradients are different in the two cells adjacent to the face, there is only one relaxation state, and one corresponding distribution function f_0, Eq. (3.12). On the other hand, in the initial state, *both* the state variables (and corresponding distribution functions) and the gradients are discontinuous. A schematic depiction of this situation is shown in Fig. 3.1, which serves to illustrate the postulated relaxation process from a discontinuous to a continuous state.

For notational convenience we denote the spatial gradients corresponding to the new relaxation state by

$$\overline{\partial_x f_0}(u) = \partial_x^l \bar{f}_0 H(u) + \partial_x^r \bar{f}_0(1 - H(u)) \ . \tag{3.15}$$

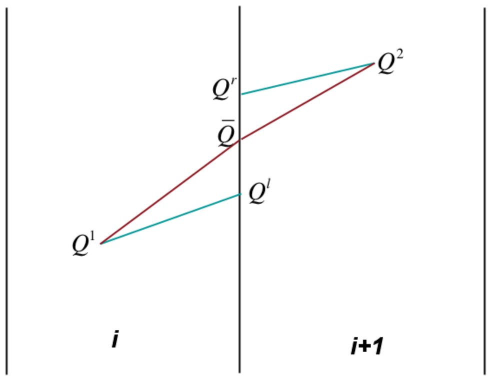


Figure 3.1: Macroscopic variables and gradients based on the initial flow variables, and the postulated relaxation state $\bar{Q}$.

We discuss the time derivative of $\bar{f}_0$ appearing in Eq. (3.12) in section 3.2.3. Given Eq. (3.12), the integral in Eq. (3.5) can be evaluated analytically.

3.2.3 The Time Derivative of the Relaxation State

The time derivative $\widetilde{\partial_t f_0}$ in Eq. (3.12) is finally computed using the following constraint:

$$\int_0^{\Delta t} \int \phi (f(0,u,t) - \bar{f}_0(0,u,t))\, d\Xi\, dt = 0 \ . \tag{3.16}$$

This constraint seeks to mimic condition (2.21), i.e. $\int \phi\, f_k\, d\Xi = 0$ for $k > 0$, in a time-averaged fashion. Eq. (3.16) can be solved to produce

$$\int \phi\, \widetilde{\partial_t f_0}\, d\Xi = \int \phi \left(\alpha_1 u\, \widehat{\partial_x f_0} + \alpha_2 u\, \overline{\partial_x f_0} + \alpha_3 \widehat{\partial_t f_1} \right) d\Xi \ , \tag{3.17}$$

where $\int \phi \left(\bar{f}_0 - \widehat{f}_0 \right) d\Xi = 0$, by construction, see Eq. (3.13), has been used. The coefficients α_i can be obtained by straight forward time integration, see [73].

Although the abbreviated nomenclature used here perhaps does not make it obvious, the computation of this term is extremely expensive. In the original (one-dimensional) BGK scheme it accounts for 40% of the moments that have to be computed. It is surmised here that the term $\widetilde{\partial_t f_0}$ has been formulated in this complicated manner in the original BGK scheme for lack of a more obvious way of computing it. Furthermore, it must be understood that Eq. (3.16) is merely a rough approximation to the constraint (2.21). This point is addressed in section 3.3, where it will become clear that the term can be computed in a more straightforward way, which is not only much cheaper, but also uses the constraint (2.21) directly.

3.2.4 The Distribution Function and Some Comments

Integrating the relaxation state in time, according to Eq. (3.5), we can finally write the distribution function as

$$\begin{aligned} f(0,u,t) = \left(1 - e^{-\frac{t}{\tau}}\right) \{\bar{f}_0 - ut\,\overline{\partial_x f_0}\} + e^{-\frac{t}{\tau}} \left\{\widehat{f}_0 - ut\,\widehat{\partial_x f_0}\right\} - \\ \tau \left(1 - e^{-\frac{t}{\tau}}\right) \left\{\widetilde{\partial_t f_0} + u\,\overline{\partial_x f_0}\right\} + \tau e^{-\frac{t}{\tau}} \widehat{f}_1 + \\ t \left\{\widetilde{\partial_t f_0} + u\overline{\partial_x f_0}\right\} \ . \end{aligned} \tag{3.18}$$

One can easily identify the corresponding terms in Eq. (3.3). The first two terms approximate the equilibrium distribution along the characteristic line ut using a nonlinear average between the continuous and discontinuous reconstruction techniques. The second and third term effectively model the nonequilibrium distribution f_1, which gives rise to the viscous fluxes, using a similar average. Finally the last term approximates the last term in Eq. (3.3). Finally, the fluxes can be computed using the one-dimensional analog of Eq. (2.60), i.e. $F = \int u\phi f(0,u,t)\, d\Xi$.

Consider the third and fourth term in Eq. (3.18). These two terms form a consistent approximation of the first nonequilibrium Chapman-Enskog state, based on the BGK equation. We thus define the viscous fluxes of the BGK scheme as

$$F_{BGK,v} = \int u\phi \left\{\tau \left(1 - e^{-\frac{t}{\tau}}\right) f_1^a + \tau e^{-\frac{t}{\tau}} f_1^b\right\} d\Xi \ . \tag{3.19}$$

where

$$f_1^a = -\left\{\widetilde{\partial_t f_0} + u\,\overline{\partial_x f_0}\right\} \ , \tag{3.20}$$

$$f_1^b = \widehat{f}_1 \ . \tag{3.21}$$

For moderate Reynolds numbers, i.e. $\tau \ll \Delta t$, it is actually the term (3.20) which dominates, while the term (3.21), which comes from the approximation of the initial f_1 state, Eq. (3.6), is negligible. The dominating term (3.20), however, was actually *not* derived to be an approximation to an f_1 state. The terms appearing there come together from different parts of the approximation to the solution of the BGK equation. It is believed here that failure to realize that these terms effectively form the viscous fluxes is a deficiency in the derivation of the original BGK scheme. It leads to the complicated way of computing $\widetilde{\partial_t f_0}$ in Eq. (3.17), because it was not realized that $\int \phi f_1^a \, d\Xi = 0$ can and should be used directly to compute $\widetilde{\partial_t f_0}$, as has been done for the state $\widehat{f_1}$, see Eq. (3.8). As a consequence, this important constraint is *not* satisfied for the term f_1^a.

3.3 The new BGKgg Scheme

Two modifications to the classical BGK scheme are proposed here, which address the construction of the nonequilibrium terms, Eqns. (3.20) and (3.21), and in particular the computation of the time derivative of the relaxation state, i.e. the term $\widetilde{\partial_t f_0}$ discussed in section 3.2.3.

In the BGK scheme the discretization of the initial state in Eq. (3.5) is identical to the splitting used in the viscous KFVS scheme proposed by Chou and Baganoff [8], which is based on the collisionless Boltzmann equation $\partial_t f + u \partial_x f = 0$. This discretization contributes the term (3.21) to the viscous fluxes. It has often been found that the KFVS scheme does not give very good results for viscous flow [73]. The BGK scheme's superior modeling of viscous flow ultimately comes from the right-hand side of the BGK equation, leading to the relaxation state, which is contributes Eq. (3.20) to the viscous terms, which is blended with the KFVS approximation in Eq. (3.19). In fact, the KFVS term is blended out for most flow situations, as discussed in section 3.2.4.

Hence, one may ask if this term is needed. Apparently, the success of the BGK scheme is due to the fact that its presence does not contribute significantly to the flux, and it has often been found to be a poor choice in viscous discretization when appearing alone (see Xu's analysis [73] and Ohwada's results in [52]). This, as well as cost considerations, which will be addressed below, provide sufficient motivation to rethink the entire computation of the nonequilibrium terms. We start by writing instead of Eq. (3.21):

$$f_1^b = \overline{f_1} = \overline{\partial_t f_0} + u\,\overline{\partial_x f_0}\ . \tag{3.22}$$

The KFVS discretization is thus replaced with the averaging $\overline{\partial_x f_0}$, and the state $\overline{f_1}$ is obtained by enforcing $\int \phi \overline{f}_1\, d\Xi = 0$:

$$\int \phi\, \overline{\partial_t f_0}\, d\Xi = - \int \phi\, u \overline{\partial_x f_0}\, d\Xi\ , \tag{3.23}$$

where it is understood that the right-hand side has to be modified in the same way as outlined in section 3.2.1 and Appendix B to correct values of the transport coefficients. The viscous terms of the initial state are thus more akin to Eq. (3.20), which dominates except when τ is of the order of t or larger. This is the case only for very low Reynolds numbers or whenever τ is artificially increased for the purpose of adding dissipation, e.g. for shock capturing as discussed in section 3.5.3 below. Validation of the new formulation should reflect this by including low Reynolds number computations and test cases with shocks.

In the modified formulation the nonequilibrium terms now cancel by construction when computing $\widetilde{\partial_t f_0}$ according to Eq. (3.16). One can write

$$\int \phi\, \widetilde{\partial_t f_0}\, d\Xi = \int \phi u \left(\alpha_1 \overline{\partial_x f_0} + \alpha_2 \widehat{\partial_x f_0} \right)\, d\Xi\ . \tag{3.24}$$

Furthermore, the first term on the right-hand side has already been computed for Eq. (3.23), so that only the two moments resulting from the averaging $\widehat{\partial_x f_0}$ need to be computed. At this point one can write the distribution function as

$$\begin{aligned} f(x,u,t) = \left(1 - e^{-\frac{t}{\tau}}\right) \left\{ \bar{f}_0 - ut\, \overline{\partial_x f_0} \right\} + e^{-\frac{t}{\tau}} \left\{ \hat{f}_0 - ut\, \widehat{\partial_x f_0} \right\} - \\ \tau \left(1 - e^{-\frac{t}{\tau}}\right) \left\{ \widetilde{\partial_t f_0} + u\, \overline{\partial_x f_0} \right\} + \tau e^{-\frac{t}{\tau}} \overline{f_1} + \\ t \left\{ \widetilde{\partial_t f_0} + u \overline{\partial_x f_0} \right\}\ . \end{aligned} \tag{3.25}$$

This is only an intermediate step, and Eq. (3.25) is not recommended for general use. However, we include this function for validation purposes in order to isolate the effects of the proposed modifications. We shall refer to the scheme based on this distribution function as the BGKg scheme.

Consider now Eq. (3.20). Any f_1 state should satisfy Eq. (2.21), i.e. $\int \phi f_1 \, d\Xi = 0$. It has been pointed out in section 3.2.4 that this is not the case for the viscous terms in the original BGK scheme. Hence, it is not only computationally less expensive, but also appears physically reasonable to replace $\widetilde{\partial_t f_0}$ by $\overline{\partial_t f_0}$, i.e. using Eq. (3.23). We can thus further rewrite the distribution function as

$$f(x,u,t) = \left(1 - e^{-\frac{t}{\tau}}\right) \left\{\bar{f}_0 - ut\,\overline{\partial_x f_0}\right\} + e^{-\frac{t}{\tau}} \left\{\widehat{f}_0 - ut\,\widehat{\partial_x f_0}\right\} + \tau \overline{f_1} - t\overline{f_1} \ . \qquad (3.26)$$

The scheme based on this formulation will be denoted BGKgg, to reflect the fact that the overbar averaging, Eq. (3.23), is now used for both viscous terms, which collapse into one. The BGKgg scheme is the new method proposed in this thesis.

The constraint $\int \phi f_1 \, d\Xi = 0$ means that the *current* time-dependent equilibrium state f_0 is the only part which survives an integration against ϕ. In the new formulation, it has been ensured that all approximated f_1 states vanish individually at all times. The only terms which survive integration against ϕ are the first two terms in Eq. (3.26), which are indeed the complete approximation to the most current f_0 state. In the original formulation, only the relaxation state $\bar{f}_0$ has been considered the "current" f_0 state (omitting the equilibrium portion of the initial state), and has been used in Eq. (3.16) to mimic Eq. (2.21).

In order to compute the macroscopic fluxes, moments of the distributions function, i.e. integrals over the phase space, have to be computed, see Eq. (2.60). Section 3.5.1 below gives a more detailed overview of the practical implementation. To appreciate the reduction in computational cost, consider Table 3.1. Most of the computational effort lies in computing the moments of terms involving spatial or temporal expansions, see Appendix B. We have included here the cost of a full three-dimensional reconstruction for the original BGK scheme, according to [74]. We discuss the multidimensional extension in section 3.4 below, and point out how the additional 3D overhead can be avoided. The cost of the new BGKgg scheme thus remains the same as in Table 3.1. The simplified formulation reduces the computational effort significantly, in particular in comparison with the multidimensional version of the scheme.

Moments for...	BGK3D [74]	BGK [73] Eq. (3.18)	BGKg Eq. (3.25)	BGKgg Eq. (3.26)
$\widetilde{\partial_t f_0}$	12	6	2	-
Noneq. Terms (initial state)	6	2	2	2
Fluxes	13	7	6	5
Additional Terms for 3d	4	-	-	-
Total	35	15	10	7

Table 3.1: The number of moments coming from terms involving spatial and temporal expansions, which have to be taken for one flux computation for the standard BGK scheme and the two modified versions, BGKg and BGKgg.

3.4 Multidimensional Scheme and Viscous Terms

The schemes outlined above are applicable to higher dimensions, using Eq. (2.60) to compute normal fluxes for a finite-volume scheme, provided that at each face the local coordinate x is aligned with the face normal (which suggests the use of a locally rotated coordinate system). However, a slight error in viscous terms is incurred. While the heat flux vector can be accurately represented by normal gradients alone, for the viscous terms in the momentum equation both normal and tangential derivatives are present even in the normal component of the viscous flux. One might consider a more elaborate multidimensional formulation, such as the one proposed recently by Xu et al. [74]. For simplicity we consider only the two-dimensional case. The extension to three dimensions is straight forward. In rotated coordinates with normal and tangential directions x and y, and corresponding macroscopic velocities U and V, the stress tensor is given by

$$\sigma = \begin{pmatrix} 2\mu\partial_x U + \lambda(\partial_x U + \partial_y V) & \mu(\partial_x V + \partial_y U) \\ \mu(\partial_x V + \partial_y U) & 2\mu\partial_y V + \lambda(\partial_x U + \partial_y V) \end{pmatrix} . \tag{3.27}$$

In the rotated frame the face normal is simply $\underline{n} = (1, 0)^T$. The viscous fluxes thus depend only on the components in the first column of the stress tensor:

$$F_{\text{vis,mom}} = \sigma \cdot \underline{n} = \begin{pmatrix} 2\mu\partial_x U + \lambda(\partial_x U + \partial_y V) \\ \mu(\partial_x V + \partial_y U) \end{pmatrix} . \tag{3.28}$$

We point out here that the second coefficient of viscosity is fixed at $\lambda = -2/5\mu$ by the Chapman-Enskog expansion of the BGK equation for diatomic molecules. This deviates from the value $\lambda = -2/3\mu$, which is usually used, but is itself only an approximation for monoatomic gases. In practice this difference is of minor importance. The corresponding first nonequilibrium term in the Chapman-Enskog expansion of the BGK equation is given by

$$\tau f_1 = -\tau \left\{ \partial_t f_0 + u\partial_x f_0 + v\partial_y f_0 \right\} . \tag{3.29}$$

In the continuous limit, the gas-kinetic formulation is entirely equivalent to the macroscopic one, so that

$$\tau \int u \begin{pmatrix} u \\ v \end{pmatrix} f_1 \, d\Xi = -\sigma \cdot \underline{n} \tag{3.30}$$

holds exactly. If the tangential derivative $v\partial_y f_0$ is omitted we have

$$\tau \int u \begin{pmatrix} u \\ v \end{pmatrix} \left\{ \partial_t f_0 + u\partial_x f_0 \right\} \, d\Xi = \begin{pmatrix} 2\mu\partial_x U + \lambda\partial_x U \\ \mu\partial_x V \end{pmatrix} . \tag{3.31}$$

By comparing to Eq. (3.28), it is obvious that for the quasi-1D expansion even in the continuous limit the full stress tensor is not recovered, and the viscous flux at a given face depends on the local rotation. It is important to note that this holds for any type of mesh. In [74] it was stressed that a multidimensional reconstruction has to be considered for unstructured meshes. However, even for Cartesian meshes a quasi-1D reconstruction does not produce the exact stress tensor in the continuous limit. It should be pointed out, that unlike conventional schemes, where the difference between Eq. (3.31) and (3.28) is computationally insignificant, the BGK model becomes almost prohibitively expensive, certainly in the three-dimensional case, if the gas-kinetic reconstruction is carried out in all dimensions, which leads to the number of moments shown in Tab. 3.1.

For inviscid flow, a quasi-1D formulation should certainly suffice, since the difficulties involve only the viscous terms. For viscous flow in higher dimensions the difference between the multidimensional reconstruction and the quasi-1D reconstruction is small for a wide class of problems, even on highly stretched unstructured meshes. Accordingly the use of a three-dimensional gas-kinetic reconstruction seems excessive. We propose instead to augment the normal reconstruction by a simple central average of the tangential components of the stress tensor, i.e.

$$F_{\mathrm{mom}} = F_{\mathrm{mom}}^{BGK} + \begin{pmatrix} \lambda \partial_y V \\ \mu \partial_y U \end{pmatrix} . \tag{3.32}$$

The necessary gradients are readily available from the reconstruction, and the stencil for the viscous discretization remains unchanged.

A similar rationale has been used for conventional viscous discretization as well. It is well known that a simple central averaging of the gradients for evaluation of the stress tensor leads to odd-even decoupling, and possibly instability. For cell-vertex schemes a discretization technique, which replaces the normal component of the central average with a directional difference approximation, has been proposed and used successfully [29]. This correction provides an effective re-coupling of the discretization stencil, which is given here by the gas-kinetic normal expansion, while tangential components are simply averaged. Very few additional floating point operations are needed to add those components.

The only remaining matter of concern for the multidimensional formulation is the fact that Eq. (3.14) is no longer valid, as it assumes that the cell-centroid lies in a direction normal to the face, which is not the case for general meshes. We use a correction involving a simple projection, which has been proposed and validated in [62].

3.5 Algorithms

3.5.1 Practical Implementation

Using the method of lines to separate the spatial discretization from the time integration, the generic finite-volume scheme of Eq. (2.58) may be considered as a starting point for the practical implementation of the finite-volume BGK scheme. Fig. 3.2 shows a flowchart of the solution algorithm. Firstly, given the initial conditions or a previous iterate, the flow gradients are computed, as they are needed in the evaluation of the Chapman-Enskog state, which depends on the gradients of the Maxwellian, which in turn depends on the gradients of the flow solution (see also Appendix B).

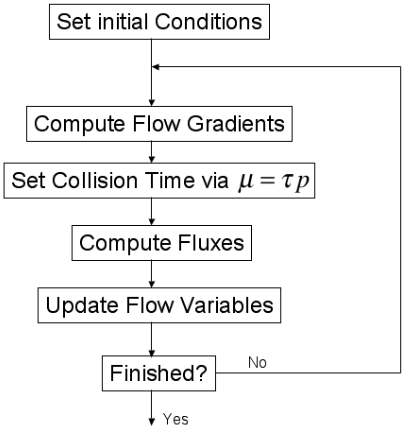


Figure 3.2: Algorithmic Implementation of BGK-type schemes.

The collision time τ appearing in the distribution function is a function of the viscosity, according to Eq. (2.31). The viscosity itself can be set to a constant value, corresponding to the fluid under consideration, or can be computed using the current solution and an appropriate model, such as Sutherland's law, Eq. (2.32), which makes the viscosity a function of the temperature.

Compared to a conventional finite-volume scheme, the main difference lies in the flux evaluation, which is given for kinetic schemes by Eq. (2.60) in generic terms as a moment of the distribution function, i.e. an integration over phase space. In sections 3.2 and 3.3 it has been shown how to discretize the distribution function f as a function of only a Maxwellian distribution and its derivatives, corresponding to the first order Chapman-Enskog expansion. Either Eq. (3.18), which corresponds to the Xu-Prendergast BGK scheme, Eq. (3.25), the BGKg scheme, or (3.26), the BGKgg scheme, may be used. Only moments of the distribution function are needed, all of which can be expressed as

$$\int_I u^l v^m w^n f_0 \, d\Xi \tag{3.33}$$

for some nonnegative integer numbers l, m, n, and can thus be computed by exact integration. The integration is carried out over $I = \widetilde{I} \times \mathbb{R} \times \mathbb{R}$. Depending on which integral is computed, we have either $\widetilde{I} = \mathbb{R}$, $\widetilde{I} = \mathbb{R}^+$, or $\widetilde{I} = \mathbb{R}^-$, where limits corresponding to the positive and negative real axis result from the kinetic upwinding on the normal phase-space velocity outlined in section 3.2. The integrals can be computed using an efficient recursion formula [73], and their contributions are simply added to evaluate Eq. (2.60).

3.5.2 Multigrid

For large-scale computations of steady flow multigrid techniques are almost indispensable. Following Jameson [32], a nonlinear multigrid method for nonlinear equations can be outlined as follows. Consider the general nonlinear homogeneous PDE

$$\mathcal{L}(u) = 0 \ . \tag{3.34}$$

A numerical estimate of the solution, v_h say, will not satisfy this equation exactly. Assume a correction δv_h can be found such that

$$\mathcal{L}_h(v_h + \delta v_h) = 0 \ , \tag{3.35}$$

which, after linearization, leads to

$$\mathcal{A}_h \delta v_h + R_h = 0 \ , \tag{3.36}$$

where $R_h = L_h(v_h)$ is the residual and A_h is the Jacobian of the nonlinear operator L_h. On the coarse grid the above equation can be replaced by

$$\mathcal{A}_{2h} \delta v_{2h} + I_{2h}^h R_h = 0 \ , \tag{3.37}$$

where I_{2h}^h represents the aggregation or restriction operator. To avoid using the Jacobian explicitly add

$$\mathcal{L}_{2h}(v_{2h}) - R_{2h} = 0 \tag{3.38}$$

to Eq. (3.37) to get

$$\mathcal{L}_{2h}(v_{2h} + \delta v_{2h}) + I_{2h}^{h} R_h - R_{2h} = 0 \ . \tag{3.39}$$

This is formally identical to the original equation, except for an additional source term involving the residuals on the coarse and fine mesh. The solution can thus be advanced on the coarse mesh using the same routines as on the fine mesh. Let $v_{2h} + \delta v_{2h} = v_{2h}^{+}$. After v_{2h}^{+} has been computed on the coarse grid, the corrected solution on the fine grid can be written

$$v_h^{+} = v_h + I_h^{2h}(v_{2h}^{+} - v_{2h}) \ , \tag{3.40}$$

where I_h^{2h} is an interpolation operator. We combine this multigrid strategy with Runge-Kutta time stepping, as proposed by Jameson [32]. For computations on structured meshes we use the nonlinear symmetric Gauss-Seidel scheme with multigrid, as proposed by Jameson and Caughey [37]. Both approaches yield a very efficient computational framework.

In principle the same numerical methods that are used on the finest mesh in a multigrid sequence can be used on coarser meshes as well. However, the numerical schemes used on coarse meshes do not affect the converged solution, but only the rate of convergence. It has been found, in particular for irregular unstructured meshes, that a more diffusive flux computation on coarse meshes has advantages. In fact, it often pays to smooth the residual correction in Eq. (3.40) before adding it to the next finer mesh. Keeping these constraints in mind, it seems unreasonable to employ the full gas-kinetic formulation on coarse meshes. Instead we have chosen a simple, more diffusive, conventional method for the convective fluxes, using central averages with scalar diffusion according to

$$F_{Riemann} = \frac{1}{2}\left(F^L + F^R\right) - \alpha\left(q^R - q^L\right) . \tag{3.41}$$

Here F^L and F^R are the projected inviscid fluxes evaluated with cell-averaged values from the left and right cells (no reconstruction is carried out for coarse meshes), and α is a diffusion coefficient proportional to the spectral radius of the projected flux Jacobian. The procedure described in [29], which was referred to in section 3.4, may be used for viscous discretization. This method is only first order accurate for general meshes, which is consistent with the overall first order approach used for coarse meshes, and does not change the discretization stencil.

3.5.3 Adaptive Diffusion

For viscous flow the collision time is proportional to the dynamical viscosity, according to Eq. (2.31), i.e. $\tau = \mu/p$, as outlined in section 3.5.1. For inviscid flow one can set a lower value corresponding to numerical diffusion, $\tau = \beta_1 \Delta t$, where usually $\beta_1 \ll 1$, to provide a small amount of background diffusion [73].

For the BGK scheme it has been common practice, however, to augment the collision time by a pressure dependent term, designed to make the physical viscosity proportional to the mesh width, whenever shocks are present [73]:

$$\tau = \widetilde{\mu}/p + \beta_2 \left| \frac{p^L - p^R}{p^L + p^R} \right| \Delta t \ . \tag{3.42}$$

Here $\widetilde{\mu}$ is given by Eq. (2.31) or by $\widetilde{\mu} = \beta_1 \Delta t p$ for viscous and inviscid flow, respectively. The variables can be evaluated with the relaxation state at the face. The rationale behind the additional term is that the numerical viscosity should scale as the characteristic mesh length in under-resolved regions. Eq. (3.42) indeed makes the *physical* viscosity proportional to this characteristic length, provided the time step scales linearly with it. This has two main effects at shock locations: Firstly the physical viscous terms are proportional to the increased coefficient of viscosity. Secondly, the discontinuous reconstruction is more heavily weighted, providing additional dissipation.

It is not at all obvious whether the additional increase of the physical viscosity is necessary or sufficient for adequate shock capturing. For some transonic flow cases, in particular in three dimensions, it was not used in our implementations, while for most two-dimensional computations we generally found beneficial. This simply suggests that the best treatment of dissipation is still an open problem for this class of scheme.

3.6 Results

Extensive validation of the new BGKgg scheme, which is based on the distribution function given in Eq. (3.26), is presented in this section. Furthermore, the new scheme is compared to the original BGK scheme, Eq. (3.18), the intermediate BGKg scheme, Eq. (3.25), and various conventional methods.

3.6.1 Isentropic Nozzle Flow: A Convergence Study

The governing equations for isentropic, quasi one-dimensional nozzle flow can be written

$$\partial_t Q + \partial_x F + S = 0 \ , \tag{3.43}$$

where

$$Q = \begin{pmatrix} \rho \\ \rho U \\ E \end{pmatrix} , \qquad F = \begin{pmatrix} \rho U \\ \rho U^2 + p \\ U(E+p) \end{pmatrix} , \qquad S = \frac{1}{A}\frac{d\,A}{dx} \begin{pmatrix} \rho U \\ \rho U^2 \\ U(E+p) \end{pmatrix} . \tag{3.44}$$

The flow variables are defined as in section 2.6. For the purposes of this convergence study we have defined a cross section of the nozzle on the interval $[0, x_{\text{out}} = 10]$ by

$$A(x) = \begin{cases} d_{\text{in}} + \xi(x)^2(3 - 2\xi(x))(d_t - d_{\text{in}}) \ , & x < x_t \\ d_t + \eta(x)^2(3 - 2\eta(x)(d_{\text{out}} - d_t) \ , & x \geq x_t \end{cases} , \tag{3.45}$$

where $\xi(x) = x/x_t$, and $\eta(x) = (x - x_t)/(x_{\text{out}} - x_t)$. The coordinate $x_t = 3.75$ is the location of the nozzle throat, while the nozzle diameters at the inlet, outlet and throat are given by $d_{\text{in}} = 1$, $d_{\text{out}} = 1.25$, and $d_t = 0.875$, respectively.

We choose a nozzle exit Mach number of $M = 0.3$, which leads to a maximum Mach number in the nozzle throat of $M \approx 0.46$ for the chosen nozzle geometry. At the inflow we extrapolate the outgoing Riemann invariant and fix the total enthalpy and entropy function $s = \rho^\gamma/p$, while at the outflow the entropy and total enthalpy are extrapolated and the pressure is fixed. The exact solution for the Mach number can be obtained from the well known area-Mach number relation for isentropic nozzle flow, which can be found in many standard text books, e.g. [1]. Table 3.2 displays the maximum error in the Mach number for the BGKgg scheme in mesh refinement, where we have denoted the number of mesh cells with N. The data is compared to a conventional finite volume scheme, which uses the CUSP flux [36]. It can be seen that the BGKgg scheme reaches second order of accuracy in the asymptotic region. A graphical comparison between the two schemes is shown in Fig. 3.3, where the error is plotted on a logarithmic scale against the number of cells. Here it is clearly visible that both schemes converge at approximately the same rate.

	BGKgg		CUSP	
N	L_∞(Error)	Order	L_∞(Error)	Order
20	5.920739e-04		7.531804e-04	
30	3.093076e-04	1.601359	4.037851e-04	1.537548
40	1.740776e-04	1.998161	2.375062e-04	1.844706
50	1.044453e-04	2.289281	1.513558e-04	2.019152
70	5.049497e-05	2.160028	8.068882e-05	1.869495
100	2.409334e-05	2.074545	3.744039e-05	2.152801

Table 3.2: Comparison of the maximum Error in the Mach number between the BGKgg scheme and the CUSP scheme for isentropic nozzle flow.

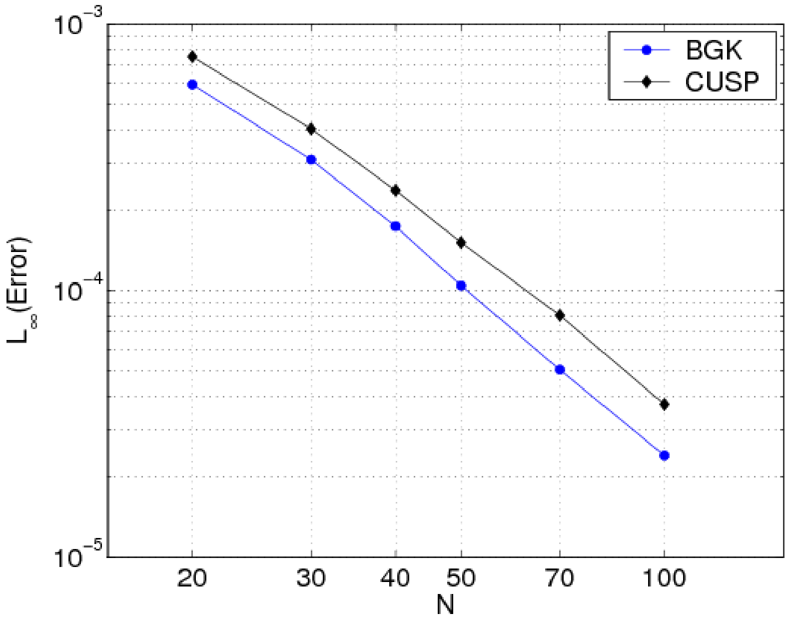
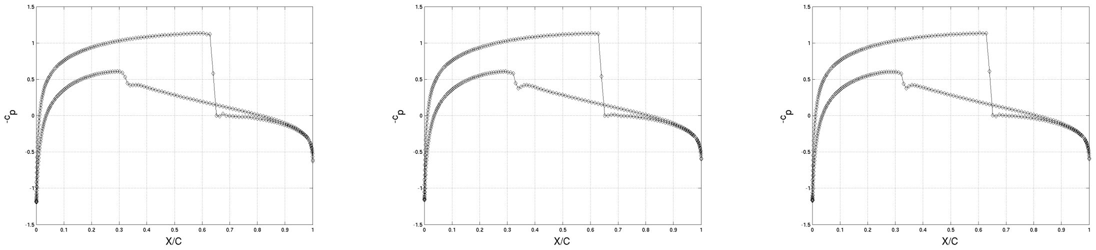


Figure 3.3: Comparison of the maximum Error in the Mach number between the BGKgg scheme and the CUSP scheme for isentropic nozzle flow.



(a) Solution after 3 multigrid cycles

(b) Solution after 10 multigrid cycles

(c) Solution after 50 multigrid cycles

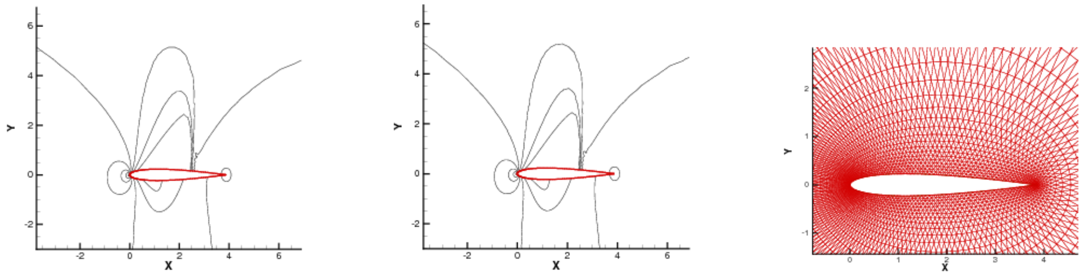
Figure 3.4: The BGKgg scheme with the nonlinear SGS method.

3.6.2 2D Results

For a demonstration of the convergence properties of the BGKgg scheme with the nonlinear SGS method consider inviscid transonic flow. Fig. 3.4 shows the distribution of the pressure coefficient for flow around the NACA0012 airfoil at $M = 0.8$ and $\alpha = 1.25^o$, computed on a 320x64 structured mesh. The solution strategy uses a full-approximation multigrid method, where the solution process is started on the coarsest mesh. The figure shows the solution on the finest mesh, using seven levels of multigrid. It can be seen that the solution is almost converged after three cycles. The multigrid overhead becomes negligible after the first coarse mesh, since the number of cells decreases by a factor of 4 for each mesh level. The asymptotic overhead due to multigrid is approximately 33% for computations on the coarse meshes plus the costs for the extra residual computation to compute the forcing function for the coarser meshes in Eq. (3.39). The total cost for the multigrid computations is thus higher by a factor of about 1.6, regardless of the number of meshes used, compared to a single-grid calculation in two dimensions.

The remaining computations in this section use triangular meshes in cell-centered discretization, i.e. the primary mesh elements are used as control volumes. On numerous occasions we compare the kinetic schemes to a conventional finite-volume method, for which we have chosen the CUSP scheme [36] for convective terms, and a central discretization for the viscous terms along the lines proposed in [38]. Both the BGK scheme and the CUSP scheme use the same reconstruction method, based on a least-squares estimate of the gradient along with Venkatakrishnan's limiter [67]. In fact, aside from the flux computation, the numerical codes used for comparison are identical.

We first present two-dimensional results for inviscid transonic flow. Fig. 3.5 shows the computed dimensionless pressure contours p/p_∞, where p_∞ is the free stream pressure, for a NACA0012 profile in transonic flow at $M = 0.8$ and incidence $\alpha = 1.25^o$, computed with the CUSP scheme and the BGKgg scheme. It is not necessary to include the intermediate BGKg scheme here, which is identical to the original BGK scheme for inviscid flow (which



(a) Contour plot of the pressure (CUSP scheme)

(b) Contour plot of the pressure (BGKgg scheme)

(c) Mesh

Figure 3.5: Inviscid transonic flow around the NACA 0012 airfoil at $M = 0.8$, $\alpha = 1.25^o$.

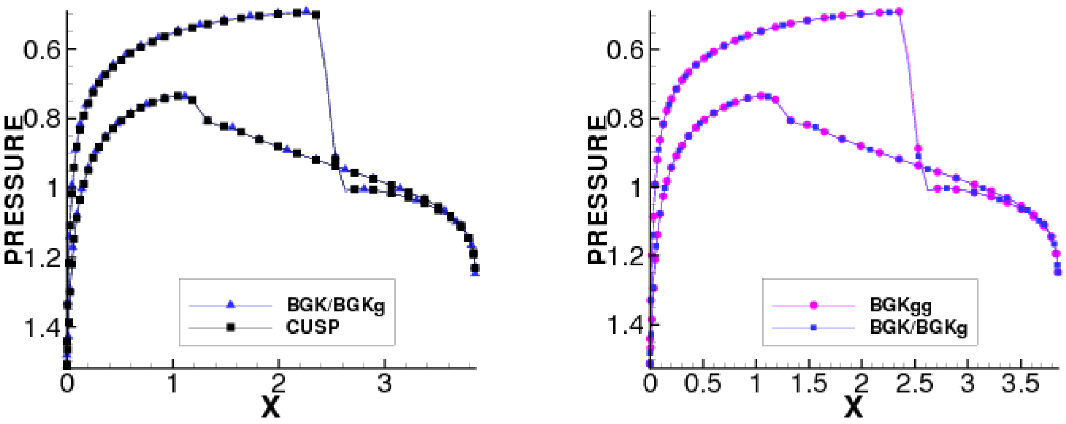
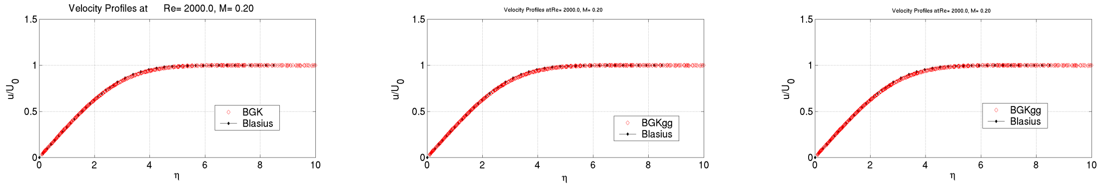


Figure 3.6: Inviscid transonic flow around the NACA0012 airfoil at $M = 0.8$, $\alpha = 1.25^o$. Comparison of the dimensionless pressure p/p_∞ along the airfoil.



(a) BGK

(b) BGKgg with tangential gradients

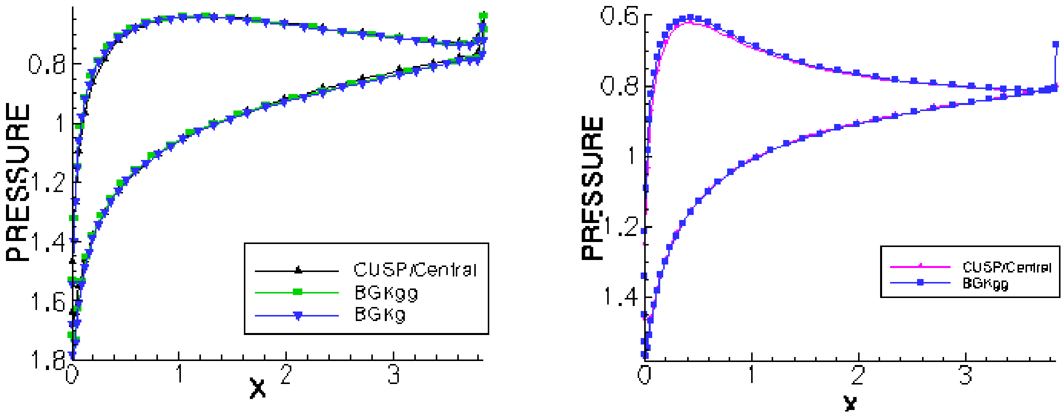
(c) BGKgg without tangential gradients

Figure 3.7: BGKgg vs. BGK: Zero-pressure-gradient boundary layer on a stretched triangular mesh.

does not require the initial nonequilibrium term in Eq. (3.6) to be included). Fig. 3.6 compares the pressure distributions along the airfoil for the same test case. It can be seen that the shock capturing capabilities are virtually identical for all schemes. These computations have been carried out on a mesh with $10,240$ triangles, shown in Fig. 3.5(c), using $\beta_1 = 10^{-2}$ and $\beta_2 = 1$ in Eq. (3.42).

For viscous validation consider a zero-pressure gradient boundary layer. Fig. 3.7 shows the streamwise velocity at $Re = 2000$, $M = 0.2$ computed on a stretched triangular mesh using the standard BGK and the new BGKgg scheme. The figures overplot the entire solution along the plate (excluding the first 10% length) in similarity coordinates. The BGK scheme is shown in Fig. 3.7(a). The original formulation from [73], based on Eq. (3.18) has been used, which includes only the normal expansions outlined in section 3.2. The BGKgg scheme is presented with tangential gradients added, see Eq. (3.32) in section 3.4, and without them, in Figs. 3.7(b) and 3.7(c), respectively. It can be seen that the influence of the "2d" correction is small, and that the new scheme gives a very similar result compared to the original BGK scheme (at much reduced cost).

We now focus on transonic low-Reynolds-number computations to both amplify the contribution of the viscous terms, and lend some importance to the generation of artificial dissipation as well. Some of the test cases we present have also been considered in [3], and have been previously used as benchmarks for viscous flow [6]. Fig. 3.8 shows the dimensionless pressure distribution p/p_∞ for the NACA0012 profile at $M = 0.8$ and $\alpha =$

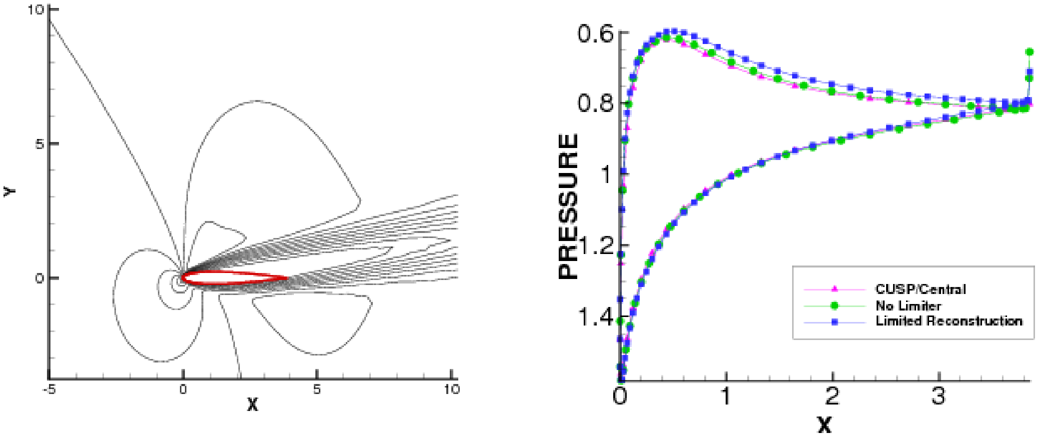


(a) $Re = 73$ (b) $Re = 500$

Figure 3.8: Viscous flow around the NACA0012 profile at, $M = 0.8$ and $\alpha = 10^o$. Comparison of different schemes.

10^o for Reynolds numbers $Re = 73$ and $Re = 500$ and various schemes. It can be seen that the kinetic schemes give very similar results compared to the very well validated conventional method. Recall that, compared to the BGK scheme, the BGKg scheme replaces the KFVS nonequilibrium state $\widehat{f}_1$, Eq. (3.11), with Eq. (3.22). In addition to this modification, the BGKgg scheme replaces the time derivative (3.17) with the one based on the constraint (3.23). It may be seen that these modifications do not significantly affect the results. However, as has been outlined in section 3.3, they are conceptually much cleaner, and reduce the computational cost dramatically.

Fig. 3.9 demonstrates the influence of limiters on the solution for the $Re = 500$ test case. A contour plot of the Mach number, computed with the BGKg scheme, is shown in Fig. 3.9(a). Fig. 3.9(b) shows a comparison of the pressure distribution between the conventional CUSP/central scheme, and the BGKg scheme with and without limiters. The influence of data limiting on the BGK scheme has already been pointed out by Xu [73]. Although not shown, the same behavior is observed for the BGK and BGKgg schemes. An advantage of the BGKgg scheme is that the option of not using data limiting for the nonequilibrium terms can be easily included at only slightly increased computational cost: The term $\overline{\partial_x f_0}$ has to be computed twice, once with, and once without limiting. In the



(a) Mach number contours computed with the BGKg scheme without limiters

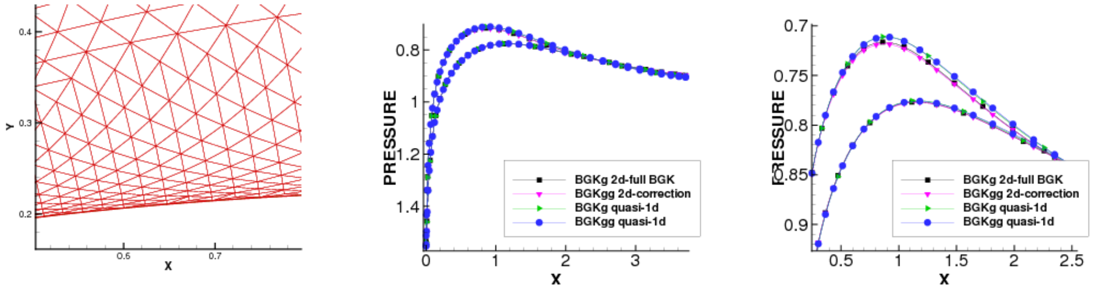
(b) Pressure distribution for the CUSP/central scheme and the BGKg scheme with and without limiter

Figure 3.9: NACA0012 profile at $Re = 500$, $M = 0.8$, and $\alpha = 10^o$. Influence of the limiters.

original scheme, such a separation between the gradients for inviscid and viscous terms is virtually impossible, because of the intricate nature of the formation of the nonequilibrium terms.

Fig. 3.10 shows the solution for the NACA 0012 profile at $M = 0.8$, $Re = 1500$, and incidence $\alpha = 1.25^0$. To resolve the boundary layer, the mesh becomes increasingly stretched. This testcase has been computed mainly with the aim to assess the differences between the full 2d kinetic reconstruction along the lines of [74], the 2d correction of Eq. (3.32), and the quasi-1d expansion without any correction of the viscous terms using tangential derivatives. The difference between these techniques is remarkably small for both the BGKg and the BGKgg scheme. These results are representative, and it seems therefore unwarranted to prefer the extremely expensive multi-dimensional reconstruction over the 2d correction of Eq. (3.32), which is computationally insignificant.

Laminar steady flow around a cylinder is investigated in Fig. 3.11 for low-Reynolds-number flow up to $Re = 50$. For these Reynolds numbers a steady separation bubble is obtained, the length of which has been measured and compared to an empirical correla-

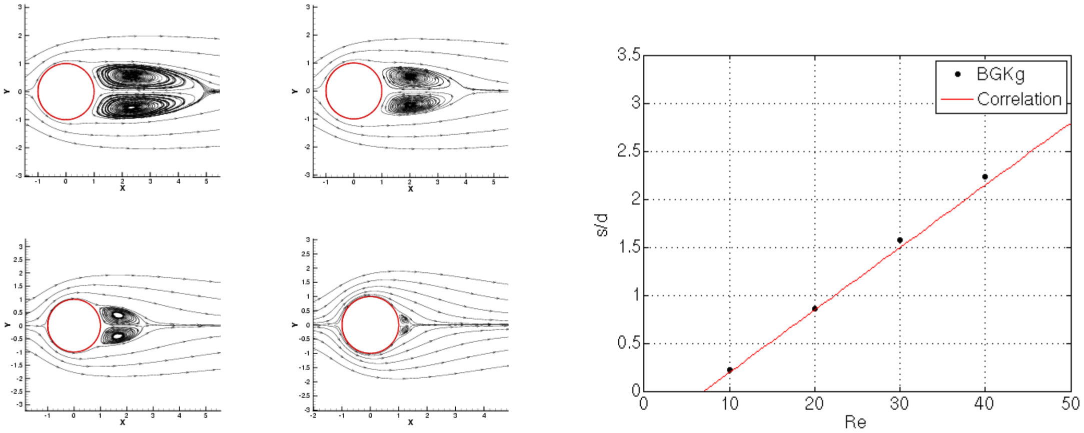


(a) Magnified view of the mesh inside the boundary layer.

(b) Pressure distribution along the profile.

(c) Magnified view of the pressure distribution.

Figure 3.10: NACA0012 profile at $Re = 1500$, $M = 0.8$, and $\alpha = 1.25^o$.



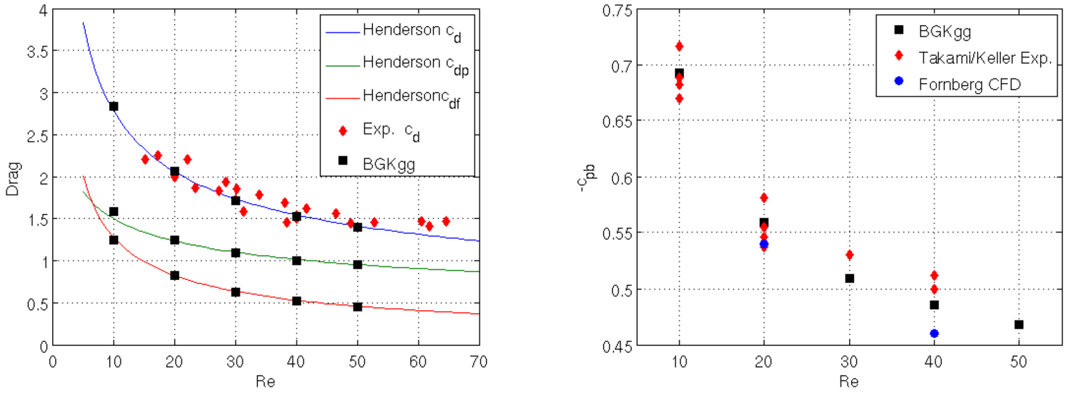
(a) Stream lines for Reynolds Numbers $Re = 40, 30$ (top row) and $Re = 20, 10$ (bottom row)

(b) Nondimensional length of the separation bubble

Figure 3.11: Laminar Separation from a cylinder.

tion [70] obtained by a fit through experimental data. The length s of the separation zone has been nondimensionalized by the cylinder diameter d. Generally the deviation from this correlation is below 4% for all Reynolds numbers considered. While the correlation cannot be considered the exact solution (the experimental data from which it is obtained shows some scatter), the higher deviations for higher Reynolds numbers are likely caused by the fact that the mesh resolution deteriorates rapidly away from the cylinder, which causes larger separation zones to be underresolved.

Fig. 3.12 shows a plot of the drag coefficient and the pressure coefficient at the rear stagnation point (the "base pressure" coefficient). The data obtained with the BGKgg scheme is compared to numerical computations by Henderson [30] and Fornberg [14], as well as experimental results [64, 66]. It may be seen that the results for the drag coefficient agree very well with the computations by Henderson, and fit the experimental data nicely. The base pressure is notoriously sensitive, and both experimental and computational results show some scatter. Still, the results obtained with the BGKgg scheme compare favorably with the experimental results and other computations. Deviations are of the same magnitude as the observed scatter.



(a) Drag Coefficient.

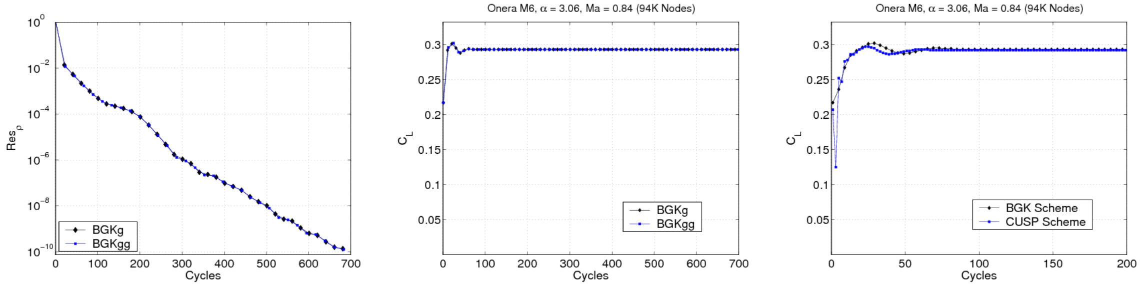
(b) Base suction ($-c_{pb}$).

Figure 3.12: Laminar steady flow around a cylinder. Drag coefficient and base suction vs. Reynolds number.

3.6.3 3D Results

We focus mainly on the multigrid implementation on unstructured meshes in this section. The BGK method has been embedded into the Flo3xx computational aerodynamics architecture, documented in [49], which solves the the three-dimensional Euler and Navier-Stokes equations on arbitrary meshes in finite-volume formulation.

The Onera M6 testcase at $M = 0.84$, $\alpha = 3.06^o$ has been chosen to validate three-dimensional inviscid transonic flow with multigrid. Fig. 3.13 shows the convergence history of the lift and the density residual on a tetrahedral mesh with $94,000$ nodes. Also shown is a comparison of the convergence history for the lift coefficient between the BGK scheme and the CUSP scheme, which serves to show that the BGK scheme is certainly competitive with conventional schemes in terms of convergence to a steady state. We have used $\beta_1 = 10^{-2}$ and, in fact, $\beta_2 = 0$ for these computations. This means that no artificial dissipation is added, and we rely solely on the dissipation generated by the gas-kinetic upwinding. The shock capturing behavior is excellent, as is shown in Fig. 3.14, where the distribution of the pressure coefficient is shown for the wing section $y/b = 0.2$. Both the CUSP method and the BGK schemes are shown.



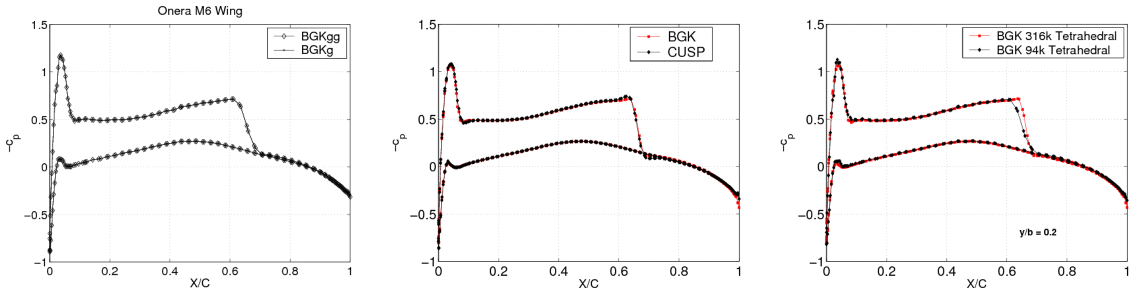
(a) Density Residual: BGK vs. BGKgg

(b) Lift: BGK vs. BGKgg

(c) Lift: BGK vs. CUSP

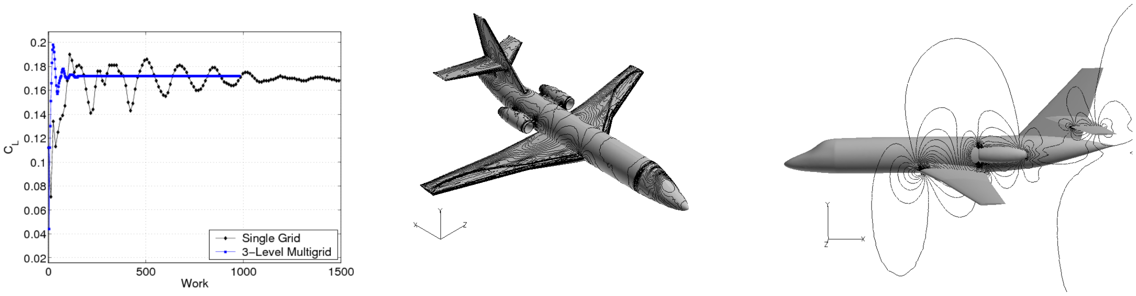
Figure 3.13: Comparison of convergence history for the BGK, BGKgg and CUSP schemes.

Fig. 3.15 shows the convergence history of a single-grid and a three-level multigrid solution for a computation of a Business Jet configuration in transonic inviscid flow, along with contour plots of the density. The multigrid overhead of approximately 40% has been taken into account in the work units used to compare the convergence. The speedup due



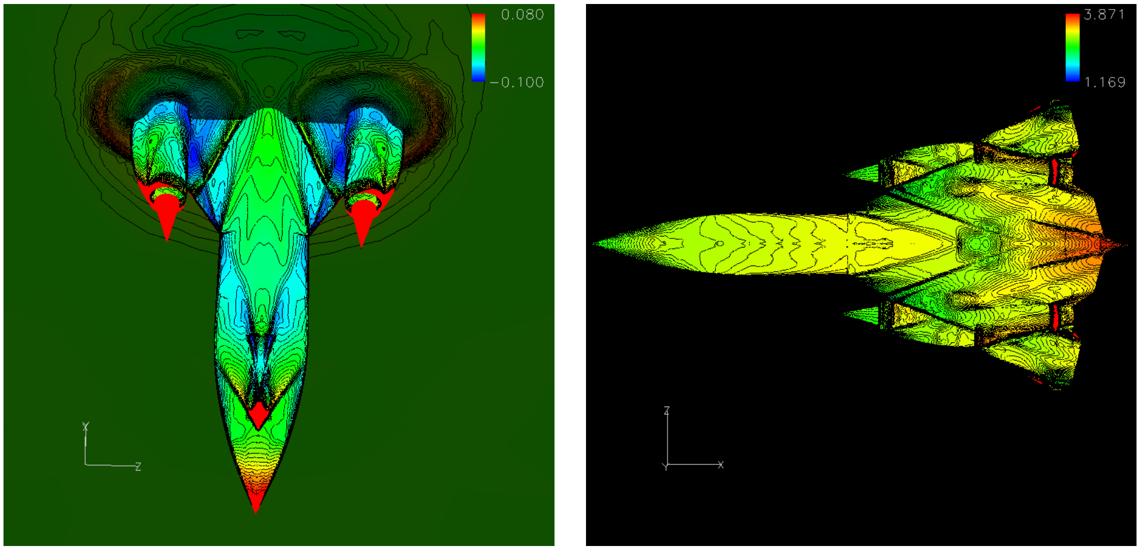
(a) 94k Node Tetrahedral Mesh (b) 316k Node Tetrahedral Mesh (c) Mesh Refinement

Figure 3.14: Pressure Coefficient for the Onera M6 Wing at $M = 0.84$, $\alpha = 3.06$ at wing section $y/b = 0.2$.



(a) Convergence history of the coefficient of lift (b) Density Contours (c) Density Contours

Figure 3.15: Falcon Business Jet at $M = 0.8$ and $\alpha = 2^o$.

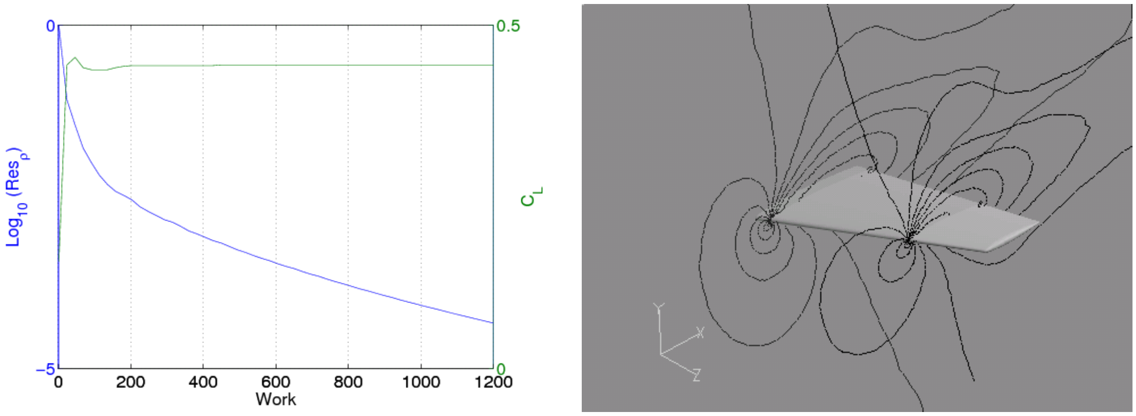


(a) Density Contours (b) Mach Number Contours

Figure 3.16: SR71-based Configuration with $M_\infty = 3.2$, $\alpha = 5^o$.

to multigrid is most dramatic in integrated quantities, such as the coefficient of lift, shown in Fig. 3.15. We point out that solutions for the BGKg scheme and the BGKgg scheme are virtually identical. All schemes are stable at $CFL = 10$ with an explicit 5 Stage Runge-Kutta time stepping technique with residual averaging and local time stepping. Fig. 3.16 demonstrates the shock capturing capabilities for high-Mach number flow about complex configurations, where density and Mach number contours are shown for a test case involving a configuration based on the SR71 Blackbird aircraft. The free stream Mach number is given by $M_\infty = 3.2$, and the angle of attack is $\alpha = 5^o$.

Basic viscous validation for a low Reynolds number is shown in Fig. 3.17, where the solution for the Onera M6 Wing at $M = 0.8$, $\alpha = 3.06^0$ and $Re = 73$ is shown. Note that the convergence history has been plotted against work units, which introduces a factor of about 1.4 in increase of cost compared to single grid solutions, made up in equal parts by the computations on coarser meshes and the cost of performing one extra residual evaluation on the fine mesh for the coarse mesh forcing function in Eq. (3.39). Note how the coefficient of lift is essentially converged when the residual has been reduced by about three orders of magnitude. The mechanism of convergence acceleration is an effective



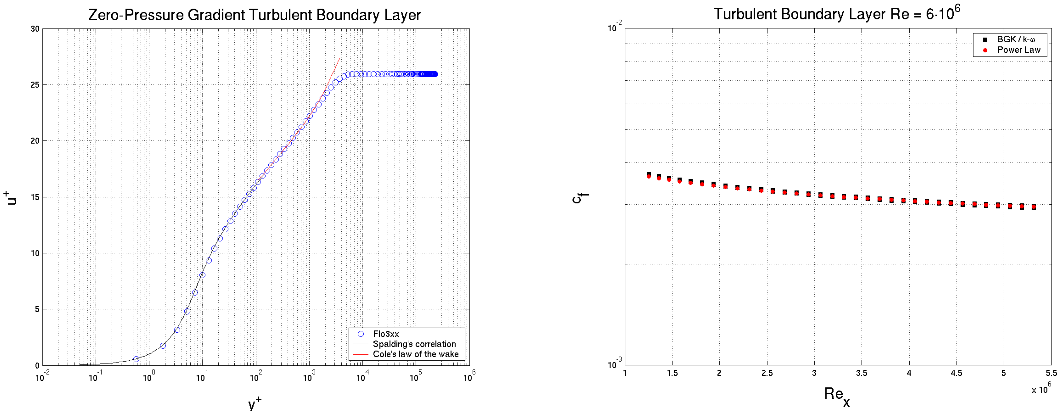
(a) Convergence history of density residual and coefficient of lift.

(b) Contours of pressure coefficient

Figure 3.17: Onera M6 Wing at $Re = 73$, $M = 0.8$, and $\alpha = 3.06^o$.

increase of the speed with which information is propagated in the flow field. Typically, solutions in multigrid computations do not tend to creep, and the steady state is reached at much higher levels of residuals compared to single grid computations.

Finally, consider the case of turbulent flow. Although it is not the scope of this work to validate the methodology for such flows, we nevertheless wish to illustrate that the kinetic route is viable for computation of turbulent flow using eddy viscosity via the Boussinesq hypothesis to model turbulent momentum transfer. The eddy viscosity can be determined using a standard turbulence model, and is subsequently added to the laminar viscosity. This gives the effective viscosity, which is used in Eq. (2.31) to compute the relaxation time τ. Figure 3.18 shows results for a turbulent boundary layer computed on a hexahedral mesh with roughly $90,000$ nodes in terms of the velocity profile and the friction coefficient. It can be seen that both the friction coefficient and the streamwise velocity in the viscous sublayer and the log layer region match theoretical results rather well. For the friction coefficient we have used the $1/5^{th}$ power law, which is reasonably accurate for this Reynolds number [70]. For the streamwise velocity we compare the solution to the the sublayer solution $y^+ = u^+$ and Spalding's correlation [70].



(a) Velocity profile

(b) Skin friction coefficient

Figure 3.18: Turbulent boundary Layer at $Re = 6 \cdot 10^6$, $M = 0.2$.

Chapter 4

Contributions to the Development of the Spectral Difference Method

4.1 Introduction

Most industrial applications of computational physics have been dominated by low order schemes. The simulation of compressible fluid flow for complex geometry is one notable example. In many disciplines, however, extremely high accuracy is required, for example in LES to avoid the masking of subgrid scale dissipation by numerical dissipation, and in problems of aeroacoustics and electromagnetics, where one needs to track wave propagation over long distances. The resulting mesh requirements for low-order methods are often prohibitive, while high-order methods provide higher accuracy with fewer degrees of freedom. Traditionally this has been achieved for modeling complex physics in relatively simple domains. On the other hand, low order methods have become remarkably widespread in the simulation of fluid flow in complex domains on unstructured meshes, and are generally regarded as more efficient in this context. This stems in part from the fact that in industrial applications the accuracy requirements are often not excessive ("engineering accuracy"). Furthermore, low-order schemes are very robust, while pertaining convergence acceleration techniques for steady-state problems are mature and well understood.

However, the promise of high-order methods to provide superior performance at lower computational cost is compelling and motivates further research to extend the range of efficient applicability to complex computational domains. In particular unstructured meshes, which facilitate the meshing of complex domains enormously, are of special interest here. The extension of such algorithms that have made low order schemes so efficient over the past thirty years, such as multigrid, to high-order methods is crucial if their theoretical potential is to be realized.

In this context, "high-order" refers mainly to the spatial accuracy, with the understanding that for certain time-dependent problems time accuracy is also important. However, time accuracy is not the main focus of research in this work, as it is more easily achieved, and is not as heavily dependent on complexity of geometry and the mesh topology. For steady-state problems, which shall receive considerable attention here, time accuracy during the relaxation process is of no importance.

Recently the high order Spectral Difference (SD) method for unstructured meshes, which has been proposed by Liu et al. [46], has been extended to the Euler equations by Wang et al. [69] and May et al. [48]. May et al. have also presented a formulation for advection-diffusion systems and the Navier-Stokes equations on unstructured meshes [50]. The scheme has been established as a viable alternative to such schemes as the Discontinuous Galerkin Method [10, 9] or the Spectral Volume Method [76]. The SD Method is particularly attractive because of its simple formulation, which is shared with finite-difference techniques, while such properties as local conservation, arbitrary order of accuracy on unstructured mesh elements, are similar to Discontinuous Galerkin methods.

This chapter contains contributions made towards the development of the SD scheme, some of which were reported in the above-mentioned references by the present author. The topics presented here include stability analysis, data limiting methodology, and formulation for viscous problems, such as advection-diffusion systems and the Navier-Stokes equations. Furthermore, convergence acceleration for steady-state problems via h/p-multigrid is developed, which addresses one of the major shortcomings of the current state-of-the-art. The present research thus not only contributes to the development of the SD scheme in particular, but also to the advancement of high order schemes in general.

4.2 The Spectral Difference Method

4.2.1 Basic Formulation

The present discussion focuses on the solution of hyperbolic conservation laws of the form

$$\frac{\partial u}{\partial t} + \nabla \cdot F = 0 \ , \qquad (x,t) \in \Omega \times [0,T] \ , \tag{4.1}$$

where $\Omega \subset \mathbb{R}^d$, and $u = (u_1, \ldots, u_p)$. We consider one and two dimensional problems ($d = 1$ and $d = 2$) subject to suitable initial and boundary conditions. Suppose a triangulation of $\mathbb{R}^d$ is given, which is assumed, for the sake of simplicity, to consist of simplexes. The Spectral Difference method uses a pseudo-spectral collocation-based reconstruction for both the dependent variables $u(x)$ and the flux function $F(u)$ inside a mesh element, T_i, say. The reconstruction for the dependent variables can be written

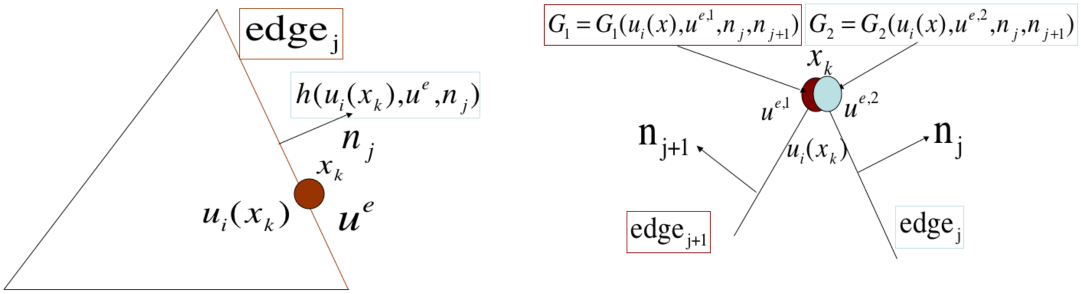
$$\hat{u}_i(x) = \mathcal{I}_i^m(u)(x) = \sum_{j=0}^{N(m)} L_j(x) u_{ij} \ , \qquad x \in T_i \ , \tag{4.2}$$

where $u_{ij} = u(x_{ij})$. Throughout this chapter a double subscript notation refers to a node (second index) inside a cell (first index). Here x_{ij} is the j^{th} solution collocation node in the i^{th} mesh element. Henceforth we shall frequently omit the cell index i, whenever there is no risk of confusion. The interpolation operator $\mathcal{I}_i^m$ denotes a local collocation in cell i using multivariate polynomials of total degree m. The $L_j(x)$ are the cardinal basis functions for the chosen set of collocation nodes x_j, where $j = 0, \ldots, N(m)$, and

$$N(m) + 1 = \frac{\Pi_{j=1}^{d}(m+j)}{d!} \ . \tag{4.3}$$

This leads to an asymptotic order of accuracy of order $n = m + 1$ for smooth functions, according to standard interpolation theory. We shall use the general notation "p_m-SD Scheme" to define a scheme of order $m + 1$.

If the discrete solution u_h is defined as the union of all the interpolated functions on the mesh elements, $u_h = \hat{u}_i(x)$ for $x \in T_i$, $i = 1 \ldots Ne$, where Ne is the number of elements, it will be discontinuous across the element boundaries. The treatment of the element boundaries is one of the key ingredients of the spectral difference method and is



(a) Flux computation for points on edges (b) Flux computation for points on corners

Figure 4.1: Illustration of flux computation for nodes on element boundaries.

discussed below. The reconstruction of the flux function in T_i reads

$$\hat{F}_i(u(x)) = \mathcal{I}_i^{m+1}(F(u))(x) = \sum_{k=0}^{N(m+1)} M_k(x)F(u(x_k)) \ , \qquad x \in T_i \ , \tag{4.4}$$

where the M_k are the cardinal basis functions corresponding to the collocation nodes x_k , $k = 0, \ldots, N(m+1)$. Again the interpolation will be discontinuous at element boundaries. The numerical flux function F_h on the triangle T_i and its boundary ∂T_i is defined as

$$F_h = \begin{cases} \hat{F}_i & , \quad x \in T_i \\ G(\hat{u}_i(x), u^{e,j_1}, \ldots, u^{e,j_k} \ldots) & , \quad x \in \partial T_i \end{cases} \ , \tag{4.5}$$

where the $u^{e,j_1}, \ldots, u^{e,j_k} \ldots$ are "external" solutions on triangles T_{j_k} with $j_k \neq i$ such that x is shared by T_i and T_{j_k}, and the value is given by $u^{e,j_k} = \lim_{y \to x} \hat{u}_{j_k}(y)$. For nodes on edges of triangles there is only one such external solution u^e. It is necessary for discrete conservation that the normal flux component be continuous across the edge, which suggests the use of numerical flux functions, as discussed in section 2.3, such that the normal flux component $F_n = \hat{F}_i \cdot n$, for the node x_k, say, is replaced by the numerical flux $h(\hat{u}_i(x_k), u^e, n)$, where n is the edge normal, and h is the numerical flux function approximating F_n. A schematic illustration is shown in fig. 4.1(a). For flux nodes on corners there is more than one external solution. One may compute the corner fluxes from normal fluxes on the two incident edges of a triangle by imposing

$$\begin{aligned} F_h \cdot n_1 &= h_{n_1} \\ F_h \cdot n_2 &= h_{n_2} \ , \end{aligned} \tag{4.6}$$

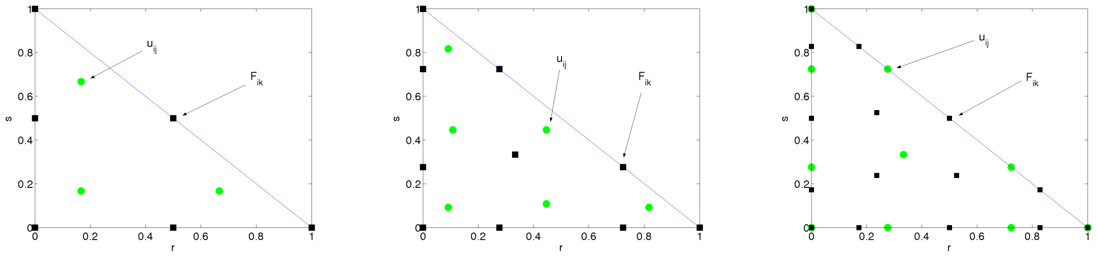
where n_1 and n_2 are the normals on the incident edges, and h_{n_1}, h_{n_2} are the associated numerical fluxes. This treatment was suggested by Wang et al. [69] and makes the flux unique, while allowing for conservation. The linear system, Eq. (4.6), can be solved analytically to give modified numerical fluxes on corners that can be split into two parts, which are associated with the two incident edges. This is shown in Fig. 4.1(b). All numerical fluxes may thus be restricted to three-point flux functions, i.e. fluxes of the form $H(u^r, u^l, n)$, defined by a logical left and right state and a normal n. The subsequent discussion will be restricted to such flux functions. Any of the fluxes discussed in sections 2.4 and 2.5 may be used. In particular, the Lax-Friedrich flux, Jameson's CUSP flux [36], and the Roe-flux [58] have been used in this work. Tangential flux components can either be evaluated in each cell and left unchanged or averaged across cell interfaces using an arithmetic average to produce a flux function F_h which is single-valued everywhere.

The baseline scheme is now readily defined in ODE form, with the degrees of freedom given by the values of the solution at the collocation nodes, and the solution operator is given by the exact differentiation of the reconstructed flux function:

$$\frac{du_{ij}}{dt} + (\nabla \cdot F_h)(x_{ij}) = 0 \ , \tag{4.7}$$

where $u_{ij} = u_h(x_{ij})$, and again x_{ij} is the j^{th} solution collocation node in the i^{th} mesh element. The SD method is related to staggered grid multidomain spectral methods, proposed by Kopriva [42] for semi-global approximations on quadrilateral domains.

The discussion in this introductory section has focused on the essential ingredients of the method. A more detailed discussion of the implementation of the scheme, i.e. the computation of the space discretization operator for general triangular meshes, is deferred to Appendix C.2.



(a) Linear (b) Quadratic (c) Cubic

Figure 4.2: Schematic depiction of collocation nodes for triangles.

4.2.2 Collocation Nodes

Any combination of collocation nodes may be used, provided that the following assumption is satisfied

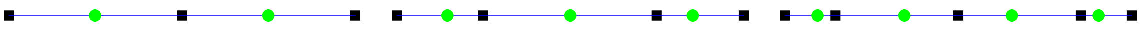
Assumption 4.2.1. *The chosen set of collocation nodes*

- *supports a volume quadrature of the order of the interpolation* n
- *the restriction of the flux nodes to the boundaries supports a* $d-1$*-dimensional quadrature of order* $n+1$

This ensures discrete conservation in the sense that

$$\frac{d}{dt}\int_{T_i} u_h \, dx = -\int_{\partial T_i} F_h \cdot n \, dA \ , \tag{4.8}$$

can be numerically satisfied exactly for the solution and reconstructed flux function, such that the scheme for the averages is conservative, a point that will become clear in section 4.4. Hesthaven proposed nodes based on the solution of an electrostatics problem for simplexes [31], which support both a volume and a surface integration to the required degree of accuracy, and have very good interpolation properties. These nodes can be used for both flux and solution collocation. Fig. 4.2 shows examples of nodes for elements of various orders. For lower and medium orders the asymptotic interpolation properties of the nodal set are not extremely important, and one may simply choose Gauss quadrature points for the solution nodes.



(a) Linear (b) Quadratic (c) Cubic

Figure 4.3: Reconstruction in 1D: The squares depict flux nodes, the circles are solution nodes.

In one dimension one may use Lobatto nodes for the flux collocation, and Gaussian nodes for the solution, based on a suitable family of orthogonal polynomials, such as Chebyshev or Legendre. An example based on Chebyshev polynomials is shown in Fig. 4.3.

4.3 Time Stepping and Relaxation

The Spectral Difference Scheme, when written in the form of Eq. (4.7), is amenable to a variety of time stepping and relaxation techniques. It is shown in sections 4.4.3, and 4.4.4 that linear and nonlinear stability may be established under a mild cfl condition for explicit schemes, such as Runge-Kutta schemes. Several Runge-Kutta (RK) schemes are considered in this work, all of which can be written as

$$\begin{aligned} u^{(0)} &= u^n \\ u^{(k)} &= \sum_{l=0}^{k-1} \alpha_{kl} u^{(l)} + \Delta t \beta_{kl} R^{(l)} \ , \qquad k = 1, \ldots, m \\ u^{n+1} &= u^{(m)} \ , \end{aligned} \tag{4.9}$$

where $R = -\nabla \cdot F_h$ is the residual, see Eq. (4.7). Among the schemes that have been used in this work, two general classes of RK schemes may be identified: Firstly, the class of high-order accurate TVD RK schemes proposed by Shu [60, 61], and secondly minimum storage schemes, such as the ones proposed by Jameson [33, 32] and Shu [60]. Of the former class, the following two schemes have found use:

$$\text{Shu}-\text{RK2}: \qquad \alpha = \begin{pmatrix} 1 & \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} , \qquad \beta = \begin{pmatrix} 1 & \\ 0 & \frac{1}{2} \end{pmatrix} \tag{4.10}$$

$$\text{Shu}-\text{RK3}: \qquad \alpha = \begin{pmatrix} 1 & & \\ \frac{3}{4} & \frac{1}{4} & \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix} , \qquad \beta = \begin{pmatrix} 1 & & \\ 0 & \frac{1}{4} & \\ 0 & 0 & \frac{2}{3} \end{pmatrix} , \tag{4.11}$$

where the coefficients have been summarized in convenient matrix notation, with the columns corresponding to l and the rows corresponding to k in (4.9). The schemes belonging to the class of minimum-storage RK schemes are specifically designed for higher stability regions, often sacrificing higher-order accuracy. They are thus most often used for steady-state problems. Minimal storage means that $\alpha_{k0} = 1$, $\alpha_{kl} = 0$ for $l \neq 0$, and $\beta_{kl} = 0$ for $l \neq k-1$ in (4.9). The family of minimum-storage schemes proposed by Shu, denoted Shu-RK$_s$ here, is designed to preserve the TVD property of the spatial operator for large cfl numbers [60], and can be written

$$\beta_{k,k-1} = \frac{k}{m(m-k+1)} , \qquad k = 1, \ldots , m . \tag{4.12}$$

Jameson's schemes, denoted Jameson-RK$_s$ [32, 33], have been designed to achieve large cfl numbers for a linear model problem. They can be written:

$$\beta_{k,k-1} = \frac{1}{m-k+1} , \qquad k = 1, \ldots , m . \tag{4.13}$$

The Jameson-RK4$_s$ scheme is fourth order accurate in time for the linear problem [32].

Generally the time step restrictions for linear stability are more severe for higher-order schemes, compared to first order accurate model problems, see section 4.4.3 and [2]. Often only marginal linear stability can be established, in the sense that the amplification factor can be bound only by $1 + \mathcal{O}(\text{cfl}^p)$ for some $p > 0$. In practice this may not be of great concern, as this poses severe restrictions mainly in the limit of mesh refinement, and the nonlinear stability estimates are much more benign, see section 4.4.4. It is nevertheless worthwhile to consider implicit formulations, which may have the potential to speed up relaxation to steady-state solutions by allowing much higher cfl numbers. Furthermore, it is felt that the simple formulation of the SD scheme is a distinct advantage, which carries over to the implicit version.

Consider the scheme in semi-discrete form (4.7). Using a standard first order linearization a baseline implicit time discretization can be written as

$$\left(I + \Delta t \left.\frac{\partial R}{\partial u}\right|_n\right) \Delta u^n = -\Delta t R(u^n) \ , \tag{4.14}$$

where $\Delta u^n = u^{n+1} - u^n$. Several paths can be taken to solve Eq. (4.14). A good compromise between computational efficiency and storage requirements is given by the general class of symmetric Gauss-Seidel schemes. A special case is the LU-SGS scheme due to Jameson and Yoon [34, 39]:

$$(D+L)\Delta u^{(1)} = -R(u^n) \ , \tag{4.15}$$

$$(D+U)\Delta u^{n+1} = D\Delta u^{(1)} \ . \tag{4.16}$$

As an example consider the one-dimensional problem. The system Jacobian matrix will be block tridiagonal, and may be written for the p_m-SD Scheme for each cell as

$$D_{rs} \quad = I + \lambda \textstyle\sum_{k=0}^{m+1} d_{rk} \widetilde{A}(p_k) \frac{\partial p_k}{\partial u_s} = I + \lambda \sum_{k=0}^{m+1} d_{rk} \widetilde{A}_k l_{ks} \ , \tag{4.17}$$

$$L_{rs} \quad = \lambda \, d_{r0} \widetilde{A}_{m+1} l_{m+1,s} \ , \tag{4.18}$$

$$U_{rs} \quad = \lambda \, d_{r,m+1} \widetilde{A}_0 l_{0s} \ . \tag{4.19}$$

where $r, s = 0 \ldots , m$ and $\lambda = \Delta t / \Delta x$. The matrices L and U operate on the solution in cells $i-1$ and $i+1$, respectively, and the reconstruction coefficients are defined by

$$p_k = \sum_{j=0}^{m} l_{kj} u_{ij} \ , \qquad l_{kj} = L_j(x_k) \tag{4.20}$$

$$\frac{1}{\Delta x} \sum_{k=0}^{m+1} d_{jk} f_k = \frac{d}{dx} \mathcal{I}^{m+1}(f_h)(x_j) \ . \tag{4.21}$$

The local Jacobian $\widetilde{A}$ is simply the flux Jacobian evaluated at the flux collocation points for interior nodes, and the differentiation of the numerical flux function h for nodes on element boundaries. In practice some suitable approximation for the derivative to the numerical flux function is used (e.g. the diffusion coefficients are treated as local constants, etc.). In essence the system Jacobian is thus a simple superposition of local flux Jacobians, which can be assembled very efficiently in a loop over the local flux nodes using the known universal reconstruction coefficients.

4.4 Properties of the One-Dimensional Scheme

Studying the one-dimensional scheme can give valuable insights into the nature and basic properties of the proposed methodology. Results regarding linear and nonlinear stability give important guidelines for the extension of the scheme to multidimensional systems, such as the Euler and Navier-Stokes equations. Throughout this section, a subdivision of the real line into equal intervals $K_i = (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$, which serve as collocation cells, and a sequence $t^n = n\Delta t,\ n = 0, \dots, NT$ is assumed [1]. Furthermore, $\lambda = \Delta t/\Delta x$, where $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. Boundary conditions are not considered. Consider the one-dimensional version of Eq. (4.1):

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = 0\ . \tag{4.22}$$

Let the local solution nodes be defined by $x_j\ ,\ j = 0, \dots, m$ and the flux nodes by $x_k\ ,\ k = 0, \dots m+1$. Define the degrees of freedom as $u^n_{ij} = u_h(t^n, x_{ij})$. The p_m-SD Scheme with a forward-Euler time discretization reads:

$$u^{n+1}_{ij} = u^n_{ij} - \lambda \sum_{k=0}^{m+1} d_{jk} f_{ik}\ , \tag{4.23}$$

where $f_{ik} = f_h(p_k)$, and p_k and d_{jk} are defined as in Eq. (4.20). Note that the reconstruction coefficients are universal, in the sense that they do not depend on the cell i. Hence they need only be stored for a reference element. [2]

4.4.1 The Scheme for the Means: Conservation Property

Conservation is ensured by construction, and verified numerically in the scheme for the means, i.e. the scheme for the means can be written in the form of Eq. (2.48). This property allows one to establish nonlinear stability and convergence as a consequence of

---

[1] Restriction to a uniform mesh is by no means necessary, and is done here only for notational convenience

[2] This is true in higher dimensions as well, as long as the element is straight-sided

TVD stability theory in much the same way as has been done for many high-resolution finite-difference and high-order methods, provided certain criteria in the numerical flux computation are satisfied and appropriate data limiting procedures are used, which is discussed in section 4.4.4.

The equation for the volume averages $\bar{u}_i$ can be written

$$\sum_{j=0}^{m} w_j u_{ij}^{n+1} = \bar{u}_i^{n+1} = \bar{u}_i^n - \lambda \sum_{j=0}^{m} w_j \sum_{k=0}^{m+1} d_{jk} f_{ik} = \sum_{k=0}^{m+1} \widetilde{w}_k f_{ik} \ , \tag{4.24}$$

where the w_j are quadrature weights corresponding to the chosen set of solution nodes, which by assumption 4.2.1 support exact quadrature of polynomials at least up to degree m. Also,

$$f_k = \begin{cases} h_{i-\frac{1}{2}} & k = 0 \\ h_{i+\frac{1}{2}} & k = m+1 \\ f(u_h(x_k)) & \text{otherwise} \end{cases} \ . \tag{4.25}$$

The modified quadrature weights $\widetilde{w}_k$ are given by

$$\widetilde{w}_k = \sum_{j=0}^{m} w_j d_{jk} \ . \tag{4.26}$$

This, however, can be simplified by noting that $d_{jk} = \Delta x M_k'(x_j)$, where $M(x)$ is the cardinal basis function for the flux collocation, whose derivative is a polynomial of degree m by construction, and is hence integrated exactly, which means that

$$\sum_{j=0}^{m} w_j d_{jk} = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} M_k'(x)\, dx = M_k(x_{i+\frac{1}{2}}) - M_k(x_{i-\frac{1}{2}}) = \begin{cases} -1 & k = 0 \\ 1 & k = m+1 \\ 0 & \text{otherwise} \end{cases} \ , \tag{4.27}$$

using the fact that $M_k(x)$ is an interpolation polynomial for which $M_k(x_l) = \delta_{kl}$ holds. Note that the Interval endpoints must be flux collocation nodes for the scheme to be well defined, since numerical flux functions must be used there to provide the inter-element coupling. Upon substitution of this result into equation (4.24), the conservation form for the volume averages, Eq. (2.48), is recovered. The same is true in higher dimensions, although the algebra is a bit more tedious, see Appendix C.1.

4.4.2 Basic Concepts of Data Limiting

Weak solutions to hyperbolic conservation laws may contain discontinuities, which poses a challenge for the development of numerical schemes. According to a famous theorem due to Godunov, linear-coefficient schemes of order greater than one will not be monotone (and not TVD as discussed in section 2.4). This has led to the development of nonlinear schemes, which introduce limiting functions that depend on the solution. The development of such limiters is discussed in a more general context in section 4.5. However, in order to establish the fundamental stability properties of the scheme, it is necessary to introduce a baseline one-dimensional limiting methodology, which will subsequently be used to guide the development of the more general multidimensional procedures.

Consider a three-point numerical flux $h_{i+\frac{1}{2}} = h(u_{i+1}^{-}, u_i^{+})$, where the states $u_i^{\pm}$ may be evaluated using the collocation polynomial:

$$u_i^{+} = p_{i,m+1} \ , \qquad u_i^{-} = p_{i0} \ . \tag{4.28}$$

To ensure monotonicity, the reconstruction is modified as follows. First define the quantities

$$\widetilde{\delta_+ u_i} = m(\delta_+ u_i, \Delta_+ \bar{u}_i, \Delta_+ \bar{u}_{i-1}) \ , \tag{4.29}$$

$$\widetilde{\delta_- u_i} = m(\delta_- u_i, \Delta_+ \bar{u}_i, \Delta_+ \bar{u}_{i-1}) \ . \tag{4.30}$$

where Δ_+ is the usual difference operator $\Delta_+ u_i = u_{i+1} - u_i$, and $\delta_+ u_i = p_{i,m+1} - \bar{u}_i$, $\delta_- u_i = \bar{u}_i - p_{i0}$. The function $m(a,b,c)$ is a minmod function, defined by

$$m(a,b,c) = \begin{cases} \operatorname{sign}(a)\min(|a|,|b|,|c|) & \operatorname{sign}(a) = \operatorname{sign}(b) = \operatorname{sign}(c) \\ 0 & \text{otherwise} \end{cases} \ . \tag{4.31}$$

Eq. (4.28) is then replaced by

$$u_i^{\pm} = \begin{cases} (4.28) \ , & \text{if} \quad \widetilde{\delta_+ u_i} = \delta_+ u_i \quad \text{and} \quad \widetilde{\delta_- u_i} = \delta_- u_i \\ \bar{u}_i \pm \widetilde{u}_i \ , & \text{otherwise} \end{cases} \tag{4.32}$$

where

$$\widetilde{u}_i = \min\left(\widetilde{\delta_+ u_i}, \widetilde{\delta_- u_i}\right) \ . \tag{4.33}$$

This methodology ensures that the reconstruction at the cell interfaces remains bounded by the volume averages in the cell i and the two neighboring cells. This may be used to ensure monotonicity and establish the TVDM property (total-variation-diminishing in the means), see section 4.4.4.

For the scheme to be well defined, one needs to extend the limiting procedure to all degrees of freedom. Although this can be accomplished in several ways, for the purposes in this section it is enough to extend the linear variation defined in Eq. (4.32) and (4.33) to all nodes in a cell, in the case of active limiters.

4.4.3 Linear Stability

In this section we study the stability properties of the (linear) baseline scheme for linear equations of the form

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0 \ , \tag{4.34}$$

with periodic boundary conditions. A straightforward computation reveals that the p_m-SD scheme with an upwind discretization of the interface fluxes, i.e. $h_{i+\frac{1}{2}} = u_i^+$, with u_i^+ as defined in Eq. (4.28), can be written in ODE form

$$\frac{d\mathbf{u}_i}{dt} + \frac{1}{\Delta x}\left(A\mathbf{u}_i + B\mathbf{u}_{i-1}\right) = 0 \ , \qquad i = 1, \ldots, N \ , \tag{4.35}$$

where $\mathbf{u}_i = (u_{i,0}\,, \ldots, u_{i,m})$ is the vector of degrees of freedom in cell i. The $m \times m$ matrices A and B are given by

$$a_{rs} = \sum_{k=0}^{m+1} d_{rk} l_{ks} \ , \qquad b_{rs} = d_{r0} l_{m+1,s} \ , \qquad r, s = 0, \ldots, m \ . \tag{4.36}$$

The elements of the vector-valued solution may be subjected to a discrete Fourier transformation. This leads to a the scheme for the (decoupled) Fourier modes, which can be written generically:

$$\Delta t \frac{d\hat{\mathbf{u}}}{dt} = Z\hat{\mathbf{u}} \ , \qquad \text{with} \quad Z = -\lambda(A - \exp(\sqrt{-1}\xi)B) \ , \tag{4.37}$$

where Z is the Fourier symbol of the space discretization, and $\lambda =$ cfl, since the wave speed is unity. Ultimately one aims to find an expression for $\hat{u}(t^{n+1}) = G(\lambda, \xi)\hat{u}(t^n)$, the exact form of which depends on Z and the time discretization. It is necessary for L_2-Stability that the spectral radius $\rho(G)$ be bound by unity for all $\xi \in [0, 2\pi]$, or more precisely, that

$$\rho(G(\lambda, \xi)) \leq 1 + C\Delta t \ , \qquad \forall \xi \in [0, 2\pi] \ , \tag{4.38}$$

which is thus formulated as a condition on λ. Consider the general Runge-Kutta scheme, Eq. (4.9) for Eq. (4.35). Going to frequency domain, one may define an intermediate amplification factor $\hat{u}_k^{(k)} = G_k^{(k)} \hat{u}_k^{(0)}$ for each stage. The following recursion presents itself:

$$\hat{u}^{(0)} = \hat{u}^n \qquad\qquad G^{(0)} = I \tag{4.39}$$

$$\hat{u}^{(1)} = (\alpha_{10} I + \beta_{10} Z)\, u^{(0)} \qquad\qquad G^{(1)} = \alpha_{10} I + \beta_{10} Z \tag{4.40}$$

$$\vdots \qquad\qquad \vdots \tag{4.41}$$

$$\hat{u}^{(k)} = \left(\sum_{l=0}^{k-1} \alpha_{kl} I + \beta_{kl} Z \right) G^{(l)} u^{(0)} \qquad\qquad G^{(k)} = \left(\sum_{l=0}^{k-1} \alpha_{kl} I + \beta_{kl} Z \right) G^{(l)} \ . \tag{4.42}$$

Setting $k = m$ gives the amplification factor for the update from time level t^n to t^{n+1}. The analytical evaluation of Eq. (4.38), which entails finding the Eigenvalues of the amplification matrix as a function of λ and ξ becomes extremely complicated for higher orders of accuracy, but can be easily accomplished numerically. First, however, we note the following proposition

Proposition 4.4.1. *The Eigenvalues of the amplification matrix are independent of the choice of (admissible) solution nodes.*

Proof. Consider two sets of solution nodes, $\mathbf{x}_j$ and $\widetilde{\mathbf{x}}_j$, where $j = 0, \ldots, m$. In both cases the solution is given as interpolation polynomials of degree m. Hence, the local solution vectors can be transformed as $\widetilde{\mathbf{u}}_i = M\mathbf{u}_i$, where M is an $(m+1) \times (m+1)$ matrix. When the scheme is written as in Eq. (4.35) it is a matter of straight forward computation to show that the iteration matrices corresponding $\widetilde{\mathbf{x}}_j$ and $\mathbf{x}_j$ are related to each other by a similarity transformation, e.g. $\widetilde{A} = M^{-1}AM$ and $\widetilde{B} = M^{-1}BM$. Hence the eigenvalues are the same for both sets of matrices, and the eigenvalues of G remain unchanged as well. $\square$

According to proposition 4.4.1 linear stability for the p_m-SD Scheme depends only on the choice of flux collocation nodes, not on the solution nodes. Here we analyze SD

(a) SD_l / JamesonRK4_s

Order	cfl	cfl · DOF
2	0.696	1.392
3	0.363	1.089
4	0.226	0.904
5	0.156	0.780
6	0.115	0.690
7	0.089	0.623

(b) SD_l / Shu-RK3

Order	cfl	cfl · DOF
2	0.595	1.190
3	0.322	0.966
4	0.201	0.804
5	0.139	0.695
6	0.103	0.618
7	0.0799	0.559

(c) DG / Shu-RK3

Order	cfl	cfl · DOF
2	0.409	0.818
3	0.209	0.627
4	0.130	0.520
5	0.089	0.445
6	0.066	0.396
7	0.051	0.357

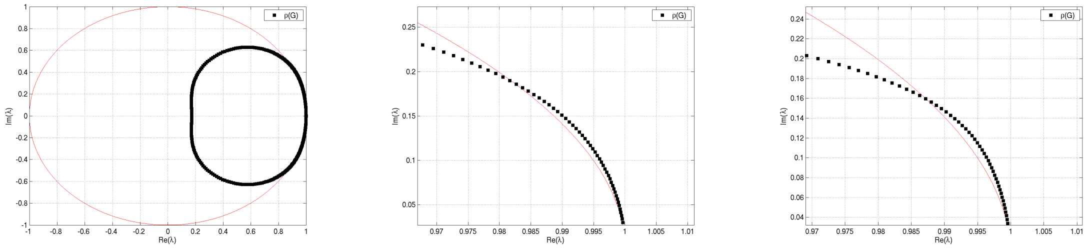
Table 4.1: Linear Stability Limits for SD_l and DG Schemes with Runge-Kutta Time Stepping.

schemes with flux nodes based on Legendre-Lobatto nodes, and Chebyshev-Lobatto nodes, denoted by p_m-SD_l and p_m-SD_c, respectively, with the Runge-Kutta schemes discussed in section 4.3. Furthermore, The SD schemes have also been compared to 2^{nd} and 3^{rd} order Discontinuous Galerkin (DG) schemes, based on Legendre polynomials. Table 4.1 shows the numerically determined stability limits for SD_l and DG schemes of up to order 7 with the Shu-RK3 scheme and the Jameson-RK4_s scheme. The DG results for order 4 through 7 have been taken from [2]. Following [2], one may argue that the low cfl values observed in Tab. 4.1 are in part a consequence of the definition of the cfl number, which does not account for the fact that an increasing number of degrees of freedom are present in each cell, as the order of approximation is increased. If one insists on comparing the cfl number to standard difference schemes with the same number of degrees of freedom, the cfl number must be multiplied with the degrees of freedom, which gives an alternate and perhaps better suited definition. This is also shown in Tab. 4.1.

A feature that appears in many combinations of the RK schemes tested with both the DG and SD schemes is marginal stability, in the sense, that the amplification factor can be bound not strictly by unity, but rather in the form

$$\rho(G) = 1 + \mathcal{O}(\mathrm{cfl}^{\mu}) \tag{4.43}$$

for some finite power μ. This, in conjunction with Eq. (4.38), means that asymptotically linear stability can only be realized by reducing the cfl number as a power of the mesh width. Fig. 4.4 shows an example of this behavior for the DG3 scheme and the p_2-SD



(a) DG3/ShuRK4$_s$ (b) DG3/ShuRK4$_s$ (Zoom) (c) SD3/ShuRK4$_s$ (Zoom)

Figure 4.4: Loci of the Eigenvalues for DG3 and SD3 with the Shu-RK4$_s$ for cfl =0.06.

Scheme with Shu-RK4$_s$ time stepping for cfl = 0.06. In 4.4(a) the loci of Eigenvalues appear to be within the unit circle for this low cfl number. However, Figs. 4.4(b) and 4.4(c) reveal that in fact a portion lies outside the unit circle by a small margin (note that the scale in these figures has been distorted for better visibility). In fact, this behavior is observed for *all* combinations of SD/DG schemes (of order greater than two) and RK schemes tested here, except the ones shown in table 4.1.

We thus wish to note two points. Firstly, it seems that the SD scheme based on the Legendre-Lobatto nodes shows better linear stability characteristics than the one based on the Chebyshev Lobatto nodes, since it is linearly stable with the Shu-RK3 and the Jameson-RK4$_s$ scheme at fixed cfl numbers (which are slightly higher than the ones that have been determined for (Legendre) DG schemes). Secondly, the linear stability of DG schemes has not been very often examined. In [2] a fixed linear stability limit is reported for such schemes of order greater than two with the Shu-RK2 scheme as well as the Shu-RK3 scheme. To the best of our knowledge this is incorrect in the case of the Shu-RK2 scheme at least for the third order scheme tested here. The amplification factor can only be bound as in (4.43), although with a high power of $\mu \approx 10$. It is believed that this result carries over to higher orders. On the other hand the stability limits for the DG3 scheme reported in [2] can be confirmed, at least for the DG2 and DG3 schemes tested here.

4.4.4 Nonlinear Stability

Consider the one-dimensional scheme for the means. The flux function is evaluated with the reconstructed left and right states at the cell interfaces as described in section 4.4.2. It was seen that the scheme for the means can be written as

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \lambda \left(h_{i+\frac{1}{2}} - h_{i-\frac{1}{2}} \right) \ . \tag{4.44}$$

We have the following proposition

Proposition 4.4.2. *Assume the numerical flux is Lipschitz continuous, and monotone, i.e. nonincreasing in its first argument and nondecreasing in its second. Furthermore, assume that the limiting procedure outlined in section 4.4.2 is utilized. Then the scheme for the means (4.44) is TVD under the condition* $\text{cfl} = \lambda\alpha \leq 0.5$, *where* $\alpha \sim f'$ *depends on the numerical flux.*

Proof. Since h is Lipschitz continuous, the scheme can be rewritten as in (2.50). A straight forward computation reveals the coefficients $C^+_{i+\frac{1}{2}}$ and $C^-_{i-\frac{1}{2}}$ with

$$C^+_{i+\frac{1}{2}} = -\lambda\alpha_1 \left\{ 1 - \frac{\widetilde{u}^-_{i+1} - \widetilde{u}^-_i}{\bar{u}_{i+1} - \bar{u}_i} \right\} \ , \qquad C^-_{i-\frac{1}{2}} = \lambda\alpha_2 \left\{ 1 + \frac{\widetilde{u}^+_i - \widetilde{u}^+_{i-1}}{\bar{u}_i - \bar{u}_{i-1}} \right\} \ , \tag{4.45}$$

where

$$\widetilde{u}^+_i = u^+_i - \bar{u}_i \ , \qquad \widetilde{u}^-_i = \bar{u}_i - u^-_i \tag{4.46}$$

with $u^\pm_i$ defined as in Eq. (4.32). Furthermore,

$$\alpha_1 = \frac{h(u^-_{i+1}, u^+_i) - h(u^-_i, u^+_i)}{u^-_{i+1} - u^-_i} \ , \qquad \alpha_2 = \frac{h(u^-_i, u^+_{i-1}) - h(u^-_i, u^+_i)}{u^+_{i-1} - u^+_i} \ . \tag{4.47}$$

The properties of the limiting procedure ensure that the expression in brackets in (4.45) is nonnegative, while the monotonicity of the flux function ensures that $\alpha_1 \leq 0$ and $\alpha_2 \geq 0$, which gives $C^+_{i+\frac{1}{2}} \geq 0$ and $C^-_{i-\frac{1}{2}} \geq 0$. Furthermore we have

$$C^+_{i+\frac{1}{2}} + C^-_{i+\frac{1}{2}} \leq \lambda K \left(\alpha_2 - \alpha_1 \right) \ , \tag{4.48}$$

where

$$K = \max\left(\left\{ 1 - \frac{\widetilde{u}^-_{i+1} - \widetilde{u}^-_i}{\bar{u}_{i+1} - \bar{u}_i} \right\}, \left\{ 1 + \frac{\widetilde{u}^+_{i+1} - \widetilde{u}^+_i}{\bar{u}_{i+1} - \bar{u}_i} \right\} \right) \ , \tag{4.49}$$

and $K \in [0,2]$ by construction of the limiting procedure. Setting $\alpha = \alpha_2 - \alpha_1$ the proposition follows from Harten's theorem (2.4.1). $\square$

Remark 4.4.1. *Note that the proof is very similar to TVDM proofs given for the DG method [7, 10], and TVD proofs other for high-resolution schemes [54]. The essential steps depend only on the conservation form and the limiting methodology.*

Fluxes that satisfy the assumptions of proposition 4.4.2 include the Godunov flux (2.54), and the Lax-Friedrichs flux (2.55), among others.

Once the TVD property is established the stability region can be extended by using Runge-Kutta schemes instead of the simple forward-Euler time discretization in Proposition 4.4.2. The m-stage Shu-RK$_s$ schemes in Eq. (4.12) are designed to preserve the TVD property under the condition $(C^+_{i+\frac{1}{2}} + C^-_{i+\frac{1}{2}}) \leq m$ [60], i.e. $\lambda\alpha \leq m/2$ in the nomenclature of proposition 4.4.2.

The TVD property is restrictive in the sense that it will reduce the order of accuracy not only at discontinuities, but also at smooth extrema. This can be avoided by utilizing the concept of total-variation boundedness (TVB) [59], which potentially deactivates the limiter at smooth extrema. The implementation merely requires a slight modification of the minmod function defined in Eq. (4.31),

$$\hat{m}(a,b,c) = \begin{cases} a & |a| < M(\Delta x)^2 \\ m(a,b,c) & \text{otherwise} \end{cases} . \tag{4.50}$$

The constant M depends on the problem, but is usually unproblematic in adjustment. Values between, say, $M = 5$ and $M = 40$ give good results with the solution not being very heavily dependent on the exact value of the constant, see section 4.7. TVB proofs are generally very similar to the ones for TVD, and basically state the allowed variation of $\mathcal{O}(\Delta x)^2$, while otherwise enforcing TVD. the details are omitted here. Many similar TVB-adjusted proofs have been published for high-resolution and high-order schemes, e.g. [9].

To pass the TVD or TVB property from the means to the entire solution one merely needs to note that in smooth monotone regions the variation $|\hat{u}_i(x) - \bar{u}_i|$ is asymptotically of $\mathcal{O}(h)$ for both methodologies. At smooth extrema and in nonsmooth regions the variation is bound by $\Delta = C_1 \min(|\Delta_+ \bar{u}_i|, |\Delta_+ \bar{u}_{i-1}|)$ in the TVD case the, by $\Delta + C_2 h^2$ in the TVB case.

Remark 4.4.2. *The condition $C^{\pm} \geq 0$ for all j is enough for the spatial operator to be a TVD operator in the semi-discrete sense. This also implies the LED property.*

4.5 Limiting Procedure in Higher Dimensions

Limiting of the high-order reconstruction is essential for shock capturing. For one-dimensional schemes results regarding TVD stability can be obtained, as in section 4.4.4, based on limiting procedures, such as the ones described in section 4.4.2. The challenge is a successful translation to an implementation on unstructured meshes for higher dimensional systems of equations, i.e. $p, d > 1$ in Eq. (4.1).

Consider first the scalar equation, i.e. $p = 1$. The basic steps of the limiting procedure, to be carried out at each time instance in each cell of the triangulation, can be identified in generic terms as

1. compute a number of local reference states
2. compare the current reconstruction to the local reference states
3. if necessary, modify the reconstruction to comply with a local maximum/minimum principle established by a predefined criterion which limits the admissible range of the reconstructed solution relative to the reference states

As an example consider the one-dimensional procedure outlined in section 4.4.2. Step one is represented by computing the volume averages in adjacent cells. The minmod function combines steps 2 and 3 by comparing the slopes of the reconstruction and reference states and returning the original reconstruction in smooth monotone regions and zero otherwise.

The main task in devising the multidimensional limiting procedure is to select the right method for each of these steps. For the local reference state the cell-averaged solutions in a local neighborhood of a cell, T_i say, may be used. Consider the quantities

$$\begin{aligned} u_{i,max} &= \max(\overline{u}_i, \max\nolimits_{k\in\mathcal{N}_i}(\overline{u}_k)) \ , \\ u_{i,min} &= \min(\overline{u}_i, \min\nolimits_{k\in\mathcal{N}_i}(\overline{u}_k)) \ , \end{aligned} \tag{4.51}$$

where $\mathcal{N}_i = \{k : T_k \cap T_i \in E\}$, and E is the set of edges of the triangulation. Alternatively a nodal neighborhood can be defined, for a node x_r, say: $\mathcal{N}_r = \{k : x_r \text{ is a node of } T_k\}$. In this case the reference state becomes

$$\begin{aligned} u_{r,max} &= \max_{k\in\mathcal{N}_r}(\overline{u}_k) \ , \\ u_{r,min} &= \min_{k\in\mathcal{N}_r}(\overline{u}_k) \ . \end{aligned} \tag{4.52}$$

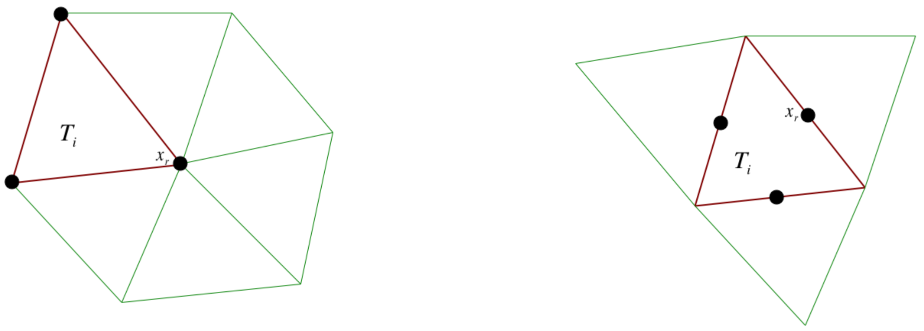
Fig. 4.5 shows a schematic depiction of the local neighborhoods defined in Eq. (4.52). and (4.51). Step 1 is thus complete. Step 2 depends to some extend on how the reference state has been chosen and the way in which the solution is compared to it. In [12], a multidimensional limiting method for high-order schemes has been proposed. The authors suggest to use the linear portion of the reconstruction for comparison, under the assumption that limiting is only necessary if it is necessary for the linear restriction. In a slightly more general fashion, one may define the quantity Δu at new collocation nodes, x_r, $r = 1, 2, 3$, which we introduce in anticipation of a linear reconstruction for the case that it should be necessary to limit the reconstruction:

$$\Delta u(x_r) = (\mathcal{P}(u_i)(x_r) - \overline{u}_i) \ , \qquad r = 1, 2, 3 \ . \tag{4.53}$$

Here $\mathcal{P}$ may be the projection onto the space of linear function or the identity if the full reconstruction is used in the comparison. If a cellwise neighborhood has been used in step 1 the nodes x_r may be chosen at edge midpoints, whereas for the nodal neighborhood the nodes of the triangle are used. This is also depicted in Fig. 4.5. If one insists that the reconstructed solution must not exceed the bounds established by the maximum and minimum volume averages of the chosen local neighborhood, the limiter will be activated whenever the magnitude of Δu exceeds the magnitude of the allowed variation at any of the nodes x_r. Note that in the case of Fig. 4.5(b), i.e. Eq. (4.52), the local reference state is different for each node x_r, while it is the same for all x_r for Eq. (4.51).

The quantity $\phi(x_r)$ may be defined as

$$\phi(x_r) = \begin{cases} 1 & \Delta u^- \leq \Delta u(x_r) \leq \Delta u^+ \\ \frac{\Delta u^{\text{ref}}}{\Delta u(x_r)} & \text{otherwise} \end{cases} \ , \tag{4.54}$$



(a) Eq. (4.51) (b) Eq. (4.52)

Figure 4.5: Local neighborhoods for slope comparison and new collocation nodes for linear reconstruction.

where

$$\Delta u^+ = u_{i,\max} - \overline{u}_i \ , \qquad \text{or} \qquad \Delta u^+ = u_{r,\max} - \overline{u}_i \ , \tag{4.55}$$

$$\Delta u^- = u_{i,\min} - \overline{u}_i \ , \qquad \text{or} \qquad \Delta u^- = u_{r,\min} - \overline{u}_i \ , \tag{4.56}$$

depending on whether Eq. (4.51) or (4.52) is used, and

$$\Delta u^{\mathrm{ref}} = \begin{cases} \Delta u^+ \ , & \Delta u(x_r) > 0 \\ \Delta u^- \ , & \Delta u(x_r) < 0 \end{cases} \ . \tag{4.57}$$

Whenever $\phi < 1$ the local reconstruction may be modified according to

$$\hat{u}_i(x) = \overline{u}_i + \mathcal{I}^1(\widetilde{\Delta u})(x) \ , \tag{4.58}$$

where $\mathcal{I}^1$ is the linear interpolation operator defined by the nodes x_r, and

$$\widetilde{\Delta u(x_r)} = \phi(x_r)\,(\mathcal{P}(u_i)(x_r) - \overline{u}_i) \ . \tag{4.59}$$

The procedure is sufficient to ensure that the reconstruction stays within the prescribed bounds, but it is not conservative in the sense that in general

$$\int_{T_i} \widetilde{u}_i \, dx \neq \int_{T_i} u_i \, dx \ . \tag{4.60}$$

A slight modification of the linear solution is proposed in [12], which ensures conservation in a way that prevents an increase of the slope. We adopt this strategy for the present work, omitting the technical details for the sake of brevity, and refer instead to [12].

Furthermore, a modification may be introduced by rejecting any limiting whenever $\Delta u < Mh^2$, where h is a characteristic local mesh length, for instance the distance $\|x_r - x_c\|$, where x_c is the centroid of the cell. This will prevent limiting of the reconstruction near smooth extrema and is related to the theory of total-variation boundedness (TVB) [10], as outlined in section 4.4.4. In practice values between $M = 5$ and $M = 40$ usually give good results.

The extension to systems is straight forward. In principle a characteristic decomposition could be carried out, and the limiting could be applied to the characteristic variables, which has the advantage that the 1D stability results for the scalar equation carry over to 1D systems. For higher dimensions, however, these results do not hold, and the additional expense of a characteristic decomposition seems unwarranted. The procedure is thus applied in a component-wise fashion.

4.6 h/p-Multigrid Formulation

An outline of a generic Multigrid procedure using a sequence of meshes has been given in section 3.5.2 in the context of kinetic finite-volume schemes. The extension to higher order spatial approximations can be cast into a similar framework.

Assume that a step with any of the relaxation techniques discussed in section 4.3 has been carried out using the p_m-SD scheme on a mesh of characteristic length h. Such a mesh/accuracy combination is combined in a tuple (p_m, h), which is simply called a "level". A strategy of transferring the solution to successively lower orders of accuracy and/or successively coarser meshes needs to be defined, in order to compute corrections for the solution, as outlined in section 3.5.2.

Define a discrete solution operator that on the highest level (p_m, h) is simply given by

$$\mathcal{L}_h^{p_m}(u_h^{p_m}) = 0 \; . \tag{4.61}$$

One has several orders of approximation p_n, p_{m-1}, p_{m-2}, …, p_0 to work with as well as the possibility of using coarser meshes $h, 2h, 4h, \ldots$ [3]. One possibility is to transfer the solution to successively lower orders of accuracy on the finest mesh, using a modified relaxation step on the lower level, analogous to Eq. (3.39):

$$\mathcal{L}_h^{p_{m-1}}(u_h^{p_{m-1}} + \delta u_h^{p_{m-1}}) + I_{p_{m-1}}^{p_m} R_h^{p_m} - R_h^{p_{m-1}} = 0 \; . \tag{4.62}$$

Once p_0 has been reached, the solution is transferred to successively coarser meshes, where the coarse mesh modified solution operator, Eq. (3.39), is used to advance the solution, while the order of approximation is kept at p_0. After the coarsest mesh has been reached, the corrections are interpolated to successively finer meshes using Eq. (3.40). Similarly, after the finest mesh has been reached one has for the corrections that are transferred to higher orders of accuracy:

$$u_h^{+,p_m} = u_h^{p_m} + I_{p_m}^{p_{m-1}}(u_h^{+,p_{m-1}} - u_h^{p_{m-1}}) \; . \tag{4.63}$$

One may depict this strategy in a V-shaped diagram, as shown in Fig. 4.6. The curved arrows indicate corresponding mesh levels for which a correction is computed from the lower levels on the way up to the highest level. Note that no relaxation step is carried out on the way up. Alternatively, one may use a so-called W-cycle: A correction is only interpolated to the next higher level if the current level has already been visited twice,

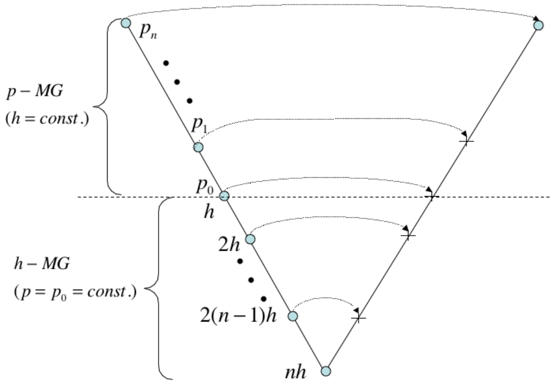


Figure 4.6: h/p-Multigrid V-Cycle

---

[3] The notation 2^nh where n is an integer does not imply that the characteristic mesh length has to be exactly doubled at each step. It is merely a convenient notation.

otherwise it is advanced on the current level, and the solution is transferred to a lower level again. For a three-level strategy this leads to a W-shaped cycling diagram, hence the name.

The best cycling strategy is not obvious. For hyperbolic equations multigrid may be viewed as an effective increase of the wave speed at which information is propagated on the way to reaching a steady-state. A coarser mesh connects information from larger portions of the computational domain, and furthermore, for the same cfl number, the physical time step is larger on a coarser mesh. Against this background it is questionable whether the best strategy is to go down the accuracy ladder step-by-step. It must be stated quite clearly that any utilization of higher order reconstruction on levels below (p_m, h) is wasted in the sense that the steady-state solution is not affected by it. The only matter of concern is convergence. Any cycling strategy, and any order of approximation used on a lower level is acceptable if and only if it speeds up convergence. In light of these thoughts it is anticipated here that a minimum level of polynomial accuracy p_m with $m > 1$ exists, for which a direct transfer to p_0 is carried out. In section 4.7 validation is presented for the p_2-SD Scheme which omits the level p_1 in the multigrid cycle.

4.7 Results for Hyperbolic Problems: Inviscid Fluid Flow

Consider first isentropic flow through a nozzle, governed by the quasi-1D Euler equations, see section 3.6.1, Eqns. (3.43) and (3.44). The nozzle exit Mach number is set to $M_{\text{out}} = 0.3$, which leads to subsonic flow with a maximum Mach number of $M \approx 0.46$ in the throat. The nozzle cross-section and the boundary conditions have been chosen as in section 3.6.1. Table 4.2 shows the steady-state entropy production in mesh (h-) refinement for the p_m-SD Scheme and various m. Here and in the following the entropy production is defined as deviation from the freestream entropy,

$$\Delta s = p/p_\infty(\rho_\infty/\rho)^\gamma - 1 \ . \tag{4.64}$$

For shock-free inviscid flow the solution should approach $\Delta s = 0$. It can be seen that the p_m-SD Scheme converges at the nominal rate for all m in mesh refinement until machine zero is approached. Since the flow is smooth no limiters are necessary, although it has been verified that their use does not compromise the accuracy if a TVB condition is imposed. This is illustrated in Fig. 4.7, where the entropy production for the 3^{rd} order scheme is shown for the TVD limiter and the TVB modification (see sections 4.4.2 and 4.4.4). It can be clearly seen that strictly enforcing the TVD property leads to a local reduction of accuracy at the nozzle throat, and also contaminates the solution downstream of the throat. On the other hand, the TVB limiter with $M = 40$ in Eq. (4.50) produces lower entropy levels and avoids the spike at the nozzle throat, enabling convergence at the nominal rate, shown in table 4.1(b). While both the CUSP and Roe flux have found use in this study, the results do not depend on the choice of flux function. Fig. 4.8 demonstrates the convergence characteristics of the scheme with explicit and implicit methods. The LU-SGS scheme is a rather efficient smoother and shows good convergence characteristics. The explicit Runge-Kutta time stepping is restricted by relatively low cfl numbers. It is significantly cheaper per time step, although not quite enough to become more efficient than the LU-SGS scheme (see Fig. 4.8(b), which plots the density residual against time). The implicit scheme can be run at any cfl number, but for cfl numbers above approximately cfl = 500, there is no significant speedup, and the scheme becomes essentially a Newton iteration.

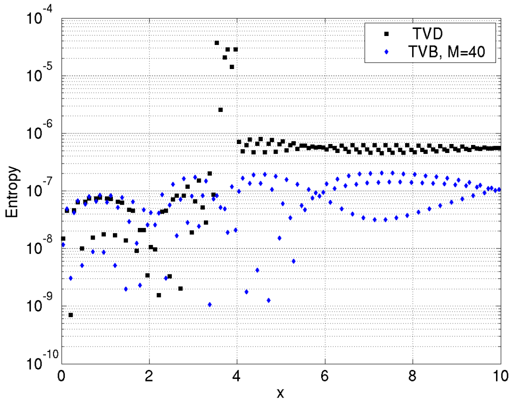


Figure 4.7: TVD vs. TVB limiter in the computation of smooth nozzle flow using the 3^{rd} order p_2-SD scheme.

(a) 2^{nd} order scheme, p_1-SD

Cells	L_∞(Error)	Order	L_2(Error)	Order
10	7.736700e-05		7.736700e-05	
50	1.868200e-06	2.313593e+00	1.868200e-06	2.313593e+00
100	4.968100e-07	1.910883e+00	4.968100e-07	1.910883e+00
1000	5.233300e-09	1.977415e+00	5.233300e-09	1.977415e+00
10000	5.209600e-11	2.001971e+00	5.209600e-11	2.001971e+00

(b) 3^{rd} order scheme, p_2-SD

Cells	L_∞(Error)	Order	L_2(Error)	Order
10	1.328636e-05		6.918978e-06	
20	1.693499e-06	2.971867e+00	8.735518e-07	2.985594e+00
50	1.027596e-07	3.058156e+00	5.234414e-08	3.071872e+00
100	1.238252e-08	3.052895e+00	6.261137e-09	3.063532e+00
500	9.544221e-11	3.023118e+00	4.805796e-11	3.025721e+00
1000	1.181044e-11	3.014565e+00	5.943157e-12	3.015474e+00

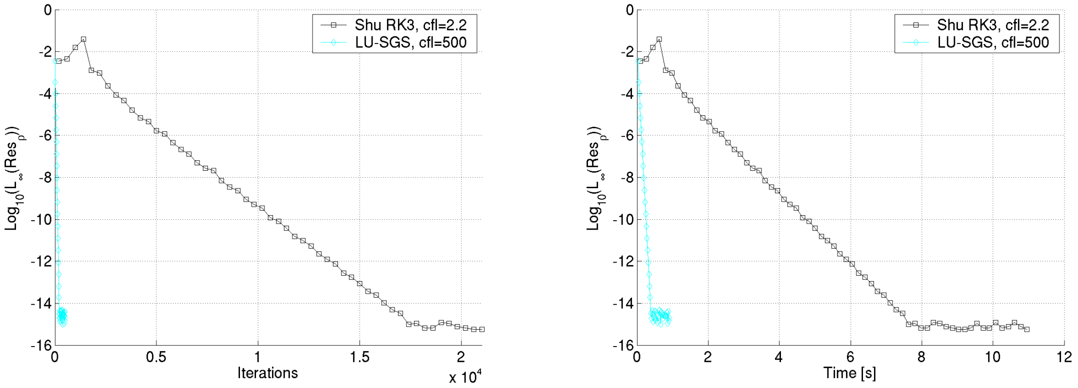
(c) 4^{th} order scheme, p_3-SD

Cells	L_∞(Error)	Order	L_2(Error)	Order
10	1.069145e-06		4.082450e-07	
20	9.141652e-08	3.547858e+00	2.796467e-08	3.867758e+00
30	1.828542e-08	3.969077e+00	5.562735e-09	3.982715e+00
50	2.165135e-09	4.176809e+00	8.162414e-10	3.756928e+00
100	1.502647e-10	3.848879e+00	5.302680e-11	3.944202e+00
200	9.763967e-12	3.943895e+00	3.599578e-12	3.880822e+00

(d) 5^{th} order scheme, p_4-SD

Cells	L_∞(Error)	Order	L_2(Error)	Order
6	2.368234e-06		8.662295e-07	
10	1.734733e-07	5.116963e+00	5.265576e-08	5.482055e+00
20	5.942866e-09	4.867411e+00	1.614429e-09	5.027496e+00
40	2.162333e-10	4.780499e+00	7.421668e-11	4.443136e+00
50	7.531975e-11	4.726174e+00	2.842774e-11	4.300476e+00
100	5.998313e-12	3.650400e+00	1.224929e-12	4.536530e+00

Table 4.2: SD Scheme: Convergence in mesh refinement for quasi-1D nozzle flow.



(a) LU-SGS vs. Shu-RK3 (Residual vs. Iteration)

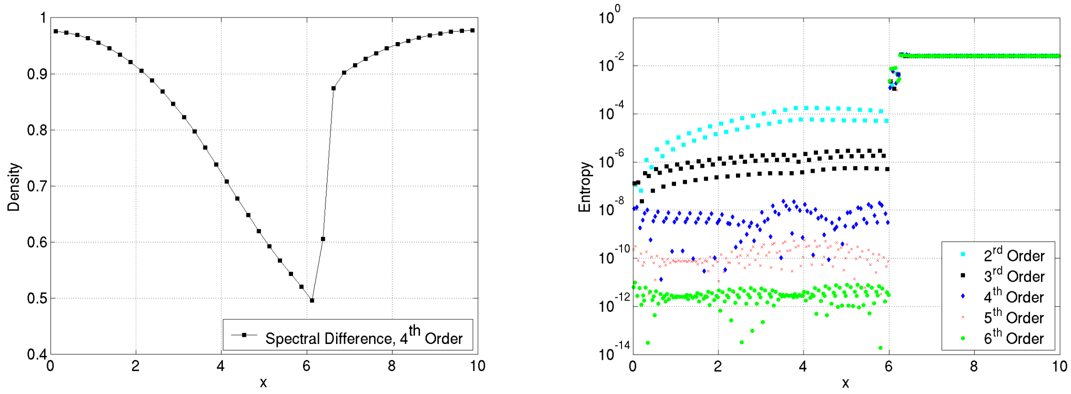
(b) LU-SGS vs. Shu-RK3 (Residual vs. Time)

Figure 4.8: Convergence for smooth inviscid flow using the 5th order p_4-SD scheme on a mesh with 10 cells with implicit and explicit relaxation techniques. Density residuals.

Consider now the case of shocked nozzle flow. An exit Mach number of $M = 0.6$ has been chosen, which leads to a shock in the diverging part of the nozzle. Fig. 4.9(a) shows the solution in terms of the density volume averages for the 4th order scheme on a mesh with 40 cells. Fig. 4.9(b) shows a plot of the pointwise entropy production, as defined in Eq. (4.64), for schemes of various orders. Note that all degrees of freedom are shown in this plot. Behind the shock there is a constant entropy production, fixed by the shock jump conditions, while in front of the shock it decreases rapidly with the order of the scheme. It can be seen that the jump is very cleanly captured, within one mesh cell for all orders. The CUSP flux has been used for all these transonic computations, along with a TVB limiter.

The natural limit of the SD scheme, when operating on only one cell, is a nodal spectral method on a staggered grid. It can be demonstrated that spectral convergence can be attained for a fixed mesh and increasing polynomial accuracy (p-refinement). Fig. 4.10 shows such a refinement study for a mesh with 16 cells and polynomial orders $m = 2, \ldots, 15$. For this case a nozzle geometry on the interval $[0, 1]$ has been defined, where on $[0, 1/2]$ the nozzle area is given by

$$A(x) = \begin{cases} 1 & , \ x \leq 0.375 \\ 1 - \left(a_1\xi^5 + a_2\xi^6 + a_3^7 + a_4\xi^8 + a_5\xi^9\right) & , x > 0.375 \end{cases} . \tag{4.65}$$



(a) Density solution for 4th order scheme.

(b) Entropy Production.

Figure 4.9: SD scheme: Shocked nozzle flow.

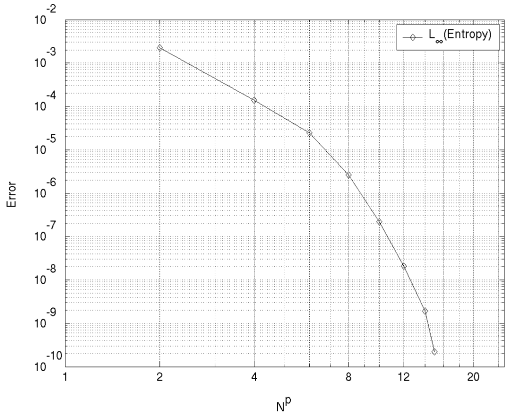
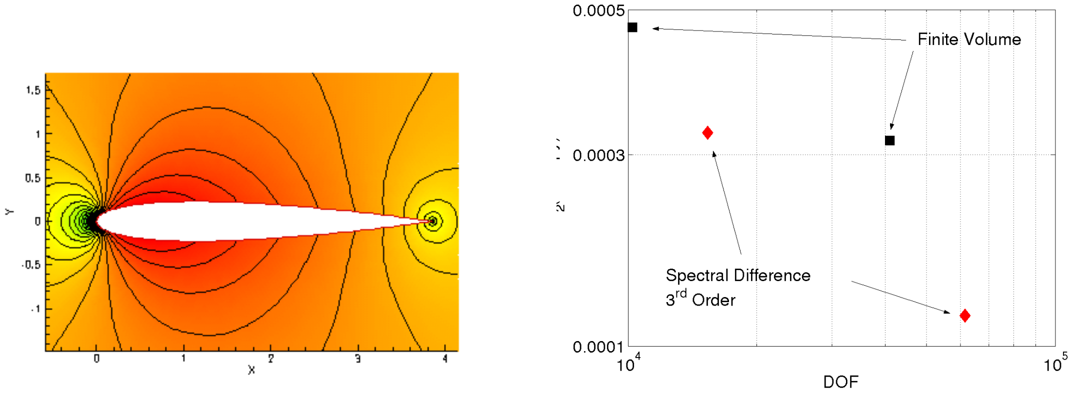


Figure 4.10: Maximum norm of the steady-state entropy production for isentropic nozzle flow in p-Refinement.

Here $\xi = (x - 0.375)/0.125$ and the area on the interval $[1/2, 1]$ is obtained by symmetry. Obviously the first four derivatives of the polynomial expression for the nozzle throat vanish at the intersection with the constant part of the geometry definition. The remaining coefficients are chosen such that the nozzle throat is given by $d_t = 0.85$, and all derivatives up to order four correspond to a circle with radius $r = 1 - d_t$ at the throat. The reason for this geometry definition is increased smoothness, compared to the definition in Eq. (3.45), which is beneficial for the higher orders of approximation, encountered in this spectral convergence study.

We now shift our attention to the 2D Euler equations. A first order finite-volume scheme is the natural low-order limit of the SD scheme. However, when run in first order (p_0) mode, an additional piecewise linear reconstruction has been added in the 2D implementation, which actually makes the method identical to the finite-volume method described in section 3.6.2, where it has been compared to the BGK scheme (see also [38]). In this way one can accommodate a standard finite-volume solver and the high-order SD scheme in the same code without much computational complexity. This finite-volume method has been used for comparison in this section. Normally, one would label such a finite-volume method 2$^{\text{nd}}$ order accurate. However, to distinguish it from the 2$^{\text{nd}}$ order accurate p_1-SD Scheme, and keeping in mind that often times the accuracy of such solvers is slightly below 2$^{\text{nd}}$ order, the finite-volume solver mode is labeled $p_{0.7}$ here. This terminology is also useful in identifying multigrid levels.

Consider inviscid subsonic flow around the NACA0012 airfoil. Fig. 4.11 shows results for the 3$^{\text{rd}}$ order p_2-SD Scheme at flow conditions $M = 0.3$ and zero angle of attack. In Fig. 4.11(a) the solution is displayed in terms of Mach number contours, while Fig. 4.11(b) compares the RMS value of entropy production on the airfoil surface between the SD scheme and the finite-volume scheme described above. The data is plotted against total degrees of freedom (DOF). For the third order scheme two meshes with $2,560$ and $10,240$ triangles have been used, which corresponds to $15,360$ and $61,440$ degrees of freedom. For the finite volume scheme meshes with $10,240$ and $40,961$ triangles have been used, where the number of triangles is equal to the degrees of freedom. Note that the entropy error in Fig. 4.11(b) for the SD scheme with $15,360$ DOF is roughly equal to the entropy error



(a) Mach number contours

(b) RMS value of the entropy production on the airfoil surface

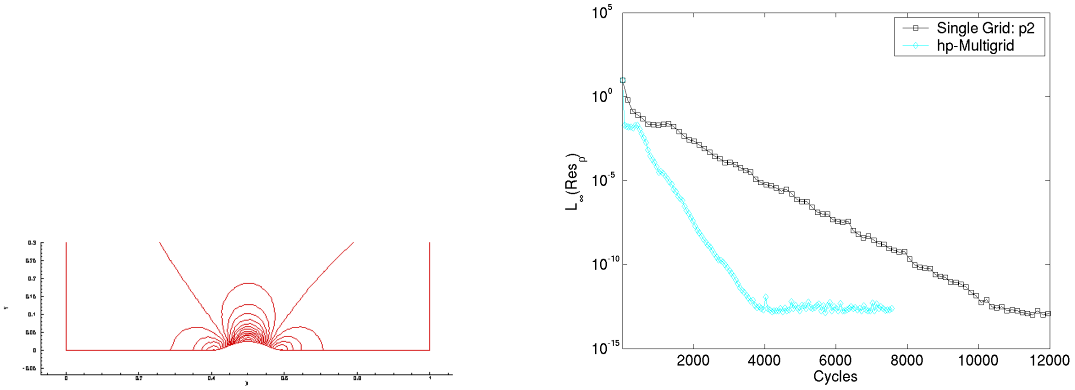
Figure 4.11: Subsonic flow around the NACA0012 airfoil: 3^{rd} order SD scheme vs. Finite Volume for various mesh sizes and degrees of freedom (DOF).

of the finite-volume scheme with approximately $40,000$ DOF. The CUSP flux along with Shu-RK3 time stepping has been used for this testcase.

We now turn to the h/p-multigrid method outlined in section 4.6. A validation for SD schemes of order up to three is presented. The cycling strategy omits the p_1 level, i.e. the second order SD stage, as discussed in section 4.6. Instead, the solution is transferred directly to the p_0 level, and subsequently to coarser meshes. On the finest mesh, the p_0 level is augmented by a piecewise linear reconstruction, whereas on coarser meshes, a first-order finite-volume methodology is employed. Three levels of h-multigrid have been used. According to the nomenclature discussed in section 4.6, the cycling strategy is thus

$$(p_2, h) \longrightarrow (p_{0.7}, h) \longrightarrow (p_0, 2h) \longrightarrow (p_0, 4h) \ .$$

The transfer coefficients (see the nomenclature in section 4.6) between the first two levels are given simply by the volume average for the restriction to the lower level, and by the unity operator in the prolongation to the higher level. For the transfer between h-levels a volume-weighted interpolation may be used, which can be precomputed and saved for all mesh levels. The results shown here use the Jameson RK4_s scheme. Fig. 4.12(b) shows the convergence of the maximum density residuals for inviscid flow over a bump using



(a) Density Contours

(b) Max. Density Residuals

Figure 4.12: Inviscid subsonic flow over a bump: Density solution and convergence for single grid and h/p-multigrid.

(a) Bump

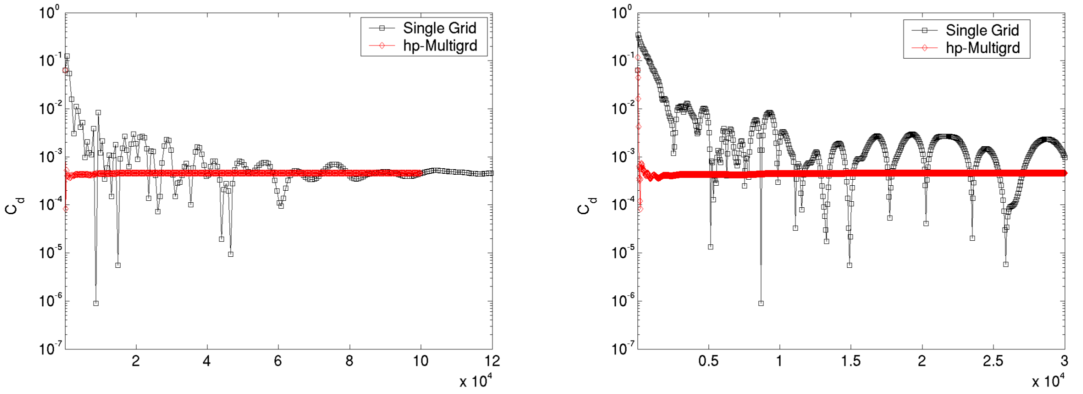
Level	DOF	p	Cells	cfl
4	4872	2	812	1
3	812	0.7	812	6
2	210	0	210	6
1	42	0	42	6

(b) NACA0012

Level	DOF	p	Cells	cfl
4	74472	2	12412	1
3	12412	0.7	12412	6
2	2917	0	2917	6
1	661	0	661	6

Table 4.3: Meshes and Degrees of freedom (DOF) used with h/p-multigrid.

this h/p-multigrid strategy. The solution in terms of the density contours can be seen in Fig. 4.12(a). The cycling strategy is summarized in Tab. 4.2(a), showing polynomial order, and degrees of freedom for each level. The speedup due to the multigrid methodology is clearly visible. While one may expect higher reduction rates of the residual from best-practice finite-volume solvers, one has to keep in mind that the fine grid cfl number is substantially lower for the high-order SD scheme, compared to standard finite volume techniques. Here the cfl number on the finest grid is cfl $= 1$. As can be seen in table 4.2(a) it is possible to use substantially higher cfl numbers (cfl $= 6$) on the levels below (p_2, h), i.e. the finite-volume levels. This is so, because, in addition to multigrid, residual smoothing is used, allowing an increase of the cfl number. The increase in cfl number on lower levels



(a) C_d vs. cycles (b) Close-up view

Figure 4.13: Convergence of the drag coefficient for inviscid flow around the NACA0012 airfoil using the 3^{rd} order SD scheme.

contributes significantly to convergence acceleration.

The speedup provided by the multigrid methodology is best demonstrated in the convergence of integrated quantities, such as drag coefficients (see also the discussion in section 3.6.3). Fig. 4.13 demonstrates this quite clearly. Here the convergence of the drag coefficient is shown for subsonic flow around the NACA0012 profile with the same flow conditions as in Fig. 4.11. The convergence for this mesh with roughly $75,000$ degrees of freedom is painfully slow for single-grid computations with explicit time stepping. For the h/p-Multigrid method using the sequence of levels shown in Tab. 4.2(b), the convergence of the drag coefficient can be accelerated by approximately two orders of magnitude. The use of coarser meshes along with the increased cfl number on finite-volume levels, speeds up the propagation of bulk information dramatically, so that the convergence of the actual flow solution progresses even faster than can be seen from the reduction rate of the residuals.

Figure 4.14 compares the convergence of the drag coefficient with a finite-volume method, which uses pure h-multigrid, obtained by simply omitting the highest level in the h/p-multigrid strategy used with the SD method, see Tab. 4.2(b). The cfl number for the finite-volume method has been set to cfl $= 1$ on the finest grid. This is certainly well below the maximum possible cfl number, but allows better assessment of the effect of multigrid by using the same cfl number on the highest level for both solver modes. It can be seen the convergence is very similar, despite the fact that the SD-h/p-multigrid method adds the third order scheme on top of the 3-level mesh sequence used by the finite-volume method. Also note that the value of the drag coefficient, which should be zero for two-dimensional smooth inviscid flow, is significantly lower for the third-order method, compared to the finite-volume scheme. We have $c_d = 0.00046$ for the p_2-SD Scheme and $c_d = 0.0015$ for the finite-volume scheme.

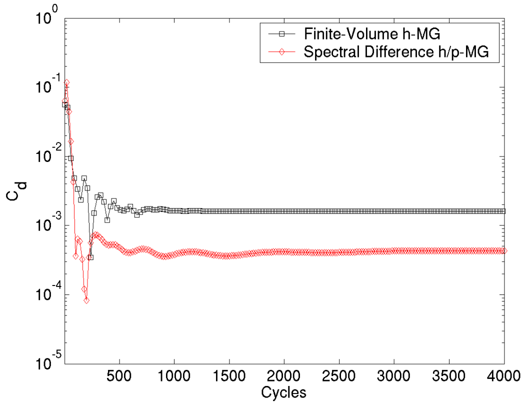
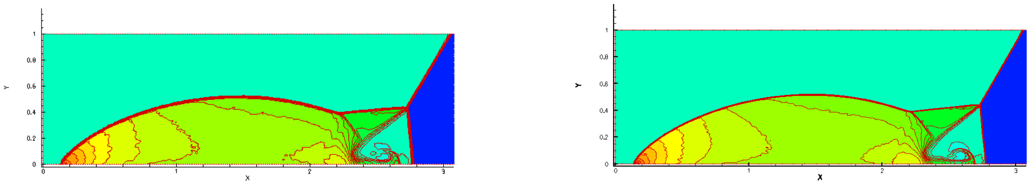


Figure 4.14: Convergence for steady inviscid flow around the NACA0012 airfoil. Comparison between p_3-SD h/p-Multigrid and finite-volume h-multigrid.

We now turn to 2D shocked flow, in order to demonstrate the viability of the limiting procedure outlined in section 4.5. A well known testcase for unsteady gas dynamics with shockwaves and contact discontinuities is the shock reflection case studied by Woodward and Colella [71], and subsequently many other researchers. Figure 4.15 shows results for the third-order SD scheme at time $t = 0.2$ in terms of density contours for two different triangular meshes. It can bee seen that shocks are very well captured. For a better illustration of the limiting procedure, consider Fig. 4.16, which corresponds to the numerical solution in fig. 4.15(a). The plot shows contour lines of the values of the limiting variable ϕ, defined in Eq. (4.54). The values are clipped above $\min_l \phi_l = 0.99$ where the l stands for the l^{th} conserved variable. This effectively shows the regions of active limiters, which can be seen to be largely confined to the vicinity of the shock and the contact discontinuity in



(a) Contour lines of the density. Mesh: 133,480 Triangles

(b) Contour lines of the density. Mesh: 533,920 Triangles

Figure 4.15: The shock reflection testcase using the third order Spectral Difference scheme.

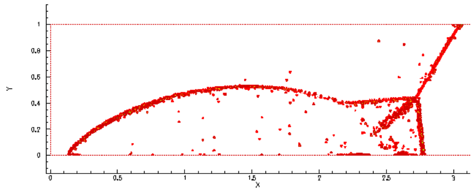


Figure 4.16: Shock reflection case: Cells with active limiters.

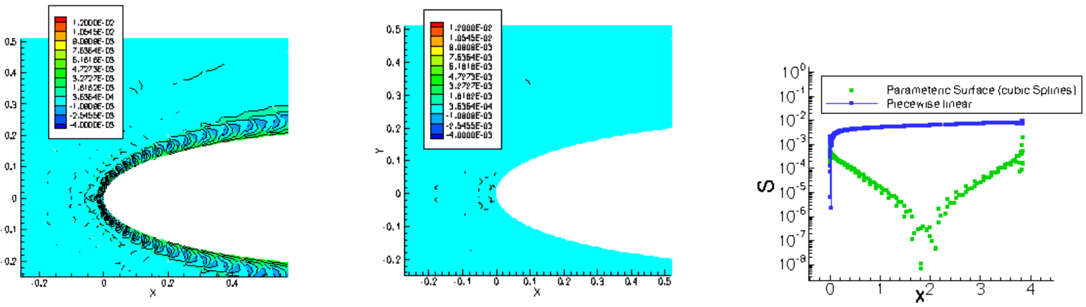
the solution.

4.7.1 Practical Details

The Spectral Difference Scheme has been at the center of a new high-order flow solver architecture that was initiated as part of this research. Although it is the primary goal of this thesis to analyze the scheme, as well as evaluate and validate fundamental contributions, some practical aspects are worthy of consideration.

Parametric Surface Representation

Firstly consider the approximation of curved surfaces. A piecewise linear approximation is standard in second order accurate numerical schemes, which is sufficient, because such a scheme cannot tell the difference between the actual surface and a piecewise linear approximation. For higher-order schemes, however, such a simplified surface representation can have disastrous effects. This is demonstrated in Fig. 4.17, which analyzes spurious entropy production for inviscid subsonic flow ($M_\infty = 0.3$) around the NACA 0012 airfoil at zero



(a) Piecewise linear

(b) Parametric (cubic splines)

(c) Entropy error along the airfoil surface

Figure 4.17: Entropy production for the 3rd order SD Scheme and different Methods of Surface Representation for smooth inviscid flow around the NACA0012 airfoil.

angle of attack. In 4.17(a) contour lines of the entropy error are shown for a piecewise linear surface approximation computed with the 3rd order SD scheme, while in 4.17(b) the same contour plot is shown for a parametric surface representation using cubic splines and 80 knots (i.e. the original surface points given as input parameters). The spurious entropy production for the piecewise linear approximation is clearly visible and is even more obvious in Fig. 4.17(c), where the entropy error is plotted along the airfoil on a logarithmic scale. For the parametric surface treatment there is significant entropy production only at the leading edge stagnation point and at the sharp trailing edge. Otherwise the entropy error is practically negligible. For the piecewise linear surface representation, however there is significant entropy production along the entire airfoil. The reason for this behavior is that if a piecewise linear airfoil is specified, a high order numerical scheme will compute precisely that, a flow over a piecewise linear surface, because it is able to discern the difference. This can dramatically deteriorate the quality of the solution.

We use a parametric surface representation for all curved surfaces. This means that some cells will have curved boundaries, which has consequences for the storage requirement of the scheme, as differentiation matrices and other data relevant to the solution algorithm have to be stored for all elements individually. See appendix C.2 for details.

Adaptive Meshing

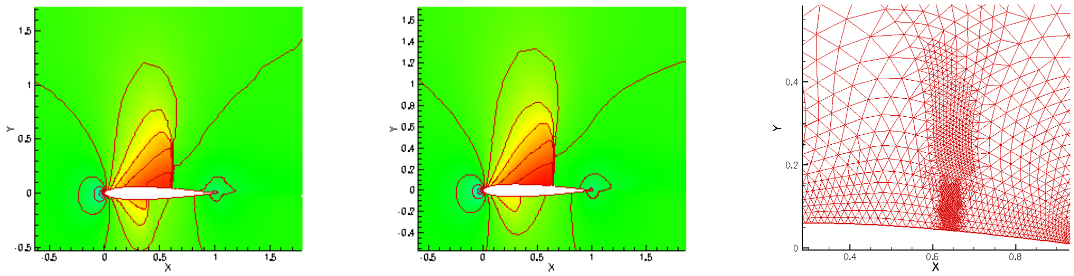
To replace mesh resolution by increased order of approximation is the main rationale behind using high-order numerical methods. However, for hyperbolic equations, one has to admit discontinuous solutions, for example shocks in gas dynamics. A significant challenge is thus introduced: If a coarser mesh is used to reap the benefits of a high-order numerical approximation, what is to become of the shock capturing capabilities? Putting all questions of robustness and stability aside, it is evident that shock resolution must necessarily deteriorate if the limiting procedures outlined in section 4.5 are used, as the shock thickness will be determined by the local characteristic mesh length. Adaptive meshing is a possible solution, and particularly attractive in conjunction with high-order approximations when shock waves have to be expected. One may use this technique to locally refine the mesh near a discontinuity in order to improve shock capturing capabilities.

Furthermore, local limiting procedures are rather intrusive for high-order methods, where a local high-order reconstruction is usually replaced by a carefully limited linear one (see section 4.5). For large mesh elements this will be an extremely poor approximation and at the same time a very massive change in the solution representation, likely to lead to limit cycles and impede convergence. For smaller elements near a discontinuity, on the other hand, the linear approximation will be more fitting.

The solution-adaptive mesh refinement used in the present work has been patterned after the methodology documented by Kim et al. [40], and we refer the interested reader to their paper instead of reproducing the technical details. The criterion for adaptation are gradients in the solution. We have used both gradients in the pressure and the entropy.

As a test of the solution-adaptive procedure, consider transonic flow over a NACA 0012 airfoil. A rather popular testcase for inviscid aerodynamics is given by the flow conditions $M = 0.8$ at an angle of attack $\alpha = 1.25^o$. Fig. 4.18 shows contour lines of the Mach number for this test case for a finite-volume scheme and the p_2-SD Scheme. Both solutions have been computed on the same mesh, which can be seen in Fig. 4.18(c).

Although not used here, adaptive coarsening may also prove useful in conjunction with a locally increased order of approximation in smooth regions to reduce the computational



(a) Contour lines of the Mach number for the third-order SD scheme

(b) Contour lines of the Mach number for the second order Finite-Volume scheme

(c) Close-up view of refined area on the suction side

Figure 4.18: The NACA 0012 airfoil in transonic flow. $M = 0.8$, $\alpha = 1.25$.

cost.

4.8 Viscous Discretization

The Spectral Difference Method is designed specifically for hyperbolic conservation laws. However, many physical phenomena, even if they are dominated by wave propagation, have dissipative features as well. While those are small for many problems, and the resulting equations are still nearly hyperbolic, viscous effects often cannot be neglected. A suitable way of treating dissipative terms must be found, while we would like to keep the treatment of the convective terms intact.

Consider first the one-dimensional linear advection-diffusion equation:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2} \ . \tag{4.66}$$

One way of solving Eq. (4.66) is to rewrite it as a system:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \nu \frac{\partial q}{\partial x} \tag{4.67}$$

$$\frac{\partial u}{\partial x} - q = 0 \ . \tag{4.68}$$

This general approach has been taken by many researchers for viscous discretization in a Discontinuous Galerkin context [3, 11, 47].

If the discretization of the variable q is "local", i.e. it depends only on the latest solution of u, the system can be solved in a segregated manner. This means that the approximation for the gradient q can be computed first, using the latest estimate of u, and a suitable discretization of Eq. (4.68). Subsequently this approximation is used in Eq. (4.67) to advance the solution u. Consider Eq. (4.67) of the new system in semi-ODE form and p_m-SD discretization, i.e. using polynomials of order m, on a uniform mesh, as in section 4.4:

$$\Delta t \frac{du_{ij}}{dt} + \lambda \sum_{k=0}^{m+1} d_{jk} f_{h,ik} \ , \tag{4.69}$$

where $\lambda = \Delta t/\Delta x$, the d_{jk} are the differentiation coefficients defined in Eq. (4.21), and $f_h = f_{h,c} - f_{h,v}$. The flux functions $f_{h,c}$ and $f_{h,v}$ discretize the convective and dissipative terms, $f_c = cu$ and $f_v = \nu q$, respectively. For the convective part one may choose an upwind discretization:

$$f_{h,c} = c\, v_{ik} = c \begin{cases} p_{ik} & k \neq 0 \\ p_{i-1,m+1} & k = 0 \end{cases} \ , \tag{4.70}$$

where p_k is the reconstructed solution defined as in Eq. (4.20). The concepts of wave propagation that guide the discretization of the convective terms do not apply to dissipation, which is not associated with a preferred direction of propagation. In principle a central discretization of dissipative terms is therefore appropriate, and has (in one form or another) often been applied to both scalar equations and systems such as the Navier-Stokes equations. However, stability concerns have to be addressed. We define a more general weighted discretization of the form

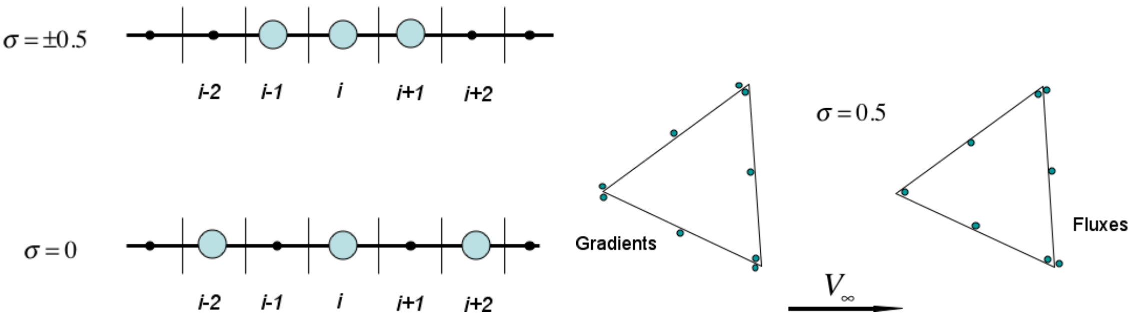
$$f_{v,k} = \frac{\nu}{\Delta x} \widetilde{q}_k \ , \tag{4.71}$$

where

$$\widetilde{q}_k = \Delta x \cdot \begin{cases} \frac{1}{2}\left(q_{i-1,m+1} + q_{i,0}\right) - \sigma\left(q_{i-1,m+1} - q_{i,0}\right), & k = 0 \\ q_{ik}, & k \neq 0, m+1 \\ \frac{1}{2}\left(q_{i,m+1} + q_{i+1,0}\right) - \sigma\left(q_{i,m+1} - q_{i+1,0}\right), & k = m+1 \end{cases} \ . \tag{4.72}$$

Using $\mathrm{cfl} = \lambda c$, and introducing the cell Peclet number $\mathrm{Pe} = c\Delta x/\nu$, the scheme becomes

$$\Delta t \frac{du_{ij}}{dt} + \mathrm{cfl} \sum_{j=0}^{m} d_{jk} \left(v_{ik} - \frac{\widetilde{q}_{ik}}{\mathrm{Pe}} \right) \ . \tag{4.73}$$



(a) First order SD discretization

(b) 2D discretization of the Navier-Stokes viscous terms

Figure 4.19: Spectral Difference discretization stencils for viscous terms.

The remaining task of discretizing the second equation of the new system, Eq. (4.68), is equivalent to precomputing an estimate of the gradient if the scheme is local. In the same spirit as Eq. (4.71), a weighted discretization of the form

$$q_k = \frac{1}{\Delta x} \sum_{l=0}^{m+1} \mu_{kl} \widetilde{p}_l \tag{4.74}$$

is introduced, where μ_{kl} are elements of a differentiation matrix, evaluating the derivative of $\widetilde{p}$ at the flux nodes x_k, i.e. $\frac{1}{\Delta x} \sum_{l=0}^{m+1} \mu_{kl} \widetilde{p}_l = \frac{d}{dx} \widetilde{p}(x)(x_k)$ in each cell i, and $\widetilde{p}$ is given by

$$\widetilde{p}_{il} = \begin{cases} \frac{1}{2}\left(p_{i-1,m+1} + p_{i,0}\right) + \sigma\left(p_{i-1,m+1} - p_{i,0}\right) & l = 0 \\ p_{il} & l \neq 0, m+1 \\ \frac{1}{2}\left(p_{i,m+1} + p_{i+1,0}\right) + \sigma\left(p_{i,m+1} - p_{i+1,0}\right) & l = m+1 \end{cases} . \tag{4.75}$$

Note the change of sign between Eq. (4.72) and Eq. (4.75) in the term augmenting the central average. We may characterize this discretization as having a complementary upwind/downwind bias. It can be easily verified that in the case of a first order scheme, i.e. $m = 0$, the above discretization collapses to the standard three-point stencil for the second derivative if $|\sigma| = 0.5$. The central discretization is recovered for $\sigma = 0$, which leads to a decoupled five-point stencil for the first order scheme (to avoid such a decoupled stencil is the primary motivation behind introducing this weighted form of discretization). These two limits are shown in Fig. 4.19(a). The method is related to the *LDG* method [11] , which utilizes a very similar upwind/downwind splitting for the viscous terms in the context of a Galerkin formulation, and reduces to the same first order discretization, i.e. a three-point stencil.

The asymptotic convergence in p refinement must be expected to be suboptimal in general, in the sense that convergence no better than order $n = m$ can be guaranteed if polynomials of order m are used. This is due to the extra derivative for the dissipative terms. This property is also shared by the LDG method. In practice, however, optimal order of convergence is often achieved. In order to avoid confusion we continue to refer to the "accuracy" of the scheme as $n = m + 1$, keeping in mind that optimal convergence for viscous problems may not always be achieved.

Consider the numerical solution of Eq. (4.66) using the scheme defined by (4.73) and (4.74). Several tests for varying order of accuracy, cell Peclet number, and discretization parameter σ have been carried out. We show a few illustrative examples (tables 4.4 through 4.7). The accuracy is maintained until machine zero is approached. The Jameson RK4$_s$ scheme has been used to advance the solution. However, time accuracy plays no role in this test, since the time step has been chosen small so that the error is dominated by the spatial discretization. The equation has been integrated to $T = 1$ for $Pe = 0.01$. For $Pe = 0.1$ we have chosen $T = 0.1$ as the solution will decay much more quickly. Fig. 4.20 summarizes the convergence study for $\sigma = -0.5$ and $Pe = 0.1$. In Fig. 4.20(a) the error is measured as the mesh is refined for the 2nd through 5th order scheme. Fig. 4.20(b) demonstrates convergence in p-refinement as the order of accuracy is increased, while the number of mesh elements is kept constant (40 in this case).

# Elements	L_∞(Error)	Order	L_2(Error)	Order
10	6.547057e-02		3.888390e-02	
20	1.749077e-02	1.904253	1.002766e-02	1.955188
30	7.804592e-03	1.990211	4.479220e-03	1.987590
40	4.398009e-03	1.993730	2.526923e-03	1.989858
50	2.828825e-03	1.977608	1.621178e-03	1.989075
75	1.266920e-03	1.981114	7.239641e-04	1.988251
100	7.168179e-04	1.979693	4.086552e-04	1.987854

Table 4.4: Second Order scheme, $\sigma = -0.5$, $Pe = 0.01$.

# Elements	L_∞(Error)	Order	L_2(Error)	Order
10	2.111335e-03		1.188727e-03	
20	2.386526e-04	3.145172	1.460845e-04	3.024542
30	7.050337e-05	3.007283	4.264994e-05	3.036413
40	2.974449e-05	2.999897	1.789916e-05	3.018165
50	1.515007e-05	3.023340	9.170556e-06	2.996976
75	4.611315e-06	2.933648	2.746238e-06	2.973785
100	1.985667e-06	2.928782	1.172301e-06	2.959043

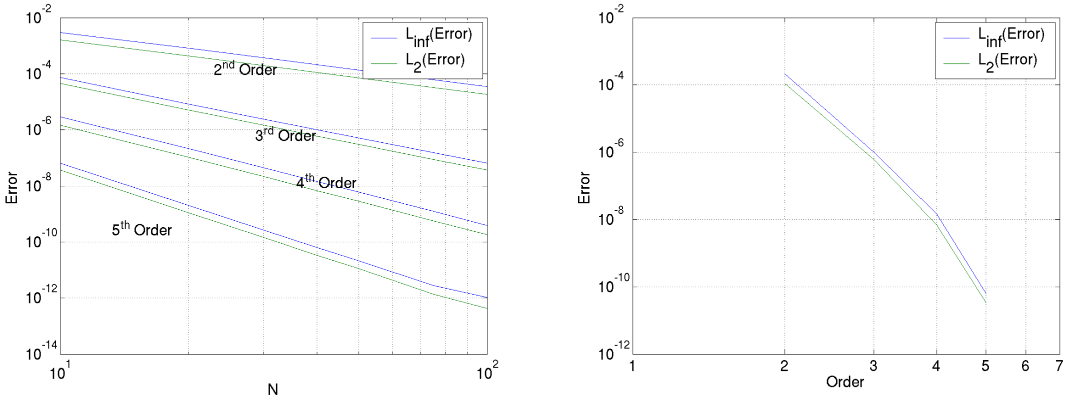
Table 4.5: Third Order scheme, $\sigma = 0$, $Pe = 0.01$.

# Elements	L_∞(Error)	Order	L_2(Error)	Order
10	2.896343e-06		1.477824e-06	
20	2.137082e-07	3.760518	1.036278e-07	3.833991
30	4.458241e-08	3.865369	2.124086e-08	3.908792
40	1.452898e-08	3.897337	6.844647e-09	3.936498
50	6.049254e-09	3.926667	2.833991e-09	3.951631
75	1.219807e-09	3.949149	5.680424e-10	3.963954
100	3.897074e-10	3.966364	1.808876e-10	3.977725

Table 4.6: Fourth Order scheme, $\sigma = -0.5$, $Pe = 0.1$.

# Elements	L_∞(Error)	Order	L_2(Error)	Order
10	6.517415e-08		3.716928e-08	
20	2.042238e-09	4.996077	1.109798e-09	5.065742
30	2.675080e-10	5.013137	1.439591e-10	5.037188
40	6.346705e-11	5.000759	3.396990e-11	5.019619
50	2.088936e-11	4.980117	1.112197e-11	5.003740
75	2.818483e-12	4.940107	1.375349e-12	5.155104
100	1.060079e-12	3.399083	4.199459e-13	4.123778

Table 4.7: Fifth Order scheme, $\sigma = -0.5$, $Pe = 0.1$.



(a) Accuracy in h refinement for the SD scheme, 2^{nd} through 5^{th} order scheme.

(b) Accuracy in p refinement with $\sigma = -0.5$ and $Pe = 0.1$ at $N = 40$

Figure 4.20: Convergence of the p_m-SD scheme for the advection-diffusion equation.

For the extension to higher dimensions consider the Navier-Stokes equations. The viscous fluxes can be written as

$$F_v = F_v(u, v, q_1, q_2, q_3) \tag{4.76}$$

$$q_1 = \nabla u \tag{4.77}$$

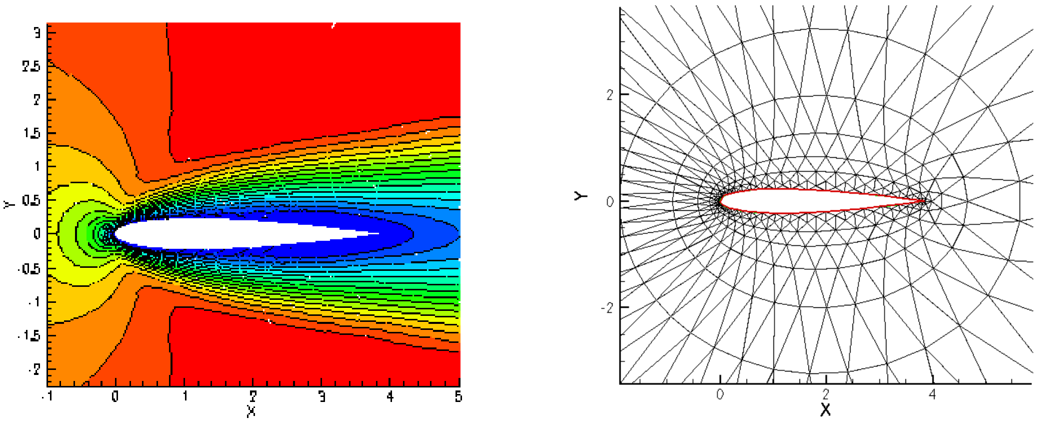
$$q_2 = \nabla v \tag{4.78}$$

$$q_3 = \nabla T\ , \tag{4.79}$$

where the velocity components are denoted by u and v, and T is the temperature. We can write $F_v = (g_v, h_v)$, where

$$g_v = \begin{pmatrix} \tau_{11} \\ \tau_{21} \\ u\tau_{11} + v\tau_{12} - \Xi_1 \end{pmatrix} \qquad h_v = \begin{pmatrix} \tau_{12} \\ \tau_{22} \\ u\tau_{21} + v\tau_{22} - \Xi_2 \end{pmatrix} . \tag{4.80}$$

Here $\tau(q_1, q_2)$ is the stress tensor, Eq. (2.30), and $\Xi = -\kappa q_3$ is the heat flux vector (κ is the thermal conductivity). Both are linear functions of the new variables q_i, and hence so is the viscous flux. The ideas behind the discretization discussed above thus carry over to the Navier-Stokes equations. We introduce the weighted discretization first in the computation of the gradients q_i in Eqns. (4.77) to (4.79) that form the stress tensor and the heat flux



(a) Contour lines of the Mach number for the third Order SD scheme

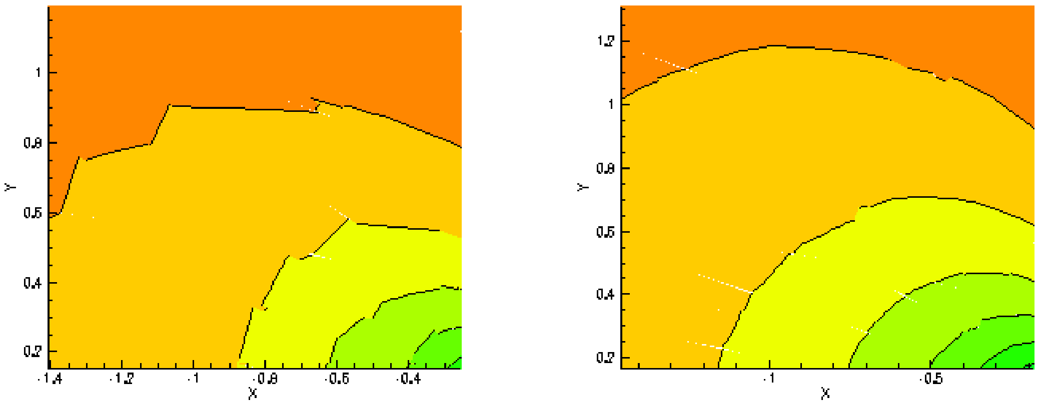
(b) Numerical grid with 960 triangles

Figure 4.21: 3^{rd} order SD scheme: The NACA 0012 profile at $M = 0.3$, $Re = 100$, $\alpha = 0^o$.

vector, and subsequently a complementary discretization is used when discretizing the flux in a manner analogous to Eq. (4.72) and Eq. (4.75). Several techniques can be conceived to introduce the splitting, such as switching the sign of the coefficient σ based on the local normal velocity across a face, i.e. $q_n = v \cdot n$, where v is the velocity vector and n is the edge normal. This would in fact mimic the standard concept of upwinding. However, for the discretization of the viscous fluxes this splitting is not necessarily more meaningful than other techniques. The important property is the complementary character of the upwind/downwind bias. In the present work the free stream direction is used to split the discretization, as depicted in Fig. 4.19(b)

Consider subsonic viscous flow around the NACA 0012 profile at flow conditions $M_\infty = 0.3$, $Re = 100$ and zero angle of attack. Fig. 4.21(a) shows a contour plot of the Mach number distribution computed with the 3^{rd} order SD scheme on a coarse triangular mesh with 960 elements. For the 3^{rd} order scheme this results in 5760 degrees of freedom (6 nodes to each triangle). The weighting parameter $\sigma = -0.5$ has been used. The computational mesh is shown in figure 4.21(b).

Discontinuous rendering of the solution for all high-order computations is used in this section, such as the one shown in Fig. 4.21(a). This helps assess the resolution



(a) Second Order SD scheme

(b) Third Order SD scheme

Figure 4.22: The NACA 0012 profile at $M = 0.3$, $Re = 100$, $\alpha = 0^o$. Contour lines of the Mach number. Close-up view of the solution near the leading edge.

of the scheme. Each triangle is rendered individually which makes the contour lines discontinuous across elements, as is the solution. Fig. 4.22 illustrates this using a close-up view of a region upstream of the leading edge for the 2^{nd} and 3^{rd} order SD scheme. The highly curved streamlines can be seen to be discontinuous across the triangles for the 2^{nd} order scheme, but become almost continuous for 3^{rd} order computations. which suggests adequate resolution, even on this relatively coarse mesh (Fig. 4.21(b)). For a standard finite-volume scheme this is not true, which is confirmed by Fig. 4.23, where the same testcase is shown, computed on the same mesh (Fig. 4.21(b)) with a standard cell-centered finite volume scheme using a linear reconstruction of the solution variables (i.e. the SD solver in $p_{0.7}$ mode), and a central discretization of the viscous terms, as in section 3.6.2 and documented in [38]. It is clear that the mesh resolution is not adequate. To compute a comparison for the 3^{rd} order SD solution the finite-volume scheme requires a finer mesh. A solution on a mesh with 10,240 triangles is shown in Fig. 4.24, where a contour plot of the Mach number distribution is shown along with the computational mesh. Note that the finite volume scheme uses roughly twice as many degrees of freedom, compared to the 3^{rd} order SD scheme.

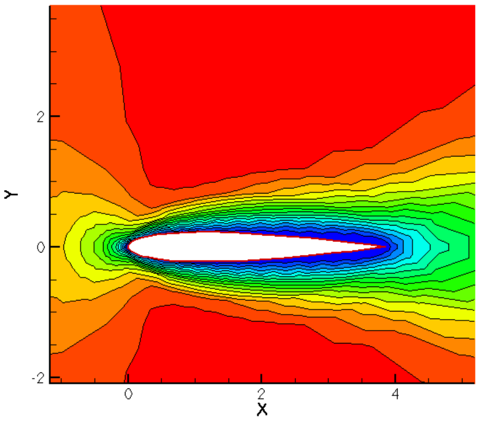
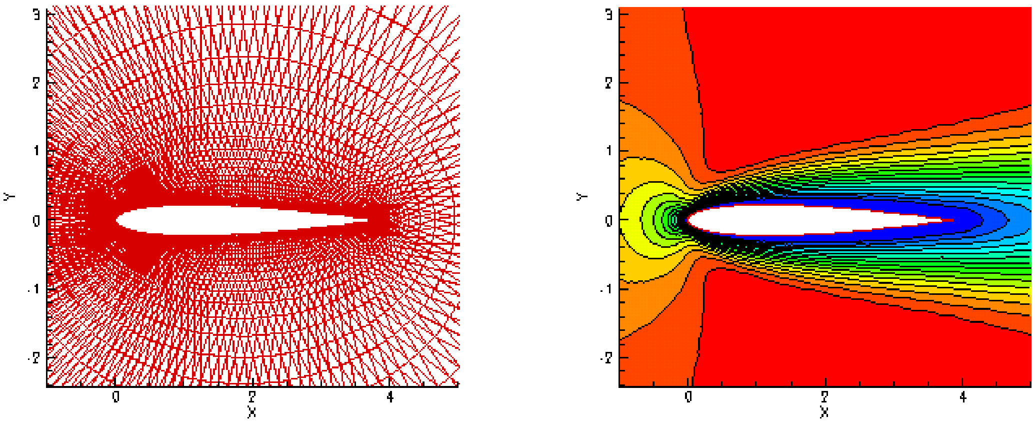


Figure 4.23: Finite Volume Scheme: The NACA 0012 profile at $M = 0.3$, $Re = 100$, $\alpha = 0^o$. Coarse mesh (960 triangles).



(a) Numerical grid with 10240 triangles

(b) Contour lines of the Mach number for the second order finite-volume scheme

Figure 4.24: Finite Volume Scheme: The NACA 0012 profile at $M = 0.3$, $Re = 100$, $\alpha = 0^o$. Fine mesh (10240 triangles).

Chapter 5

Advanced Postprocessing: Gibbs-Complementary Reconstruction

5.1 Introduction

It is well-known that pointwise exponential convergence for spectral approximations can only be obtained for infinitely differentiable functions [19]. This is true for both Galerkin and interpolation-based approximations. Generally the rate of convergence depends on the smoothness of the function. For functions that are only piecewise continuous convergence in the maximum norm fails completely, and the approximated function exhibits non-decaying oscillations around the discontinuities, which is known as the Gibbs phenomenon.

The history of the discovery and the subsequent scientific consideration of the Gibbs phenomenon is an interesting one (a short account can be found in [24]). One question that has been the subject of debate is whether the presence of oscillations actually means that information about the approximated function is irretrievably lost, or if it is possible, by means of a suitable filtering procedure or projection method, to extract pointwise exponential convergence from the approximation.

In [24] an account of the effort of overcoming the Gibbs phenomenon is given. These efforts have led to a theory, developed by Gottlieb and coworkers in a series of papers [25, 20, 23, 21, 22], that has established that, in fact, one can recover pointwise exponential convergence by a suitable re-projection of the reconstructed function, dubbed Gibbs-complementary reconstruction. This is true for a wide variety of approximation techniques.

In this section we are concerned with exploiting this theory in the context of numerical schemes for hyperbolic conservation laws. The theory of Gibbs-comlementary reconstruction as summarized in [24] applies to a truncated, but otherwise exact series of spectral coefficients. It is attempted to extend the method to a case where the spectral coefficients themselves are only approximations, in addition to being truncated, corresponding to the numerically computed solution of the conservation law using a high-order accurate numerical method [1].

It is known that high-order methods cannot capture discontinuities in a monotone fashion unless nonlinear limiting functions are used, and invariably the order of accuracy (at least locally) is compromised. Often the concept of limiting the solution is also an integral part of nonlinear stability analysis for (fixed order) high-resolution schemes, for example in TVD theory [26, 27] as outlined in section 2.4. In the realm of high-order methods other methods of stabilizing the solution are known, for example spectrally vanishing viscosity [63] or high-order filters. It is well known that such methods can stabilize high-order numerical schemes, although they do not render the solution monotone.

Based on these preliminary thoughts, the route taken here, can be summarized as the following two-step procedure:

- Compute the solution to a nonlinear hyperbolic conservation law using a high-order method without enforcing monotonicity (but possibly using high-order stabilization techniques, such as filtering)
- Extract pointwise accuracy from the oscillatory solution by a suitable re-projection.

Convergence proofs have been established for some high-order stabilization methods,

---

[1] It is to be understood that spectral methods as well as fixed order methods of “high” order of accuracy are called high-order methods here.

and the aforementioned theory of Gibbs-Complementary reconstruction is also rigorous. The connection, or rather concatenation of the two theories, however, remains somewhat informal. Nevertheless, interesting results are presented in the following that highly suggest the viability of the concept. Some obstacles and areas of concern are also mentioned. The organization of this chapter is as follows. Firstly, the basic theory of Gibbs-Complementary reconstruction is recapitulated in section 5.2. A reproduction of the convergence proofs is omitted, and rather cited from the literature. Some practical considerations are given as well. Subsequently, some known Gibbs-complementary families of basis functions are discussed in sections 5.3 and 5.4. The problem of identifying discontinuities is discussed in section 5.5, and finally numerical results are presented in section 5.6.

5.2 Basic Theory

In this section we summarize the concept of Gibbs-Complementary reconstruction rather informally. Discussion of the rigorous approach is deferred to section 5.3, where the Gegenbauer polynomials are introduced as a specific example, and some results from the literature are recalled. The theory for the Gibbs-complementary reconstruction is presented for one-dimensional functions. Tensor product forms of basis functions can be used for the multidimensional extension. Assume that $N+1$ spectral coefficients b_k, for $k = 0, \ldots, N$ of a *piecewise-smooth* function $f(x)$, defined on an interval $[a, b]$, are known, such that they define the projection

$$P_N(f) = f_N(x) = \sum_{n=0}^{N} b_k \phi_k(x) \ . \tag{5.1}$$

The function f_N is thus an approximation to the original function f. The ϕ_k represent the chosen basis, which spans a certain functional space, for instance the space of polynomials of degree at most N. The coefficients b_k can be obtained by Galerkin projection, i.e $b_k \sim (\phi_k, f)$, with a suitable definition of a (weighted) scalar product, or collocation for which $b_k \sim f(x_k)$, where the x_k belong to a set of collocation nodes. It is known that the approximation (5.1) converges exponentially in N for infinitely differentiable functions. However, for piecewise continuous functions convergence in maximum norm fails completely.

In a series of papers Gottlieb and coworkers have established the possibility of recovering high-order accuracy in maximum norm for spectral approximations of piecewise continuous functions, i.e. the removal of Gibbs' phenomenon (see [22, 24] and references therein). Results can be obtained for quite general approximations of the type (5.1), such as Fourier or Chebyshev expansions. In fact, given the first N (exact) spectral coefficients exponential convergence in *maximum norm* as $N \to \infty$ has been proved for any subdomain $[x_1, x_2]$ where the function is analytic, i.e.

$$\max_{x\in[x_1,x_2]} |f(x) - f_N(x)| < e^{-\alpha N} \; , \qquad \alpha > 0 \; , \tag{5.2}$$

so that by juxtaposing the domains of analyticity exponential convergence can be recovered in the complete interval. Formally the method can be summarized as follows. Suppose a projection $P_N(f)$ of the form of Eq. (5.1) is given. The approximation f_N is re-projected onto another space, spanned by basis functions which shall be denoted by Ψ_m^λ (it is anticipated here that the new basis functions will be two-parameter families). We seek conditions under which the new projection, denoted by G_M leads to an exponentially small error in maximum norm. Consider

$$||G_M P_N f - f||_\infty \leq ||G_M P_N f - G_M f||_\infty + ||G_M f - f||_\infty \; . \tag{5.3}$$

The second term on the right-hand side measures how well the new basis is suited to approximate the original function f. This step could normally be handled by standard approximation theory. However, for the known Gibbs complementary bases one actually needs to show that this approximation converges while *both* parameters of Ψ_m^λ are increased simultaneously. This is a result of proving that the first component of the error in Eq. (5.3) vanishes exponentially fast. Write this term as

$$||G_M P_N f - G_M f||_\infty = ||G_M (P_N - I) f||_\infty \; . \tag{5.4}$$

In a sense we thus have a measure of the orthogonality of the new projection and the part of the spectrum of the original projection which is not resolved. If these are nearly orthogonal, or become orthogonal at an exponential rate as N is increased, the high modes have very little effect on the representation in the new projection. The error will decrease exponentially as $N \to \infty$, assuming the second term in Eq. (5.3) converges exponentially as well.

Consider now a solution of a hyperbolic conservation law, computed with a high-order numerical method. Assume that high-order stabilizing diffusion or filtering has been added to the spectral or pseudo-spectral method, so that instead of Eq. (5.4) one may formally write

$$||G_M \widetilde{P_N} f - G_M f||_\infty \leq ||G_M(\widetilde{P_N} - P_N)f||_\infty + ||G_M(P_N - I)f||_\infty \ , \tag{5.5}$$

Thus a third error term is introduced, which comes from the numerical scheme. Both the projection P_N and the numerical approximation thereto, $\widetilde{P_N}$, will be oscillatory. Despite the fact that we generally assume that the numerical approximation is convergent in some sense, it will certainly not converge pointwise if discontinuities are present. The Gibbs-Complementary reconstruction thus has to reduce this leftover exponentially fast as well. In assuming that it does, it is implied that the numerical approximation is exponentially accurate in the *information* that is contained in the solution, despite the presence of oscillations. This assumption can be traced back to Lax, who argued this way [44], based on some considerations inspired by information theory. Numerical evidence is presented below that it is indeed possible to recover pointwise accuracy from the numerical approximation to the spectral coefficients of the unknown solution.

5.3 The Gegenbauer Polynomials

The first basis which has been identified as Gibbs complementary to, in fact, a whole host of spectral approximations, are the Gegenbauer polynomials, $\Psi_m^\lambda = C_m^\lambda$. Here m is the degree of the polynomial, and λ is a second parameter. The Gegenbauer polynomials can be defined as the polynomials that are orthogonal on the interval $[-1, 1]$ with the weight function $w(x) = (1 - x^2)^{\lambda - \frac{1}{2}}$, i.e.

$$\int_{-1}^{1} C_n^\lambda(x) C_m^\lambda(x) (1 - x^2)^{\lambda - \frac{1}{2}} \, dx = \delta_{nm} h_n^\lambda \ . \tag{5.6}$$

The Gegenbauer polynomials belong to the family of Jacobi polynomials, and they include the Legendre and Chebyshev polynomials as special cases for fixed $\lambda = 1/2$ and $\lambda = 0$, respectively. In the most common normalization [4], the constant h_n^λ is given by

$$h_n^\lambda = 2^{1-2\lambda}\pi \frac{\Gamma(n+2\lambda)}{(n+\lambda)\Gamma^2(\lambda)\Gamma(n+1)} \ , \tag{5.7}$$

for $\lambda > -1/2$, where $\Gamma(x)$ is the Gamma function.

The re-projection of a spectral representation using Gegenbauer polynomials will be referred to as the Gegenbauer reconstruction procedure. The Gegenbauer reconstruction operates on each subinterval on which the solution is smooth. Assume for the sake of simplicity that $[-1, 1]$ is such an interval (general intervals can be handled by appropriate scaling). It can be shown that the projection $G_M P_N f$ converges exponentially fast to f in maximum norm, even at discontinuities. This can be condensed in a general theorem that is cited here from [24] in a slightly weaker form to state the main point and avoid technical details related to different kinds of spectral approximations.

Theorem 5.3.1. *(Gottlieb et al.) Assume a function f is given, which is piecewise analytic on $[a, b] \subset \mathbb{R}$, and in particular analytic on $[-1, 1] \subset [a, b]$. Furthermore assume that f possesses an analytic extension into the complex plane such that the closest pole is at least a distance $\rho > 0$ away from the interval $[-1; 1]$. Assume the first $N + 1$ coefficients of an approximation f_N of the form (5.1), coming from a Fourier or Chebyshev spectral approximation, either Galerkin or collocation are know. For the pointwise error of the Gegenbauer reconstruction*

$$E = \max_{[-1:1]} \left| f - \sum_{n=0}^{M} \widetilde{g}_m^\lambda C_m^\lambda(x) \right| \ , \tag{5.8}$$

where

$$\widetilde{g}_m^\lambda = \frac{1}{h_m^\lambda} \int_{-1}^{1} (1-x^2)^{\lambda-\frac{1}{2}} f_N(x) C_n^\lambda(x)\, dx \ , \tag{5.9}$$

there exists $\beta > 0$ such that if

$$M = \lambda = \beta N \ , \tag{5.10}$$

the error E converges exponentially fast, i.e. $E < q^N$ as $N \to \infty$ with $q < 1$.

Proof. see [24] □

Remark 5.3.1. *Theorem 5.3.1 can be formulated more strongly. In particular sharp estimates on the constants of proportionality in Eq. (5.10) can be obtained depending on the type of spectral approximation used. Furthermore, the result can be obtained for more general classes of approximation, not only Fourier or Chebyshev. Those two, however, are of special interest here.*

While the details of the proof depend on the spectral representation, one can roughly argue that the necessity of varying the parameters as in Eq. (5.10) stems from the trade-off between the obvious fact that the the order of the Gegenbauer polynomials M has to be increased along with the order of the original projection to achieve spectral convergence, and the need to maintain a reasonable degree of separation between the complement of the original projection and the Gegenbauer space to keep the orthogonality error, Eq. (5.4), small.

5.3.1 Remarks on the Gegenbauer Basis

Despite the fact that exponential convergence can be proved there are a few well-known problems associated with the Gegenbauer basis, which can have a profound effect in numerical implementations. We briefly summarize a few points.

- The success of the reconstruction depends on the smoothness of the unknown function. More precisely, the constants of proportionality in Eq. (5.10) depend on how far the function can be extended onto the complex plane. In the absence of such information these constants have to be estimated, and once fixed values are chosen, the reconstruction might fail for certain functions.

- The Gegenbauer polynomials assume very large numerical values very quickly as the parameters are increased. This potentially amplifies the effect of round-off errors. The function to be approximated is usually of order unity, so that for large N some coefficients in the truncated sum may need to be smaller than machine accuracy, but at the same time highly accurate to represent the solution with small absolute errors. This may lead to catastrophic failure.

- The nature of the Gegenbauer projection can lead to difficulties, since it becomes more and more extrapolatory as λ is increased (the weight function approaches a delta function). In the limit of $\lambda \to \infty$ only the function value at the midpoint of the domain contributes to the expansion.

5.4 An Alternative Basis: The Freud Polynomials

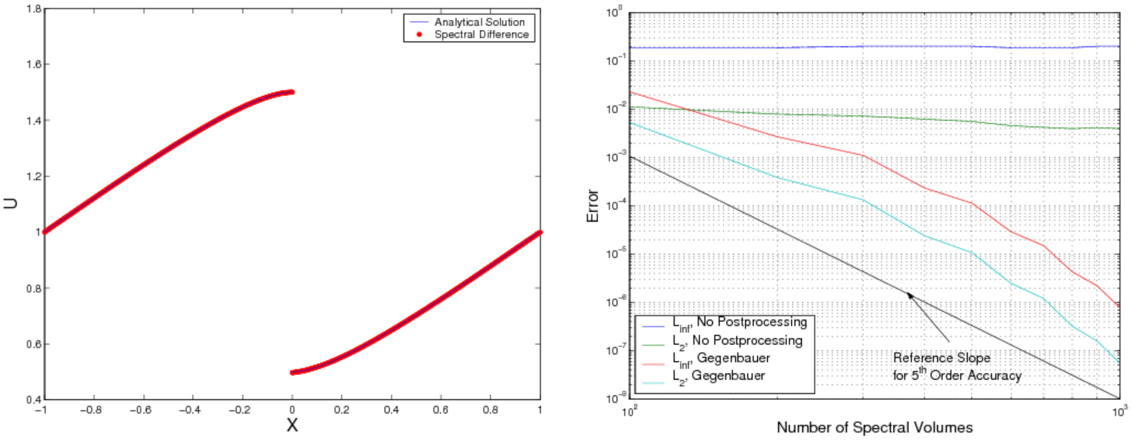
Partially in response to the numerical difficulties associated with Gegenbauer polynomials, the concept of robust Gibbs complements has been proposed by Tanner and Gelb [65]. The additional requirement is that the associated weight of the new basis converge to a limit which still has an associated set of polynomials that can serve as a basis for an exponentially convergent projection (this is not the case for the Gegenbauer polynomials, since the weight converges to a delta function). The Freud polynomials, which are orthonormal on the real line with respect to the weight $w(x) = \exp(-cx^{2n})$, have been proposed as an alternative to the Gegenbauer basis. Unfortunately very little is known about this family of polynomials, in particular the three-term recurrence relation is not known analytically. This has prevented a rigorous proof that the Freud polynomials can serve as an alternative Gibbs complementary basis. Some elements and numerical evidence, however, have been shown [65]. The Freud polynomials are used here as an alternative to the Gegenbauer polynomials for certain test problems.

5.5 Edge Detection

A reliable method of identifying discontinuities is needed, because the Gibbs-Complementary reconstruction operates on smooth subdomains. The edge detection procedure used in this work is based on the generalized Fourier conjugate sum

$$S_N^\sigma = \sum_{k=0}^{N} \sigma\left(\frac{k}{N}\right) \left(a_k \sin kx - b_k \cos kx\right) \ , \tag{5.11}$$

as proposed in [15, 17], where $\sigma = 1$ corresponds to the classical conjugate sum. The coefficients a_k and b_k are related to the complex Fourier coefficients f_k via $f_k = a_k + ib_k$.



(a) Solution using 200 Spectral Elements

(b) Error with and without postprocessing

Figure 5.1: Globally high-order accurate solutions for Burgers' Equation using the SD Method and the Gegenbauer Procedure.

It is known that the classical Fourier conjugate sum converges to

$$-\pi/\log(N)S_N \longrightarrow [f(x)] \ , \tag{5.12}$$

where $[f(x)] = f(x^+) - f(x^-)$, and $f(x^{\pm})$ are the limits as f approaches x from the right and left, respectively. The factors $\sigma(x)$ are introduced to accelerate the convergence, see [15, 17] for details.

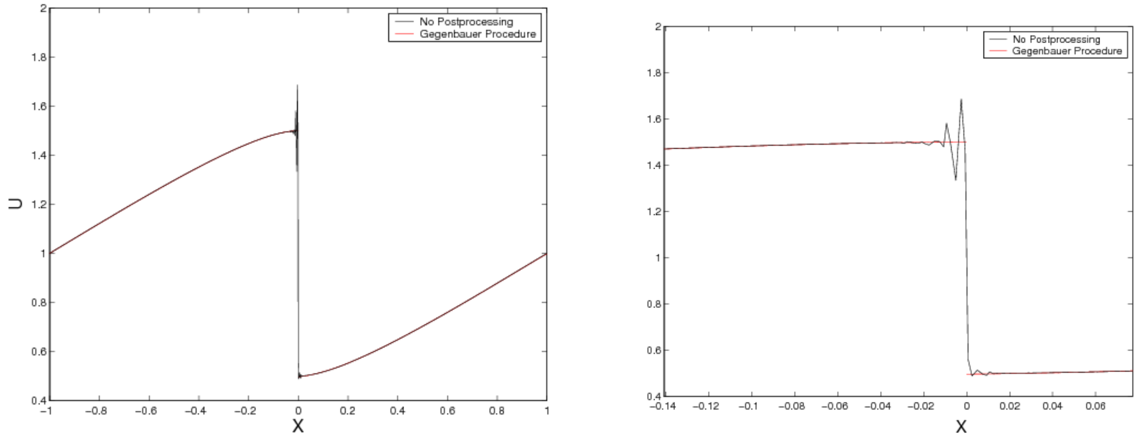
5.6 Numerical Results

As a convenient model equation consider Burgers' equation with periodic boundary conditions

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial}{\partial x}u^2 = 0 \ , \tag{5.13}$$

$$u(x,0) = 1 + \tfrac{1}{2}\sin(\pi x) \ , \qquad -1 \leq x \leq 1 \ . \tag{5.14}$$

Results for a fifth order accurate Spectral Difference method at $t = 1$, at which time a discontinuity at $x = 0$ is present, are shown in Fig. 5.1. Fig. 5.1(a) shows the solution using $N = 200$ cells after re-projection of the solution onto Gegenbauer polynomials. The effectiveness of the procedure is demonstrated by Fig. 5.2, where a comparison between

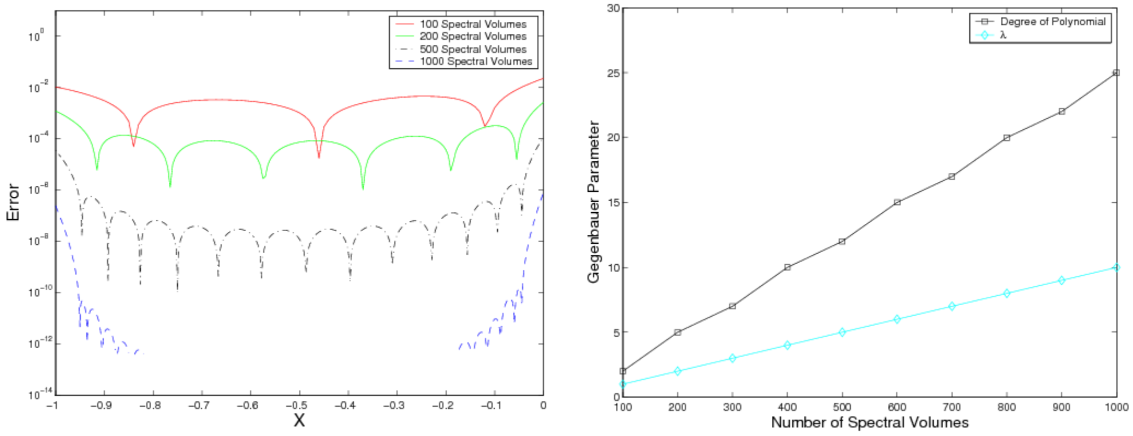


(a) Solution using 200 cells with and without postprocessing

(b) Close-up view of the solution near the discontinuity with and without postprocessing

Figure 5.2: Solutions for Burgers' Equation using the SD Method with and without the Gegenbauer Procedure.

the original numerical solution and the expansion in Gegenbauer polynomials is shown. A grid refinement study to verify the nominal order of accuracy as N is increased has been carried out. This is shown in Fig. 5.1(b) together with a reference slope indicating the required reduction rate of the error for nominal fifth order global accuracy. It can be seen that while the original solution fails to converge in maximum norm, and the L_2 norm of the Error barely reaches first order, the re-projected solution converges in both L_2 and maximum norm. It should be emphasized that error norms were computed using *all* nodes in the mesh, so that all global error measures include the discontinuity (using the left limit of the solution for $x \to 0^-$, and the right limit for $x \to 0^+$). The slope of the errors, however, is not constant, which can be explained by the fact that the parameters of the Gegenbauer polynomials, i.e. the order and second parameter λ, have to be adjusted as the mesh is refined. In theory one ought to make the number of Gegenbauer modes proportional to the number of original spectral modes. These, however, remain constant here, because the order of approximation is fixed. It seems most fitting to make the parameters proportional to the total number of collocation nodes. The dependence of the parameters on the number of spectral volumes is depicted in Fig. 5.3(b). It should be pointed out that the parameters are not optimized in any way, but merely represent "best



(a) Pointwise Error for several numbers of Spectral Volumes

(b) The Gegenbauer Order and second parameter λ as the mesh is refined

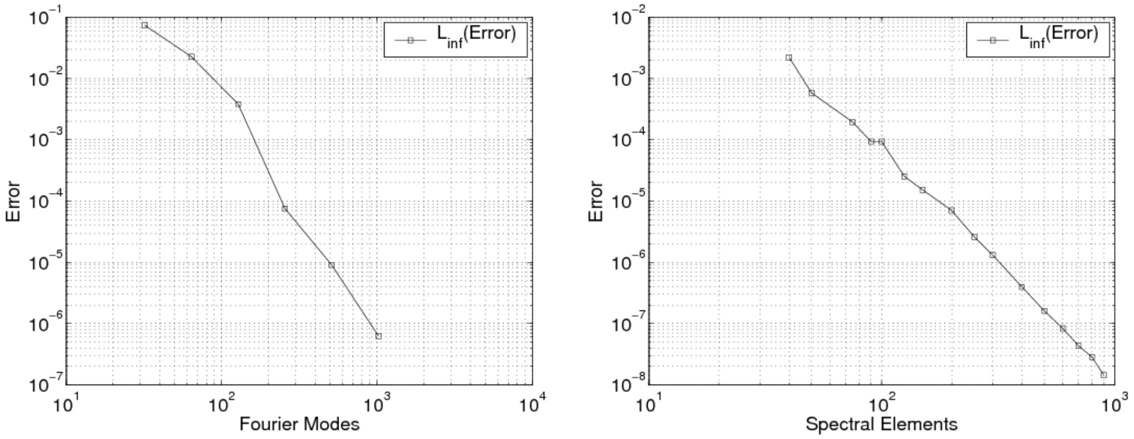
Figure 5.3: Globally high-order accurate solutions for Burgers' Equation using the SD Method and the Gegenbauer Procedure.

practice" values. Some suggestions regarding their optimization have been made in the literature [16].

Fig. 5.3(a) shows the pointwise errors for the smooth subdomain $-1 \leq x \leq 0$. It can be seen that the error decreases up to the endpoints of the domain, the right endpoint being the discontinuity at $x = 0$.

A reduction of the maximum error below roughly $1 \cdot 10^{-8}$ could not be achieved, due to the inherent numerical problems mentioned in section 5.3.1. In fact, if the mesh is further refined the error increases and diverges in the limit $N \to \infty$ if the Gegenbauer parameters are further increased (although this can be prevented by freezing them). For the extreme case of 1000 cells the order of the polynomials is $m = 25$ and $\lambda = 10$. The value of the corresponding Gegenbauer polynomial at $x = 1$ is $C_{25}^{10}(1) = 1.4 \cdot 10^{+12}$. It becomes clear how it is increasingly difficult to achieve an absolute precision of at least twenty orders of magnitude below the computed values of the Gegenbauer polynomials.

No limiters, filters, or spectral viscosity have been used for these computations, which means that we have relied solely on the dissipation from the fluxes at element boundaries to stabilize the solution. It is shown below that for higher orders of approximation spectral filtering can be used for stabilization purposes.



(a) Maximum error in p-refinement for a Fourier Spectral Method.

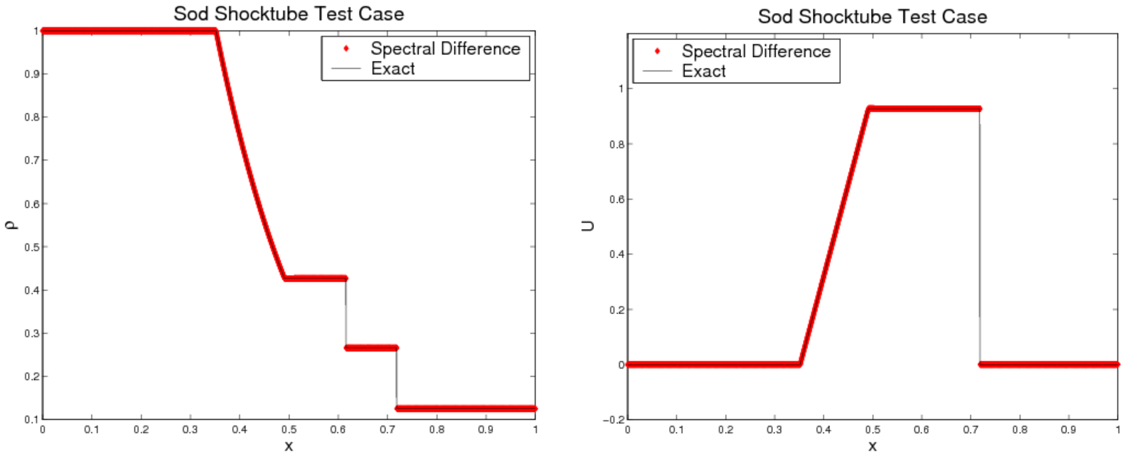
(b) Maximum error in h-refinement for 5^{th}-order SD Method.

Figure 5.4: The Freud reconstruction method for Burgers' Equation.

As an alternative to Gegenbauer polynomials the Freud polynomials can be used to perform Gibbs-complementary reconstruction. Two tests have been carried out for the Burgers Equation. Firstly the equations are solved by a Fourier nodal spectral method with an exponential filter to stabilize the solution. Secondly we use the 5^{th} order Spectral Difference method. These two approaches allow us to test the Freud reconstruction procedure in p-refinement and h-refinement. The results of the refinement studies are shown in Fig. 5.4. It can be seen that the Freud polynomials can be used to extract pointwise high-order accuracy from both the collocation-based spectral and the SD fixed order method.

As a preliminary testcase for the Euler equations the Sod shocktube problem has been considered. Fig. 5.5 shows the solution obtained with a fourth-order SD Method with scalar diffusion at element boundaries on a mesh with 800 cells. It can be seen that after Gegenbauer reconstruction the shocks and the contact discontinuity are perfectly captured. No limiter or spectral diffusion has been used, which means that the dissipation from the element boundaries must suffice to stabilize the solution.

Fig. 5.6 shows an example of a typical distribution of the maximum order of the polynomials and value of the parameter λ in each domain of analyticity. For better reference, the Gegenbauer reconstruction of the density solution is shown in the plot as



(a) Density

(b) Velocity

Figure 5.5: Solution for the Sod shocktube case using the SD Scheme and Gegenbauer reconstruction.

well. One can easily identify the domains of analyticity corresponding to the solution of the Riemann problem. According to Fig. 5.6, the maximum order of the polynomial

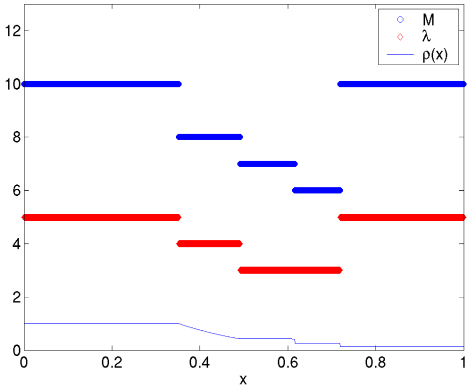


Figure 5.6: Shocktube case: Gegenbauer parameter (M,λ) for the domains of analyticity for 300 cells.

in the leftmost and rightmost domains is $M = 10$. Nevertheless, the solution is very nearly constant. Tab. 5.1 shows the values of the Gegenbauer coefficients g_l^λ for these two domains. All higher modes of the Gegenbauer polynomials differ only by round-off errors from machine accuracy for double precision, while all the energy is in the zero-mode, which

(a) Left-most domain ($x < 0.352$)

l	exact	computed
0	1.0	1.0000000000371769
1	0	-3.090226125E-11
2	0	1.2051589376E-11
3	0	-1.230245018E-11
4	0	-2.841753118E-10
5	0	-3.836424587E-12
6	0	-3.013196009E-10
7	0	-1.902606656E-12
8	0	-2.636347148E-10
9	0	-1.143471536E-12
10	0	-2.225742876E-10

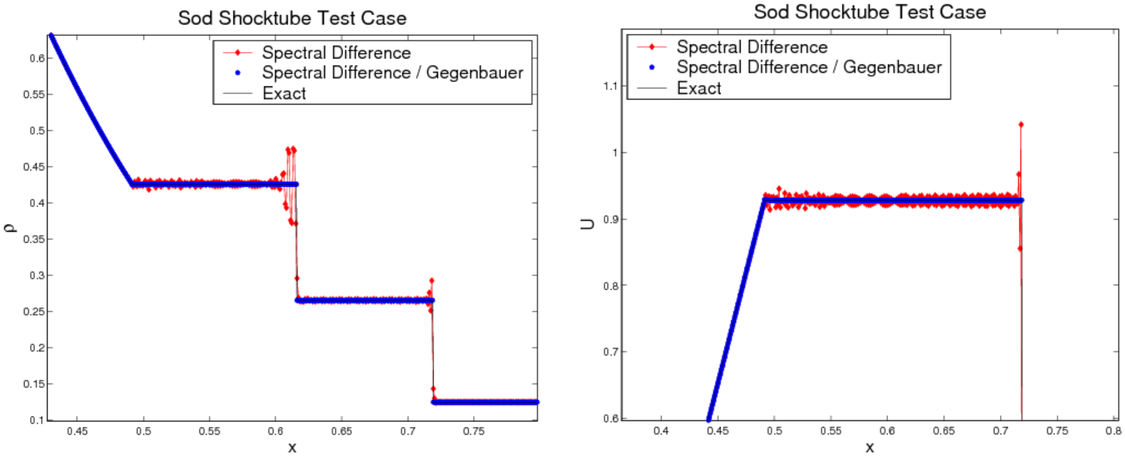
(b) Right-most domain ($x > 0.719$)

l	exact	computed
0	0.125	0.12500000000112668
1	0	7.180037463E-12
2	0	-3.287720660E-12
3	0	5.607469282E-12
4	0	-9.349026141E-11
5	0	5.806738414E-12
6	0	-9.924890331E-11
7	0	6.572377112E-12
8	0	-8.845402656E-11
9	0	7.018368055E-12
10	0	-7.634299126E-11

Table 5.1: Representative example of Gegenbauer coefficients $\widetilde{g}_l^\lambda$ for the Sod test case.

has the exact value up to machine accuracy. The fact that the maximum error in density could not be reduced beyond approximately $L_\infty(\rho - \rho_{exact}) \approx 1 \cdot 10^{-3}$, illustrates once again how numerically difficult the Gegenbauer basis is: The coefficients corresponding to the higher modes would have to be reduced below machine accuracy to reduce the error further. However, note that the level of error achieved represents the maximum error in the whole domain, including the discontinuities.

To demonstrate the effect of the reconstruction we show the unprocessed solution along with the reprojected one in Fig. 5.7. The unlimited original solution, computed with the 4^{th} order spectral difference method, exhibits oscillations of the order of what one would expect from Gibb's phenomenon, if the analytical solution were to be expanded in smooth basis functions.



(a) Density

(b) Velocity

Figure 5.7: Solution for the Sod shocktube case using the SD scheme with and without Gegenbauer reconstruction.

Chapter 6

Conclusions and Outlook

An new kinetic scheme for the Euler and Navier-Stokes equations has been developed, based on the Xu-Prendergast BGK scheme. The new scheme is not only computationally less expensive but also clarifies and improves the formulation with respect to the underlying kinetic gas theory. Along with the proposed multigrid methodology, the new scheme is capable of large-scale computations of practical engineering interest, while providing a completely mesh-independent discretization of the Navier-Stokes equations.

Such a mesh-independent discretization is desirable for next-generation flow solvers that are designed to operate on arbitrary polyhedral meshes. In particular the trivial extension to turbulent flow discussed in section 3.6.3 offers tremendous potential for the application of general-mesh solvers to high-Reynolds-number flow. Building on the initial validation shown here, estimates of eddy viscosity can be used to include the effects of turbulence in much the same way as in conventional finite volume schemes. Validation for such flow regimes is the subject of further research.

The compact discretization stencil also makes the new kinetic scheme attractive for high-order schemes, such as the Spectral Difference Method, and it may find use in this context in future research.

Significant contributions to the development of the Spectral Difference Method have been reported in this thesis, including stability analysis, limiting methodology, viscous formulation, and convergence acceleration via h/p-Multigrid. A major push for a new high-order solver architecture has been initiated. The presented results, in particular as regards robustness and convergence acceleration, show promise for achieving the ultimate goal of being able to compute problems of practical engineering interest faster with high-order methods compared to low order methods.

Future work on the Spectral Difference method is aimed at producing a hybrid formulation for triangular/quadrilateral meshes. The low-order limit of the Scheme is a first-order finite-volume scheme, while the high-order limit on quadrilateral meshes is a nodal multi-domain spectral method, based on a tensor product polynomial basis, along the lines discussed in [42]. With very little computational complexity a software package can thus be created that covers the range from low-order finite volume to high-order spectral. First steps toward this goal have been undertaken, and have produced some of the results shown in this thesis.

Another important future project is the extension of the scheme to three dimensions, which is not anticipated to produce new conceptual challenges, but requires extensive technical programming.

The theory of Gibbs-complementary reconstruction has been demonstrated to be a viable path to extending high-order methods to problems involving discontinuous solutions, using the Gegenbauer or Freud polynomial basis. The extension to multiple dimensions is the most obvious next step. Tensor product forms of basis functions, such as the Gegenbauer polynomials may be used with (multi-domain) spectral methods for this purpose.

Theoretical difficulties in connecting Gibbs-complementary reconstruction with convergence of high-order stabilized numerical methods also remain and need to be addressed. Furthermore, the Gegenbauer basis is numerically problematic, while very little is known about the Freud polynomials. It would be highly desirable to find other, more benign, basis functions that are suitable for Gibbs-complementary reconstruction with respect to standard spectral expansions.

Appendix A

Chapman-Enskog Expansion of the Boltzmann Equation

The Chapman-Enskog expansion, when applied to the Boltzmann equation, follows the same general concept that was outlined in section 2.1.2 for the BGK equation. The analog to Eq. (2.17) is given by

$$\epsilon^0 : \qquad J(f_0, f_0) = 0 \; , \tag{A.1}$$

$$\epsilon^1 : \qquad \frac{\partial f_0}{\partial q}\Phi_0 + u_i \frac{\partial f_0}{\partial x_i} = 2J(f_0, f_1) \; , \tag{A.2}$$

$$\vdots$$

$$\epsilon^k : \sum_{l+m=k-1} \left(\frac{\partial f_l}{\partial q}\Phi_m + \frac{\partial f_l}{\partial \nabla q}\nabla\Phi_m + \ldots \right) + u_i \frac{\partial f_{k-1}}{\partial x_i} = \sum_{l+m=k} J(f_l, f_m)\, d\Xi \; . \tag{A.3}$$

The left-hand side for $k \geq 1$ thus remains unchanged compared to Eq. (2.17). The term corresponding to the zeroth order gives again the Maxwellian distribution, as was noted in section 2.1.1. This means that all the constraints that were introduced in section 2.1.2 remain the same, i.e. Eqns. (2.20), (2.21), (2.22), and (2.22). The key difference compared to the Chapman-Enskog expansion for the BGK equation is thus that each equation of the form (A.3) is an inhomogeneous integral equation. It may be shown that the homogeneous problem in Eq. (A.2), (A.3) admits only functions of the collisional invariants ϕ. According to standard theory of integral equations this in turn means that the left-hand-sides must be orthogonal to ϕ, which is again the same result that has been obtained for the BGK-based expansion. Thus, one arrives again at Eq. (2.24):

$$\Phi_k = -\frac{\partial}{\partial x_i} \int \phi u_i f_k d\Xi \ . \tag{A.4}$$

The equations for the computation of the macroscopic equations from the noneqilibrium distribution fucntions remains the same, and hence the only difference between the Chapman-Enskog expansion of the BGK and the Boltzmann equation is that the computation of the f_k makes it necessary to solve the integral equations in (A.3). The algebra is extremely tedious, but at least for the f_1 state, many standard textbooks devote a number of pages to it, e.g. [41]. The general form of f_1 is given by

$$f_1 = -f_0 \left\{ \eta(y)\, c_j \frac{\partial T}{\partial x_j} + \zeta(y) \left(c_i c_j - \frac{1}{3} c^2 \delta_{ij} \right) \frac{\partial U_i}{\partial x_j} \right\} \ , \tag{A.5}$$

where $c_i = u_i - U_i$, and $y = c^2/2RT$. The functions $\eta(y)$ and $\zeta(y)$ are polynomial expansions coming from the solution of the integral equations, which determine the transport coefficients, since they multiply the velocity and temperature gradients, which integrate to the stress tensor and heat flux. Thus each f_k again contains an infinite number of coefficients. In practice, of course, the polynomial series is truncated after very few terms, leaving some freedom in the choice of transport coefficient, corresponding to different kinetic models. In fact, in this way it is possible to recover empirically determined transport coefficients for many fluids over a wide range of operating conditions. On the other hand, for the f_1 state based on the BGK equation, the only free parameter is the expansion parameter ϵ. This allows adjustment of the model to recover only one empirically determined transport coefficient, e.g. the coefficient of viscosity, while the second coefficient of viscosity and the coefficient of heat conduction are fixed relative to that value.

Appendix B

Reconstruction for Kinetic Schemes

In this section explicit formulas are given for the reconstruction procedures needed for the spatial gradients and time derivatives of the distribution functions for the kinetic schemes discussed in chapter 3. Very similar formulas have been given by Xu for the two-dimensional case [73]. The three-dimensional formulas are given merely for convenience. The spatial gradients of the Maxwellian distribution can be evaluated by direct differentiation. Write the first order Taylor series expansion of the distribution function about $x = 0$ as

$$f_0 + x\partial_x f_0 = f_0\,(1 + ax)\ , \tag{B.1}$$

where

$$a = \frac{1}{f_0}\frac{\partial f_0}{\partial x} = \frac{\partial(\ln f_0)}{\partial x}\ . \tag{B.2}$$

We can then write

$$\begin{aligned} \partial_x(\ln f_0) &= \partial_x \left\{ \ln\rho + \frac{5}{2}\ln\left(\frac{\lambda}{\pi}\right) - \lambda\,(\mathbf{u}-\mathbf{U})\cdot(\mathbf{u}-\mathbf{U}) \right\} \\ &= \frac{1}{\rho}\partial_x\rho + \frac{5}{2\lambda}\partial_x\lambda - \partial_x\lambda\,(\mathbf{u}-\mathbf{U})\cdot(\mathbf{u}-\mathbf{U}) - 2\lambda\,(\mathbf{u}-\mathbf{U})\cdot\partial_x\mathbf{U}\ , \end{aligned} \tag{B.3}$$

where the differentiation of the velocity vectors is to be taken componentwise.

Since in practical implementations only moments are needed it pays to group the expansion coefficients with the collisional invariants, i.e. the elements of ϕ:

$$a = a_0 + a_1 u + a_2 v + a_3 w + \frac{a_4}{2}(u^2 + v^2 + w^2) \ . \tag{B.4}$$

Suppose the gradients of the conservative variables Q are known, where we assume that the gradient of density, momentum components and energy are stored in this order in a five-element vector. The following recursive action can be used to compute the expansion coefficients from Eq. (B.3)

$$\begin{aligned}
\partial_x U_i &= \frac{1}{\rho}\left(\partial_x Q_{i+1} - U_i \partial_x Q_1\right) \qquad i = 1,2,3 && \text{(B.5)}\\
\partial_x p &= (\gamma - 1)\left\{\partial_x Q_5 - \frac{1}{2}\partial_x Q_1 (U^2 + V^2 + W^2) - \frac{1}{2}\mathbf{U}\cdot\partial_x \mathbf{U}\right\} && \text{(B.6)}\\
\partial_x \lambda &= \partial_x\left(\frac{\rho}{2p}\right) = \frac{\lambda}{\rho}\partial_x Q_1 - \frac{2\lambda^2}{\rho}\partial_x p && \text{(B.7)}\\
a_4 &= -C_\lambda \partial_x \lambda && \text{(B.8)}\\
a_3 &= 2\lambda \partial_x W - 2W a_4 && \text{(B.9)}\\
a_2 &= 2\lambda \partial_x V - 2V a_4 && \text{(B.10)}\\
a_1 &= 2\lambda \partial_x U - 2U a_4 && \text{(B.11)}\\
a_0 &= -2U a_1 - 2V a_2 - 2W a_3 + \frac{1}{\rho}\partial_x Q_1 - \frac{5}{2\lambda} a_4 - (\mathbf{U}\cdot\mathbf{U}) a_4 && \text{(B.12)}
\end{aligned}$$

Note that in the equation for a_4 we have by default $C_\lambda = 1$. However, if an f_1 state is computed using the method described below, the value is set to $C_\lambda = 1/\mathrm{Pr}$. This ensures that the heat conduction coefficient is set according to the specified Prandtl number Pr, if the collision time τ is used to set the coefficient of viscosity [62].

This routine can be used to compute f_1 states by enforcing $\int \phi(\partial_t f_0 + u\partial_x f_x)\, d\Xi = 0$, where it is understood that the spatial expansion is re-scaled to set the correct Prandtl number [1]. It can be appreciated that the time derivative of the Maxwellian is given by Eq. (B.3) with the space derivatives replaced by time derivatives. All time derivatives are computed by the conservation constraint

$$\int \phi \partial_t f_0 \, d\Xi = \int \phi u \partial_x f_0 \, d\Xi \ . \tag{B.13}$$

---

[1] which means that it is strictly speaking not any longer the derivative of the maxwellian. However we continue to use this nomenclature for convenience

According to Eqns (2.26) and (A.4) the right-hand side is the term that gives the right-hand side of $\partial_t q = \Phi_0$ in the Chapman-Enskog expansion, which means that it must be equal to the corresponding time derivatives of the macroscopic variables on the left-hand side. It can thus be input to Eq. (B.5) through (B.12) to compute the time derivatives of the Maxwellian. The constraint $\int \phi f_1 \, d\Xi = 0$ represented by Eq. (B.13) thus eforces the Euler equations as the macroscopic equations corresponding to the Maxwellian distribution. Note that the scaling of the temperature gradients with the Prandtl number effectively means, that the f_1 state will correspond to the Chapman-Enskog expansion of the Boltzmann Equation.

Appendix C

Implementation of the SD Scheme

C.1 SD Scheme for the Means in Higher Dimensions

Suppose u can be represented by a d-dimensional polynomial of degree at most n, and F is represented by a d-dimensional polynomial of degree at most $n+1$. The semi-discrete analog for Eq. (4.24) can be written:

$$\mathcal{I} = \frac{d}{dt}\int_{T_d} u\, dV = -\int_{T_d} \sum_{k=1}^{N_f} \nabla M_k \cdot \mathbf{F}_h(x_k)\, dV = -V \sum_{j=1}^{N_u} w_j \sum_{k=1}^{N_f} \nabla M_k(\mathbf{x}_j) \cdot \mathbf{F}_h(x_k) \ . \tag{C.1}$$

Note that all the equalities are exact by construction, provided the collocation points for u support a quadrature with weights w_j, which is exact for polynomials of degree n.

Denote the elements of the differentiation matrices $\nabla M_k(\mathbf{x}_j) = \mathbf{m}_{kj}$, where each $\mathbf{m}_{kj}$ is a vector with d components. We have

$$\mathcal{I} = -V \sum_{k=1}^{N_f} \sum_{j=1}^{N_u} w_j \mathbf{m}_{kj} \cdot \mathbf{F}_h(x_k) = -\sum_{k=1}^{N_f} \widetilde{\mathbf{w}_k} \cdot \mathbf{F}_h(x_k) \ . \tag{C.2}$$

We thus arrive at a modified quadrature using the flux points. The integral, however, must not depend on interior flux points. Indeed, we can write using the definition of the new weights $\widetilde{w_k}$:

$$\widetilde{\mathbf{w}_k} = V \sum_{j=1}^{N_u} w_j \mathbf{m}_{kj} = \int_{T_d} \nabla M_k(\mathbf{x})\, dV = \int_{\partial T_d} M_k \nu\, dA \ , \tag{C.3}$$

where ν is the outward pointing face normal on ∂T_d. The second equality is exact, because ∇M is a polynomial of degree n. Suppose the collocation points of F restricted to the boundaries of the simplex support a $d-1$-dimensional quadrature of degree $n+1$ for each boundary $l = 1 \dots d$, with weights w_m^l, $m = 1 \dots N_e$, where N_e is the number of points on the boundaries. We then have the following exact relationship:

$$\widetilde{\mathbf{w}_k} = \int_{\partial T_d} M_k \nu \, dA = \sum_{l=1}^{d} \mathbf{S}^l \sum_{m=1}^{N_e} w_m^l M_k(\mathbf{x}_m) \ , \tag{C.4}$$

where S^l is the area of face l. But since M_k is an interpolation polynomial with the property $M_k(\mathbf{x}_m) = \delta_{km}$, this will simply pick out the weight w_m^l corresponding to the flux node k. In particular it is clear that the weight $\widetilde{w}_k$ for interior nodes vanish, since $k \neq m$ for all such nodes. We finally arrive at

$$\mathcal{I} = -\sum_{k=1}^{N_f} \widetilde{\mathbf{w}_k} \cdot \mathbf{F}_k = -\sum_{l=1}^{d} \sum\nolimits_{m=1}^{N^e} \mathbf{w}_m^l \mathbf{F}_h(x_{k(l,m)}) \cdot \mathbf{S}^l = -\int_{\partial T_n} \mathbf{F}_h \cdot d\mathbf{S} \ , \tag{C.5}$$

where it is understood that the m^{th} node on the l^{th} edge, has a corresponding index $k(l,m)$ in the flux collocation.

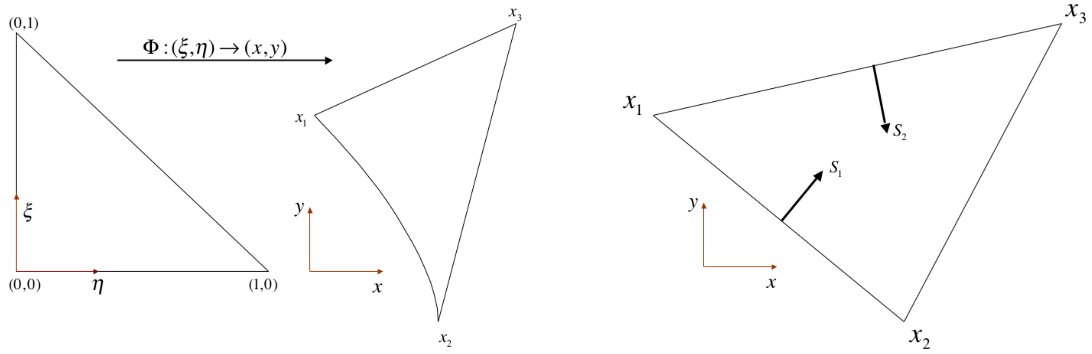
C.2 The SD Scheme on General Triangular Meshes

Consider the semidiscrete scheme, Eq. (4.7), which may be written as

$$\frac{du_{ij}}{dt} + R_{ij} = 0 \ , \tag{C.6}$$

where the residual is given by the exact differentiation of the reconstructed flux function $R_{ij} = (\nabla \cdot F_h)(x_{ij})$. The differentiation is first computed for a fixed reference element. The operation for general mesh elements is then obtained by using a transformation between the straight sided reference element and elements in the mesh. Such a transformation Φ is schematically depicted in Fig. C.1(a). The advantage of this approach is that the differentiation operator is universal for all mesh elements with straight edges, up to simple metric scaling terms, which will be outlined in the following. Consider a differentiation matrix for an interpolated function on the reference element, i.e.

$$\sum_{k=0}^{N(m)} D_{jk}^{\xi} f(x_k) = \frac{\partial}{\partial \xi} (\mathcal{I}^m f)(x_j) \ . \tag{C.7}$$



(a) Transformation between the reference element and a general mesh element.

(b) Definition of metric terms for straigh-sided mesh element.

Figure C.1: Implementation of the SD scheme on triangular meshes: Definitions for general mesh elements.

Suppose such differentiation matrices for ξ and η are known for interpolations of degree $m+1$. Then the divergence of the flux function $F_h = (f_h, g_h)^T$ may be written for a general mesh by applying the chain rule:

$$(\nabla \cdot F_h)(x_{ij}) = \sum_{k=0}^{N(m+1)} D^{\xi}_{jk}\, \nabla\xi|_{x_{ij}} \cdot F_h(x_{ik}) + D^{\eta}_{jk}\, \nabla\eta|_{x_{ij}} \cdot F_h(x_{ik}) \ , \tag{C.8}$$

where the gradients $\nabla\xi$ and $\nabla\eta$ are the rows of the inverse Jacobian of the transformation $\Phi(\xi,\eta)$. Consider a straight-sided mesh element, for which the transformation may be written

$$x = x_1 + (x_2 - x_1)\xi + (x_3 - x_1)\eta \tag{C.9}$$

$$y = y_1 + (y_2 - y_1)\xi + (y_3 - y_1)\eta \ , \tag{C.10}$$

where the (x_j, y_j) are the vertex coordinates of the triangle, ordered counter-clockwise. In this case it is evident that the elements of inverse Jacobian are given by the inward pointing directed face areas, depicted in Fig. C.1(b), divided by twice the volume (as $2V = (x_2 - x_1) \times (x_3 - x_1)$), and the residual may be written:

$$(\nabla \cdot F_h)(x_{ij}) = \frac{1}{2V_i} \sum_{k=0}^{N(m+1)} D^{\xi}_{jk} S_2 \cdot F_h(x_{ik}) + D^{\eta}_{jk} S_1 \cdot F_h(x_{ik}) \ . \tag{C.11}$$

Note that the Jacobian of Φ is thus constant in each element, which means that it does not have to be stored for each node of the element. The differention operator depends on the mesh element only through the volume and face areas. For elements with curved edges, however, higher order terms in Eq. (C.9) appear, which means that the elements of the Jacobian depend on (ξ, η) and need to be stored for all solution nodes, which increases the storage requirement quadratically with the order of accuracy in two dimensions, and cubically in three dimensions. Fortunately, elements with curved edges are normally only needed at boundaries, where it is necessary to have an accurate representation of the geometry, see section 4.7.1. Some authors suggest that it is viable to use straight sided elements everywhere, and respect the geometry at the boundaries only through modified normals. In this work isoparametric elements have been used, meaning that a transformation is computed using polynomials of the same degree as those of the solution representation.

The local differentiation matrices D^ξ and D^η may be computed using suitable basis functions, such as the Dubiner basis for triangles, which uses local tensor coordinates, given by the singular transformation

$$a = \frac{2\xi}{1-\eta} - 1 \quad , \qquad b = 2\eta - 1 \ , \tag{C.12}$$

to write the basis as a tensor product of Jacobi polynomials. Define

$$\begin{aligned} \widetilde{\psi}^a_n &= P^{(0,0)}_n(a) & \text{(C.13)} \\ \widetilde{\psi}^b_{nl} &= \left(\frac{1-b}{2}\right)^n P^{(2n+1,0)}_l(b) \ , & \text{(C.14)} \end{aligned}$$

where $P^{(\alpha,\beta)}_i$ are Jacobi polynomials of degree i, which are orthogonal with respect to the weight $w = (1-x)^\alpha (1+x)^\beta$. The triangle basis may then be written

$$\psi_{(n,l)} = \widetilde{\psi}^a_n(a) \widetilde{\psi}^b_{nl}(b) \ , \tag{C.15}$$

where $n + l \leq m$. Since the derivatives of the Jacobi Polynomials can be computed analytically using recursion formulas [4], the differentiation matrices may be computed by solving linear systems for the rows. For example, for the j^{th} row of D^ξ, corresponding to the evaluation of the derivative at the j^{th} solution node one can write

$$\sum_{k=0}^{N(m+1)} D^\xi_{jk} \psi_{(n,l)}(a_k, b_k) = \frac{\partial}{\partial \xi} \psi_{(n,l)}(a_j, b_j) \ , \qquad n + l \leq m \ , \quad j = 0, \ldots, N(m) \ , \tag{C.16}$$

where the subscripts of the tensor coordinates correspond to the k^{th} flux node, and the j^{th} solution node, which may be evaluated using the transformation (C.12), while the derivative with respect to ξ follows by chain rule. A similar expression holds for D^{η}. The advantage of using this basis, as opposed to a monomial basis, for example, is a much more benign numerical conditioning of the linear systems that arise.

The procedures outlined in this section may be used to evaluate the residual at the solution nodes using the collocation based on the flux nodes. The solution may then be advanced using a time integration scheme. Subsequently, the solution at the next time step is evaluated at the flux nodes to compute the next residual. The transfer matrix, $A = \{a_{kj}\}_{k=1,\ldots,N(m+1),j=1,\ldots,N(m)}$, for the evaluation of the solution at the flux nodes is again universal for straight sided elements. One may write

$$\sum_{j=0}^{N(m)} A_{kj}\psi_{(n,l)}(a_j, b_j) = \psi_{(n,l)}(a_k, b_k)) \ , \qquad n+l \leq m+1 \ , \quad k = 0, \ldots, N(m+1) \ . \quad \text{(C.17)}$$

For curved elements, the transfer matrices are computed using an explicit evaluation of isoparametric mapping Φ instead of the triangle basis, which is by definition of the same order as the polynomial basis. In this case the reconstruction is no longer universal, and the transfer matrix has to be stored for all mesh elements with curved edges.

Bibliography

[1] J. D. Anderson. *Fundamentals of Aerodynamics.* McGraw Hill, 2001.

[2] H. L. Atkins and C. W. Shu. Quadrature-free implementation of discontinuous galerkin method for hyperbolic equations. *AIAA Journal*, 36(5):775–782, 1998.

[3] F. Bassi and S. Rebay. A high-order accurate discontinuous finite-element method for the numerical solution of the compressible navier-stokes equations. *J. Comp. Phys.*, 131:267–279, 1997.

[4] H. Bateman. *Higher Transcendental Function, Vol 2.* McGraw Hill, 1953.

[5] P.L. Bhatnagar, E.P. Gross, and M. Krook. A model for collision processes in gases I: Small amplitude processes in charged and neutral one-component systems. *Phys. Rev.*, 94:511, 1954.

[6] M. O. Bristeau, R. Glowinski, J. Periaux, and H. Viviand, editors. *Numerical simulation of compressible Navier-Stokes Flow: A GAMM Workshop*, volume 18 of *Notes on Numerical Fluid Mechanics.* Vieweg Verlag, Braunschweig; Wiesbaden, 1987. The workshop was held on 4-6 December 1985 at Nice, France by INRIA with the sponsorship of the GAMM Committee on Numerical Methods in Fluid Mechanics.

[7] G. Chavent and B. Cockburn. The local projection P^0-P^1 Discontinuous-Galerkin finite element method for scalar conservation laws. M^2AN, 23(4):565–592, 1989.

[8] S. Y. Chou and D. Baganoff. Kinetic flux-vector splitting for the Navier-Stokes equations. *J. Comp. Pys.*, 130:217–230, 1996.

[9] B. Cockburn and S. Y. Lin. TVB Runge-Kutta local projection Discontinuous Galerkin finite element method for conservation laws III: One dimensional systems. *J. Comp. Phys.*, 84:90–113, 1989.

[10] B. Cockburn and C. W. Shu. TVB Runge-Kutta local projection Discontinuous Galerkin finite element method for conservation laws II: General framework. *Math. Comp.*, 52(186):411–435, 1988.

[11] B. Cockburn and S. W. Shu. The Local Discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM. J. Numer. Anal.*, 35(6):2440–2463, 1998.

[12] B. Cockburn and S. W. Shu. The Runge-Kutta Discontinuous Galerkin method for conservation laws V: Multidimensional systems. *Math. Comp.*, 141:199–224, 1998.

[13] S. M. Deshpande. A second order accurate, kinetic-theory based, method for inviscid compressible flows. Tech. paper no. 2613, NASA Langley., 1986.

[14] B. Fornberg. A numerical study of steady viscous flow past a circular cylinder. *J. Fluid Mech.*, 98(4):819–855, 1980.

[15] A. Gelb. Detection of edges in spectral data. *Appl. Comp. Harm. Anal.*, 7:101–135, 1999.

[16] A. Gelb. Parameter optimization and reduction of round-off error for the gegenbauer reconstruction method. *J. Sci. Comp.*, 20(3):433–459, 2004.

[17] A. Gelb and E. Tadmor. Detection of edges in spectral data ii. nonlinear enhancement. *SIAM J. Numer. Anal.*, 38(4):1389–1408, 2000.

[18] E. Godlewski. *Numerical Approximation of Hyperbolic Systems of Conservation Laws.* Springer Verlag, 1996.

[19] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications.* CBMS-NSF Regional Conference Series in Applied Mathematics, 1977.

[20] D. Gottlieb and C. W. Shu. Resolution properties of the fourier method for discontinuous waves. *Comput. Meth. Appl. Mech. Eng.*, 116:27–37, 1994.

[21] D. Gottlieb and C. W. Shu. On the Gibbs phenomenon IV: Recovering exponential accuracy in a subinterval from the gegenbauer partial sum of a piecewise analytic function. *Math. Comp.*, 64:1081–1095, 1995.

[22] D. Gottlieb and C. W. Shu. On the gibbs phenomenon V: Recovering exponential accuracy from collcoation point values of a piecwise analytic function. *Numer. Math.*, 71:511–526, 1995.

[23] D. Gottlieb and C. W. Shu. On the Gibbs phenomenon III: Recovering exponential accuracy in a subinterval from the spectral partial sum of a piecewise analytic function. *SIAM J. Numer. Anal.*, 33:280–290, 1996.

[24] D. Gottlieb and C. W. Shu. On the gibbs phenomenon and its resolution. *SIAM Rev.*, 39(4):644–668, 1997.

[25] D. Gottlieb, C. W. Shu, A. Solomonoff, and H. Vandeven. On the Gibbs phenomenon I: Recovering exponential accuracy from the fourier partial sum of a nonperiodic analytic function. *J. Comput. Appl. Math.*, 43:81–92, 1992.

[26] A. Harten. High-resolution schemes for hyperbolic conservation laws. *J. Comp. Phys.*, 49(3):357–393, 1983.

[27] A. Harten. On a class of high-resolution total-variation-stable finite-difference schemes. *SIAM J. Numer. Anal.*, 21(1):1–23, 1984.

[28] A. Harten and P. D. Lax. A random choice finite-difference scheme for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 18:285–315, 1981.

[29] A. Haselbacher and J. Blazek. Accurate and efficient discretization of Navier-Stokes equations on mixed grids. *AIAA Journal*, 38(11):2094–2102, 2000.

[30] R. D. Henderson. Details of the drag curve near the onset of vortex shedding. *Phys. Fluids*, 7(9):2102–2104, 1995.

[31] J. S. Hesthaven. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. *SIAM J. Num. Anal.*, 35(2):655–676, 1998.

[32] A. Jameson. Solution of the Euler equations for two dimensional transonic flow by a multigrid method. *Appl. Math. Comp.*, 13:327–356, 1983.

[33] A. Jameson. A non-oscillatory shock capturing scheme using flux limited dissipation (Princeton University MAE Report 1653). In B. E. Engquist, S. Osher, and R. C. J. Sommerville, editors, *Lectures in Applied Mathematics, Vol. 22, Part 1, Large Scale Computations in Fluid Mechanics*, pages 345–370. AMS, 1985.

[34] A. Jameson. Multigrid algorithms for compressible flow calculations (Princeton University MAE Report 1743). In W. Hackbusch and U. Trottenberg, editors, *Proceedings of the* 2^{nd} *European Conference on Multigrid Methods, Cologne, Germany*, pages 166–201. Springer Verlag, 1986.

[35] A. Jameson. Analysis and design of numerical schemes for gas dynamcics 1: Artificial diffusion, upwind biasing, limiters and their effect on accuracy and multigrid convergence. *Int. J. Comp. Fluid. Dyn.*, 4:171–218, 1995.

[36] A. Jameson. Analysis and design of numerical schemes for gas dynamcics 2: Artificial diffusion and discrete shock structure. *Int. J. Comp. Fluid. Dyn.*, 5:1–38, 1995.

[37] A. Jameson and D. A. Caughey. How many steps are required to solve the Euler equations of steady compressible flow: In search of a fast solution algorithm. AIAA Paper 01-2673, 2001.

[38] A. Jameson and J. C. Vassberg. A vertex-centroid (V-C) scheme for the gas-dynamics equations. In N. Satofuka, editor, *Computational Fluid Dynamics 2000: Proceedings of the First International Conference on Computational Fluid Dynamics, Kyoto, Japan*, pages 37–52. Springer Verlag, 2001.

[39] A. Jameson and S. Yoon. Lower-upper implicit schemes with multiple grids for the Euler equations. *AIAA Journal*, 25(7):929–935, 1987.

[40] C. A. Kim and A. Jameson. A robust and accurate LED-BGK solver on unstructured adaptive meshes. *J. Comp. Phys.*, 143:598–627, 1998.

[41] M. N. Kogan. *Rarefied Gas Dynamics.* Plenum Press, 1969.

[42] D. A. Kopriva. A conservative staggered-grid Chebyshev multidomain method for compressible flows II: Semi-structured method. *J. Comp. Phys..*, 128:475, 1996.

[43] P. D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves.* SIAM Regional Series on Applied Mathematics, 1973.

[44] P. D. Lax. Accuracy and resolution in the computation of solutions of linear and nonlinear equations. In *Recent Advances in Numerical Analysis, Proc. Symp.*, pages 107–117. Mathematical Research Center, University of Wisconsin, Academic Press, 1978.

[45] R.J. LeVeque. *Finite-Voume Methods for Hyperbolic Problems.* Cambridge University Press, 2002.

[46] Y. Liu, M. Vinokur, and Z. J. Wang. Discontinuous spectral difference method for conservation laws on unstructured grids. In *Proceedings of the 3rd International Conference on Computational Fluid Dynamics, July 12-16, 2004, Toronto, Canada.* Springer, 2004.

[47] I. Lomtev and G. E. Karniadakis. A discontinuous galerkin method for the Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 29:587–603, 1999.

[48] G. May and A. Jameson. High-order accurate methods for high-speed flow. AIAA Paper 05-5251, 2005.

[49] G. May and A. Jameson. Unstructured algorithms for inviscid and viscous flows embedded in a unified solver architecture: Flo3xx. AIAA Paper 05-0318, 2005.

[50] G. May and A. Jameson. A Spectral Difference Method for the Euler and Navier-Stokes equations on unstructured meshes. AIAA Paper 06-0304, 2006.

[51] T. Ohwada. On the construction of kinetic schemes. *J. Comp. Phys.*, 177:156–175, 2002.

[52] T. Ohwada and S. Kobayashi. Management of discontinuous reconstruction in kinetic schemes. *J. Comp. Phys.*, 197:116–138, 2004.

[53] S. Osher. Riemann solvers, the entropy condition, and difference approximations. *SIAM J. Numer. Anal.*, 21(2):217–235, 1984.

[54] S. Osher. Convergence of generalized MUSCL schemes. *SIAM J. Numer. Anal.*, 22(5):947–961, 1985.

[55] B. Perthame. *Kinetic Formulation of Conservation Laws.* Oxford Lecture Series in Mathematics, 2002.

[56] K.H. Prendergast and K. Xu. Numerical hydrodynamics from gas-kinetic theory. *J. Comp. Phys.*, 109:53–66, 1993.

[57] D. I. Pullin. Direct simulation methods for compressible inviscid gas flow. *J. Comp. Phys.*, 34:231–244, 1980.

[58] P. L. Roe. Approximate riemann solvers, parameter vectors, and difference schemes. *J. Comp. Phys.*, 43:357–372, 1981.

[59] C.W. Shu. Tvb uniformly high-order schemes for conservation laws. *Math. Comp.*, 49(179):105–121, 1987.

[60] C.W. Shu. Total-variation-diminishing time discretizations. *SIAM J. Sci. Stat. Comput.*, 9(6):1073–1084, 1988.

[61] C.W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock capturing schemes. *J. Comp. Phys.*, 77:439–471, 1988.

[62] B. Srinivasan. *The BGK and LRS Schemes for Computing Euler and Navier-Stokes Flows.* PhD thesis, Stanford University, Stanford, CA 94305, September 2005.

[63] E. Tadmor. Convergence of spectral methods for nonlinear conservation laws. *SIAM J. Num. Anal.*, 26(1):30–44, 1989.

[64] H. Takami and H. B. Keller. Stead two-dimensional viscous flow of an incompressible fluid past a circular cylinder. *Phys. Fluids*, 12(II):51–56, 1969.

[65] J. Tanner and A. Gelb. Robust reprojection methods for the resolution of Gibbs phenomenon. *Appl. Comput. Harm. Anal. (preprint)*, 2004.

[66] D. J. Tritton. Experiments on the flow past a circular cylinder at low Reynolds numbers. *J. Fluid Mech.*, 6:547–567, 1959.

[67] V. Venkatakrishnan. Convergence to steady state solutions of the Euler equations on unstructured grids with limiters. *J. Comp. Phys.*, 118:120–130, 1995.

[68] W. G. Vincenti. *Introduction to Physical Gas Dynamics.* Krieger Publishing Company, 2002.

[69] Z. J. Wang and Y. Liu. The Spectral Difference Method for the 2D Euler equations on unstructured grids. AIAA Paper 05-5112, 2005.

[70] F. M. White. *Viscous Fluid Flow.* McGraw Hill, 1974.

[71] P. Woodward and P. Colella. The numerical simulation of two-dimensional fluid flow with strong shocks. *J. Comp. Phys.*, 54:115–173, 1984.

[72] K. Xu. Gas-kinetic schemes for unsteady compressible flow simulations. Von Karman Institute for fluid dynamics lecture series, 1998-03, 1998.

[73] K. Xu. A gas-kinetic BGK scheme for the Navier-Stokes equations and its connection with artificial dissipation and godunov method. *J. Comp. Phys.*, 171(48):289–335, 2001.

[74] K. Xu, M. Mao, and L. Tang. A multidimensional gas-kinetic BGK scheme for hypersonic viscous flow. *J. Comp. Phys.*, 203:405–421, 2005.

[75] K. Xu, L. Martinelli, and A. Jameson. Gas-kinetic finite-volume methods, flux vector splitting and artificial diffusion. *J. Comp. Phys.*, 120(48), 1995.

[76] L. Zhang Z. J. Wang and Y. Liu. High-order spectral volume method for 2D Euler equations. AIAA Paper 03-3534, 2003.